



vrije Universiteit

Het modelleren en voorspellen van het landelijke weekverkeer

Het modelleren en voorspellen van het landelijke weekverkeer

Auteur: Inge Thiescheffer,
Studente Bedrijfswiskunde en Informatica aan de
VU

Managementuittreksel

Probleemstelling:

Kan een mathematisch model bijdragen aan de efficiency?

Vanwege de teruglopende hoeveelheid geadresseerde post wordt TPG Post geconfronteerd met het probleem dat het moeilijker wordt om de doelstelling 'het verbeteren van de efficiency' te verwezenlijken. Door middel van het goed prognosticeren van de wekelijkse hoeveelheid post wordt er door TPG Post getracht een goed instrument te leveren die de vestigingen van TPG Post moet ondersteunen in het halen van de doelstelling 'het verbeteren van de efficiency'.

Naar aanleiding van het belang van goede prognoses is het onderzoek 'modelleren en voorspellen van het landelijke weekverkeer' gestart. Door verschillende mathematische modellen te ontwikkelen, kan er geanalyseerd worden of de huidige prognoses verbeterd kunnen worden. Een mathematisch model zal een meerwaarde voor TPG Post hebben, indien de prognoses een maximale relatieve afwijking van 2% hebben in plaats van de huidige maximale relatieve afwijking van 10%.

Beschrijving van de werkzaamheden:

Ontwikkeling van 10 modellen en analyse.

Gezien de breedte van het onderzoeksterrein en de beperkte duur van een stage, is er voor gekozen om de volgende modellen te ontwikkelen:

Type mathematisch model	Normaal	Uitbreiding		
		Gebruik makend van de informatie over grote partijen	Gebruik makend van de door de TPG verwachte grootte van invloed van PHP ¹	Gebruik makend van de Invloed van de omliggende weken
tijdreeks model	3.1	3.2 en 3.4	3.3 en 3.4	
log-lineair model	4.1	4.2		
log-quasi-likelihood model	4.1	4.2		
Poisson model	5.1	5.2 en 5.4		5.3 en 5.4

Tabel 0.1: Overzicht van de ontwikkelde modellen (naam+eventuele uitbreiding)

Nadat deze modellen ontwikkeld zijn, is er geanalyseerd of deze modellen toepasbaar zijn voor het maken van prognoses van de wekelijkse hoeveelheid post. Tevens is de verwachte invloed van bijvoorbeeld feestdagen en maandwisselingen op de hoeveelheid post geanalyseerd.

Resultaten en conclusies:

3 modellen kunnen waarschijnlijk bijdragen aan de efficiency

De ontwikkelde modellen zijn niet in staat prognoses te maken met een maximale relatieve afwijking van 2%. De kwaliteit van de prognoses van log-quasi-likelihood model 4.1 en Poisson model 5.4 nadert echter de kwaliteit van de huidige prognoses. Met behulp van meer data wordt er verwacht dat de kwaliteit van de prognoses van log-quasi-likelihood model 4.2 en Poisson model 5.4 beter zal zijn dan de kwaliteit van de huidige prognoses.

De verwachte invloed van feestdagen, maandwisselingen en andere variabelen komen niet overeen met de verwachtingen van medewerkers van TPG Post. Uitzondering hierop is de verwachte invloed van Pasen.

De log-quasi-likelihood modellen 4.1 en 4.2 en Poisson-model 5.4 blijken toepasbaar te zijn voor het prognosticeren van de wekelijkse hoeveelheid post. Het log-quasi-likelihood model is tevens toepasbaar voor een analyse van de grootte van de invloeden van de verschillende variabelen.

¹ PHP = Pasen, Hemelvaart en Pinksteren

Deze modellen kunnen ook gebruikt worden voor het maken van prognoses van de wekelijkse hoeveelheid post per vestiging of voor prognoses van de hoeveelheid post die per dag of per week gesorteerd moet worden. Indien deze modellen verbeteringen in de kwaliteit van de prognoses zullen opleveren, zal er een beter instrument zijn voor het behalen van de doelstelling 'het verbeteren van de efficiency'.

Tot slot

Omdat de prognoses van de log-quasi-likelihood modellen 4.1 en 4.2 als ondersteuning gebruikt zullen worden tijdens het maken van prognoses, is er een gebruikershandleiding geschreven. Deze gebruikershandleiding beschrijft hoe er met de log-quasi-likelihood modellen prognoses van de wekelijkse landelijke hoeveelheid post gemaakt kunnen worden. Nu deze gebruikershandleiding ontwikkeld is, kan TPG Post verder met het proces 'het verbeteren van de prognoses'.

Inhoud

Begrippenlijst	13
Voorwoord	15
1. Bedrijfsbeschrijving	17
1.1 TPG N.V.	17
1.2 Processen TPG Post Nederland	18
1.3 Organisatie TPG Post Nederland	19
1.4 Afdeling Distributie	19
2. Probleembeschrijving	21
2.1 Huidige methode prognosticeren werkaanbod	21
2.2 Probleemdefinitie	24
2.3 Aanpak afstudeeronderzoek	25
2.4 Betrokkenen tijdens het afstudeeronderzoek	26
3. Tijdreeks model	27
3.1 Introductie	27
3.2 Beschrijving model	31
3.2.1 Trend	32
3.2.2 Seizoensinvloed	34
3.2.3 Controle op afhankelijkheid tussen residuen Y_t	40
3.3 Stationaire tijdreeks model	43
3.3.1 Keuze stationaire tijdreeks model	43
3.3.2 Bepaling orde	45
3.3.3 Schatten parameters	45
3.3.4 Controle op afhankelijkheid tussen residuen R_t	46
3.4 Analyse fitting	48
3.5 Voorspellingen	49
3.5.1 Voorspellingmethode model 3.2 en 3.4	49
3.5.2 Voorspellingsmethode model 3.1 en 3.3	50
3.5.3 Analyse voorspellingen	50
3.6 Conclusie	55
4. Log-linear model & Log-quasi-likelihood model	57
4.1 Introductie	57
4.2 Beschrijving model	60
4.3 Analyse fitting	62
4.3.1 Response residuen	62
4.3.2 Deviance residuen	64
4.3.3 Pearson residuen	66
4.3.4 Conclusie n.a.v. analyse fitting	69
4.4 Voorspellingen van log-quasi-likelihood modellen	70
4.4.1 Voorspellingsperiode van 48 weken	70
4.4.2 Voorspellingsperiode van 8 weken	73
4.4.3 Voorspellingsperiode van 1 week	76
4.4.4 Voorspellingsintervallen	77
4.5 Conclusie	78
5. Poisson model	79
5.1 Introductie	79
5.2 Beschrijving model	79
5.3 Fitting model	82
5.4 Voorspellingen	85
5.5 Conclusie	88
6. Conclusie	89
6.1 Conclusies wat betreft de voorspellingen	89

6.2	Conclusies wat betreft de variabelen die van invloed zijn	90
6.3	Aanbevelingen	91
Bijlage III		
A	Extremen binnen de data	93
B	Toename zomerinvloed	94
C	State-space model & Kalman filter	95
D	Cijfers fitting tijdreeks model	96
E	Cijfers voorspellingen	100
Bijlage IV		
A	Analyse effect verschillende voorspelperiodes	104
B	De regressiewaarnemingen x en de regressiecoëfficiënt β	105
C	Cijfers fitting log-lineair model	111
D	Schatten van de regressiecoëfficiënt β voor 2003	115
E	Cijfers voorspellingen	117
F	Toets voor uitbijters	123
G	Klassiek lineair regressiemodel	124
H	Kernschatter	126
I	Cijfers fitting log-quasi-likelihood model	127
Bijlage V		
A	Meest aannemelijke schatter	129
B	Cijfers fitting Poisson model	133
C	Schatten van de hoeveelheid post voor 2003	137
D	Cijfers voorspellingen	139
Bijlage V1		
A	Informatie over de invloed van tijd gerelateerde variabelen	143
Referenties		152

Begrippenlijst

Afgiftepunt BOP	<ul style="list-style-type: none"> - Een punt waar je post kan afgeven - Afdeling Beheer en Ondersteuning Processen die medeverantwoordelijk is voor het doelgericht en efficiënt werken bij Distributie
Briefpost aan huis	<ul style="list-style-type: none"> - Alle poststukken die in een brievenbus passen en die zonder een extra dienst bezorgd worden.
Business Balie	<ul style="list-style-type: none"> - Een collectiepunt waar de gecollecteerde poststukken gewogen en gecontroleerd worden op bijvoorbeeld het voorgemelde aantal stuks, en daarna worden getransporteerd naar het dichtstbijzijnde sorteercentrum.
Collectie	<ul style="list-style-type: none"> - De collectie van post geschiedt op een viertal manieren: het halen van de post in de brievenbussen, aanlevering van post van het midden- en klein bedrijf bij een Business Point, aanlevering van post op een postkantoor en aanlevering van post op een business balie. De post die door TPG gehaald wordt bij onder andere bedrijven en kantoren, wordt aangeleverd op een business balie, waardoor dit proces onder aanlevering van post op een business-balie valt.
Efficiency Fitting	<ul style="list-style-type: none"> - $(\text{Productie})/(\text{Aantal gewerkte uren})$ - Indien er een kleine afwijking is tussen de werkelijke hoeveelheid post in week t en de door een model geschatte hoeveelheid post in week t, wordt er gesproken van een goede fitting.
Grote partij	<ul style="list-style-type: none"> - Een partij bestaande uit meer dan 200.000 stukken post, waarbij het poststuk dat verspreid wordt afkomstig is van 1 klant (bedrijf).
Kwartaalcijfer Manuren	<ul style="list-style-type: none"> - Verwachte hoeveelheid te distribueren post in een kwartaal - Aantal gewerkte uren
Ongeadresseerde post	<ul style="list-style-type: none"> - Post die niet geadresseerd is en bedoeld is om bij alle afgiftepunten, behalve de huizen met een 'nee-geen-ongeadresseerde-poststukken' sticker, bezorgd te worden.
Postdichtheid Productie Productiviteit Sorteerslag	<ul style="list-style-type: none"> - Het aantal poststukken per afgiftepunt per tijdseenheid - Aanral verspreide poststukken - $(\text{Productie})/(\text{Aantal betaalde uren})$ - Een van de drie onderdelen(ook wel 'slagen' genoemd) van het sorteerproces om als eindprodukt bundels gesorteerd op straat en huisnummer op te leveren. De drie onderdelen(/slagen) zijn: (1) sorteren naar 180 geografische gebieden (2) sorteren op loop (3) sorteren op straat en huisnummer.
Sortering op loop	<ul style="list-style-type: none"> - Indien er op loop gesorteerd is, dan komen de verschillende bundels post bij de juiste postbodes terecht. De postbode ontvangt in het algemeen meerdere bundels. Om een korte route te lopen en tijd te besparen sorteert de postbode de bundels zo efficiënt mogelijk op straat en huisnummer.
Vestiging	<ul style="list-style-type: none"> - Locatie waar postbodes de post sorteren op loop (indien er geen huisnummersorteermachine aanwezig is op de vestiging), en tevens de locatie vanwaar postbodes vertrekken om de verschillende poststukken te bezorgen. Er zijn ongeveer 500 vestigingen in Nederland.
WAP	<ul style="list-style-type: none"> - Afdeling Werk Aanbod Prognose welke o.a. verantwoordelijk is voor een wekelijkse prognose over de hoeveelheid te distribueren post per week.

Voorwoord

Voorwoord: het begin, het begin van het einde, het begin van het eindverslag van mijn afstudeeronderzoek. Sollicitatiebrief: het begin van een afstudeeronderzoek, het begin van het wennen aan het bedrijfsleven, een mooi begin. Dit omdat ik na een bezoek aan de carrièrebeurs in de RAI echt 1 bedrijf op het oog had, namelijk TPG. Het bedrijf TPG wekte in de RAI mijn interesse door zijn enthousiasme, door de aanwezigheid van veel kennis met betrekking tot logistiek en door de manier waarop men binnen het bedrijf stagiaires laat kennismaken met het bedrijf. Gevolg: een sollicitatiebrief. Gevolg: een interessant afstudeeronderzoek bij het bedrijf waar ik wilde. Gevolg: een eindverslag van een afstudeeronderzoek.

Tussen het begin en het einde zijn er:

- veel informatiebronnen onderzocht
- veel dingen zonder resultaat uitgeprobeerd
- veel discussies geweest over het verschil tussen de eisen van het bedrijfsleven en het onderzoeksleven
- veel uren gemaakt
- veel gesprekken geweest met vrienden, maar vooral met mijn zus
- veel collega's geweest die mij ondersteunden en geïnteresseerd waren in mijn onderzoek
- etcetera, etcetera.

En al deze tussentijdse dingen hebben direct dan wel indirect geleid tot het gevoel dat er toch meer een onderzoeker in mij schuilt dan ik op voorhand had verwacht. Daarnaast hebben ze geleid tot dit eindverslag. Een eindverslag waarin voornamelijk de ontwikkelde modellen en de analyses van de resultaten van de modellen beschreven zijn, zodat er antwoord gegeven kon worden op de vraag of er met mathematische modellen prognoses gemaakt kunnen worden met een maximale relatieve afwijking van 2%.

De gebruikershandleiding, die geschreven is om het model met de beste resultaten te gebruiken, is niet opgenomen in dit eindverslag. Hiervoor is gekozen omdat deze gebruikershandleiding alleen geschreven is voor praktische doeleinden. Er is dus geen wetenschappelijk onderzoek beschreven in de gebruikershandleiding.

Voor het begin, het einde en alle tussentijdse dingen wil ik veel mensen bedanken. Ynes voor de thee, Ingo voor zijn opmerkingen dat studenten niet hard moeten werken, Mario voor zijn geweldige verhalen over bijvoorbeeld ganzen, Ans en Jeroen voor de goede gesprekken, Jack als quasi-Neerlandicus voor het herkennen van mijn taalfouten, Joep voor het regelen van mijn TPG-tour, zijn feedback over lay-out en leesbaarheid voor niet-wiskundigen en het kenbaar maken van het verschil tussen de het bedrijfsleven en het leven van een onderzoeker, mijn ouders en broer voor van alles en nog wat, mijn zus en Wouter voor hun adviserende rol, en mijn vrienden voor het goed benutten van mijn vrije tijd. Specifiek wil ik Geurt Jongbloed bedanken voor zijn enthousiasme om mij te begeleiden tijdens mijn afstudeeronderzoek, en zijn enthousiasme over de wereld van de stochastiek. Het werkte erg aanstekelijk. Bedankt!

En nu, nu wens ik u allen veel leesplezier. Ik ben klaar!

Inge Margriet Thiescheffer

Den Haag, maart 2004.

1. Bedrijfsbeschrijving

Het onderzoek van deze scriptie legt zich toe op het maken van goede voorspellingen voor de geadresseerde briefpost aan huis² op basis van wiskundige modellen. De divisie Post binnen TPG wordt namelijk door de dalende hoeveelheden poststukken geconfronteerd met het probleem dat het moeilijker is om de doelstelling 'het verbeteren van de efficiency' te verwezenlijken. Een instrument dat ontwikkeld is om de uitvoerende organen te ondersteunen bij het verbeteren van de efficiency is Werk Aanbod Prognose (WAP). De WAP-groep voorspelt de hoeveelheid briefpost aan huis per week per vestiging. Middels het geven van goede voorspellingen, hebben de uitvoerende organen een goed instrument om de efficiency te verbeteren. In hoofdstuk 2 zal de wijze waarop WAP voorspelt, toegelicht worden.

In de volgende paragrafen zullen de geschiedenis, de resultaten en de kernactiviteiten van TPG beschreven worden. Vervolgens zal de organisatie van TPG Post Nederland, onderdeel van de divisie Post van TPG, worden uitgelegd. Tot slot zal er dieper in worden gegaan op de organisatie van de afdeling Distributie en haar productieprocessen.

1.1. TPG N.V.

TPG is de houdstermaatschappij van koninklijke TPG Post en TNT. De oprichting van Koninklijke TPG Post vond plaats in het jaar 1799, toen de posten zich naar Frans voorbeeld omvormden tot een nationale staatsonderneming PTT Post. Tot het moment van verzelfstandiging in 1989 bracht de staat coördinatie en organisatie aan in de nationale postbezorging. Na de verzelfstandiging startte PTT Post (als onderdeel van KPN) nieuwe commerciële activiteiten zoals nationale en internationale expressdiensten en logistieke diensten. Eind 1996 vond de overname plaats van Thomas Nationwide Transport (TNT) door KPN. TNT, een Australisch bedrijf dat opgericht is in 1946, was bekend geworden om zijn diensten op het gebied van distributie en logistiek. De TNT-activiteiten die niet tot de kernactiviteiten post, express en logistiek van PTT Post behoorden, werden afgestoten. Door de overname van TNT werd PTT Post een wereldspeler op het gebied van post, express en logistiek. Deze diensten vallen vanaf halverwege 1998 onder de naam 'TNT Post Group' (TPG). Overige diensten vielen onder de naam 'KPN Telecom'. Beide bedrijven werden in 1998 als aparte fondsen genoteerd op de beurzen van Amsterdam, New York, Londen en Frankfurt. Om de internationale expansie te laten profiteren van de groeiende bekendheid en reputatie van de naam TPG, werd in mei 2002 besloten de naam PTT Post te laten verdwijnen en te vervangen door TPG.

Het netto resultaat van TPG

In het jaar 2002 werd er door TPG een omzet geboekt van 11,8 biljoen euro en werd er een netto resultaat van 599 miljoen euro bereikt. TPG heeft ruim 150 duizend werknemers en is actief in meer dan 200 landen. Een van de doelstellingen van TPG is onder andere om over excellente distributie en supply chain management te beschikken. TPG denkt dit te bereiken door:

- continu de concurrentiepositie van de klant te verbeteren (TPG heeft bijvoorbeeld de postdistributie van Rabobank in handen genomen. Hierbij wil TPG de concurrentiepositie van de Rabobank verbeteren middels een optimale postdistributie waardoor er meer klantentrouw verwezenlijkt wordt.);
- het postbedrijf in de wereld te zijn met de beste kwaliteit tegen de beste prijs;
- het snelste en meest betrouwbare express bedrijf te zijn.

Kernactiviteiten van TPG

De 3 kernactiviteiten van TPG zijn: Post, Express en Logistiek. De doelstelling van de Express divisie is om een autonome winstgevende groei te genereren met wereldwijde business-to-business express diensten, door de snelste en betrouwbaarste express diensten te leveren. De divisie Logistiek richt zich op het verwerven van erkenning als de wereldwijde leider in de logistieke sectoren. Deze divisie onderscheidt zich enerzijds door complexe toeleveringsketens te ontwerpen, in te voeren en te beheren en anderzijds door het hele logistieke proces te integreren en inzichtelijk te maken met behulp van informatie en communicatietechnologie(ICT). De divisie Post ambieert een marktleiderspositie voor diensten aan zakelijke en particuliere klanten op het gebied van communicatie, transacties en aflevering. Dit wil de divisie bereiken door zich toe te leggen op drie aspecten:

- (1) continue aandacht voor kwaliteit, efficiency en klantenservice,
- (2) het realiseren van de beste resultaten in de bedrijfstak en
- (3) het optimaal benutten van de technologieën en de liberalisering van de Europese markt.

² Zie begrippenlijst

1.2. Processen TPG Post Nederland

Het primaire proces van de TPG Post Nederland is weergegeven in figuur 1.1. Gedurende de collectie- werkzaamheden worden de te bezorgen poststukken bij de verschillende kantoren, de business-balies van TPG, en bij de bekende rode brievenbussen opgehaald en naar de verschillende verzamelpunten in Nederland gebracht. Vanaf deze plaatsen worden de poststukken naar het dichtstbijzijnde sorteercentrum vervoerd.



Figuur 1.1. Primaire proces van divisie Post³

De sortering vindt sinds 1 januari 2004 in twee of drie sorteerslagen plaats. De eerste twee sorteerronden vinden plaats op een van de zes sorteercentra in Nederland. Indien de vestiging al over een huisnummersorteermachine beschikt, vindt de derde sorteerslag plaats op de vestiging. De eerste sortering verzorgt een sortering in 180 geografische gebieden. Het geografische gebied is dan bepalend voor een eventuele transport naar een sorteercentrum welke dicht bij het afzetgebied is. In de tweede sortering wordt er doorgesorteerd tot de uiteindelijk voor distributie op te leveren sorteerprodukten (voor postbussen-post is het van belang dat er op postcode-reeksen gesorteerd is,



voor post die niet door de machines kan worden er op bestelkantoor-niveau⁴ gesorteerd, voor overige post is het van belang dat er op loop⁵ gesorteerd is). Deze sorteerprodukten worden daarna bezorgd bij de juiste vestigingen. Sinds 1 januari 2004 vindt er op verschillende vestigingen een derde sortering plaats. In de derde sortering, de huisnummersortering, wordt de post die al gesorteerd is op loop, nog gesorteerd op straat en op huisnummer. Indien de huisnummersorteermachine nog niet aanwezig is op de vestiging, dan wordt het sorteren op straat en huisnummer verzorgd door de postbode. Na het sorteren op straat en op huisnummer vindt de distributie van de poststukken plaats.



³ Sort is de afkorting van sortering.

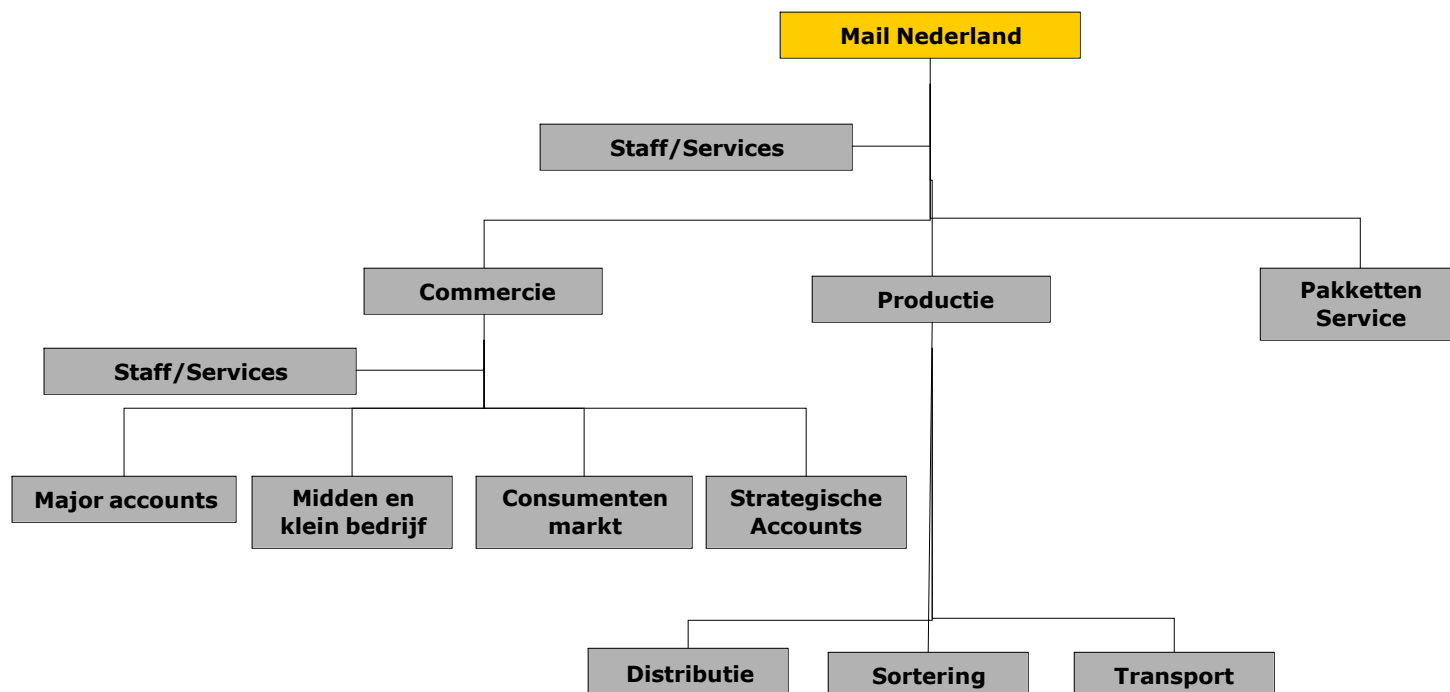
⁴ Zie begrippenlijst

⁵ Zie begrippenlijst

⁸ Er is hier geen rekening gehouden met de hoeveelheid post in de weken 49, 50, 51 en 52, omdat er in deze weken minimaal 1 keer in de jaren 1990 tot en met 2002 een extreme hoeveelheid post gerealiseerd is.

1.3. Organisatie TPG Post Nederland

Om het primaire proces en de secundaire processen goed te laten verlopen, is de organisatie van TPG Post Nederland op de volgende manier ingericht:



Figuur 1.2. TPG Post Nederland – Organisatie structuur

De afdeling Distributie legt onder andere verantwoording af aan de Productie. Productie is het overkoepelende orgaan van de afdelingen die direct gerelateerd zijn aan het primaire proces. Productie is daarom ook verantwoordelijk voor een goede aansluiting tussen de processen van deze verschillende afdelingen.

De afdeling Commercie is deels verantwoordelijk voor de secundaire processen, zoals het werven van klanten en het in stand houden van een goede relatie met de klanten. Middels een goede relatie met de klanten, is de kans groter dat TPG Post Nederland tijdig op de hoogte is van grote hoeveelheden post of een partij post die door bijvoorbeeld zijn formaat en/of gewicht meer werk en tijd vergen. Met behulp van deze informatie kan TPG Post Nederland er voor zorgen dat het primaire proces hierop ingespeeld is, en daarmee verhoogt TPG Post Nederland zijn efficiency.

1.4. Afdeling Distributie

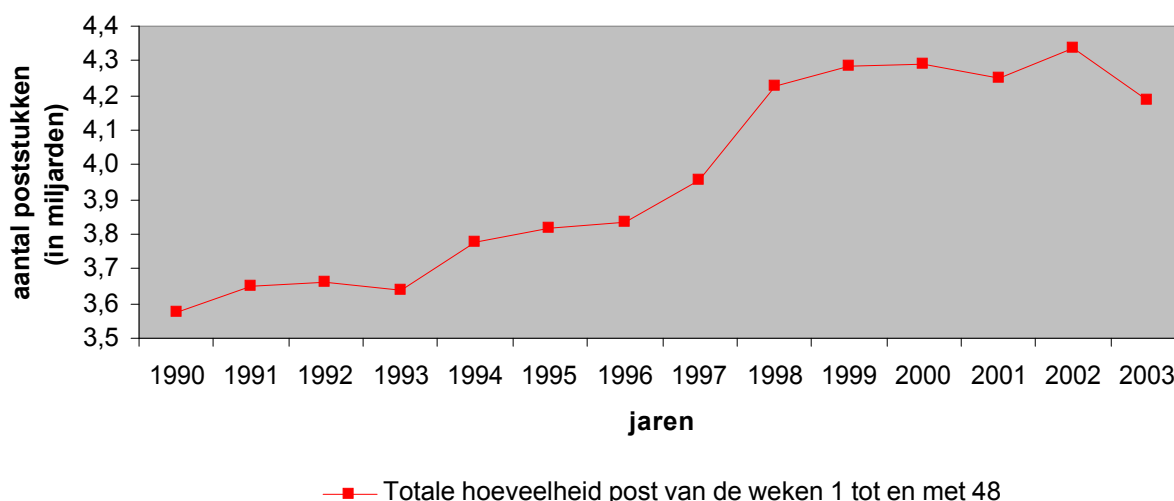
De afdeling Distributie is verantwoordelijk voor de collectie en distributie van poststukken in Nederland, oftewel het verzamelen van de post bij kantoren en brievenbussen en het afleveren van de poststukken bij de geadresseerden. De belangrijkste focus van Distributie is het leveren van een hoge kwaliteit tegen zo laag mogelijke kosten. De uitstekende positie op de postmarkt ontleent Distributie aan een uniek fijnmazig netwerk, hetgeen het mogelijk maakt geadresseerde en ongeadresseerde post te collecteren en op alle adressen in Nederland, 6 dagen per week, te bezorgen. Zowel het collecteren als het uitbouwen en in stand houden van het distributienetwerk en de daarbij behorende kennis en ervaring, rekent zij tot haar kerncompetenties.

De afdeling waar de Werk Aanbod Prognose onder valt is Beheer Ondersteuning Processen (BOP) van Distributie. BOP is een van de operationele hoofdafdelingen van Distributie. Deze afdeling stelt onder andere proces- en productinformatie beschikbaar om de uitvoeringsorganisaties te ondersteunen bij de procesvoering of het uitregelen van nieuwe ontwikkelingen. Verder is de afdeling verantwoordelijk voor het beheer van de standaardprocessen. Beheer Ondersteuning Processen is dus medeverantwoordelijk voor het doelgericht en efficiënt werken bij de verschillende

uitvoeringsorganisaties. Het onderdeel Werk Aanbod Prognose is hierbij verantwoordelijk voor een wekelijkse prognose van de hoeveelheid te distribueren post per week per vestiging. Naar aanleiding van onder andere de prognoses worden er op de verschillende vestigingen beslissingen genomen over een verhoging c.q. verlaging van de in te zetten arbeidscapaciteit om zo de efficiency te verbeteren. Daar de efficiency dus indirect van de prognoses afhangt, is het van belang om aan te tonen dat de voorspellingen van de WAP-groep goed zijn.

2. Probleembeschrijving

Vanwege de teruglopende hoeveelheid geadresseerde post, die veroorzaakt wordt door onder andere de Europese concurrentie (vanwege opheffing monopolie TPG) en de economische ontwikkelingen, wordt TPG Post geconfronteerd met het probleem dat het moeilijker wordt om de doelstelling 'het verbeteren van de efficiency' te verwezenlijken. In de jaren dat de hoeveelheid post toenam, verbeterde de efficiency namelijk automatisch indien de arbeidscapaciteit gelijk bleef of relatief minder toenam dan de hoeveelheden post. Uit onderstaand figuur is waarneembaar dat in de jaren 1993, 2001 en 2003 de totale hoeveelheid poststukken van de weken 1 tot en met 48 is teruggelopen ten opzichte van het jaar daarvoor. Doordat het verbeteren van de efficiency in de jaren dat de hoeveelheid post terugloopt niet meer automatisch is, is het instrument Werk Aanbod Prognose ontwikkeld.



Figuur 2.1: Totale hoeveelheid post van de weken 1 tot en met 48⁸ van de jaren 1990 tot en met 2003

De WAP-groep voorspelt de hoeveelheid 'Briefpost aan huis' per week per vestiging. De leiding van een vestiging kan naar aanleiding van de WAP-voorspellingen, informatie omtrent formaat en gewicht van verwachte poststukken, en hun eigen ervaringen bepalen of er extra of minder hulp ingezet dient te worden zodat de doelstelling 'het verbeteren van de efficiency' wordt behaald.

Om inzicht te krijgen in de WAP-voorspellingen, zal de manier waarop de WAP-groep voorspelt in de volgende paragraaf uitgelegd worden. Nadat er meer bekendheid is met WAP, zal in paragraaf 2.2 het doel van dit afstudeeronderzoek beschreven worden. De opzet van deze rapportage vindt u in paragraaf 2.3.

2.1. Huidige methode prognosticeren werkaanbod

De WAP-groep verzorgt de wekelijkse voorspellingen voor het hele land. Deze voorspellingen zijn gebaseerd op de volgende elementen:

- de onderstroom (de hoeveelheid post die niet toewijsbaar is aan grote partijen of seizoensinvloeden)
- seizoensinvloeden (de invloed van bijvoorbeeld de zomer op de hoeveelheid poststukken)
- grote partijen⁹(over het algemeen zijn partijen post van een klant die bestaan uit meer dan 200.000 stuks van tevoren bekend bij de WAP-groep)

⁹ Een grote partij is op het huidige moment gedefinieerd als een partij bestaande uit meer dan 200.000 stukken post, waarbij het poststuk welke verspreid wordt afkomstig is van 1 klant (bedrijf).

De relatie tussen deze drie elementen kan als volgt geformuleerd worden:

$$O_{w,j} = P_{w,j} - S_{w,j} - G_{w,j}$$

Hierbij is op dit moment

$O_{w,j}$	De onderstroom in week w jaar j
$P_{w,j}$	De hoeveelheid post in week w jaar j
$S_{w,j}$	De seizoensinvloed in week w jaar j,
$S_{w,j} = \begin{cases} > 0 & \text{indien Kerst invloed heeft op week w} \\ -14 \text{ miljoen} & \text{indien Pasen of Pinksteren invloed heeft op week w} \\ -16 \text{ miljoen} & \text{indien Hemelvaart invloed heeft op week w} \\ 0 & \text{anders} \end{cases}$	
$G_{w,j}$	De hoeveelheid post die gegenereerd worden door voorgemelde grote partijen in week w jaar j

Uit de definitie van $S_{w,j}$ blijkt dat alleen Pasen, Hemelvaart, Pinksteren, of Kerst direct invloed hebben op de onderstroom. De overige seizoeninvloeden worden middels de $O_{w,j}$ indirect in het model meegenomen.

Kerst heeft invloed op de hoeveelheid post in de weken 49 tot en met 52. Omdat in deze periode voorspellingen per dag gemaakt worden, zal het voorspellen per week beperkt blijven tot het voorspellen van week 1 tot en met 48. De invloed van Kerst is indirect echter wel van belang. Dit belang zal duidelijk gemaakt worden tijdens het beschrijven van de voorspellingsmethode.

Voorspellingsmethode

De verwachte hoeveelheid post in week w jaar (j+1) is als volgt gedefinieerd:

$$\hat{P}_{w,j+1} = \hat{O}_{w,j+1} + S_{w,j+1} + G_{w,j+1}$$

Voor het voorspellen van de hoeveelheid post in jaar (j+1) zijn de waarden van $S_{w,j+1}$ en $G_{w,j+1}$ al bekend. De verwachting van $\hat{O}_{w,j+1}$ wordt als volgt bepaald:

$$\hat{O}_{w,j+1} = \rho * O_{w,j}$$

Hierbij is

ρ	De postdichtheid
$\rho = \frac{\Delta K_q}{\Delta A * \Delta W_q}$	
ΔA	De relatieve verandering in het aantal afgiftepunten
ΔW_q	De relatieve verandering in het aantal werkdagen in kwartaal q
ΔK_q	De relatieve verandering van de hoeveelheid post in kwartaal q
$\Delta K_q = \frac{K_{q,j} - \hat{K}_{q,j+1}}{K_{q,j}}$	
$K_{q,j}$	De hoeveelheid post in kwartaal q jaar j ¹⁰

¹⁰ Indien er een ^ boven staat, dan betekent dit dat het over een verwachting i.p.v. een gerealiseerd getal gaat

De waarden van $K_{q,j}$, ΔW_q en ΔA zijn van tevoren al bekend. De verwachting $\hat{K}_{q,j+1}$ wordt als volgt bepaald:

Stap (i) Eerst wordt de onderstroom van kwartaal q jaar j als volgt vastgesteld:

$$O_{q,j} = \begin{cases} K_{q,j} - S_{q,j} - G_{q,j} & \text{als } q = 4 \\ K_{q,j} - G_{q,j} & \text{als } q = 1, 2, \text{ of } 3 \end{cases}$$

Hierbij is

$S_{q,j}$ De seizoensinvloed in kwartaal q jaar j

$G_{q,j}$ De hoeveelheid post die gegenereerd worden door voorgemelde grote partijen in kwartaal q jaar j

Stap (ii) Vervolgens wordt de verwachte onderstroom van kwartaal q jaar (j+1) bepaald

$$\text{door: } \hat{O}_{q,j+1} = \gamma * O_{q,j}$$

Hierbij is

γ De relatieve verandering van de onderstroom van kwartaal q. Deze wordt bepaald naar aanleiding van commerciële informatie en historische gegevens.

Stap (iii) De verwachte hoeveelheid post in kwartaal q jaar (j+1) is dan gelijk aan:

$$\hat{K}_{q,j+1} = \begin{cases} \hat{O}_{q,j+1} + S_{q,j+1} + G_{q,j+1} & \text{als } q = 4 \\ \hat{O}_{q,j+1} + G_{q,j+1} & \text{als } q = 1, 2, \text{ of } 3 \end{cases}$$

De invloed van Kerst heeft dus invloed op de verwachte hoeveelheid post in het 4^e kwartaal en heeft daardoor dus indirect invloed op de verwachte hoeveelheid post in de weken 37 tot en met 48. Het relatieve verschil van de postdichtheid (ρ) is door de γ indirect afhankelijk van de historische gegevens en de commerciële verwachtingen, zodat de commerciële gegevens en de historische gegevens dus ook indirect invloed hebben op de landelijke weekprognose van de hoeveelheid post.

Indien er meer kennis over grote partijen gedurende kwartaal q bekend wordt, dan kan het gebeuren dat een verwachte hoeveelheid post in week t aangepast moet worden. Deze aanpassing in week t zal omgekeerd evenredig verdeeld worden over de overige weken in kwartaal q. Met deze methode verandert de verwachte hoeveelheid post in kwartaal q dus niet.

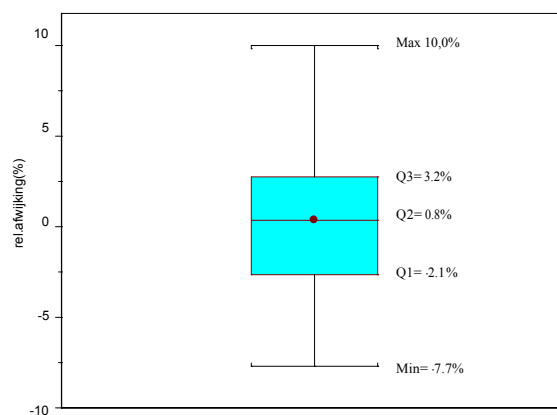
De voorlopige prognose van de wekelijkse post per vestiging, wordt verkregen middels de bepaalde postdichtheid ρ en de landelijke structuur van de weken in kwartaal q. Door de hoeveelheid post in kwartaal q jaar j van een bepaalde vestiging te vermenigvuldigen met de bepaalde postdichtheid ρ , wordt er een verwachte hoeveelheid post van kwartaal q jaar j+1 voor dezelfde vestiging bepaald. De landelijke weekstructuur bepaalt daarna hoe de verwachte hoeveelheid post van kwartaal q jaar j+1 van een vestiging verdeeld moet worden over de weken. Op deze manier worden er wekelijkse prognoses verkregen per vestiging.

2.2. Probleemdefinitie

Om meer nadruk op productiviteit en efficiëntie te leggen bij de verschillende vestigingen, is het belangrijk dat het WAP-instrument goed functioneert. Om te kijken hoe goed het instrument WAP werkt worden de volgende vragen vaak onderzocht:

- Hoe goed is de geprognosticeerde hoeveelheid 'Briefpost aan huis' per week per vestiging geweest in relatie tot de werkelijke uitkomst?
- Zou de werkelijke hoeveelheid post ook geleid hebben tot een andere werkdruk (rustig, normaal of druk)?
- Is er aansluiting tussen de theoretische berekening van de verwachte werkdruk en de beleving van de postbode?

Naar aanleiding van de eerste vraag blijkt na onderzoek dat dit jaar de landelijke voorspellingen van WAP op kwartaalbasis gemiddeld maar een procent van de werkelijke hoeveelheden post afwijken. De voorspellingen op weekbasis hebben echter grotere afwijkingen. In de boxplot in figuur 2.2 is te zien dat in de weken van 2 tot en met 48 van het jaar 2003 vijftig procent van de voorspellingen een afwijking ten opzichte van de realisaties hebben tussen de -2.1% en de 3.2%.



Figuur 2.2: Boxplot met als data de relatieve afwijkingen van de weken 2 t/m 48 in het jaar 2003

Door verschillende wiskundige modellen te ontwikkelen kan er geanalyseerd worden hoe goed de door WAP geprognosticeerde hoeveelheid 'Briefpost aan huis' is. Het doel van dit afstudeeronderzoek kan daarom beschreven worden als:

Het ontwikkelen van een model om het landelijke weekverkeer te voorspellen.

Per model zullen variabelen die invloed hebben op de resultaten van het model bepaald worden. Belangrijk is het hierbij op te merken dat het ontwikkelde model een meerwaarde voor TPG heeft, indien de bepaalde wekelijkse voorspellingen met een huidige maximale afwijking van 10% afneemt naar een maximale afwijking van 2%. Met deze maximale afwijking voldoen de voorspellingen aan de huidige TPG-wensen en wordt er verwacht dat de vestigingen met het ontwikkelde model een goed instrument hebben om hun doel 'het verbeteren van de efficiency' te behalen.

2.3. Aanpak afstudeeronderzoek

Welke modellen geschikt zijn voor het voorspellen van het wekelijkse landverkeer, is afhankelijk van:

- (i) de geschiktheid voor het maken van voorspellingen
- (ii) de beperkingen en/of de juistheid van de aannames van demodellen

Een indicatie voor de te gebruiken variabelen is verkregen middels interviews. Indien deze variabelen overeenkomst vertonen met de bij het model behorende variabelen, verwacht ik dat er een betere fitting en betere voorspellingen verkregen worden.

De variabelen die volgens de geïnterviewden invloed hebben op de wekelijkse hoeveelheid post zijn op te splitsen in tijd en niet tijd gerelateerde variabelen. De tijd gerelateerde variabelen zijn:

- maandwisseling
- kwartaaleinde
- halfjaar-einde
- zomerdaal(waaronder vakantiespreiding)
- feestdagen

Niet direct tijd gerelateerde variabelen zijn:

- opkomende concurrentie
- economische (negatieve) groei
- aantal zakelijke klanten
- aantal afgiftepunten

Naar aanleiding van deze variabelen en naar aanleiding van informatie omtrent de geschiktheid voor voorspellingen en omtrent de beperkingen, verwacht ik dat de volgende wiskundige modellen goede prognoses maken van het landelijke weekverkeer zijn:

- Tijdsreeks model: een statistisch model dat voornamelijk gericht is op het bepalen van de trend, de seizoensinvloed, andere cyclische invloeden en de overgebleven onregelmatigheden. Als de verwachtingen en covarianties van deze overgebleven onregelmatigheden bestaan en onafhankelijk zijn van het tijdstip, dan kunnen deze onregelmatigheden als stationair stochastisch proces gemodelleerd worden.
- Het log-lineair of log-quasi-likelihood model: statistische modellen waarbij de verwachting van de responsvariabele Y_i gelijk is aan het logaritme van het inproduct van de regressie waarnemingen en de geschatte regressie coëfficiënten. Bij het log-lineair model wordt er verondersteld dat de responsvariabelen Y_i Poisson verdeeld zijn. Bij het log-quasi-likelihood model wordt er geen verdeling verondersteld, maar zijn de meest aannemelijke schatters wel bepaald via de verdelingsfunctie van de Poisson-verdeling.
- Poisson model: een statistisch model waarin aangenomen wordt dat de hoeveelheid post per week Poisson verdeeld is. Met behulp van deze verdeling kunnen schattingen en voorspellingen gemaakt worden.

Bij de statistische modellen wordt er naast de analyse van de fitting en de voorspellingen, geanalyseerd of het gebruik maken van de deterministische kennis over grote partijen een verbetering oplevert in de fitting en/ of de voorspellingen. Grote partijen worden een aantal weken voordat postbodes ze bezorgen voorgemeld. Sinds 2001 wordt er gebruik gemaakt van deze informatie bij het maken van voorspellingen. De gemiddelde wekelijkse hoeveelheid poststukken die door grote partijen gegenereerd zijn in de jaren 2001 tot en met 2003 bedraagt ongeveer 16 miljoen. Aangezien de totale gemiddelde wekelijkse hoeveelheid post van de jaren 2001 tot en met 2003 89 miljoen bedraagt, zal de deterministische kennis over grote partijen een duidelijke verkleining van het stochastische gedeelte opleveren. Hierdoor wordt er een verbetering van de fitting en/of voorspellingen verwacht. Doordat er echter pas sinds 2001 data beschikbaar zijn over grote partijen, vermindert de hoeveelheid data drastisch. Dit kan een verslechtering van de fitting en/of voorspellingen veroorzaken.

Het tijdsreeks model wordt beschreven in hoofdstuk 3. Het log-lineair model wordt gezamenlijk met het log-quasi-likelihood model in hoofdstuk 4 geanalyseerd. Het laatste model, het Poisson model, wordt beschreven in hoofdstuk 5. In hoofdstuk 6 zullen de resultaten van de drie modellen vergeleken worden om vervolgens conclusies te trekken over de kwaliteit van de voorspellingen van een statistisch

model. Tevens zal er in hoofdstuk 6 geconcludeerd worden of uit de modellen blijkt dat de variabelen, die volgens de geïnterviewden invloed hebben op de hoeveelheid post, daadwerkelijk invloed hebben.

2.4. Betrokkenen tijdens het afstudeeronderzoek

Dit afstudeeronderzoek is in opdracht van en tevens onder begeleiding van de heer drs. J. Savelkoul, medewerker van Werk Aanbod Prognose, uitgevoerd. De heer M.J.S. van Dijk, tevens medewerker van Werk Aanbod Prognose, was altijd bereid om mijn vragen te beantwoorden. Namens de *Vrije* Universiteit heeft de heer dr. ir. G. Jongbloed mij frequent voorzien van feedback over mijn voortgang en uitvoering gedurende het onderzoek.

3. Tijdreeks model

3.1. Introductie

Een tijdreeks is een verzameling waarnemingen X_t die op specifieke tijdstippen t zijn waargenomen. Een tijdreeks analyse bestaat uit een statistische analyse van de afhankelijkheid tussen de waarnemingen X_t , zodat er een beter inzicht in de data wordt verkregen en er voorspellingen gemaakt kunnen worden.

In dit model zijn de waarnemingen X_t gelijk aan de hoeveelheid post in week t . Doordat de afhankelijkheden bepaald worden bij het tijdreeks model, wordt er verwacht dat het tijdreeks model inzicht geeft in de variabelen die van invloed zijn. Nadat de afhankelijkheden bepaald zijn, kunnen er voorspellingen gemaakt worden. Indien deze voorspellingen een maximale relatieve afwijking hebben van 2% t.o.v. de realisaties, dan zullen deze voorspellingen voldoen aan de huidige TPG-wensen en wordt er verwacht dat de voorspellingen een goed instrument zijn om het doel van de vestigingen 'het verbeteren van de efficiency' te behalen.

Algemene beschrijving tijdreeks model

De formele beschrijving van het tijdreeks model is beschreven in paragraaf 3.2. In het algemeen bestaat een tijdreeks model uit niet-stationaire invloeden en een (zwak) stationaire tijdreeks. De niet-stationaire invloeden zijn:

- (i) extremen,
- (ii) trend, en/of
- (iii) seizoensinvloeden

Een (zwak) stationaire tijdreeks is een tijdreeks waar de verwachting van een waarneming Y_t en de verwachting van $(Y_t Y_{t+u})$ bestaan en deze onafhankelijk zijn van het tijdstip t voor iedere u .

Door de niet-stationaire invloeden te verwijderen uit de niet-stationaire tijdreeks, kan er een (zwak) stationaire tijdreeks bepaald worden. De extremen worden hierbij helemaal uit de data verwijderd. De trend en de seizoensinvloeden kunnen pas verwijderd worden nadat de grootte van hun invloed is geschat. Als deze invloeden goed geschat zijn, blijft er een (zwak) stationaire tijdreeks over.

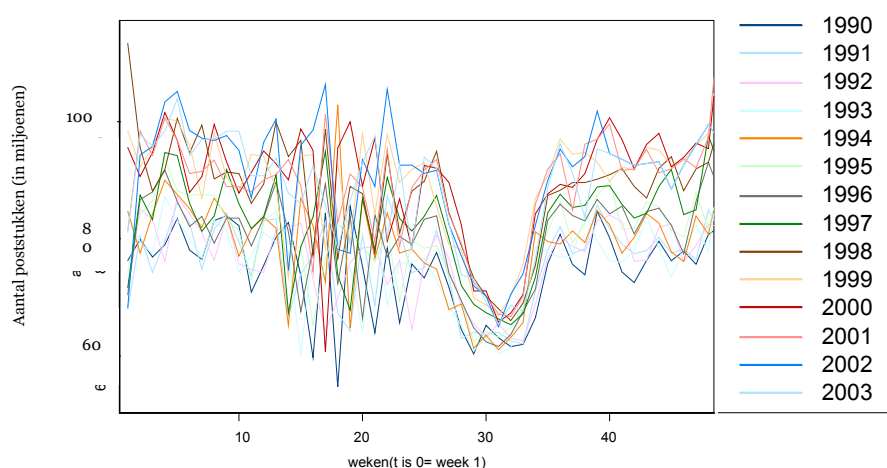
Nadat er een (zwak) stationaire tijdreeks bepaald is, wordt er indien er nog afhankelijkheid is tussen de (zwak) stationaire tijdreeks waarnemingen, een stationair tijdreeks model bepaald. Indien dit model goed bepaald is, dan wordt verwacht dat de residuen van een stationair tijdreeks model onafhankelijk en gelijk verdeeld zijn en dat het onderliggende ruisproces gelijk is aan een witte ruis. De verwachting van een witte ruis is gelijk aan nul. Uitgaande van deze verwachting van nul, kunnen er voorspellingen gemaakt worden door de waarden van het bepaalde stationaire tijdreeks model, de geschatte seizoensinvloeden en de geschatte trend bij elkaar op te tellen.

Bij sommige tijdreeksen zijn echter de trend en de seizoensinvloeden meer bepalend voor de tijdreeks dan de (zwak) stationaire tijdreeks. Dit kan veroorzaakt worden doordat het onderliggende ruisproces van de stationaire tijdreeks gelijk is aan een witte ruis. Indien verschillende toetsen aangeven dat 'de stationaire tijdreeks onafhankelijk en gelijk verdeeld is' en dus als onderliggende ruisproces een witte ruis heeft, dan heeft een modellering van de stationaire tijdreeks geen toegevoegde waarde.

Indicatie op aanwezigheid 'niet -stationaire invloeden' bij de tijdreeks die de post beschrijft

Om te bepalen of er extremen, trend en/of seizoensinvloeden verwijderd moeten worden uit de tijdreeks, dient er eerst bekeken te worden of niet-stationaire invloeden aanwezig zijn in de tijdreeks. Indien de waargenomen wekelijkse hoeveelheden post gedurende de jaren 1990 tot en met 2003 uitgezet worden tegen het tijdstip van de waarneming dan resulteert dit in de volgende figuur:

Er is sprake van seizoensinvloeden als er variaties zijn die jaarlijks terugkomen. In figuur 3.2 is een duidelijk zomerdal welke elk jaar weer terugkomt waarneembaar. Om een beter inzicht in de seizoensinvloeden te verwerven, is in figuur 3.3 de weken 1 tot en met 48 van de jaren 1990 tot en met 2003 tegen elkaar uitgezet.



Figuur 3.3: De hoeveelheid post gedurende de weken 1 tot en met 48 van de jaren 1990 tot en met 2003

Uit figuur 3.3 blijkt dat er een grote fluctuatie in de hoeveelheid post in de weken 16 tot en met 21 is. Dit wordt veroorzaakt door de verschillende feestdagen, zoals Pasen, koninginndag, hemelsvaartsdag en Pinksteren. Daarnaast is er in de zomerperiode een daling van de hoeveelheid post te zien in de weken 27 tot en met 32, waarna de hoeveelheid post langzaam uit dit dal stijgt tot week 36.

Indien de seizoensinvloed elk jaar hetzelfde is, dan levert een transformatie van de hoeveelheid post X_t geen verbeteringen in de fitting en/of voorspellingen op. De invloed van de zomervakantie lijkt echter niet constant te zijn. Na een verdere analyse blijkt dat het verschil tussen de gemiddelde hoeveelheid post in het zomerdal met het gemiddelde hoeveelheid post in de weken waar geen zomerinvloed is met de jaren toeneemt. Dit blijkt tevens uit de cijfers welke weergegeven zijn in bijlage III.B. Door de data te transformeren kan er een constantere seizoensinvloed verkregen worden. Indien de seizoensinvloed constanter is, dan zal dit een verbetering in de fitting en/of voorspellingen opleveren.

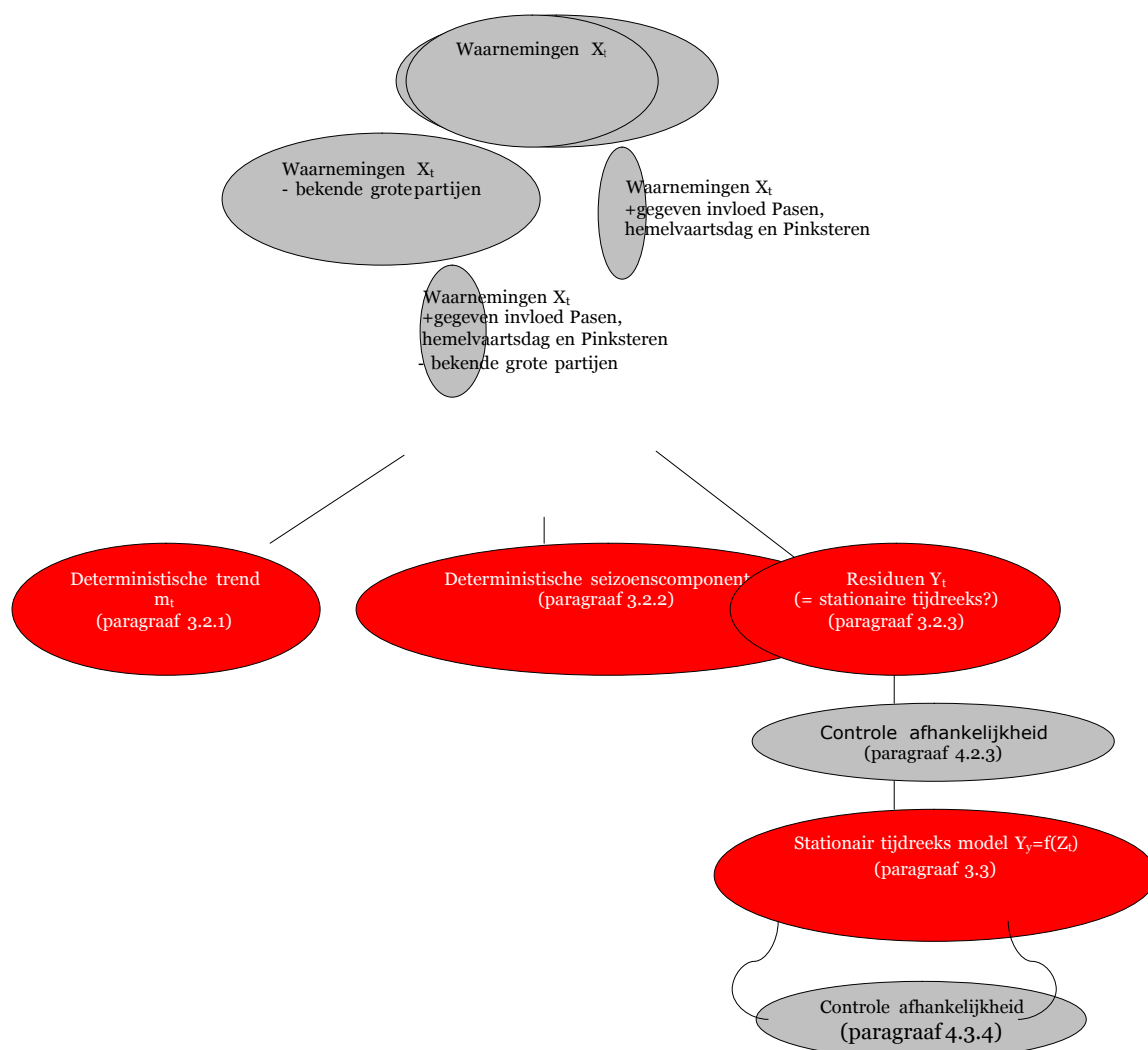
Opzet hoofdstuk

Bij de analyse van de tijdreeks wordt er, voordat de trend en de seizoensinvloeden geschat en verwijderd worden, al gebruik gemaakt van informatie omtrent gerealiseerde hoeveelheden post (grote partijen) en het door de WAP-groep verwachte verlies aan poststukken indien Pasen, Hemelvaart of Pinksteren valt in week t . Hierdoor zijn er 4 tijdreeksen die gemodelleerd worden. De formele beschrijving van deze tijdreeksen is beschreven in paragraaf 3.2.

Na de transformatie van de data en het schatten van de trend en de seizoensinvloeden per model in paragraaf 3.2.1 respectievelijk 3.2.2, zal er in paragraaf 3.2.3 gecontroleerd worden of de residuen Y_t stationair zijn en of er nog sprake is van afhankelijkheid tussen de residuen. Indien er geen afhankelijkheden zitten tussen deze residuen, dan kunnen de residuen als onafhankelijk en gelijk verdeelde variabelen met een onderliggende witte ruis proces beschouwd worden en levert verdere modellering geen toegevoegde waarde. Echter als er nog wel afhankelijkheid bestaat tussen deze residuen Y_t , dan is een complex stationair tijdreeks model nodig om deze afhankelijkheid te modelleren. Dit is voordelig voor het model, omdat afhankelijkheid vooral betekent dat de stationaire tijdreeks een bijdrage kan leveren bij het voorspellen.

In paragraaf 3.3 zal de keuze voor het stationaire tijdreeks model toegelicht worden om vervolgens na het schatten van de parameters van dit model, te toetsen of de residuen R_t van het stationair tijdreeks model een onderliggende witte ruis proces hebben. Indien dit het geval is, dan is er geen afhankelijkheid meer tussen de residuen R_t . In figuur 3.4 is een grafische overzicht gegeven van de bovenstaande stappen.

Nadat de modellen bepaald zijn, kan de fitting geanalyseerd worden. Deze analyse is beschreven in paragraaf 3.4. De voorspellingen die met het bepaalde stationaire tijdreeks model, de seizoensinvloed, de trend en/of de transformatie bepaald worden, zullen geanalyseerd worden in paragraaf 3.5. In paragraaf 3.6 vindt u de conclusies van dit tijdreeks model.



Figuur 3.4: Overzicht analyse van de tijdreeks die de post beschrijft

3.2. Beschrijving model

In de tijdreeks die de hoeveelheid post beschrijft is er sprake van een trend en van seizoensinvloeden. Omdat na verdere analyse (zie bijlage III.B) blijkt dat deze seizoensinvloeden niet constant zijn, is een transformatie van de data gewenst. Algemene transformaties zijn het nemen van de logaritme of het nemen van de wortel. Ik heb ervoor gekozen om het logaritme te nemen, omdat uit onderzoek bleek dat er een slechtere fitting was bij een tijdreeks model waarbij er voor een wortel-transformatie was gekozen. De formele beschrijving van het tijdreeks model ziet er daarom als volgt uit:

$$\text{Log}(X_t) = m_t + s_t + Y_t \quad (3.1)$$

Waarbij:

X_t	de hoeveelheid post in week t
m_t	de deterministische trend
s_t	de deterministische seizoenscomponent
Y_t	residuen (waarvan verwacht wordt dat het een stationair tijdreeks is)

Grote partijen

Van de hoeveelheid post per week is vanaf het jaar 2001 bekend welk gedeelte gegenereerd is door voorgemelde grote partijen. Doordat de hoeveelheid post welke gegenereerd wordt door grote partijen bij het maken van voorspellingen van tevoren bekend is, levert het aftrekken van dit gedeelte het voordeel op dat er een kleinere hoeveelheid post per week voorspeld moet worden. Op dit moment neemt de hoeveelheid data echter sterk af als er gebruik gemaakt wordt van voorgemelde grote partijen. Indien er voor bijvoorbeeld 2003 voorspellingen gemaakt worden van de hoeveelheid post zonder de hoeveelheid post die door voorgemelde grote partijen gegenereerd wordt, dan is er alleen data beschikbaar van de jaren 2001 en 2002. Er is data beschikbaar van de jaren 1990 tot en met 2002 als de hoeveelheid post voorspeld wordt voor het jaar 2003 zonder gebruik te maken van beschikbare informatie omtrent grote partijen. De analyse of het voordeel opweegt tegen de vermindering van de beschikbare data, wordt middels onderstaand model uitgevoerd:

$$\text{Log}(X_t - G_t) = m_t + s_t + Y_t \quad (3.2)$$

Hierbij is

G_t de hoeveelheid post die door voorgemelde grote partijen wordt gegenereerd.

Seizoenscomponent

Volgens de WAP-groep vermindert de hoeveelheid post in de weken waarin Pasen, hemelvaartsdag en Pinksteren valt met respectievelijk 14, 16 en 14 miljoen ten opzichte van de verwachte hoeveelheid in die weken waarin geen feestdag valt. Door in de weken waarin Pasen, hemelvaartsdag en Pinksteren vallen deze hoeveelheden er bij op te tellen, wordt er verwacht dat de invloed van Pasen, hemelvaartsdag en Pinksteren verwijderd is. Het verwijderen van deze invloeden zorgt ervoor dat er een andere deterministische seizoenscomponent bepaald moet worden, en leidt tot een 3^e en een 4^e model, namelijk:

$$\text{Log}(X_t + 14\text{milj} * 1_{\text{Pasen, Pinksteren} \in t} + 16\text{milj} * 1_{\text{hemelvaartsdag} \in t}) = m_t + s_t + Y_t \quad (3.3)$$

$$\text{Log}(X_t - G_t + 14\text{milj} * 1_{\text{Pasen, Pinksteren} \in t} + 16\text{milj} * 1_{\text{hemelvaartsdag} \in t}) = m_t + s_t + Y_t \quad (3.4)$$

Milj is hier een afkorting voor miljoen.

In paragraaf 3.2.1 zal de trend bepaald worden voor de 4 modellen. Vervolgens zullen in paragraaf 3.2.2 de verschillende seizoensinvloeden bepaald en geanalyseerd worden. Tijdens de analyse van de seizoensinvloeden, worden onder andere de effecten van de door de WAP-groep verwachte invloed van Pasen, Hemelvaart en Pinksteren vergeleken met de effecten van de geschatte invloeden voor deze feestdagen van model 3.1 en 3.2.

Nadat de deterministische trend en de deterministische seizoenscomponenten bepaald zijn, kan er in paragraaf 3.2.3 van de residuen Y_t bepaald worden of er sprake is van afhankelijkheid tussen de residuen. Indien dit niet het geval is, dan zal in paragraaf 3.3 deze afhankelijkheid bepaald worden middels een stationair tijdreeks model.

3.2.1. Trend

Bij het bepalen van de trend, oftewel de lange-termijn verandering in het gemiddelde niveau, zijn de 'gaten' in de tijd-as, die veroorzaakt zijn door het verwijderen van de extremen, even groot gemaakt door bij alle jaren de weken 49 tot en met 52 niet mee te nemen in het model.

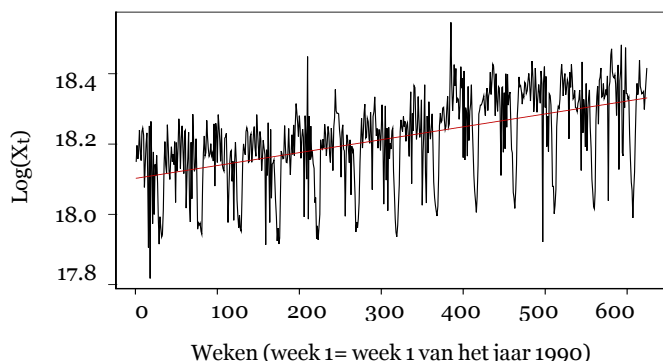
De trend is bepaald met behulp van de kleinste kwadraten schatter waarbij in dit geval een polynoom bepaald wordt. Hierbij wordt de deterministische trend m_t zo geschat, zodat het totale kwadratische verschil tussen de geschatte polynoom-waarden en de logaritme van de waargenomen hoeveelheid post X_t (in- of exclusief de door grote partijen gegenereerde hoeveelheid post, in- of exclusief de verwachte invloed van Pasen, hemelvaartsdag en Pinksteren) geminimaliseerd is. Bij elk model is de trend een lineaire functie, omdat een polynoom van een 2^e of hogere orde geen betere kleinste kwadraten schatter oplevert. Met behulp van deze lineaire functie kan de invloed van een algemene stijging of daling van de hoeveelheid post in de jaren gefilterd worden. Zodoende wordt er een tijdreeks overgehouden die dezelfde afwijking heeft van de trendlijn, waardoor er bijvoorbeeld geanalyseerd kan worden of de zomerdalen een constante seizoensinvloed zijn.

Verwachting n.a.v. geschatte trend m_t

Door de trends van de verschillende modellen te vergelijken, kan er bijvoorbeeld gekeken worden of er een indicatie is dat de hoeveelheid post die verminderd is met de bekende grote partijen en haar trend een tijdreeks als resultante heeft die constantere afwijkingen heeft. Dit is de verwachting bij model 3.2 en 3.4, omdat eventuele afwijkingen niet veroorzaakt kunnen zijn door grote partijen, maar veroorzaakt worden door seizoensinvloeden of de stationaire tijdreeks Y_t .

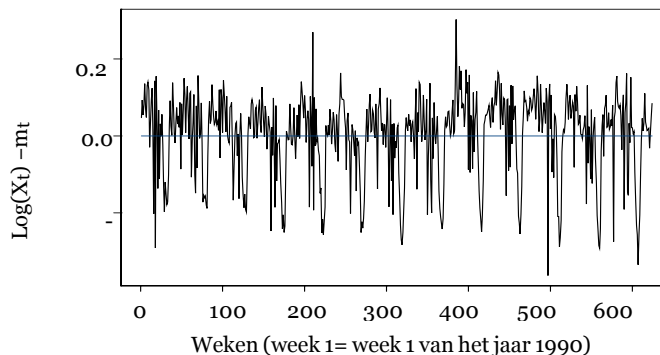
Door model 3.1 respectievelijk 3.2 te vergelijken met model 3.3 respectievelijk 3.4, is een analyse mogelijk naar de invloed van de gegeven waarden van de invloed van Pasen, hemelvaartsdag en Pinksteren. Indien er minder fluctuatie is bij model 3.3 en 3.4 in de weken rond Pasen, hemelvaartsdag en Pinksteren, dan kan er geconcludeerd worden dat de fluctuatie veroorzaakt wordt door deze feestdagen en dat de door de WAP-groep gegeven waarden voor de invloed van Pasen, hemelvaartsdag en Pinksteren een indicatie zijn voor de te schatten invloed van deze feestdagen. In de onderstaande linkerfiguren zijn de trends van de modellen weergegeven. In de rechterfiguren zijn de data zonder trend van de 4 modellen weergegeven.

Model 3.1



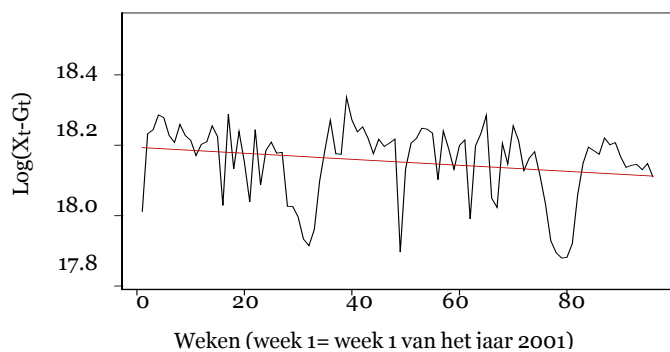
Figuur 3.5: De logaritme van "de hoeveelheid post" en de bepaalde trend gedurende de jaren 1990 tot en met 2002, waarbij de extreme hoeveelheden post uit de data verwijderd zijn. (model 3.1: trend $m_t = 18.102 + 0.00037 \cdot t$)

Model 3.1



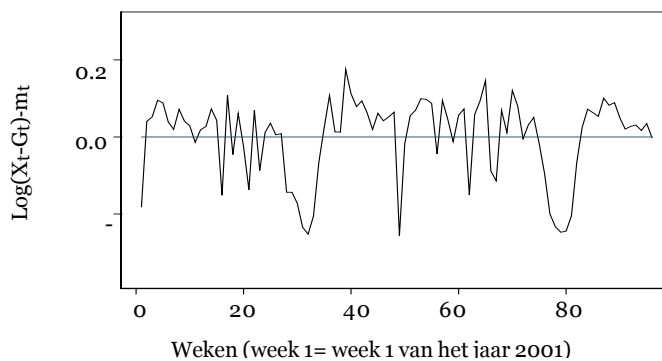
Figuur 3.6: De logaritme van "de hoeveelheid post" min de bepaalde trend gedurende de jaren 1990 tot en met 2002, waarbij de extreme hoeveelheden post uit de data verwijderd zijn. (model 3.1)

Model 3.2



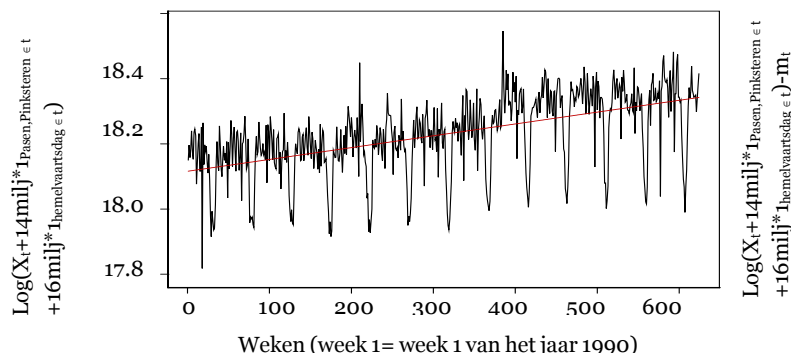
Figuur 3.7: De logaritme van “de hoeveelheid post min de grote partijen” en de bepaalde trend gedurende de jaren 2001 en 2002, waarbij de extreme hoeveelheden post uit de data verwijderd zijn. (model 3.2: trend $m_t = 18.194 - 0.00085 \cdot t$)

Model 3.2



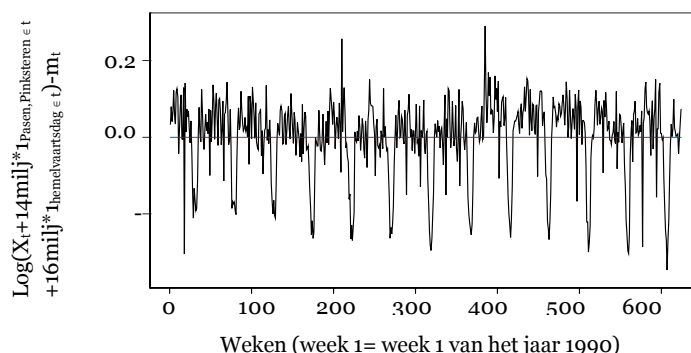
Figuur 3.8: De logaritme van “de hoeveelheid post min de grote partijen” min de bepaalde trend gedurende de jaren 2001 en 2002, waarbij de extreme hoeveelheden post uit de data verwijderd zijn. (model 3.2)

Model 3.3



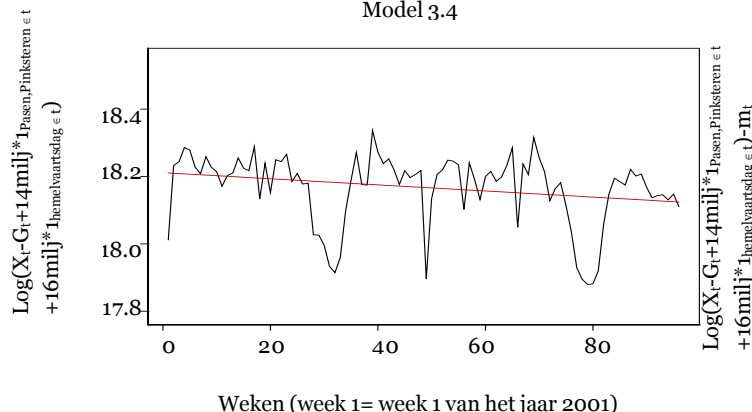
Figuur 3.9: De logaritme van de “hoeveelheid post min de invloed van Pasen, Hemelv. En Pinksteren” en de bepaalde trend gedurende de jaren 1990 t/m 2002, waarbij de extreme hoeveelheden post uit de data verwijderd zijn. (model 3.3: trend $m_t = 18.116 + 0.00036 \cdot t$)

Model 3.3



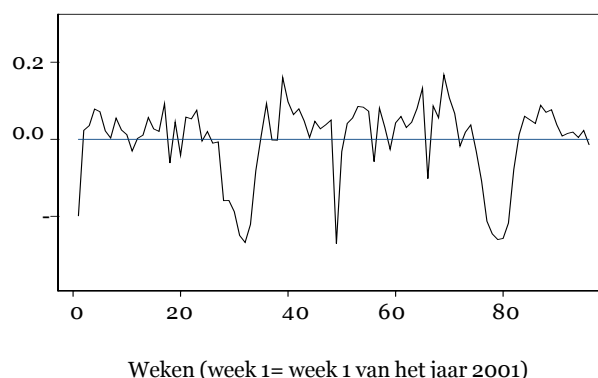
Figuur 3.10: De logaritme van de “hoeveelheid post min de invloed van Pasen, Hemelv. En Pinksteren”, min de bepaalde trend gedurende de jaren 1990 t/m 2002, waarbij de extreme hoeveelheden post uit de data verwijderd zijn. (model 3.3)

Model 3.4



Figuur 3.11: De logaritme van de “hoeveelheid post min de gr. partijen min de invloed van Pasen, Hemelv. en Pinksteren” en de bepaalde trend gedurende de jaren 1990 t/m 2002, waarbij de extreme hoeveelheden post uit de data verwijderd zijn. (model 3.4: trend $m_t = 18.211 - 0.00090 \cdot t$)

Model 3.4



Figuur 3.12: De logaritme van de “hoeveelheid post min de gr. partijen min de invloed van Pasen, Hemelv. en Pinksteren”, min de bepaalde trend gedurende de jaren 1990 t/m 2002, waarbij de extreme hoeveelheden post uit de data verwijderd zijn. (model 3.4)

Resultaten n.a.v. geschatte trend m_t

De bepaalde trends m_t zijn beschreven in de bijschriften van bovenstaande linkerfiguren.

Uit de figuren 3.9 en 3.11 blijkt dat de trend van de hoeveelheid post verminderd met de grote partijen van de jaren 2001 en 2002 een dalende lijn is. Hieruit kan geconcludeerd worden dat de hoeveelheid post zonder grote partijen aan het afnemen is. Echter of deze trend zich doorzet naar de volgende jaren, kan met de kleine hoeveelheid beschikbare waarnemingen van de hoeveelheid post die door grote partijen gegenereerd is (alleen de jaren 2001 en 2002) niet met 100% zekerheid gezegd worden. De stijging in figuur 3.8 en 3.10 geeft daarentegen de indicatie dat de hoeveelheid post stijgt, echter deze figuren laten ook zien dat de laatste jaren het aantal waarnemingen boven de trend-lijn een kleinere afstand hebben tot de trendlijn dan de jaren daarvoor. Dit geeft de indicatie dat de hoeveelheid post gaat dalen.

De verwachting dat de hoeveelheid post verminderd met de grote partijen een constantere seizoensinvloed bewerkstelligt, kan onder de vermelding dat er weinig beschikbare data is van grote partijen, bevestigd worden. In figuur 3.8 en figuur 3.12 hebben de drie minima van de drie duidelijke dalen namelijk ongeveer dezelfde waarde, terwijl de 3 minima van de laatste drie dalen in figuur 3.6 en figuur 3.10 niet dezelfde waarde hebben.

Verder kan er na vergelijking van model 3.2 en 3.4 geconcludeerd worden dat de fluctuatie rond de weken 14 tot en met 25 in de jaren 2001 en 2002 bij model 3.4 is afgenomen. Ditzelfde kan geconcludeerd worden bij vergelijking van model 3.1 en 3.3. Hierbij is er gekeken naar het verschil in het aantal waarnemingen die onder de trendlijn zitten en een relatief grote afstand hebben tot de trendlijn. Dit aantal is in model 3.3 kleiner dan het aantal in model 3.1. Doordat het enige verschil tussen deze modellen is, dat het model 3.3 inclusief de gegeven waarden is voor de invloed van Pasen, hemelvaartsdag en Pinksteren., en model 3.1 exclusief deze gegeven waarden, kan de vermindering van het aantal bij model 3.3 verklaard worden door de toevoeging van de gegeven waarden. De door de WAP-groep gegeven waarden lijken dus een goede indicatie te geven voor de waarde van de invloed van de feestdagen Pasen, Hemelvaart, en Pinksteren. Het wel/niet gebruik maken van de gegeven waarden is tevens een verklaring voor het feit dat de constanten van de trend-functie's m_t van de modellen 3.2 en 3.4 groter zijn dan bij de modellen 3.1 en 3.3.

Nadat de trendfunctie's m_t bepaald en verwijderd zijn, kunnen de seizoensinvloeden per model geschat worden.

3.2.2. Seizoensinvloed

Seizoensinvloeden zijn variaties die elk jaar terugkomen. De afhankelijkheid bij deze invloeden is dan groter tussen de variatie in jaar j en de variatie in jaar $j+1$, dan tussen de variatie in week t en de variatie in week $t+1$. Uit figuur 3.3 en uit de figuren 3.6, 3.8, 3.10 en 3.12 waar de trend gefilterd is, blijkt dat die jaarlijkse variaties niet alleen aanwezig zijn in de zomerdal en bij de officiële feestdagen, maar ook in de verschillende weken door bijvoorbeeld een maandwisseling. Onderstaand zal eerst relevante informatie omtrent de feestdagen koninginndag en bevrijdingsdag worden gegeven om vervolgens methode om de seizoensinvloeden er uit te filteren toe te lichten. De geschatte seizoensinvloeden zullen na een toelichting van de methode geanalyseerd worden.

Koninginndag

Koninginndag is in de jaren 1990 tot en met 2002 11 keer in week 18 gevallen, waarbij in het jaar 1990 bevrijdingsdag ook in week 18 viel. De overige 2 koninginndagen vielen op een zondag in de jaren 1995 en 2000 en omdat zondag al een vrije dag is, wordt er voor die jaren geen invloed verwacht van koninginndag¹¹. Door aan te nemen dat er verwacht wordt dat koninginndag in week 18 valt, is het wel van belang om te analyseren of er nu een afwijking is in de jaren 1995 en 2000 voor een eventuele voorspelling van de hoeveelheid post in een week 18 waarin koninginndag niet/ op een zondag valt. Omdat in 1995 bevrijdingsdag wel in week 18 valt, is het belangrijk om te weten wat de invloed is van bevrijdingsdag.

Bevrijdingsdag

Bevrijdingsdag is in Nederland maar 1 keer in de 5 jaar een officiële vrije dag. In de periode van 1990 tot en met 2002 is bevrijdingsdag in de jaren 1990, 1995 en 2000 een officiële vrije dag geweest. Hierbij viel bevrijdingsdag respectievelijk in week 18, week 18 en week 19. Aangenomen dat voor de bepaling van de seizoensinvloed van week 18, die datapunten (zonder trend) zijn meegenomen waarin

¹¹ Deze verwachting komt mede tot stand doordat en de wekelijkse posthoeveelheid uit ongeveer 90% zakelijke post bestaat en vooral de zakelijke post beïnvloed wordt door een officiële werkdag gedurende de werkweek. Alleen in de weken rond kerst en nieuw jaar heeft de zakelijke post een lager aandeel van de totale hoeveelheid post doordat er dan meer post is van de particulieren.

enkel en alleen koninginndag viel op een maandag tot en met zaterdag en zonder dat bevrijdingsdag ook nog viel in week 18, zal

- indien de data zonder trend en zonder seizoensinvloed van week 18 in 1990 significant hoger is dan in de overige weken dit verklaard kunnen worden door de invloed van bevrijdingsdag.
- bij week 18 in het jaar 1995 een positief dan wel negatief datapunt zonder trend zonder seizoensinvloed duiden op een kleinere dan wel grotere invloed dan koninginndag.

In week 19 in het jaar 2000 is er, onder de aanname dat bij het bepalen van de invloed van week 19 deze week niet is meegenomen omdat hier ook de invloed van bevrijdingsdag aanwezig is, een indicatie voor de invloed van bevrijdingsdag door te kijken naar het datapunt week 19 jaar 2000 zonder trend zonder seizoensinvloed. Bij het bepalen van de invloed van bevrijdingsdag zijn er helaas weinig datapunten om informatie over bevrijdingsdag in te winnen. Indien bovenstaande datapunten andere invloeden indiceren, dan kan er geen invloed van bevrijdingsdag geschat worden. Indien de bovenstaande datapunten wel dezelfde invloed voor bevrijdingsdag indiceren, dan kan deze vergeleken worden met de indicaties voor de geschatte invloed van bevrijdingsdag van het log-quasi-likelihood model (hoofdstuk 4), en de geschatte invloed van de poisson modellen(hoofdstuk 5) om daarna een conclusie te trekken.

Schattingmethode modellen 3.3 en 3.4

Doordat er bij de modellen 3.3 en 3.4 verwacht wordt dat de invloed van Pasen, hemelvaartsdag en Pinksteren gefilterd is, blijft er voor de officiële feestdagen qua filtering alleen koninginndag en bevrijdingsdag over. Bevrijdingdag wordt, zoals beschreven is in bovenstaande alinea, in eerste instantie niet meegenomen in het model. Doordat er verwacht wordt dat koninginndag en het zomerdal elk jaar in dezelfde weken vallen, zullen deze seizoensinvloeden via de seizoensinvloeden van de verschillende weken bepaald worden. De seizoensinvloed per week voor model 3.3 respectievelijk 3.4 wordt bepaald door eerst voor elke week de hulpvariabele V_k te definiëren als het gemiddelde over jaar j van respectievelijk:

$$\begin{aligned} & (\text{Log}(X_{k+jd} + 14\text{milj} * 1_{\text{Pasen, Pinksteren} \in k+jd} + 16\text{milj} * 1_{\text{hemelvaartsdag} \in k+jd}) - m_{k+jd}) \\ & (\text{Log}(X_{k+jd} - G_{k+jd} + 14\text{milj} * 1_{\text{Pasen, Pinksteren} \in k+jd} + 16\text{milj} * 1_{\text{hemelvaartsdag} \in k+jd}) - m_{k+jd}) \end{aligned}$$

Hierbij is

$k+jd$	gelijk aan t ($k+jd=t=1, \dots, n, n+1, \dots, m$).
k	gelijk aan het weeknummer ($k=1, \dots, 48$).
d	het aantal gebruikte weken per jaar ($d=48^{12}$).
j	gelijk aan het jaartal min 1990 ($j=0, \dots, 13^{13}$)
n	aantal weken van week 1 1990 tot en met week 48 2002 respectievelijk aantal weken van week 1 2001 tot en met week 48 2002 ($n=624$ respectievelijk 96)
m	gelijk aan $n+$ het aantal weken waarvoor er voorspellingen gemaakt worden
milj	een afkorting voor miljoen

De V_k bepaalt dus per week de gemiddelde afwijking van de trendlijn (zie figuur 3.10 respectievelijk figuur 3.12). Doordat de trend is bepaald met de kleinste kwadraten schatter, is de totale kwadratische afwijking van de trendlijn geminimaliseerd. Omdat de V_k 's niet vanzelfsprekend sommeren tot precies nul, zal de seizoenscomponent s_k voor de modellen 3.3 en 3.4 geschat worden door het gemiddelde van alle V_k 's van V_k af te trekken.

In formulevorm ziet dit er al volgt uit:

$$\hat{s}_k = V_k - d^{-1} \sum_{i=1}^d V_i$$

Pasen, hemelvaartsdag en Pinksteren & schattingmethode modellen 3.1 en 3.2

Bij de modellen 3.1 en 3.2 zal naast de invloed van koninginndag en bevrijdingsdag, de invloed van Pasen, hemelvaartsdag en Pinksteren bepaald moeten worden. Door aan te nemen dat de invloed van

¹² Bij het bepalen van de trend, oftewel de lange-termijn verandering in het gemiddelde niveau, zijn de 'gaten' in de tijd-as, die veroorzaakt zijn door het verwijderen van de extremen, even groot gemaakt door bij alle jaren de weken 49 tot en met 52 niet mee te nemen in het model.

¹³ $j=0, \dots, 12$ indien 2003 niet meegenomen wordt in het model. Dit is het geval bij het bepalen van de trend en de seizoensinvloeden. Tevens is dit het geval bij het analyseren van de fitting.

deze feestdagen onafhankelijk is van de week, kan de invloed van deze feestdagen als volgt bepaald worden:

Stap (i)	Eerst de weekinvloed bepalen op dezelfde manier als voor model 3.3 en 3.4, alleen hierbij de datapunten van de weken waarin Pasen, hemelvaartsdag of Pinksteren valt verwijderen uit de gegeven data. Deze weekinvloed vervolgens definiëren als \hat{w}_k .
Stap (ii)	Daarna de $V_{\text{Pasen, hemelvaartsdag en Pinksteren}}$ voor model 3.1 respectievelijk model 3.2 te definiëren als het gemiddelde van: $\begin{aligned} & (\text{Log}(X_{\text{Pasen, hemelvaartsdag, of Pinksteren} \in k+jd}) - m_{\text{Pasen, hemelvaartsdag, of Pinksteren} \in k+jd} \\ & - \hat{w}_k * 1_{\text{Pasen, hemelvaartsdag, of Pinksteren} \in k+jd}) \\ & \text{Log}(X_{\text{Pasen, hemelvaartsdag, of Pinksteren} \in k+jd} - G_{\text{Pasen, hemelvaartsdag, of Pinksteren} \in k+jd}) \\ & - m_{\text{Pasen, hemelvaartsdag, of Pinksteren} \in k+jd} - \hat{w}_k * 1_{\text{Pasen, hemelvaartsdag, of Pinksteren} \in k+jd}); \end{aligned}$
Stap (iii)	Omdat de invloed van Pasen, hemelvaartsdag en Pinksteren elk jaar uit de data zijn gehaald, kan de seizoenscomponent s_k als volgt geschat worden: $\begin{cases} \hat{s}_k = V_k - d^{-1} \left(\sum_{i=1}^d V_i + V_{\text{Pasen}} + V_{\text{hemelvaartsdag}} + V_{\text{Pinksteren}} \right) \text{ indien Pasen, Hemelvaart of Pinksteren} \notin (k + jd) \\ \hat{s}_k = V_k + V_{\text{Pasen}} - d^{-1} \left(\sum_{i=1}^d V_i + V_{\text{Pasen}} + V_{\text{hemelvaartsdag}} + V_{\text{Pinksteren}} \right) \text{ indien Pasen} \in (k + jd) \\ \hat{s}_k = V_k + V_{\text{hemelvaartsdag}} - d^{-1} \left(\sum_{i=1}^d V_i + V_{\text{Pasen}} + V_{\text{hemelvaartsdag}} + V_{\text{Pinksteren}} \right) \text{ indien hemelvaartsdag} \in (k + jd) \\ \hat{s}_k = V_k + V_{\text{Pinksteren}} - d^{-1} \left(\sum_{i=1}^d V_i + V_{\text{Pasen}} + V_{\text{hemelvaartsdag}} + V_{\text{Pinksteren}} \right) \text{ indien Pinksteren} \in (k + jd) \end{cases}$

Verwachting n.a.v. geschatte seizoen component s_k

Nadat met de trend de afhankelijkheid die de waarnemingen opeenvolgend hebben is gefilterd, verwacht ik dat er nog afhankelijkheden zijn die veroorzaakt zijn door jaarlijkse variaties en eventuele afhankelijkheden die met een stationair tijdreeks model gefilterd kunnen worden. Middels het bepalen van de seizoensinvloeden wordt er getracht de afhankelijkheden, die veroorzaakt zijn door jaarlijkse variaties, te verwijderen om een stationaire tijdreeks over te houden.

De eigenschappen van een stationaire tijdreeks Y_t is dat de verwachting van Y_t en van $(Y_t Y_{t+k})$ bestaan en onafhankelijk zijn van t . De verwachting van de figuren waarin ook nog de seizoensinvloeden zijn verwijderd, is een figuur waar geen structuur meer zichtbaar is. Indien er namelijk nog wel een structuur waarneembaar is, dan zal de verwachting van de residuen Y_t niet onafhankelijk zijn van t en zal de residuen Y_t dus geen stationaire tijdreeks zijn.

Van de bepaalde seizoensinvloeden verwacht ik (omdat er meer data beschikbaar is) dat model 3.1 en 3.3 een seizoensinvloed per week bepaalt die uiteindelijk betere voorspellingen opleveren. Om dezelfde reden wordt er verwacht dat de schatters die model 3.1 bepaalt voor de invloed van Pasen, hemelvaartsdag en Pinksteren, betere voorspellingen opleveren dan de door model 3.2 geschatte invloeden van Pasen, hemelvaartsdag en Pinksteren. Niettemin kan het zijn dat model 3.2 en model 3.4 de seizoensinvloed wel goed schatten, omdat de hoeveelheid post die door grote partijen gegenereerd verwijderd is en dus geen schattingsfout kan veroorzaken. De effecten van de geschatte seizoensinvloeden zijn ook waarneembaar bij de voorspellingen voor het jaar 2003. De analyse van de voorspellingen is beschreven in paragraaf 3.5. De effecten van de geschatte seizoensinvloeden op de fitting wordt middels de analyse van de Y_y beschreven in deze paragraaf.

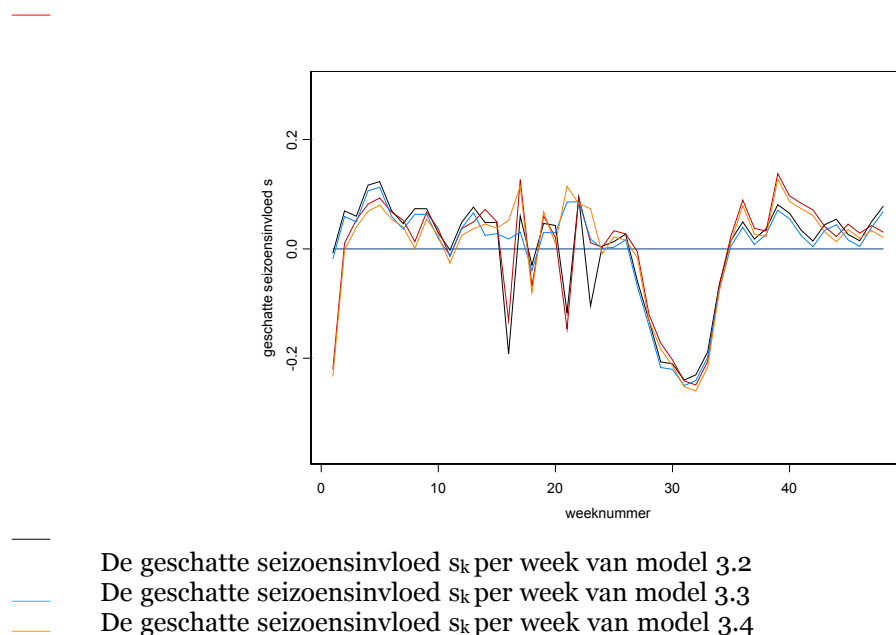
Van de fitting, welke bepaald kan worden door o.a. de residuen Y_t te analyseren, verwacht ik dat deze beter is bij de modellen 3.2 en 3.4 omdat daar met dezelfde hoeveelheid parameters (per week een V_k en een s_k) minder datapunten gefit hoeven te worden. Indien deze verwachting klopt, dan zal bij de fitting van de modellen 3.2 respectievelijk 3.4 een kleinere gemiddelde absolute afwijking bij de residuen Y_t zichtbaar zijn.

Resultaten schatting seizoenscomponent s_k

Bij de onderstaande figuren wordt naar aanleiding van de residuen Y_t een indicatie gegeven hoe goed de fitting is van de verschillende modellen. Door de residuen Y_t te analyseren voor de weken 18 van de jaren 1990, 1995 en 2000 en week 19 van het jaar 2000, kan er eventueel een indicatie voor de invloed van koninginnedag en bevrijdingsdag gegeven worden.

De geschatte seizoensinvloed per week voor de verschillende modellen is in figuur 3.13 weergegeven:

De geschatte seizoensinvloed s_k per week van model 3.1



Figuur 3.13: De geschatte seizoensinvloed per week per model

Deze seizoensinvloeden zullen gebruikt worden bij het bepalen van de fitting en de voorspellingen. Uit figuur 3.13 blijkt dat de seizoensinvloeden van model 3.1 en 3.3 redelijk synchroon lopen, behalve in de weken 14 tot en met 24. Ditzelfde geldt voor de vergelijking van de seizoensinvloeden 3.2 en 3.4. Dat de modellen niet synchroon lopen in de weken 14 tot en met 24, heeft te maken met het schatten van de invloed van de feestdagen Pasen, hemelvaartsdag en Pinksteren. In die weken lopen model 3.1 en 3.2 redelijk synchroon, omdat zij allebei bij het schatten van de s_k in eerste instantie zo goed als geen rekening houden met de invloed van de feestdagen Pasen, hemelvaartsdag en Pinksteren. De fluctuaties in de weken 14 tot en met 24 zijn dan ook groter bij de modellen 3.1 en 3.2 dan bij de modellen 3.3. en 3.4.

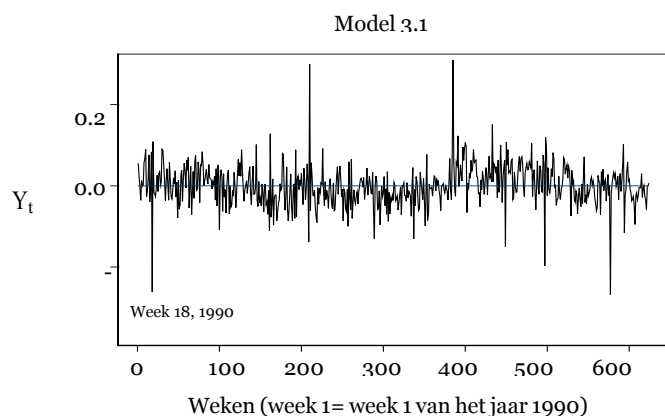
De invloed van de zomerdal is door de 4 modellen op dezelfde waarde geschat. Afwijkingen tussen het model 3.1 respectievelijk 3.3 en 3.2 respectievelijk 3.4 kan een indicatie zijn dat de seizoensinvloeden veranderen de laatste 2 jaren. Echter de afwijking kan ook veroorzaakt worden door het verschil in beschikbare data of doordat de data verminderd is met de grote partijen. Voorspellingen zullen uitwijzen of bijvoorbeeld week 1 tegenwoordig een relatief negatievere seizoensinvloed heeft. Andere duidelijke afwijkingen zijn zichtbaar in de weken 36 en 39. Hierbij zullen de voorspellingen uitwijzen of deze weken tegenwoordig een relatief hoger seizoensinvloed hebben.

Bij model 3.1 en 3.2 zijn de onderstaande geschatte waarden voor de invloed van Pasen, Hemelvaart en Pinksteren bepaald.

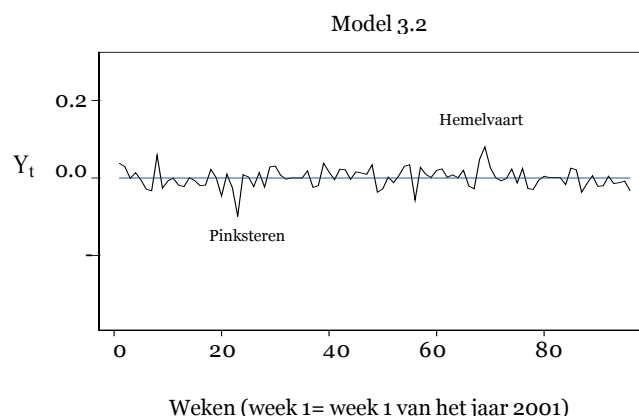
	model 3.1	model 3.2
V_{pasen}	-0.225	-0.224
$V_{\text{hemelvaartsdag}}$	-0.178	-0.146
$V_{\text{Pinksteren}}$	-0.114	-0.068

Tabel 3.1: De geschatte invloeden van Pasen, hemelvaartsdag en Pinksteren

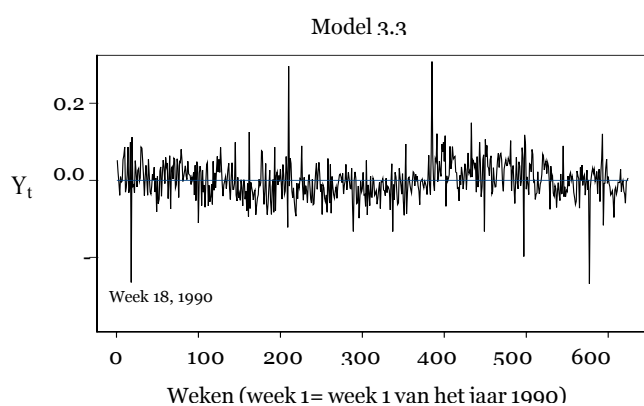
Uit tabel 3.1 blijkt dat er weinig verschil is tussen de V_{pasen} van model 3.1 en 3.2. Het verschil bij $V_{\text{hemelvaartsdag}}$ en $V_{\text{Pinksteren}}$ kan veroorzaakt worden doordat bij model 3.2 deze waarden maar met 1 datapunt bepaald worden. Hierdoor wordt er verwacht dat de $V_{\text{hemelvaartsdag}}$ en $V_{\text{Pinksteren}}$ van model 3.1 betere indicaties zijn voor de invloed van hemelvaartsdag en Pinksteren. Dit kan bevestigd worden door een analyse van de residuen Y_t van beide modellen.



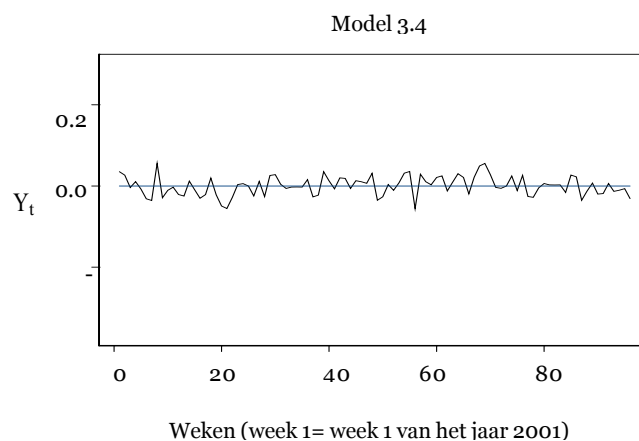
Figuur 3.14: De residuen Y_t van model 3.1, die bepaald is door de weergegeven waarnemingen in figuur 3.6 te verminderen met de voor model 3.1 geschatte seizoensinvloeden (inclusief de geschatte invloeden voor Pasen, hemelvaartsdag en Pinksteren)



Figuur 3.15: De residuen Y_t van model 3.2, die bepaald is door de weergegeven waarnemingen in figuur 3.8 te verminderen met de voor model 3.2 geschatte seizoensinvloeden (inclusief de geschatte invloeden voor Pasen, hemelvaartsdag en Pinksteren)



Figuur 3.16: De residuen Y_t van model 3.3, die bepaald is door de weergegeven waarnemingen in figuur 3.10 te verminderen met de voor model 3.3 geschatte seizoensinvloeden (inclusief de gegeven invloeden voor Pasen, hemelvaartsdag en Pinksteren)



Figuur 3.17: De residuen Y_t van model 3.4, die bepaald is door de weergegeven waarnemingen in figuur 3.12 te verminderen met de voor model 3.4 geschatte seizoensinvloeden (inclusief de gegeven invloeden voor Pasen, hemelvaartsdag en Pinksteren)

Algemeen kan er geconcludeerd worden dat de residuen Y_t stationaire tijdreeksen zijn, omdat uit bovenstaande figuren en uit paragraaf 3.2.3 blijkt dat de verwachting van $Y_t = E(Y_t)$ en de covariantie van $Y_t = E(Y_t, Y_{t+u})$ bestaan en onafhankelijk zijn van t voor iedere u .

Tevens kan er geconcludeerd worden dat de afwijking t.o.v. de nullijn is afgenomen door rekening te houden met de seizoensinvloeden per week. De kleinste maximale absolute afwijking wordt behaald

met model 3.4. Hieruit kan geconcludeerd worden dat de gegeven waarden voor Pasen, hemelvaartsdag, en Pinksteren beter zijn dan de voor model 3.2 geschatte waarden voor deze feestdagen. De extremen in figuur 3.15 zitten namelijk bij Pinksteren en hemelvaartsdag. Dit wordt waarschijnlijk veroorzaakt doordat in week 21 in het jaar 2001 hemelvaartsdag viel, en in week 21 in het jaar 2002 Pinksteren. Dit zorgt voor een minder goede schatting van de invloed van hemelvaartsdag en Pinksteren bij model 3.2. In model 3.1 zijn er meer waarnemingen beschikbaar en is er daarom ook een betere schatting van de feestdagen mogelijk. De figuren 3.14 en 3.16 lijken erg veel op elkaar, waardoor er geconcludeerd kan worden dat de voor model 3.1 geschatte invloed voor Pasen, hemelvaartsdag en Pinksteren lijkt op de door model 3.3 gebruikte gegeven waarden voor deze feestdagen. De extremen die in figuur 3.14 en 3.16 nog zichtbaar zijn, worden niet veroorzaakt door de invloed van Pasen, hemelvaartsdag en Pinksteren. Dit kan geconcludeerd worden omdat de weken waarin een extreem voorkomt, niet overeenkomt met een week waarin Pasen, hemelvaartsdag of Pasen valt. Het eerste extreem in de figuur 3.14 en 3.16 is echter wel veroorzaakt door de invloed van feestdagen, namelijk de invloed van koninginndag en bevrijdingsdag. Om een indicatie te geven van de invloed van koninginndag en bevrijdingdag, worden de volgende data geanalyseerd:

	Bevrijdingsdag	Koninginndag	Geschatte residu Y_t	
			Model 3.1	Model 3.3
Week 18,1990	Ja	Ja	-0.26	-0.27
Week 18,1995	Ja	Nee	0.059	0.057
Week 18,2000	Nee	Nee	0.12	0.12
Week 19,2000	Ja	Nee	0.10	0.094

Tabel 3.2: Analyse koninginndag en bevrijdingsdag

Aangezien er aangenomen was dat koninginndag in week 18 valt, geeft de waarde -0.26 respectievelijk -0.27 de indicatie dat dit de geschatte seizoensinvloed is van bevrijdingsdag. Als deze waarde vergeleken wordt met de geschatte seizoensinvloeden die weergegeven zijn in figuur 3.13, dan blijkt dat de geschatte seizoensinvloed -0.26 respectievelijk -0.27 erg groot is. Indien week 18 in het jaar 1995 met week 18 in het jaar 2000 vergeleken wordt, dan geeft dit de indicatie dat bevrijdingsdag maar een waarde heeft van -0.06^{14} . Dit scheelt dus 0.20 respectievelijk 0.21 met de eerdere indicatie voor de grootte van de invloed van bevrijdingsdag, waardoor er geconcludeerd kan worden dat bevrijdingsdag niet geschat kan worden, doordat de weinige informatie over bevrijdingsdag verschillende indicaties geeft.

Omdat de invloed van koninginndag niet aanwezig was in de weken 18 in het jaar 1995 en 2000, klopt het dat de Y_t 's in die weken positief zijn. De invloed van een koninginndag die niet op week 18 valt, is te bepalen door de waarde van $Y_{\text{week 18,2000}}$ te vergelijken met de geschatte seizoensinvloed voor week 18. De geschatte seizoeninvloed voor week 18 voor de modellen 3.1 en 3.3 is gelijk aan -0.029 respectievelijk -0.040 . Het verschil met de waarde van $Y_{\text{week 18,2000}}$ is dan gelijk aan 0.15 respectievelijk 0.16 . Deze waarden zijn tevens de indicatie van de grootte van de invloed als er geen koninginndag valt op week 18. Deze schatter zal vergeleken moeten worden met eventuele schatters van bijvoorbeeld het log-lineair model, omdat er geen waarde gehecht kan worden aan een schatter die bepaald is met 1 vergelijking.

¹⁴ $\hat{Y}_{\text{week 18,2000}} - \hat{Y}_{\text{week 18,1995}} = 0.06$ voor beide modellen

1.4.1. Controle op afhankelijkheid tussen residuen Y_t

Met onderstaande diagnostische toetsen kan de hypothese getoetst worden dat de residuen Y_t onafhankelijk en gelijk verdeeld zijn.

Indien de hypothese niet verworpen wordt, dan is er de indicatie dat de residuen Y_t een witte ruis heeft als onderliggend ruisproces. Uitgaande van de verwachting van nul bij een witte ruis, kunnen er voorspellingen gemaakt worden door de waarden van de geschatte seizoensinvloeden en de geschatte trend bij elkaar op te tellen.

Indien de hypothese verworpen wordt, dan is er de indicatie dat er nog afhankelijkheid is tussen de variabelen Y_t en zal er geprobeerd worden om deze afhankelijkheid met behulp van een stationair tijdreeks model te modelleren. Indien deze afhankelijkheid goed gemodelleerd wordt, dan wordt er verwacht dat de residuen R_t van het stationair tijdreeks model een witte ruis als onderliggend ruisproces hebben en dus een verwachting van nul hebben. Er kunnen dan voorspellingen gemaakt worden door de waarden van het bepaalde stationair tijdreeks model, de geschatte seizoensinvloeden en de geschatte trend bij elkaar op te tellen.

De toetsen¹⁵ zullen eerst toegelicht worden, om vervolgens de resultaten van deze toetsen te bespreken.

Autocorrelatie en autocovariantie

De autocorrelatie voor lag h is gedefinieerd als:

$$\rho_Y(h) = \text{Cor}(Y_{t+h}, Y_t) = \gamma_Y(h) / \gamma_Y(0) = \gamma_Y(h) / \text{Var}(Y_t),$$

waarbij $\gamma_Y(h)$ gelijk is aan de autocovariantie voor lag h . De autocovariantie is gelijk aan de covariantie van 2 variabelen, echter nu komen de variabelen uit hetzelfde stochastische proces. De autocovariantie bij een lag h is gedefinieerd als:

$$\gamma_Y(h) = \gamma_Y(h, 0) = \text{Cov}(Y_{t+h}, Y_t) = E[(Y_{t+h} - EY_{t+h})(Y_t - EY_t)]$$

Als Y_1, \dots, Y_n een witte ruis als onderliggend proces heeft, dan moeten de geschatte autocorrelaties van Y_t de autocorrelaties van een witte ruis benaderen. Bij een grote n is de autocorrelatie van een reeks met een witte ruis normaal(0, 1/n) verdeeld. Wanneer Y_t een witte ruis als onderliggend proces heeft,

dan zal 95% van de autocorrelaties vallen tussen de grenzen $\pm \frac{\Phi_{1-\alpha/2}}{\sqrt{n}}$.

Hierbij is

- $\Phi_{1-\alpha/2}$ gelijk aan de $(1-\alpha/2)$ kwantiel van de standaard normale verdeling, ($\Phi_{1-\alpha/2} = 1.96$)
- α significantie-niveau welke gelijk gesteld is aan 0.05

Portmanteau Ljung & Box

De toetsingsgrootte Q is gelijk aan:

$$Q = n \sum_{i=1}^h \hat{\rho}(i)^2$$

Als Y_1, \dots, Y_n een witte ruis als onderliggend proces heeft, dan moeten de geschatte autocorrelaties van Y_t de autocorrelaties van een witte ruis benaderen. Bij een grote n is de autocorrelatie van een reeks met een witte ruis normaal(0, 1/n) verdeeld. De $\sqrt{n} * (\hat{\rho}(i))$ zal daarom bij een witte ruis standaard normaal verdeeld zijn, waardoor Q een χ^2 verdeling met h vrijheidsgraden heeft. De hypothese wordt dan ook verworpen als $Q > \chi^2_{1-\alpha, h}$. Ljung en Box hebben deze toets verfijnd door Q te definiëren als

$$Q_{LB} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}(j)^2}{(n-j)}$$

¹⁵ Naast de beschreven toelichting van deze toetsen, is er meer informatie over deze toetsen beschreven in paragraaf 1.6 van "Introduction to Time Series and Forecasting" (geschreven door Peter J. Brockwell en Richard A. Davis, © 1996, drukkerij Springer-Verlag New York Inc.)

Turning point

Er is sprake van een keerpunt (turning point) als $Y_{i-1} < Y_i$ en $Y_i > Y_{i+1}$ of als $Y_i < Y_{i+1}$ en $Y_{i+1} > Y_i$. T is het aantal keerpunten in een reeks. Bij een onafhankelijk en gelijk verdeelde reeks, oftewel een reeks die een witte ruis als onderliggend proces heeft, is de verwachting van T gelijk aan

$$\mu_T = E(T) = 2(n-2)/3$$

en de variantie gelijk aan

$$\sigma_T^2 = \text{Var}(T) = (16n-29)/90.$$

Een hoge waarde van $(T - E(T))$ geeft de indicatie dat de reeks meer fluctueert dan een onafhankelijk en gelijk verdeelde reeks. Een waarde van $(T - E(T))$ die veel kleiner is dan nul, geeft de indicatie dat er een positieve correlatie aanwezig is tussen de dichtstbijzijnde waarnemingen. Als n groot is, dan is er bewezen dat T bij benadering normaal(μ_T , σ_T) verdeeld is. De hypothese wordt dan verworpen als $|T - \mu_T| / \sigma_T > 1.96$

Difference sign

Als S het aantal tijdstippen dat het verschil tussen $Y_i - Y_{i-1}$ positief is, dan is bij een onafhankelijk en gelijk verdeelde reeks, oftewel een reeks die een witte ruis als onderliggend proces heeft, de verwachting van S gelijk aan

$$\mu_S = E(S) = 1/2(n-1)$$

en de variantie van S gelijk aan

$$\sigma_S^2 = \text{Var}(S) = (n+1)/12.$$

Daarnaast geldt bij een onafhankelijk en gelijk verdeelde reeks dat S ongeveer $N(\mu_S, \sigma_S^2)$ is.

Als $|S - \mu_S| / \sigma_S > \phi_{1-\alpha/2}$ dan wordt de aanname dat er geen trend meer aanwezig is in de residuen verworpen.

The rang toets

De rang toets is vooral bedoeld om een lineaire trend in de data zichtbaar te maken. Als P het aantal combinaties(i, t) is, waarbij $Y_t > Y_i$ en $t > i$ ($i=1, \dots, n-1$), dan is de verwachte waarde van P bij een onafhankelijke en gelijk verdeelde reeks gelijk aan:

$$\mu_P = 1/4n(n-1),$$

daar de kans op $\{Y_j > Y_i\}$ een half is bij een onafhankelijk en gelijk verdeelde reeks $\{Y_1, \dots, Y_n\}$. De variantie van P is

$$\sigma_P^2 = n(n-1)(2n+5)/72.$$

Bij grote n is P ongeveer Normaal(μ_P , σ_P^2) verdeeld. Als $|P - \mu_P| / \sigma_P > \phi_{1-\alpha/2}$ dan wordt de aanname dat er geen trend meer aanwezig is in de residuen verworpen.

Resultaten toetsen

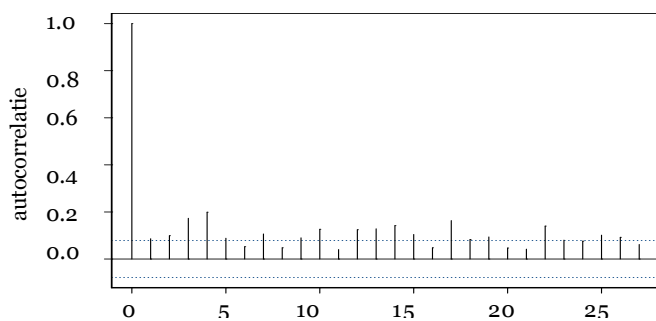
Indien bovenstaande toetsen niet verworpen worden, dan is er de indicatie dat er geen afhankelijkheid is tussen de residuen Y_t . Een stationair tijdreeks model levert dan geen toegevoegde waarde, en zal niet ontwikkeld worden in het geval dat alle toetsen niet verworpen worden. In tabel 3.3 zijn de resultaten van de toetsen weergegeven. Indien een vlak in tabel 3.3 lichtgrijs gekleurd is, dan is de hypothese voor dat model verworpen. Als het vlak niet gekleurd is, dan is de hypothese niet verworpen.

	Autocorrelatie	Portmanteau	Turning P.	Diff.sign	Rang toets
H_0 wordt verworpen als	<95% tussen de grenzen $\pm 1.96/\sqrt{n}$	$Q > \chi^2_{1-\alpha, h}$	$ T - \mu / \sigma_T > 1.96$	$ S - \mu_S / \sigma_S > 1.96$	$ P - \mu_P / \sigma_P > 1.96$
Y_t (model 3.1)	21.4% < 95%	203.95 > 40.1	2.63 > 1.96	1.04 < 1.96	0.73 < 1.96
Y_t (model 3.2)	95% = 95%	17.12 < 30.1	0.64 < 1.96	0.89 < 1.96	0.14 < 1.96
Y_t (model 3.3)	32.1% < 95%	193.77 > 40.1	2.44 > 1.96	0.21 < 1.96	0.67 < 1.96
Y_t (model 3.4)	95% = 95%	30.33 > 30.1	1.62 < 1.96	0.18 < 1.96	0.70 < 1.96

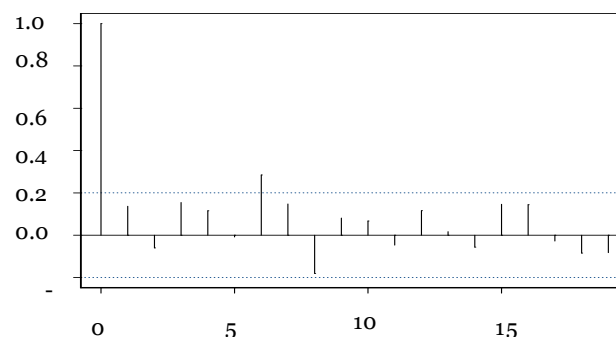
Tabel 3.3: Resultaten toetsen per model. Indien de hypothese verworpen wordt, dan is het vlak licht-grijs gekleurd

De figuren 3.18 tot en met 3.21 geven de autocorrelatie weer. Met behulp van deze figuren zijn de percentages in de 2^e kolom van tabel 3.3 berekend.

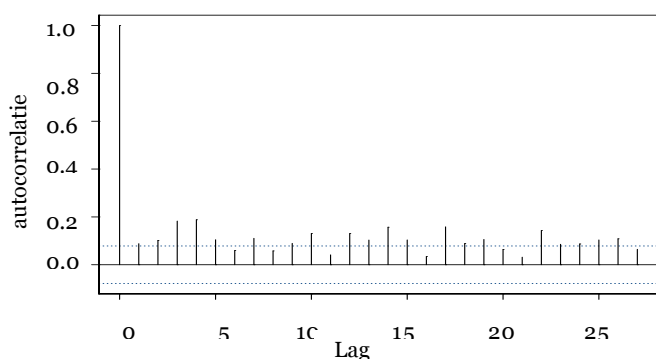
Bij model 3.4 wordt de hypothese dat de residuen Y_t onafhankelijk en gelijk verdeeld is, alleen door de Portmanteau toets verworpen. Omdat de hypothese hier niet verworpen wordt, en de hypothese bij de overige toetsen niet verworpen wordt, is er voor gekozen om aan te nemen dat de residuen Y_t van het model 3.4 tevens een onafhankelijke en gelijk verdeelde tijdreeks zijn. De resultaten van de toetsen geven tevens de indicatie dat de residuen Y_t van het model 3.2 een onafhankelijke en gelijk verdeelde tijdreeks is, en dus een witte ruis als onderliggend proces heeft. Doordat witte ruis normaal($0, \sigma^2$) verdeeld is, zal de σ^2 , oftewel de variantie van de residuen Y_t , meegenomen worden bij het bepalen van de voorspellingsintervallen van de voorspellingen. Cijfers van de fitting van model 3.2 en model 3.4 zijn weergegeven in bijlage III.D. Een korte samenvatting van deze cijfers is weergegeven in tabel 3.5. Omdat een stationair tijdreeks model geen toegevoegde waarde levert voor tijdreeksen die onafhankelijk en gelijk verdeeld zijn, zal een stationair tijdreeks model alleen ontwikkeld worden voor de modellen 3.1 en 3.3.



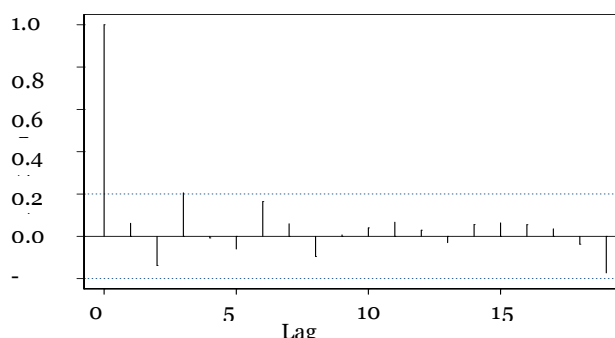
Figuur 3.18: De autocorrelatie van lag h voor de data van de residuen Y_t van model 3.1



Figuur 3.19: De autocorrelatie van lag h voor de data van de residuen Y_t van model 3.2



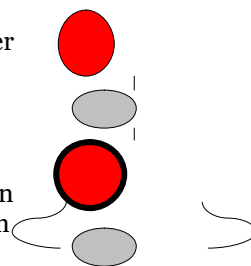
Figuur 3.20: De autocorrelatie van lag h voor de data van de residuen Y_t van model 3.3



Figuur 3.21: De autocorrelatie van lag h voor de data van de residuen Y_t van model 3.4

3.3. Stationaire tijdreeks model

Omdat een van de diagnostische toetsen verworpen wordt, is er de indicatie dat er nog afhankelijkheid is tussen de residuen Y_t bij model 3.1 en 3.3 en zal er geprobeerd worden om deze afhankelijkheid met behulp van een stationair tijdreeks model te modelleren. Indien deze afhankelijkheid goed gemodelleerd wordt, dan wordt er verwacht dat de residuen R_t van het stationair tijdreeks model een witte ruis als onderliggende proces hebben en dus een verwachting van nul hebben. Er kunnen dan voorspellingen gemaakt worden door de waarden van het bepaalde stationair tijdreeks model, de geschatte seizoensinvloeden en de geschatte trend bij elkaar op te tellen



Vaak gebruikte stationaire tijdreeks modellen zijn:

- moving average model. Dit model met orde q (waarbij q een positieve waarde heeft) is gedefinieerd als $Y_t = Z_t + \beta_1 Z_{t-1} + \dots + \beta_q Z_{t-q}$ en afgekort als MA(q). Hierbij is $\{Z_t\}$ een witte ruis, oftewel een reeks met pure onafhankelijke waarnemingen waarbij de verwachting gelijk is aan nul en de variantie gelijk aan σ^2 . (β_1, \dots, β_q zijn constanten)
- autoregressive model. Dit model met orde p (waarbij p een positieve waarde heeft) is gedefinieerd als $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + Z_t$ en afgekort als AR(p). Hierbij is $\{Z_t\}$ een witte ruis en zijn $\alpha_1, \dots, \alpha_p$ constanten
- een gemixte autoregressive/moving average model. Dit model met orde p en q is gedefinieerd als $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + Z_t + \beta_1 Z_{t-1} + \dots + \beta_q Z_{t-q}$ en afgekort als ARMA(p, q). Hierbij is $\{Z_t\}$ een witte ruis en zijn $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ constanten.

In paragraaf 3.3.1 zal geanalyseerd worden welk stationair model geschikt is voor de stationaire tijdreeksen Y_t van de modellen 3.1 en 3.3, waarna in paragraaf 3.3.2. de orde p en/of q van het stationair tijdreeks model bepaald wordt. In paragraaf 3.3.4. zal met behulp van de meest aannemelijke schatter de α_i en/of β_i geschat worden, zodat de residuen R_t van het stationaire tijdreeks model bepaald kunnen worden. Deze residuen R_t worden in paragraaf 3.3.5 getoetst met de in paragraaf 3.2.3 beschreven toetsen. Hierbij is de nulhypothese de hypothese dat de residuen R_t onafhankelijk en gelijk verdeeld zijn.

3.3.1. Keuze stationaire tijdreeks model

Om te bepalen welk stationair tijdreeks model geschikt is, zal de autocorrelatie, de partiele autocorrelatie en de autocovariantie van de verschillende stationaire tijdreeksen geanalyseerd worden.

Een toelichting van de begrippen autocorrelatie en autocovariantie is beschreven in paragraaf 3.2.3. De geschatte autocovariantie bij een MA(q) model wordt nul bij een lag die groter is dan de orde q . Bij een AR(p) model wordt de geschatte autocovariantie niet nul bij een lag die groter is dan haar orde. De geschatte autocorrelatie toont bij een MA(q) model een scherpe afloop van de significant van nul verschillende geschatte autocorrelaties bij de lag q . Na lag q is de geschatte autocorrelatie bij een MA(q) model niet significant verschillend van nul. Bij een AR(p) en een ARMA(p, q) model zakt de geschatte autocorrelatie rustig terug naar de waarde die niet significant verschilt van nul.

De partiele autocorrelatie is gelijk aan de correlatie tussen Y_t en Y_{t+h} aangepast aan de waarden tussen deze punten, oftewel aangepast aan de waarden $Y_{t+1}, \dots, Y_{t+h-1}$. De partiele autocorrelatie voor lag h is gedefinieerd als:

$$\psi(0) = 1$$

$$\psi(h) = \phi_{hh}, h \geq 1$$

Hierbij is

ϕ_{hh} = het laatste component van ϕ_h ;

$$\phi_h = \Gamma_h^{-1} \gamma_h$$

$$\Gamma_h = [\gamma(i-j)]_{i,j=1}^h$$

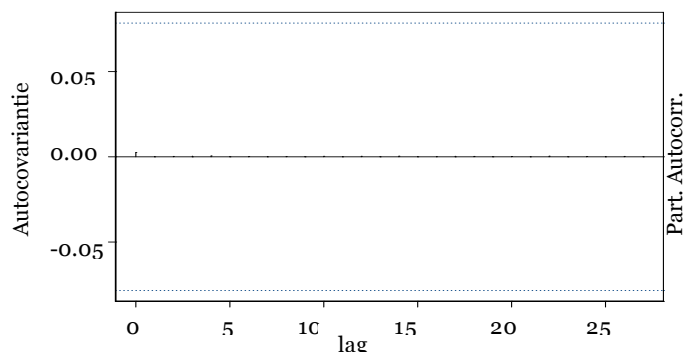
$$\gamma_h = [\gamma(1), \gamma(2), \dots, \gamma(h)]$$

$\gamma(h)$ = de autocovariantie bij lag h

De geschatte partiele autocorrelatie toont bij een AR(p) model een scherpe afloop van de significant van nul verschillende geschatte partiele autocorrelaties bij de lag p. Na lag p is de geschatte partiele autocorrelatie bij een AR(p) model niet meer significant verschillend van nul. Bij een MA(q) en een ARMA(p,q) model zakt de geschatte partiele autocorrelatie rustig terug naar de waarde die niet significant verschilt van nul.

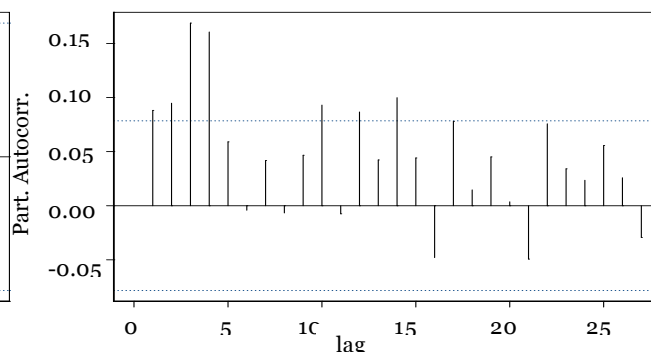
In de onderstaande figuren is van de residuen Y_t van de modellen 3.1 en 3.3 de geschatte autocovariantie en de partiele autocorrelatie weergegeven. De geschatte autocorrelatie is weergegeven in respectievelijk figuur 3.18 en 3.20. Naar aanleiding van deze figuren kan bepaald worden welk stationair tijdreeks model geschikt is.

Model 3.1



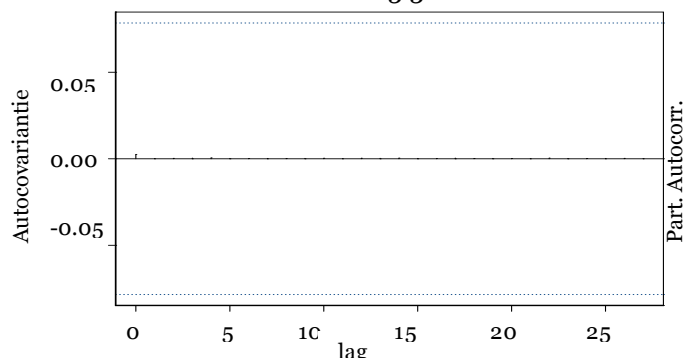
Figuur 3.20: De autocovariantie van lag h voor de data van de residuen Y_t van model 3.1

Model 3.1



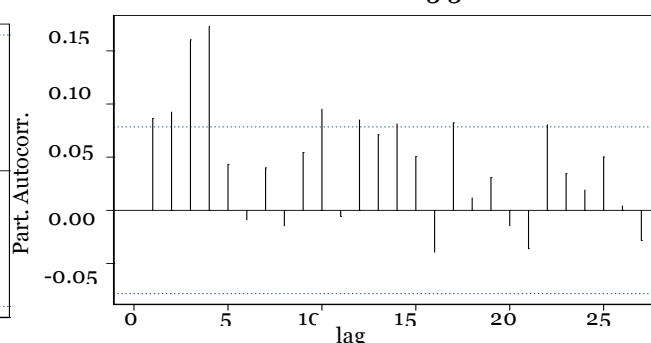
Figuur 3.21: De partiele autocorrelatie van lag h voor de data van de residuen Y_t van model 3.1

Model 3.3



Figuur 3.22: De autocovariantie van lag h voor de data van de residuen Y_t van model 3.3

Model 3.3



Figuur 3.23: De partiele autocorrelatie van lag h voor de data van de residuen Y_t van model 3.3

Aan de hand van de geschatte autocovarianties kan er aangenomen worden dat de afhankelijkheden tussen de residuen Y_t van model 3.1. en 3.3 niet met een autoregressive-model bepaald kunnen worden. Doordat de geschatte autocorrelatie van de residuen Y_t van model 3.1 en 3.3 in figuur 3.18 respectievelijk 3.20 geen scherpe afloop vertoont na een lag h maar rustig afzwakt naar nul, kan hetzelfde worden aangenomen voor de moving average model. Dat de geschatte autocorrelatie rustig afzwakt naar nul, geeft echter wel de indicatie dat de afhankelijkheden tussen de stationaire tijdsreeks-waarnemingen Y_t van model 3.1. en 3.3 bepaald kunnen worden met een ARMA-model. Omdat de geschatte partiele autocorrelaties ook rustig afzwakken naar nul, zal er geanalyseerd worden of met het ARMA-model de afhankelijkheden bepaald kunnen worden. Doordat het ARMA-model afhankelijk is van de orde p en q, zal in de volgende paragraaf deze ordes voor model 3.1 en 3.3 bepaald worden.

3.3.2. Bepaling orde

De orde van een stationair tijdreeks model, wordt hier bepaald naar aanleiding van de waarde van het Akaike criterium. Het minimum van het Akaike criterium bepaalt het aantal parameters waarbij er een balans is gevonden tussen de toename van de aannemelijkheid aan de ene kant en de toename van de complexiteit aan de andere kant.

Het Akaike criterium is als volgt gedefinieerd:

$$AIC(k) = -2 \log(L(\alpha, \beta, \sigma)) + 2k,$$

waarbij k gelijk is aan $p+q$, en $L(\alpha, \beta, \sigma)$ de aannemelijkheid van het stationair tijdreeks model ARMA(p, q). Bij Splus wordt $-2 \log(L(\alpha, \beta, \sigma))$ bepaald met een zogenaamde conditionele log-aannemelijkheids-functie. De definitie van deze functie is:

$$-2L(y_{p+1}, \dots, y_n | y_1, \dots, y_p) = (n-p) \log(2\pi\sigma^2) + \sum_{t=p+1}^n \log f_t + \frac{1}{\sigma^2} \sum_{t=p+1}^n e_t^2 / f_t,$$

waarbij

$\sigma^2 f_t$ $\text{var}(Y_1, \dots, Y_{t-1}, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ gelijk is aan de conditionele variantie.

e_t gelijk is aan $(Y_t - \hat{Y}_t^{t-1})$

\hat{Y}_t^{t-1} de conditionele gemiddelde een-stap-vooruit schatter is en geformuleerd als:

$$E(Y_t | Y_1, \dots, Y_{t-1}, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q).$$

Bij deze berekening gaat Splus er echter van uit dat het gemiddelde van de stationaire tijdreeks Y_t gelijk is aan nul. Voor de modellen 3.1 en 3.3 wordt om deze reden de Y_t vervangen door de Y_t min haar gemiddelde. In de rest van paragraaf 3.3 zal er met Y_t de stationaire tijdreeks bedoeld worden waarbij het gemiddelde gelijk is gesteld aan nul.

Het minimum van de AIC(k) voor de verschillende stationaire tijdreeksen is voor model 3.1 gevonden bij ARMA(1,6). Hierbij was de AIC(k) gelijk aan -1995. Echter doordat door Splus direct gekeken wordt naar de autocorrelatietoets en de portmanteau toets, kan bij de bepaling van de orde van het ARMA model daar direct rekening mee gehouden worden. Indien hier namelijk niet direct rekening meegehouden wordt, kunnen de residuen van het ARMA(p, q) nog afhankelijkheid vertonen. Door de minimale AIC(k) te bepalen, waarbij de hypothese dat de residuen van de ARMA(p, q) onafhankelijk en gelijk verdeeld zijn niet verworpen wordt bij de autocorrelatie toets en de portmanteau toets, wordt afhankelijkheid tussen de residuen van het ARMA(p, q) model voorkomen.

Het minimum van de AIC(k) waarbij de hypothese dat de residuen van de ARMA(p, q) onafhankelijk en gelijk verdeeld zijn niet verworpen wordt bij de autocorrelatie en de portmanteau toets, is voor model 3.1 gevonden bij ARMA(1,2). Hierbij was de AIC(k) gelijk aan -1992. De orde (p, q) bij het ARMA-model voor model 3.1 is dus gelijk aan (1,1).

Bij model 3.3 is het minimum van de AIC(k) waarbij de voorwaarden gelden, gevonden bij ARMA(1,1). Hierbij was de AIC(k) gelijk aan -1994. De orde (p, q) van het ARMA-model voor model 3.3 is dus gelijk aan (1,1).

3.3.3. Schatten parameters

Bij het bepalen van het maximum van de log-aannemelijkheidsfunctie in paragraaf 3.3.4 moesten de parameters α en β geschat worden. Bij het bepalen van het maximum van de log-aannemelijkheidsfunctie waarbij de orde p en q gegeven is, worden de \hat{Y}_t^{t-1} en de conditionele variantie $\sigma^2 f_t$ geschat. Deze hebben de α en β als parameters, waardoor de α en β al geschat zijn tijdens het bepalen van het maximum van de log-aannemelijkheidsfunctie. De ARMA-modellen kunnen met deze geschatte α en β als volgt geformuleerd worden:

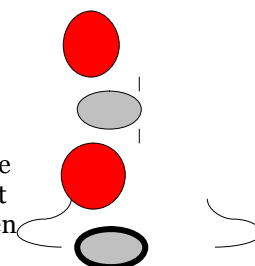
$$\text{ARMA}(1,2) \text{ voor model 3.1} = \hat{Y}_t = 0.98 \hat{Y}_{t-1} + Z_t + 0.96 Z_{t-1} - 0.05 Z_{t-2}$$

$$\text{ARMA}(1,1) \text{ voor model 3.3} = \hat{Y}_t = 0.98 \hat{Y}_{t-1} + Z_t + 0.93 Z_{t-1}$$

3.3.4. Controle op afhankelijkheid tussen residuen R_t

Met onderstaande diagnostische toetsen kan de hypothese getoetst worden dat de residuen R_t van het stationair tijdreeks model onafhankelijk en gelijk verdeeld zijn. Indien de hypothese verworpen wordt, is er de indicatie dat de residuen R_t een witte ruis als onderliggend proces hebben en dus een verwachting van nul hebben, omdat er geen afhankelijkheid bestaat tussen de residuen R_t . Er kunnen dan voorspellingen gemaakt worden door de waarden van het bepaalde stationair tijdreeks model, de geschatte seizoensinvloeden en de geschatte trend bij elkaar op te tellen.

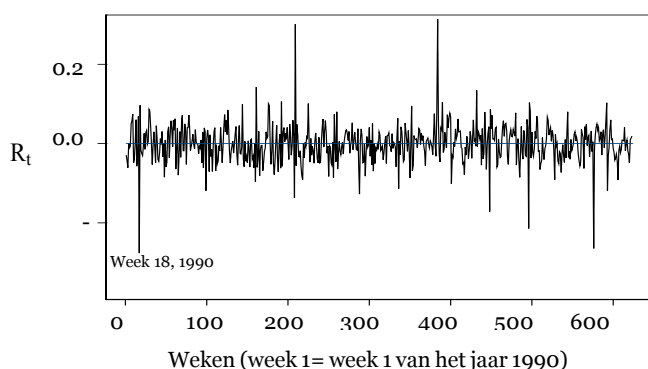
Indien een hypothese verworpen wordt, dan is er de indicatie dat er nog afhankelijkheid is tussen de residuen. Dit kan veroorzaakt worden door een verkeerde modellering van het stationaire tijdreeks, en het stationaire tijdreeks model zal dan herzien moeten worden.



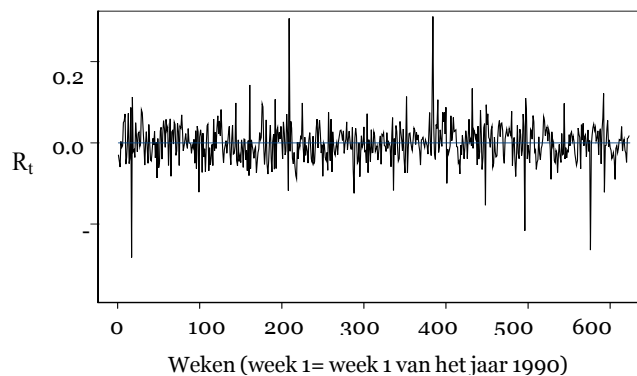
De residuen bij een ARMA(p,q)-model worden bepaald door:

$$\begin{cases} \hat{R}_t = 0 & \text{als } t \leq p \\ \hat{R}_t = \hat{Y}_t - \sum_{i=1}^p \hat{\alpha}_{i,n} \hat{Y}_{t-i} - \sum_{j=1}^q \hat{\beta}_{j,n} \hat{R}_{t-j} & \text{als } t > p \end{cases}$$

en zien er als volgt uit:



Figuur 3.24: De residuen R_t van de ARMA(1,2) van model 3.1

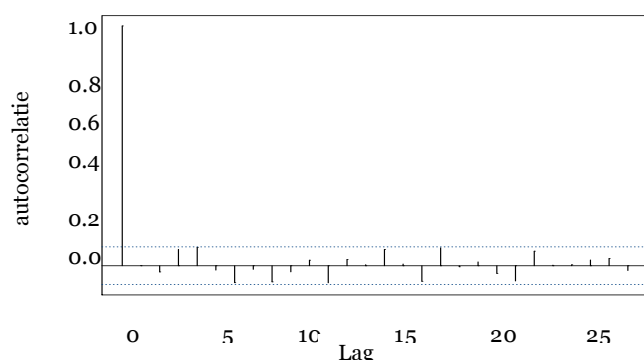


Figuur 3.25: De residuen R_t van de ARMA(1,1) van model 3.3

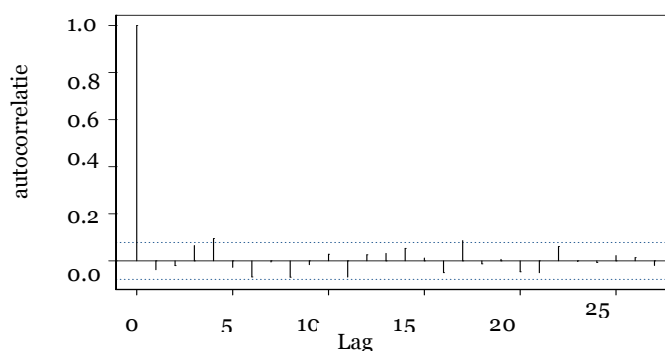
Er is geen waarneembaar verschil tussen figuur 3.24 respectievelijk 3.25 met figuur 3.14 respectievelijk figuur 3.16. Om toch te analyseren of er wel afhankelijkheid is verwijderd, zullen de residuen R_t getoetst worden op onafhankelijkheid en een gelijke verdeling. De residuen R_t kunnen met de toetsen die beschreven zijn in paragraaf 3.2.3 getoetst worden op de hypothese dat de residuen onafhankelijk en gelijk verdeeld zijn. Indien bij geen van de toetsen de hypothese verworpen wordt (vlak in tabel 3.4 heeft dan geen kleur), dan is de afhankelijkheid tussen de R_t middels de bepaalde ARMA(p,q) verwijderd.

	Autocorrelatie	Portmanteau	Turning P.	Diff.sign	Rang toets
H_0 wordt verworpen als	<95% tussen de grenzen $\pm 1.96/\sqrt{n}$	$Q > \chi^2_{1-\alpha,h}$	$ T - \mu / \sigma_T > 1.96$	$ S - \mu_S / \sigma_S > 1.96$	$ P - \mu_P / \sigma_P > 1.96$
R_t (model 3.1)	96% > 95%	34.3 < 40.1	2.76 > 1.96	1.11 < 1.96	0.91 < 1.96
R_t (model 3.3)	96% > 95%	35.5 < 40.1	2.57 > 1.96	0.42 < 1.96	0.70 < 1.96

Tabel 3.4: Resultaten toetsen voor de residuen van de bepaalde ARMA- modellen van model 3.1 en 3.3., indien de hypothese verworpen wordt dan is het vlak licht-grijs getint.



Figuur 3.26: De autocorrelatie van lag h voor de residuen van de ARMA(1,2) van model 3.1



Figuur 3.27: De autocorrelatie van lag h voor de residuen van de ARMA(1,1) van model 3.3

De figuren 3.18 tot en met 3.21 geven de autocorrelatie weer. Met behulp van deze figuren zijn de percentages in de 2^e kolom van tabel 3.4 berekend.

De resultaten van de toetsen geven de indicatie dat het aantal keerpunten in beide modellen kleiner is dan bij een onafhankelijk en gelijk verdeelde reeks. Om de toets niet te verwerpen had het aantal keerpunten tussen de 395 en 433 moeten liggen in plaats van 385 keerpunten bij model 3.1 respectievelijk 387 bij model 3.3 moeten hebben. Doordat de modellen maar 10 respectievelijk 8 keerpunten tekort komen en doordat de hypothese bij de andere toetsen niet wordt verworpen, neem ik toch aan dat de residuen R_t van de ARMA-modellen van de modellen 3.1 en 3.3 onafhankelijke en gelijk verdeelde tijdreeksen zijn. Indien deze aanname klopt, dan zal de afhankelijkheid tussen de residuen Y_t goed gemodelleerd zijn. Dit zal een verbetering in de fitting en in de voorspellingen opleveren.

3.4. Analyse fitting

De cijfers van de fitting van de 4 tijdreeksen zijn beschreven in bijlage III.D en zijn samengevat in tabel 3.5 en tabel 3.6. Bij het analyseren van de fitting wordt de vergelijking van de verwachting van de hoeveelheid post met de werkelijke hoeveelheid post gedaan aan de hand van een analyse van de relatieve afwijkingen. De relatieve afwijkingen zijn gedefinieerd als:

$$\frac{(\text{realisatie} - \text{verwachting})}{\text{verwachting}} * 100\%$$

Indien de relatieve afwijkingen rond de nul procent zitten, kan er geconcludeerd worden dat de verwachting van de hoeveelheid post de werkelijke hoeveelheid post goed benadert. Dit duidt op een goede fitting.

Interval	model 3.1		model 3.2		model 3.3		model 3.4	
	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.
[-2%,2%]	236	37.9%	55	57.3%	227	36.4%	51	53.1%
[-5%,5%]	490	78.7%	92	95.8%	491	78.8%	91	94.8%
[-10%,10%]	601	96.5%	96	100%	600	96.3%	96	100%
[-15%,15%]	616	98.9%	96	100%	617	99.0%	96	100%

Tabel 3.5: Analyse van de spreiding van de relatieve afwijkingen

Het gemiddelde van de absolute waarden van de relatieve afwijkingen geeft tevens inzicht in de fitting van een model. Het gemiddelde van de absolute waarden van de relatieve afwijkingen is gedefinieerd als het gemiddelde van:

$$\left| \frac{(\text{realisatie} - \text{verwachting})}{\text{verwachting}} \right| * 100\%$$

en wordt aangegeven als Gem*.

	model 3.1	model 3.2	model 3.3	model 3.4
Gem*	3.5%	1.9%	3.5%	1.9%

Tabel 3.6: Gemiddelde van de absolute waarden van de relatieve afwijkingen

Uit tabel 3.5 en 3.6 blijkt dat de modellen 3.2 en 3.4 zoals verwacht een betere fitting hebben. Dit was verwacht omdat er bij model 3.2 en 3.4 met dezelfde hoeveelheid parameters minder datapunten gefit hoeven te worden. Ik verwacht echter wel dat de modellen 3.2 en 3.4 slechtere voorspellingen zullen geven, omdat er relatief weinig data beschikbaar zijn bij deze modellen. Informatie over grote partijen is namelijk alleen beschikbaar vanaf het jaar 2001. Daarnaast hebben stationaire tijdreeks modellen bij de modellen 3.1 en 3.3 nog meer afhankelijkheid uit de data kunnen halen middels ARMA modellen. Voor de ARMA modellen zijn de voorspellingen met een specifieke methode bepaald. Deze methode zal in paragraaf 3.5 toegelicht worden. Tevens zullen in paragraaf 3.5 de voorspellingen geanalyseerd worden.

3.5. Voorspellingen

In deze paragraaf zullen de voorspellingsmethoden toegelicht worden, om daarna de voorspellingen voor het jaar 2003 te analyseren. Hierbij zal de analyse zich concentreren op het verschil tussen de modellen en het verschil met de voorspellingen die gemaakt zijn door de WAP-groep. Voordat de voorspellingsmethoden toegelicht worden, zal eerst een herhaling van de 4 oorspronkelijke modellen gegeven worden.

$$\text{Log}(X_t) = m_t + s_t + Y_t \quad (3.1)$$

$$\text{Log}(X_t - G_t) = m_t + s_t + Y_t \quad (3.2)$$

$$\text{Log}(X_t + 14\text{milj} * 1_{t \in \text{Pasen, Pinksteren}} + 16\text{milj} * 1_{t \in \text{hemelvaartsdag}}) = m_t + s_t + Y_t \quad (3.3)$$

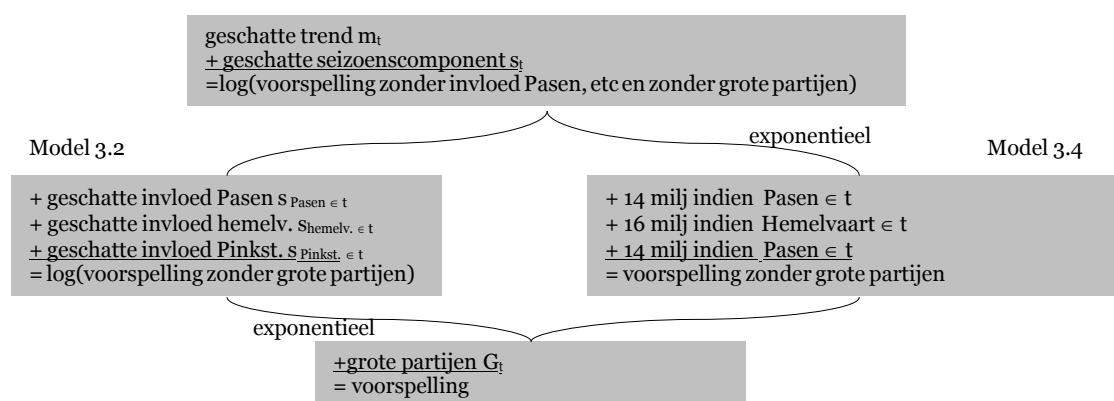
$$\text{Log}(X_t - G_t + 14\text{milj} * 1_{t \in \text{Pasen, Pinksteren}} + 16\text{milj} * 1_{t \in \text{hemelvaartsdag}}) = m_t + s_t + Y_t \quad (3.4)$$

Waarbij:

X_t	de hoeveelheid post in week t
m_t	de deterministische trend
s_t	de deterministische seizoenscomponent
Y_t	ruis (waarvan verwacht wordt dat het een stationair tijdreeks is)
G_t	de hoeveelheid post die in week t door grote partijen gegenereerd is

3.5.1. Voorspellingsmethode model 3.2 en 3.4

Bij de modellen 3.2 en 3.4 leverde een stationair tijdreeks model volgens de uitgevoerde toetsen in paragraaf 3.2.3 geen toegevoegde waarde, omdat er aangenomen kon worden dat de residuen Y_t een onafhankelijke en gelijk verdeelde tijdreeks is. Doordat de verwachting van een onafhankelijke en gelijk verdeelde tijdreeks gelijk is aan nul, kunnen de voorspellingen voor de modellen 3.2 en 3.4 op de volgende manier worden bepaald:



Figuur 3.28: Overzicht voorspellingsmethode voor de modellen 3.2 en 3.4

Hierbij zijn alle schattingen gemaakt naar aanleiding van waarnemingen van de jaren 2001 en 2002, omdat in de jaren daarvoor geen gegevens van de grote partijen zijn geadministreerd

Het voorspellingsinterval voor de verwachte hoeveelheid post, waarbij α gelijk gesteld wordt aan 0.05, is bij model 3.2 gelijk aan:

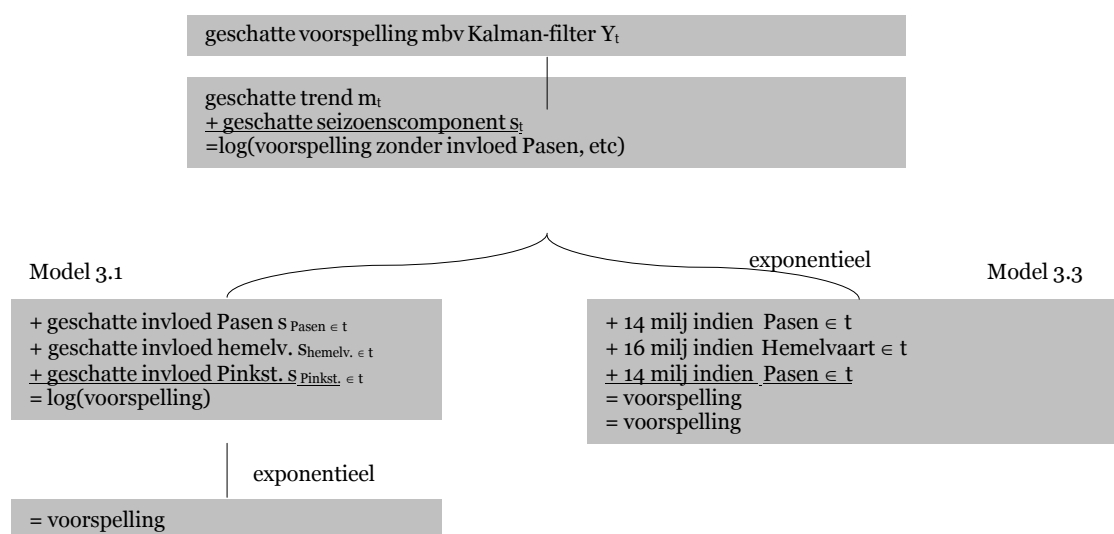
$$G_t + e^{\log(\text{voorspelling zonder grotepartijen}) \pm \sqrt{\text{var}(\hat{Y}_t) / n * t_{n-1, \alpha} / 2}}$$

Bij model 3.4 is het voorspellingsinterval gelijk aan:

$$G_t + 14\text{milj} * 1_{\text{Pasen, Pinksteren} \in t} + 16\text{milj} * 1_{\text{hemelvaartsdag} \in t} + e^{\log(\text{voorspelling zonder grote partijen en feestdagen}) \pm \sqrt{\text{var}(\hat{Y}_t) / n * t_{n-1, \alpha} / 2}}$$

3.5.2. Voorspellingsmethode model 3.1 en 3.3

Bij de modellen 3.1 en 3.3 is er wel een stationair tijdreeks model bepaald en deze heeft logischerwijs invloed op de voorspellingen. Met behulp van het Kalman filter worden de voorspellingen voor een ARMA(p,q) model bepaald door Splus. Het Kalman filter maakt hierbij gebruik van de zogenaamde state-space model. Een state-space model gaat uit van het principe dat een waarneming bestaat uit een signaal en een ruis en beschrijft het signaal vervolgens als een lineaire combinatie van de zogenaamde state-variabelen C_t . Deze beschrijving gaat via twee formule's, namelijk de waarnemings- en de state-formule. Door herformulering van de ARMA-modellen naar een waarnemings- en een state-formule kan de Kalman-filter gebruikt worden. De waarnemings- en de state-formule is voor beide ARMA-modellen beschreven in bijlage III.C. Tevens is in bijlage III.C de methode beschreven van de kalman-filter. Nadat er met de Kalman filter voorspellingen van de ARMA-modellen gemaakt zijn, kan de hoeveelheid post via de modellen 3.1 en 3.3 op de volgende manier bepaald worden:



Figuur 3.29: Overzicht voorspellingsmethode voor de modellen 3.1 en 3.3

Hierbij zijn alle schattingen gemaakt naar aanleiding van waarnemingen van de jaren 1990 en 2002.

Het voorspellingsinterval, waarbij α gelijk gesteld wordt aan 0.05, is bij model 3.1 gelijk aan:

$$e^{\log(\text{voorspelling}) \pm \sqrt{\text{var}(Y_{k+jd}) / n * t_{n-1, \alpha} / 2}}$$

Bij model 3.3 is het voorspellingsinterval gelijk aan:

$$14 \text{ milj} * 1_{k+jd \in \text{Pasen, Pinksteren}} + 16 \text{ milj} * 1_{k+jd \in \text{hemelvaartsdag}} \\ + e^{\log(\text{voorspelling zonder feestdagen}) \pm \sqrt{\text{var}(Y_{k+jd}) / n * t_{n-1, \alpha} / 2}}$$

3.5.3. Analyse voorspellingen

Prognose t.o.v. de werkelijke hoeveelheid

Van de residuen Y_t is bij model 3.2 en model 3.4 aangenomen dat deze onafhankelijk en gelijk verdeeld zijn. Omdat de verwachting bij een onafhankelijke en gelijk verdeelde tijdreeks gelijk is aan nul, zal het verschil tussen de Y_t van model 3.2 en de Y_t van model 3.4 alleen maar in de voorspellingsintervallen waarneembaar zijn.

Een verschil tussen de voor model 3.2 en 3.4 bepaalde seizoensinvloeden per week, lijkt aan de hand van figuur 3.13 alleen duidelijk aanwezig te zijn rond de periode van Pasen en Pinksteren (weken 14 tot en met 24). Voor de voorspellingen van het jaar 2003 verwacht ik dan ook in die periode verschillen te zien in de voorspellingen, waarbij ik verwacht dat model 3.4 betere voorspellingen geeft. Dit verwacht ik, omdat de residuen Y_t van model 3.4 in figuur 3.17 kleinere waarden vertoont dan de Y_t van model 3.2 (figuur 3.15).

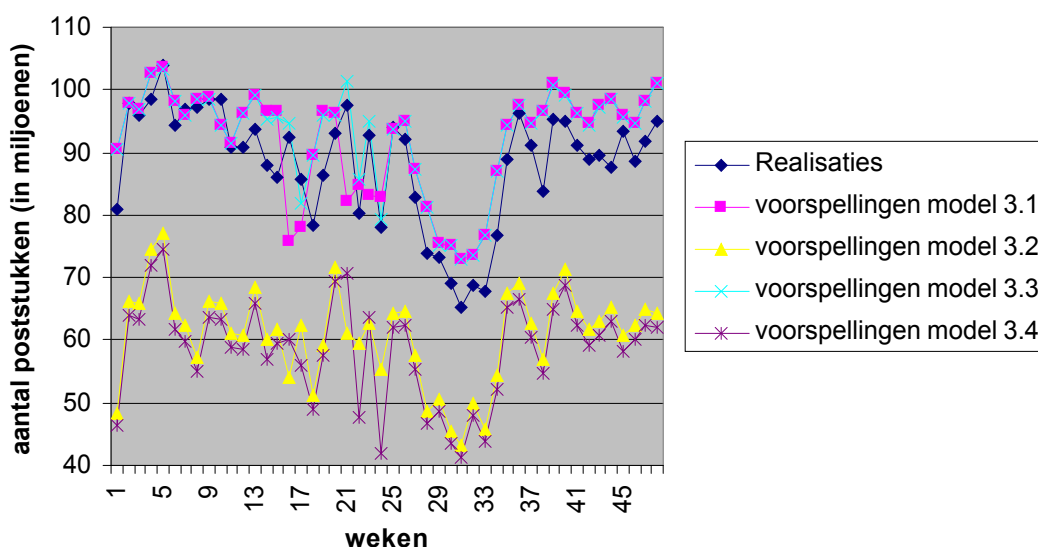
Omdat de voorspellingen, die met behulp van model 3.2 en 3.4 zijn gemaakt, tevens afhangen van de bepaalde trend, kan ik tevens een verwachting maken naar aanleiding van de trend. Doordat de twee trendlijnen elkaar raken in het jaar 1997, zou een voorspelling die berekend is met behulp van model

3.4 enkel en alleen aan de hand van de trend, lager zijn dan de voorspellingen die berekend zijn met model 3.2. Omdat de seizoensinvloeden alleen in de weken 14 tot en met 24 verschillen aangeven, verwacht ik dat de voorspellingen van model 3.4 lager zijn in de overige weken dan de voorspellingen van model 3.2.

Ik verwacht een waarneembaar verschil tussen de voorspellingen van model 3.2 en 3.4 en de werkelijke hoeveelheid post, omdat de geschatte waarden bepaald zijn met relatief weinig data. De hoeveelheid post min de grote partijen in de jaren 2001 en 2002 zal echter, door de toename van de grote partijen in 2002¹⁶ ten opzichte van het jaar ervoor, een dalende trendlijn laten zien. Deze trendlijn zou met de kennis van 2003 minder sterk dalen¹⁷. Volgens de realisaties is het totale aantal grote partijen in 2003 ongeveer gelijk aan het aantal in 2002. Aan de hand van deze gegevens verwacht ik dat er te lage voorspellingen zijn gemaakt met de tijdreeks modellen 3.2 en 3.4.

Bij de modellen 3.1 en 3.3 is er met behulp van de Kalman filter een voorspelling bepaald voor Y_t . Omdat deze voorspellingen weinig verschillen ten opzichte van elkaar zal een verschil tussen de voorspellingen van model 3.1 en 3.3 veroorzaakt worden door een verschil in de geschatte seizoensinvloeden en de geschatte trend. De geschatte seizoensinvloeden blijken alleen maar duidelijke verschillen te vertonen in de periode rond Pasen en Pinksteren. De geschatte trendlijnen raken elkaar pas in het jaar 2019 en daarom kan er verwacht worden dat de voorspellingen van model 3.3 in de weken buiten de periode rond Pasen en Pinksteren lager zijn dan de voorspellingen van model 3.1. Omdat het verschil in de trendlijnen klein is, zal de genoemde verwachting dat de voorspellingen van model 3.1 lager zijn, afhankelijk zijn van de verwachte waarden Y_t .

Omdat de hoeveelheid beschikbare data voldoende is, verwacht ik dat de voorspellingen de werkelijke hoeveelheid post goed benaderen.



Figuur 3.30: Vergelijking van de geprognosticeerde met de werkelijke hoeveelheid post in week 1 tot en met 48 in het jaar 2003

Aan de hand van figuur 3.30, waarin de werkelijke hoeveelheid post en de voorspellingen voor de weken 1 tot en met 48 zijn weergegeven, kan er geconcludeerd worden dat de voorspellingen die gemaakt zijn met behulp van model 3.2 en 3.4 te laag zijn. De voorspellingen van model 3.4 zijn volgens verwachting lager dan de voorspellingen van model 3.2. Echter het verschil tussen deze voorspellingen is, behalve in de periode rond Pasen en Pinksteren, erg klein. Door de relatieve afwijkingen te analyseren, kan er een indicatie gegeven worden of model 3.2 dan wel model 3.4 de invloed van de feestdagen beter schat.

¹⁶Volgens de WAP-groep is het verschil tussen de hoeveelheid grote partijen in 2001 en 2002 niet zo groot is als uit de cijfers blijkt. Zij denken dat het grote verschil veroorzaakt wordt door verschil in het administreren van de gegevens omtrent grote partijen.

¹⁷De trendlijn zal nog steeds een daling aangeven omdat de verwachting van de totale hoeveelheid post in 2003 lager is dan in 2002 (zie bijlage 4B).

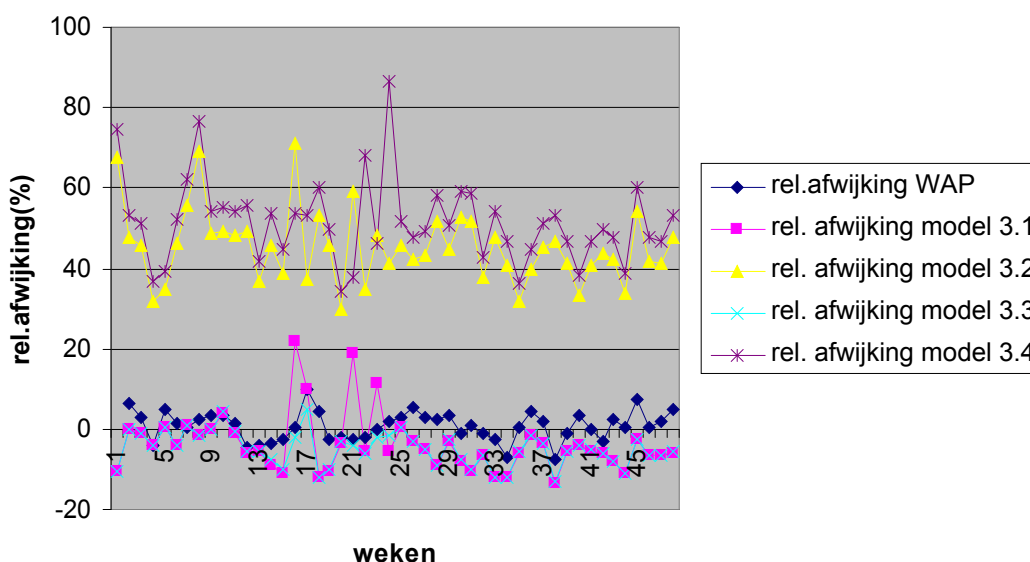
De voorspellingen van model 3.1 en 3.3 lijken in de eerste 10 weken de werkelijke hoeveelheid post goed te benaderen. De voorspellingen van model 3.3 in de periode rond Pasen en Pinksteren (weken 14 tot en met 24) lijken de werkelijke hoeveelheid post beter te benaderen dan de voorspellingen van model 3.1, echter het verschil tussen de werkelijke hoeveelheid post en de voorspellingen is nog steeds waarneembaar. Bij model 3.1 is de seizoensinvloed voor week 16, week 21 en week 23 te groot geschat. Dit blijkt uit het grote verschil van de voorspellingen van model 3.1 met de gerealiseerde hoeveelheid post. Omdat dit verschil veel kleiner is bij model 3.3, geeft dit de indicatie dat het grote verschil niet wordt veroorzaakt door abnormale hoeveelheden post in die weken in 2003, maar doordat model 3.1 de invloed van deze weken niet goed geschat heeft. Het verschil van de voorspellingen van de modellen 3.1 en 3.3 in week 17(Pasen), week 18(koninginnedag), week 19 en week 22(Hemelvaart) met de gerealiseerde hoeveelheid post, geeft de volgende indicaties:

- de invloed van Pasen neemt af, of er is een afwijkende hoeveelheid post in week 17 in het jaar 2003
- de invloed van koninginnedag neemt toe, of er is een afwijkende hoeveelheid post in week 18 in het jaar 2003
- de invloed van week 19 neemt toe, of er is een afwijkende hoeveelheid post in week 19 in het jaar 2003
- de invloed van Hemelvaart neemt toe, of er is een afwijkende hoeveelheid post in week 22 in het jaar 2003

Het zomerdal en de laatste weken zijn te hoog voorspeld door de modellen 3.1 en 3.3. Dat de voorspellingen te hoog zijn in de zomerperiode en in de laatste weken van het jaar, wordt waarschijnlijk veroorzaakt door de dalende hoeveelheid post in het jaar 2003. Dit is niet met de trend geschat en is dus niet in de voorspellingen terug te vinden.

Prognoses t.o.v. de prognoses van de WAP -groep

De vergelijking van de prognoses wordt gedaan aan de hand van een analyse van de relatieve afwijkingen. Door de lage voorspellingen van model 3.2 en 3.4 verwacht ik hoge relatieve afwijkingen. De absolute relatieve afwijkingen zullen waarschijnlijk in de weken, waarin Pasen, Hemelvaart en Pinksteren valt, groter zijn dan in de overige weken. Bij model 3.1 en 3.3 verwacht ik door de te hoge voorspellingen dat de relatieve afwijkingen negatief zijn in de zomerperiode en in de laatste weken van



het jaar 2003.

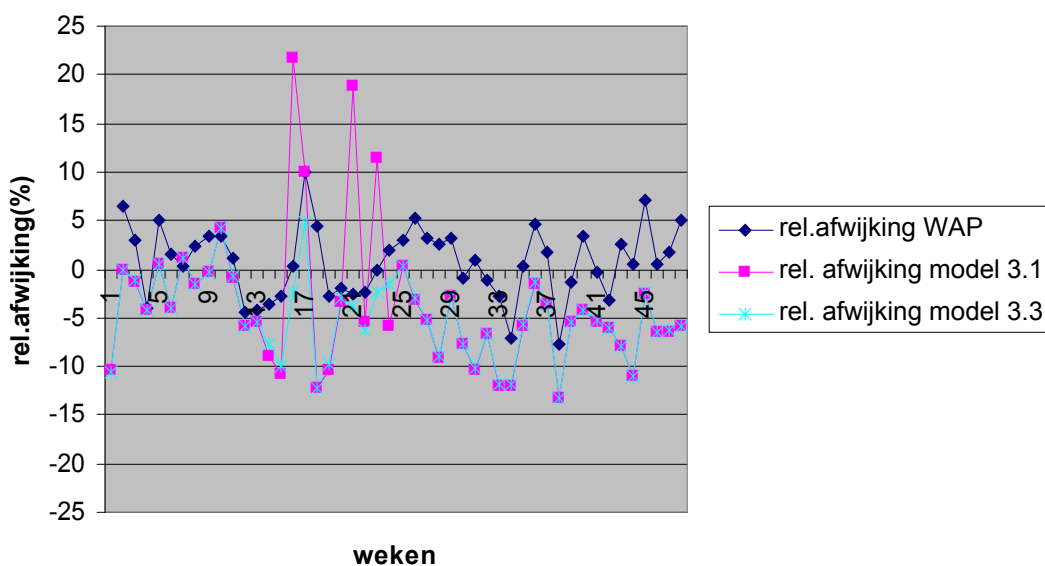
Figuur 3.31: Vergelijking van de relatieve afwijking van de realisatie t.o.v. de voorspellingen (de relatieve afwijkingen van de voorspellingen die gemaakt zijn door de WAP-groep zijn pas sinds week 12 in het jaar 2003 bijgehouden)

Het model 3.2 lijkt tegen de verwachtingen in, een betere schatting te maken van de invloed van Hemelvaart en Pinksteren. Dat de structuren van de voorspellingen van model 3.2 en 3.4 weinig overeenkomen met de structuur van de werkelijke hoeveelheid post blijkt uit de grote fluctuatie van de relatieve afwijkingen van beide modellen. Doordat deze structuren niet overeenkomen met de structuur van de werkelijke hoeveelheid post, kan er geconcludeerd worden dat een optelling van een

zelfde hoeveelheid post, als compensatie van de te hoge verwachte hoeveelheid grote partijen, geen betere voorspellingen zal opleveren dan de voorspellingen van de WAP-groep.

Naar aanleiding van de figuren 3.30 en 3.31 kan de conclusie getrokken worden dat de tijdreeks modellen 3.2 en 3.4 waarbij er rekening gehouden wordt met de kennis van de hoeveelheid voorgemelde grote partijen op dit moment geen model is dat bruikbaar is voor het maken van voorspellingen.

Om de voorspellingen van de WAP-groep, model 3.1 en model 3.3 te analyseren is er in figuur 3.32 gefocust op hun relatieve afwijkingen en is er in tabel 3.5 een overzicht gemaakt van de spreiding van de relatieve afwijkingen van de voorspellingen van de WAP-groep en van de modellen. De bijbehorende cijfers zijn te vinden in bijlage III D.



Figuur 3.32: Vergelijking van de relatieve afwijking van de realisatie t.o.v. de voorspellingen van de WAP-groep, model 3.1 en model 3.3 (de relatieve afwijkingen van de voorspellingen die gemaakt zijn door de WAP-groep worden niet voor week 1 bijgehouden)

Interval	model 3.1		model 3.2		model 3.3		model 3.4		WAP	
	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.
[-2%,2%]	9	19.1%	0	0%	10	21.3%	0	0%	17	36.2%
[-5%,5%]	18	38.3%	0	0%	23	48.9%	0	0%	41	87.2%
[-10%,10%]	36	76.6%	0	0%	41	87.2%	0	0%	47	100%
[-15%,15%]	45	95.7%	0	0%	47	100%	0	0%	47	100%

Tabel 3.7: Analyse van de spreiding van de relatieve afwijkingen

Het gemiddelde van de absolute waarden van de relatieve afwijkingen geeft tevens een indicatie over de geschiktheid van een model voor voorspellingen. Het gemiddelde van de absolute waarden van de relatieve afwijkingen is gedefinieerd als het gemiddelde van:

$$\left| \frac{(\text{realisatie} - \text{voorspelling})}{\text{voorspelling}} \right| * 100\%$$

en wordt aangegeven als Gem*.

	model 3.1	model 3.2	model 3.3	model 3.4	WAP
Gem*	6.5%	44.9%	5.2%	51.1%	3.0%

Tabel 3.8: Gemiddelde van de absolute waarden van de relatieve afwijkingen

Van model 3.2 en 3.4 is er al geconcludeerd dat deze modellen op dit moment niet bruikbaar zijn voor voorspellingen. Dit wordt bevestigd door de cijfers in tabel 3.5 en tabel 3.6.

Aan de hand van figuur 3.32, tabel 3.5 en tabel 3.6 kan er geconcludeerd worden dat model 3.3 beter prognosticeert dan model 3.1. Vooral in de weken tussen Pasen, Hemelvaartsdag en Pinksteren wordt er door model 3.3 beter geprognosticeerd. Echter de voorspellingen van de WAP-groep zijn beter dan de voorspellingen van alle modellen. Omdat de voorspellingen van de tijdreeks modellen ook niet voldoen aan de TPG-wens om prognoses te verwezenlijken met een maximale relatieve afwijking van 2%, zullen deze modellen op dit moment niet gebruikt worden voor voorspellingen. Omdat er bij de modellen 3.1 en 3.3 data beschikbaar is vanaf het jaar 1990, verwacht ik geen grote verbeteringen in de voorspellingen met meer data. Voor de modellen 3.2 en 3.4 is er op dit moment weinig data beschikbaar en zijn er slechte voorspellingen. Met meer data is er grote kans dat de voorspellingen verbeteren. Of er met deze modellen een maximale relatieve afwijking van 2% wordt bereikt, kan op dit moment niet gezegd worden.

Voorspelling intervallen

De realisaties vallen alleen in het voorspellingsinterval¹⁸ als de voorspellingen een relatieve afwijking hebben die kleiner is dan 0,5%. Bij model 3.1 en 3.3 zijn er hierdoor erg weinig realisaties die in het voorspellingsinterval zitten. Bij model 3.2 en 3.4 zijn er geen realisaties die in het voorspellingsinterval zitten. De voorspellingsintervallen geven dus geen goede indicatie voor de te verwachten hoeveelheid post. Dit wordt veroorzaakt door de relatief kleine intervallen.

¹⁸ zie bijlage D

3.6. Conclusie

Voor dit onderzoek is het doel gesteld om een model te ontwikkelen die het landelijke weekverkeer voorspelt. Een model heeft een meerwaarde voor TPG als de huidige maximale afwijking van 10% afneemt naar een maximale afwijking van 2%.

Tijdens het ontwikkelen van een model, worden de waarden van de variabelen die bij het model horen geschat. De grootte van de invloed van de seizoensinvloeden kan het beste met behulp van model 3.1 en 3.3 bepaald worden, omdat bij deze modellen er meer data beschikbaar is. Hierbij moet er echter wel vermeld worden dat dit een indicatie is van de grootte van de invloed. De juistheid van de schattingen van de seizoensinvloeden kan namelijk betwist worden, aangezien de voorspellingen van de tijdreeks modellen niet beter zijn dan de voorspellingen van de WAP-groep. Naar aanleiding van de analyse van de fitting kan er echter wel geconcludeerd worden dat het percentage relatieve afwijkingen dat in het interval $[-2\%, 2\%]$ bij de tijdreeks modellen groter is. Als het log-lineair model (hoofdstuk 4) en/of het Poisson-model een zelfde indicatie geven voor de seizoensinvloeden dan kan er aangenomen worden dat er een goede schatting gemaakt is van de invloed, maar dat de afwijkingen zijn in de hoeveelheid post in het jaar 2003 niet zijn opgevangen met de tijdreeks modellen. Dit wordt tevens ondersteund door de duidelijke afname van grote afwijkingen ten opzichte van de trendlijn indien de geschatte seizoensinvloeden verwijderd worden. De vergelijking van de geschatte invloeden van de tijdreeks modellen, het log-lineair model en de Poisson-modellen zal in hoofdstuk 6 beschreven worden.

Doordat de voorspellingen van tijdreeks modellen geen maximale afwijking hebben van 2%, hebben de tijdreeks modellen op dit moment geen meerwaarde voor TPG. Deze doelstelling wordt bij de modellen 3.2 en 3.4 niet gehaald door het tekort aan data. Dat de modellen 3.1 en 3.3 deze doelstelling niet behalen, wordt waarschijnlijk veroorzaakt doordat de tijdreeks modellen “onwetend” zijn over de daling van de hoeveelheid post in het jaar 2003. Deze daling is dan ook niet terug te vinden bij de bepaalde trend. Of dit een beperking van het tijdreeks model is, kan aan de hand van een case niet geconcludeerd worden. Bij deze case, het voorspellen van de hoeveelheid post per week, kan het ‘missen’ van de daling van de hoeveelheid post in het jaar 2003 namelijk veroorzaakt worden door het gegeven dat de daling per week relatief niet groot is en pas begint na het eerste kwartaal van 2003.

4. Log-lineair model & Log-quasi-likelihood model

4.1. Introductie

Gegeneraliseerde lineaire modellen

Het log-lineair model is een voorbeeld van een gegeneraliseerd lineair model. De gegeneraliseerde lineaire modellen vloeien voort uit het klassieke lineaire regressiemodel.

De meest gebruikte gegeneraliseerde lineaire modellen zijn het:

- logistisch regressie model
- probit regressie model
- log-lineair model

Het verschil tussen het klassieke lineaire regressiemodel en de gegeneraliseerde lineaire modellen is dat de gegeneraliseerde lineaire modellen andere verdelingen dan de normale verdeling voor de responsvariabele toestaan.

Een ander verschil is dat het schaal-probleem¹⁹ van het klassieke lineaire regressiemodel vermeden wordt met behulp van de linkfunctie. De linkfunctie specificeert het verband tussen de voorspeller η_i en de verwachting van de responsvariabele Y_i . Hierbij is de zo genoemde voorspeller η_i gelijk aan het inproduct van de regressie waarnemingen x_i en de regressie coëfficiënten β . De linkfunctie kan ervoor zorgen dat de geschatte verwachting van de responsvariabele niet negatief is, of tussen de nul en een ligt. In tabel 4.1 is een kort overzicht van de 4 genoemde modellen weergegeven.

Model	Klassieke lin	Logist. regr.	Probit regr.	Log-lineair
Verdeling responsvariabele Y_i	$Y_i \sim N(\mu_i, \sigma^2)$	$n_i Y_i \sim \text{Bin}(n_i, \mu_i)$	$n_i Y_i \sim \text{Bin}(n_i, \mu_i)$	$Y_i \sim \text{Poisson}(\mu_i)$
Linkfunctie	$g(\mu_i) = \mu_i$ $= \eta_i = x_i^T \beta$	$g(\mu_i) = \log[\mu_i / (1 - \mu_i)]$ $= \eta_i = x_i^T \beta$ (logit functie)	$g(\mu_i) = \Phi^{-1}(\mu_i)$ $= \eta_i = x_i^T \beta$ (probit functie)	$g(\mu_i) = \log(\mu_i)$ $= \eta_i = x_i^T \beta$ (log functie)
Schaal responsvariabele Y_i	$(-\infty, \infty)$	$(0, 1)$	$(0, 1)$	$(0, \infty)$

Tabel 4.1: Overzicht van het klassieke lineaire model, het logistische regressie model, het probit regressie model en het log-lineair model.

Log-lineair model

De verdeling die het meest in aanmerking komt voor het model waarbij de responsvariabele Y_i gelijk is aan het aantal poststukken week i , is de Poisson-verdeling. Een Poisson proces is een telproces die het aantal momenten, dat een specifieke gebeurtenis plaatsvindt gedurende een bepaalde periode, telt. Deze momenten worden gekenmerkt door een hoge mate van willekeur.

Eigenschappen van een Poisson-proces zijn:

1. het aantal momenten waarop een specifieke gebeurtenis plaatsvindt in een periode T_k , is onafhankelijk van het aantal momenten dat dezelfde specifieke gebeurtenis plaatsvindt in een disjuncte periode T_l
2. het aantal momenten dat een specifieke gebeurtenis plaatsvindt in een tijdsinterval is Poisson verdeeld met een verwachting die evenredig is met de lengte van het tijdsinterval (periode).

De relatie van het Poisson proces met de hoeveelheid poststukken per week, zit hem in de momenten dat een poststuk gecreëerd wordt. Met behulp van de eigenschappen van een Poisson-proces kan er aangenomen worden dat:

- het aantal poststukken dat wordt aangeboden in week k onafhankelijk is van het aantal poststukken dat wordt aangeboden in week l . Hierbij is l niet gelijk aan k . Dit is overeenkomstig met mijn verwachtingen naar aanleiding van een eerste analyse van de data. Hieruit blijkt dat het aantal poststukken in de ene week veel meer kan zijn dan in een andere week.

¹⁹ de schaal van de responswaarden is onbeperkt

- het aantal poststukken dat aangeboden wordt afhankelijk is van de lengte van het tijdsinterval (periode). Dit komt ook overeen met mijn verwachtingen naar aanleiding van een eerste analyse van de data en de interviews. Hieruit blijkt dat er verwacht wordt dat het aantal poststukken vermindert, indien er een werkdag minder is in een bepaalde week door bijvoorbeeld een officiële feestdag.

De linkfunctie van het log-lineaire model, het gegeneraliseerde lineaire model waarbij de responsvariabele een Poisson-verdeling heeft, zorgt ervoor dat de verwachting van de responsvariabele Y_i niet negatief is. Er hoeft echter geen rekening gehouden te worden met het schaalprobleem, doordat het aantal poststukken per week in de jaren 1990 tot en met 2003 tussen de 54,7 en de 113,5 miljoen zit. Door deze grote getallen is de kans dat er bij bijvoorbeeld het klassieke lineaire model een negatieve verwachting van de responsvariabele Y_i bepaald wordt erg klein.

Log-quasi-likelihood model

De verschillen tussen het log-lineair model en een log-quasi likelihood model worden veroorzaakt door een verschil in de waarde van de dispersion parameter ϕ . De dispersion parameter ϕ is gelijk aan:

$$\phi = \frac{\text{variantie van } Y_i}{\text{verwachting van } Y_i} = \frac{\text{var}(Y_i)}{E(Y_i)} = \frac{\text{var}(Y_i)}{\mu_i}$$

De verwachting van een stochastische variabele Y_i met een Poisson-verdeling is gelijk aan de variantie van Y_i . Als de aanname dat Y_i Poisson verdeeld is klopt, wordt er verwacht dat de dispersion-parameter ϕ gelijk is aan 1. Indien de dispersion-parameter niet gelijk is aan 1, dan klopt de aanname van de Poisson-verdeling niet en zal er naar een ander model gekeken moeten worden.

Omdat bij het log-quasi-likelihood model geen verdeling verondersteld wordt, is het een mogelijkheid om bij een incorrecte aanname van bijvoorbeeld de Poisson-verdeling uit te wijken naar het log-quasi-likelihood model. Bij een log-quasi-likelihood model kan er aangegeven worden welke link-functie gewenst is en hoe de variantie meebeweegt met de verwachting, Indien er afgeweken wordt van een log-lineair model naar een log-quasi-likelihood model, zal

- de variantie van Y_i gelijk gesteld worden aan een geschatte dispersion-parameter $\hat{\phi}$ maal de verwachting van Y_i . Indien van tevoren bekend is welk mechanisme een dispersion parameter die ongelijk is aan 1 veroorzaakt, zal de variantie van Y_i op een andere manier bepaald moeten worden.
- de link-functie overeenkomstig zijn aan de link-functie van een log-lineair model

Indien van tevoren niet bekend is welk mechanisme een dispersion parameter die ongelijk is aan 1 veroorzaakt, dan zullen de geschatte waarden voor de regressie coëfficiënten β gelijk zijn aan de de geschatte waarden voor de regressie coëfficiënten β bij een log-lineair model. De waarde van de deviance zal gelijk zijn aan de deviance van het log-lineair model gedeeld door de dispersion-parameter.

Opzet hoofdstuk

In paragraaf 4.2 wordt het log-lineair en het log-quasi-likelihood model voor de responswaarden Y_i beschreven. Tevens zal in paragraaf 4.2 de schattingsmethode, die bij het log-lineaire en het log-quasi-likelihood model hoort, beschreven worden. De resultaten van het log-lineair model (die bepaald is met als data de jaren 1990 tot en met 2002) worden in paragraaf 4.3 geanalyseerd. Uit deze analyse blijkt dat de aanname dat Y_i Poisson verdeeld is, niet juist is. In paragraaf 4.4 worden de voorspellingen voor perioden van het jaar 2003 van het log-quasi-likelihood model vergeleken met de WAP-voorspellingen en de gerealiseerde hoeveelheden post. Hierbij wordt er gekeken naar het verschil in de voorspellings-resultaten indien er voor:

- a) een periode van 48 weken voorspeld wordt, oftewel de weken 1 tot en met 48 van 2003
- b) een periode van 8 weken voorspeld wordt, en wel voor de perioden:
 - week 17 tot en met 24 omdat daar grote fluctuaties in de wekelijkse hoeveelheden post door pasen, pinksteren, etc.
 - week 41 tot en met 48 omdat daar weinig fluctuaties in de wekelijkse hoeveelheden post
- c) elke week opnieuw voorspeld wordt. Uit eerdere analyse²⁰ blijkt dat de resultaten van deze voorspellingsmethode voor de weken 9 tot en met 48 vergelijkbaar zijn met de resultaten van de

²⁰ zie bijlage IV.A

voorspellingsmethode waarbij er per 8 weken geprognosticeerd wordt. Deze voorspellingsmethode zal daarom voor de weken 1 tot en met 8 van 2003 worden geanalyseerd.

4.2. Beschrijving model

In onderstaand model is het log-lineair en het log-quasi-likelihood model beschreven:

Model 4.1

- i. $Y_i \sim \text{Poisson}(\mu_i)$ respectievelijk $Y_i \sim \text{Quasi}(\mu_i, \phi\mu_i)$, $i = 1, \dots, n, n+1, \dots, m$

Hierbij is

Y_i	het aantal poststukken in week i.
μ_i	verwachting van het aantal poststukken in week i.
ϕ	gelijk aan de dispersion-parameter
$\mu_i \phi$	variantie van het aantal poststukken in week i.
n	het totaal aantal weken in de jaren 1990 tot en met 2002, zonder de weken waarin een extreme hoeveelheid post gegenereerd is ($n=647$)
m	(het aantal weken waarvoor een voorspelling gegenereerd wordt)- n

- ii. $\eta_i = x_i^T \beta$

$$\begin{aligned}
 &= \beta_0 + \beta_1 \{ \text{Valentijnsdag} \in i \} + \beta_2 \{ \text{Pasen} \in i \} + \beta_3 \{ \text{Hemelvaartsdag} \in i \} + \beta_4 \{ \text{Pinksteren} \in i \} \\
 &+ \beta_5 \{ \text{Koninginnedag} \in i \} + \beta_6 \{ \text{Bevrijdingsdag} \in i \} + \beta_7 \{ \text{Mailmoment} \in i \} \\
 &+ \beta_8 \dots 201 \{ \text{Jaar } 1991 \dots \text{Jaar } 2003^{21} \in i \} + \beta_{21 \dots 70} \{ \text{Week } 2 \dots \text{Week } 50, \text{Week } 52 \in i \} + \beta_{71, 72} \{ \text{Kerst } 1, \text{Kerst } 2 \in i \} \\
 &+ \beta_{73 \dots 78} \{ \text{Vakantie } 1 \dots \text{Vakantie } 6 \in i \} \\
 &+ \beta_{79 \dots 87} \{ \text{Eind jan, eind maart, eind april, eind juni} \dots \text{eind sept, eind nov, eind dec} \in i \}
 \end{aligned}$$

- iii. $\eta_i = g(\mu_i) = \log(\mu_i)$

De regressie waarnemingen x_i zijn bepaald naar aanleiding van de te verwachten invloedsfactoren op de hoeveelheid post per week. Een toelichting waarom er bijvoorbeeld gekozen is om deze regressie waarnemingen op te nemen in de data is te vinden in bijlage IV.B. Tevens is in deze bijlage een voorbeeld en een validatie van de regressie waarnemingen x_i te vinden in bijlage IV.B.

Sinds 2001 is er wekelijks bijgehouden hoeveel poststukken de voorgemelde grote partijen genereren. De door de voorgemelde grote partijen gegenereerde hoeveelheden poststukken bedraagt tussen de 4,7 miljoen en de 26,7 miljoen in de jaren 2001 tot en met 2003. Door deze hoeveelheden van het aantal poststukken per week af te trekken, blijft er een kleiner gedeelte over dat voorspeld moet worden. Dit zou de voorspellingen normaliter verbeteren, echter de hoeveelheid beschikbare data neemt af. Door tevens onderstaand log-lineair en log-quasi-likelihood model te onderzoeken, kan er geanalyseerd worden of het model verbeterd indien er rekening gehouden wordt met de invloed van de voorgemelde grote partijen.

Model 4.2

- i. $Y_i \sim \text{Poisson}(\mu_i)$ respectievelijk $Y_i \sim \text{Quasi}(\mu_i \phi \mu_i, \mu_i)$, $i = 1, \dots, n, n+1, \dots, m$

Hierbij is

Y_i	het aantal poststukken in week i min de voorgemelde grote partijen in deze week.
μ_i	verwachting van het aantal poststukken in week i min de voorgemelde grote partijen in deze week.
ϕ	gelijk aan de dispersion-parameter
$\mu_i \phi$	variantie van het aantal poststukken in week i.
n	het totaal aantal weken in de jaren 2001 tot en met 2002, zonder de weken waarin een extreme hoeveelheid post gegenereerd is ($n=96$)
m	(het aantal weken waarvoor een voorspelling gegenereerd wordt)- n

- ii. $\eta_i = x_i^T \beta$

$$\begin{aligned}
 &= \beta_0 + \beta_1 \{ \text{Pasen} \in i \} + \beta_2 \{ \text{Hemelvaartsdag} \in i \} + \beta_3 \{ \text{Pinksteren} \in i \} \\
 &+ \beta_{4,5} \{ \text{Jaar } 2002, \text{Jaar } 2003^{22} \in i \} + \beta_{6 \dots 52} \{ \text{Week } 2 \dots \text{Week } 48 \in i \} +
 \end{aligned}$$

²¹ het jaar 2003 wordt alleen bij het maken van voorspellingen voor het jaar 2003 als regressie waarneming meegenomen in het model.

²² het jaar 2003 wordt alleen bij het maken van voorspellingen voor het jaar 2003 als regressie waarneming meegenomen in het model.

$$\begin{aligned}
 & +\beta_{53,54}1_{\{\text{Vakantie 1, vakantie 2} \in i\}} \\
 & +\beta_{55}1_{\{\text{eind sept} \in i\}}
 \end{aligned}$$

$$\text{iii. } \eta_i = g(\mu_i) = \log(\mu_i)$$

Schattingsmethode

Om te bepalen wat de verwachting van Y_i is, wordt de meest aannemelijke schatter bepaald van de regressie coëfficiënten β . Dit omdat:

- van tevoren is bepaald in welke weken en in welke jaren een waarneming x_i de waarde 1 heeft. De x_i is dus niet stochastisch.
- de verwachting van Y_i is gelijk aan μ_i , welke weer gelijk is aan de exponentieel van η_i . Daar de waarde van η_i alleen verandert als de waarde van β verandert, is de verwachting van Y_i direct afhankelijk van β .

De meest aannemelijke schatter van β wordt verkregen via het oplossen van de vergelijking:

$$\delta l(\beta, Y) / \delta l(\beta) = 0$$

De log likelihood ($=l(\beta, Y)$) is bij de modellen gelijk aan

$$\sum_{i=1}^n (y_i (x_i^T \beta) - e^{x_i^T \beta}) = \sum_{i=1}^n (y_i \log(\mu_i) - \mu_i)$$

Bij het oplossen van de vergelijking wordt er gebruik gemaakt van een iteratieve gewogen kleinste kwadraten methode²³.

Nadat de regressie coëfficiënten β geschat zijn, kan $\hat{\mu}_i$ (de verwachting van Y_i) bepaald worden via de linkfunctie.

De dispersion-parameter kan als volgt geschat worden:

$$\hat{\phi} = \frac{\sum_{i=1}^{n-p} \left[\frac{(y_i - \hat{\mu}_i)^2}{\mu_i} \right]}{n-p}$$

Hierbij is

p gelijk aan het aantal regressie coëfficiënten β .

²³ Zie paragraaf 4.4 van 'An introduction to Generalized Linear Models' (geschreven door Annette J. Dobson, © 1990, drukkerij Chapman and Hall)

4.3. Analyse fitting

Het bepalen van hoe goed het model fit, wordt over het algemeen gedaan via het analyseren van de residuen. De residuen zijn bepaald met behulp van de gerealiseerde waarden $Y_i=y_i$, de verwachte waarden μ_i , en de dispersion parameter ϕ . In deze paragraaf zullen de volgende residuen geanalyseerd worden:

- response residuen
- deviance residuen
- pearson residuen

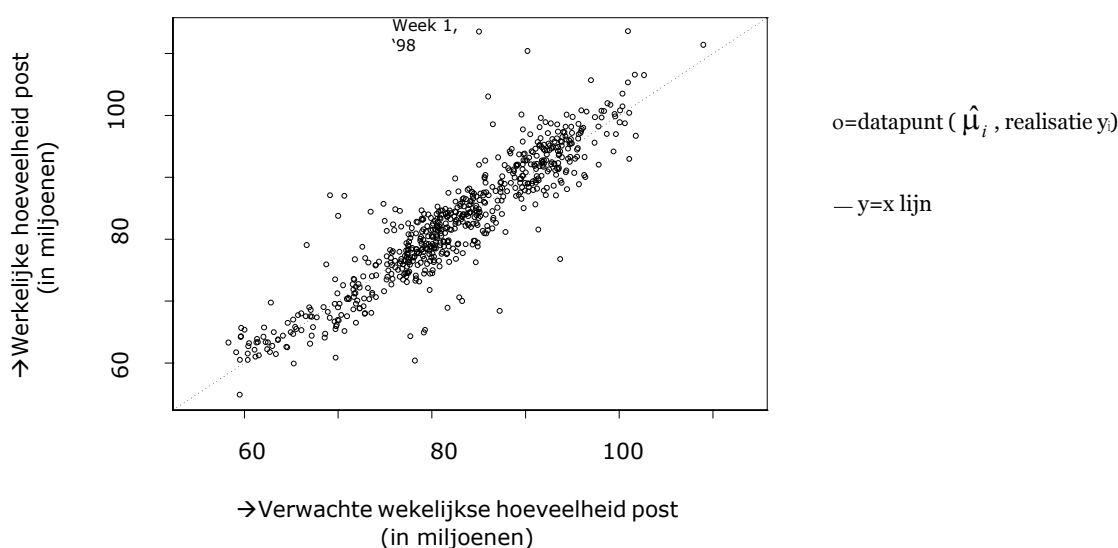
4.3.1. Response residuen

De response residuen zijn gelijk aan het verschil tussen de werkelijke hoeveelheid post in week i en de verwachte post μ_i in week i , oftewel $y_i - \hat{\mu}_i$. Indien de werkelijke hoeveelheid post de verwachte post benadert, zal in een grafiek waar deze waarden tegen elkaar uitgezet zijn de datapunten sterk gecentreerd zijn rond de lijn $y=x$. Dit geeft de indicatie dat er sprake is van een goede fitting. Indien er enkele waarden afwijken van de lijn, dan kan er sprake zijn van uitbijters. Uitbijters zijn waarnemingen die lijken af te wijken van de rest van de waarnemingen.

De response residuen van een log-lineair model zijn gelijk aan de response residuen van een log-quasiliikelihood model, doordat $\hat{\mu}_i$ hetzelfde is bij beide modellen. De dispersion parameter ϕ die wel per model verschilt. Het verschil tussen de modellen zijn.

Model 4.1

Indien de werkelijke hoeveelheden post Y_i afgezet worden tegen de verwachte hoeveelheden post (waarbij er geen rekening wordt gehouden met de voorgemelde grote partijen) wordt de volgende figuur verkregen:



Figuur 4.1: de verwachte²⁴ response Y_i

t.o.v. de realisaties

In figuur 4.1 is te zien dat het grootste deel van de datapunten niet sterk gecentreerd zijn rond de lijn $y=x$, wat betekent dat er toch nog een redelijk verschil is tussen de werkelijke hoeveelheid wekelijkse post en de verwachte wekelijkse hoeveelheid post. Dit is tevens de conclusie als er naar de cijfers van de relatieve afwijkingen gekeken wordt. De relatieve afwijking is gedefinieerd als:

²⁴ verwachting bepaald met als data de jaren 1990 tot en met 2002. Waar de responswaarde Y_i gelijk is aan het aantal poststukken in week i , is de data van de jaren 1990 tot en met 2002 zonder de specifieke extreme waarden in het model meegenomen.

$$\frac{(y - \hat{\mu})}{(\hat{\mu})} * 100\% = \frac{\text{response residu}}{(\hat{\mu})} * 100\%$$

De relatieve afwijkingen zijn weergegeven in bijlage IV.C(deel I en II) en kort samengevat in onderstaand tabel.

Interval	model 4.1	
	Aantal	Perc.
[-2%,2%]	260	40.2%
[-5%,5%]	510	78.8%
[-10%,10%]	617	95.3%
[-15%,15%]	631	97.5%
[-20%,20%]	641	99.1%

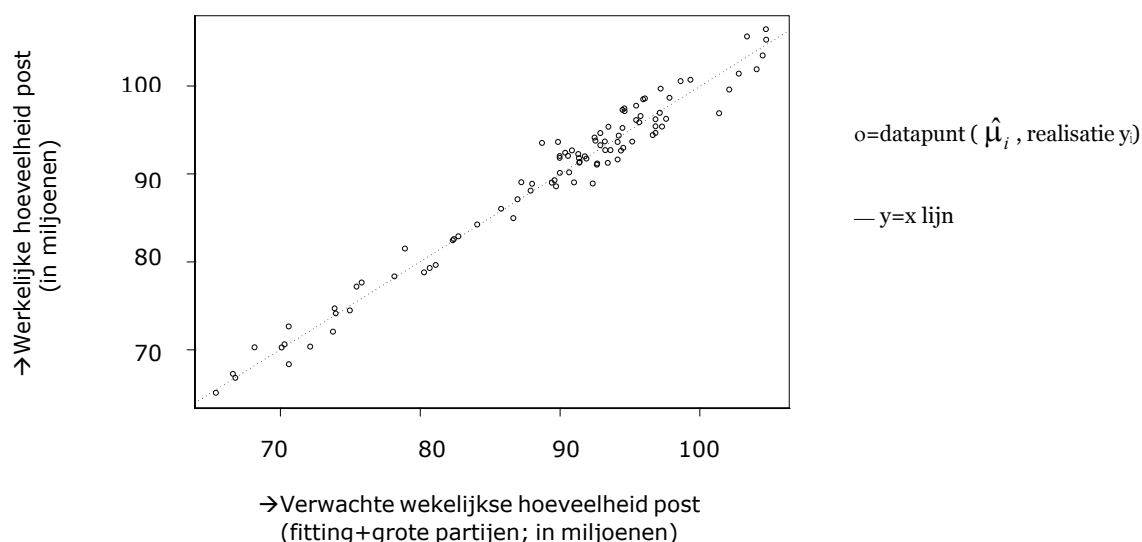
Tabel 4.2: Analyse van de relatieve afwijkingen van model 4.1

Het maximum van de absolute waarde van de relatieve afwijkingen is gelijk aan 33.1%. Deze grote afwijking is waargenomen bij week 1 van het jaar 1998. Het gemiddelde van de absolute waarde van de relatieve afwijkingen is gelijk aan 3,4%

Uit de cijfers van de response residuen blijkt dat maar 40,2% procent van de bepaalde verwachtingen een maximale relatieve afwijking heeft van 2%. Daar een wiskundig model voor TPG Post alleen een meerwaarde heeft indien alle prognoses worden gegeven met een maximale afwijking van 2%, geeft dit een indicatie dat het log-lineair model (waarbij er geen rekening wordt gehouden met de voorgemelde grote partijen) niet gaat voldoen aan deze eis.

Model 4.2

Indien de werkelijke hoeveelheden post afgezet worden tegen de verwachte hoeveelheden post waarbij er wel rekening gehouden wordt met de voorgemelde grote partijen, dan wordt het volgende figuur verkregen:



Figuur 4.2: de verwachte response Y_i + de voorgemelde grote partijen t.o.v. de realisaties

In figuur 4.2 is te zien dat de datapunten sterker gecentreerd zijn rond de lijn $y=x$ dan bij model 4.1. Dit geeft de indicatie dat er een betere fitting is bij model 4.2. Dit is tevens de conclusie als er naar de cijfers van de relatieve afwijkingen gekeken wordt. De relatieve afwijking bij model 4.2 is gedefinieerd als:

$$\frac{((y + G) - (\hat{\mu} + G))}{(\hat{\mu} + G)} * 100\% = \frac{(y - \hat{\mu})}{(\hat{\mu} + G)} * 100\% = \frac{\text{response residu}}{(\hat{\mu} + G)} * 100\%$$

Deze relatieve afwijkingen zijn weergegeven in bijlage IV.C(deel I en II) en kort samengevat in onderstaand tabel.

Interval	model 4.2	
	Aantal	Perc.
[-2%,2%]	63	65.6%
[-5%,5%]	95	99.0%
[-10%,10%]	96	100 %

Tabel 4.2: Analyse van de relatieve afwijkingen van model 4.2

Het maximum van de absolute waarde van de relatieve afwijkingen is gelijk aan 5,2%. Deze afwijking is waargenomen bij week 9 van het jaar 2001. Het gemiddelde van de absolute waarde van de relatieve afwijkingen is gelijk aan 1,7%. De kleine verschillen tussen de verwachte en de werkelijke wekelijkse hoeveelheid post, is te verklaren door het feit dat met 58 parameters 96 waarnemingen gefit worden.

4.3.2. Deviance residuen

De mate van fitting kan tevens bepaald worden via een analyse van deviance residuen. Deviance residuen kunnen een indicatie geven dat sommige observaties een te grote invloed hebben bij het schatten van de regressie coëfficiënten β . De deviance residuen worden bepaald door:

$$r_i^D = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{2(y_i \ln(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i))}$$

waarbij d_i de contibutie voor de deviance is van de responswaarde Y_i . De sommatie van d_i over i is dan ook gelijk aan de deviance $D(\mu, y)$.

In formulevorm is de relatie tussen de deviance $D(\mu, y)$ en de residuen deviance residuen als volgt gedefinieerd:

$$D(\mu, y) = \sum_{i=1}^n i(r_i^D)^2$$

De deviance $D(\mu, y)$ meet het verschil tussen het model 4.1 respectievelijk het model 4.2 en het model waarbij het aantal variabelen gelijk is aan het aantal waarnemingen. Bij dit laatste model zal de fitting logischerwijs perfect zijn.

De deviance $D(\mu, y)$ is gelijk aan

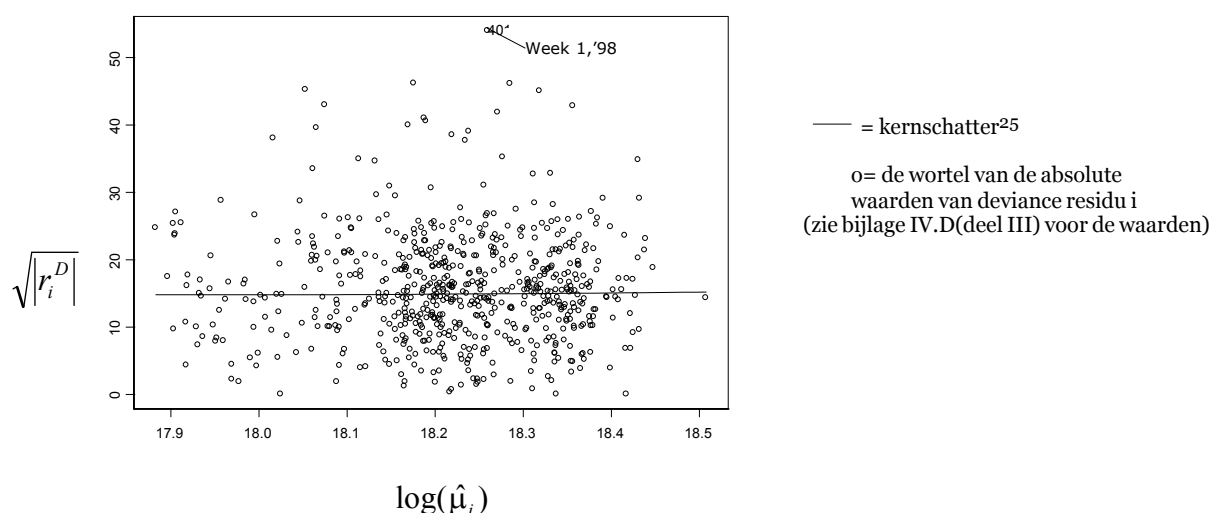
$$\frac{D}{\phi} = 2[l(\tilde{\beta}, y) - l(\hat{\beta}, y)]$$

Hierbij is

- ϕ gelijk aan de dispersion-parameter
- $l(\tilde{\beta}, y)$ gelijk aan de meest aannemelijke functie van het model waarbij het aantal variabelen gelijk is aan het aantal waarnemingen
- $l(\hat{\beta}, y)$ gelijk aan de meest aannemelijke functie van model 4.1 respectievelijk model 4.2

Indien de wortel van de absolute waarde van de deviance residu r_i een relatief grote waarde heeft, dan geeft dit een indicatie dat er sprake is van een uitbijter.

De deviance residuen zullen grafisch weergegeven voor de log-lineaire modellen. Omdat de waarde van de deviance van het log-quasi-likelihood model gelijk is aan de deviance van het log-lineair model gedeeld door de dispersion-parameter van het log-quasi-likelihood model, zullen de verhoudingen tussen de deviance redsiduen gelijk blijven. Bij de grafieken van de log-quasi-likelihood modellen zullen daarom dezelfde uitbijters waarneembaar zijn. De grafieken van de deviance residuen van de log-quasi-likelihood model zijn daarom overbodig en worden dan ook niet in deze paragraaf weergegeven.

Log-lineair model 4.1

Figuur 4.3: de wortel van de absolute waarde van de deviance residuen t.o.v. de logaritme's van de verwachte²⁶ responsewaarde Y_i

Bovenstaande figuur geeft de indicatie dat de responswaarde van week 1 van het jaar 1998 (punt 401) een uitbijter is, oftewel een waarnemingspunt waarbij de respons niet lijkt te corresponderen met het model. Het model lijkt echter voor een groot gedeelte van de punten wel te passen. Na een analyse van de regressie waarnemingen x_i blijkt dat de β van het jaar 1998 en/of nieuwjaarsdag te hoog kan zijn geschat vanwege deze uitbijter. De relatieve afwijkingen geven deze indicatie alleen voor nieuwjaarsdag. Indien namelijk de relatieve afwijkingen van de weken waarin nieuwjaarsdag²⁷ valt gesommeerd worden, resulteert dit in een percentage van 8,1%. Dit percentage kan een indicatie geven dat de weken met nieuwjaarsdag te laag geschat worden. Gezien de grote fluctuatie in de weken waarin nieuwjaarsdag valt, zal echter het gemiddelde relatieve verschil van 0.73% maar een summiere verbetering bij het fitting-proces opleveren.

Uit de toets voor uitbijters beschreven in bijlage IV.F, wordt bevestigd dat week 1 van het jaar 1998 een uitbijter is. Doordat de uitbijter geen meetfout is, is dit een indicatie dat een ander model eventueel beter is.

Log-lineair model 4.2

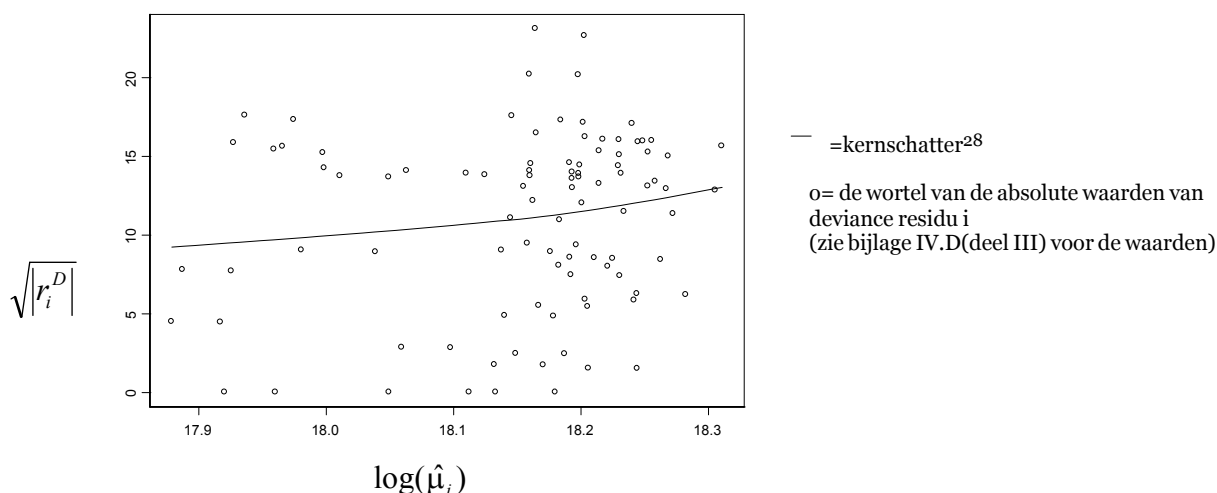
Uit figuur 4.4, waar de wortel van de absolute waarden van de deviance residuen van het log-lineair model 4.2 zijn weergegeven, blijkt dat de mate van fitting beter is dan bij het log-lineair model 4.1.

Het maximum van de wortel van $\sqrt{|r_i^D|}$ is namelijk een stuk kleiner dan bij het vorige log-lineaire model. De betere fitting kan verklaard worden door het feit dat bij model 4.2 96 waarnemingen met 54 parameters gefit wordt. Bij model 4.1 zijn er 86 parameters die moeten zorgen voor een goede fitting van 647 waarnemingen. De verhouding tussen het aantal waarnemingen en het aantal parameters verschilt dus aanzienlijk.

²⁵ Zie bijlage IV.H

²⁶ verwachting bepaald met als data de jaren 1990 tot en met 2002. Hierbij is er bij de jaren 2001 en 2002 geen rekening gehouden met de voor die jaren beschikbare hoeveelheden post die door voorgemelde grote partijen gegenereerd werden.

²⁷ Nieuwjaarsdag valt in de volgende weken: week 1 van de jaren 1990 tot en met 1994, en de weken 1 van de jaren 1996 tot en met 2002. In het jaar 1995 valt nieuwjaarsdag namelijk op een zondag en is er dus geen officiële vrije dag, waardoor de invloed hiervan op de wekelijkse hoeveelheid post wegvalt. Zie bijlage IV.D(deel I) voor de bijbehorende cijfers.



Figuur 4.4: de wortel van de absolute waarde van de deviance residuen t.o.v. de logaritme's van de verwachte²⁹ responsenwaarde Y_i

Een andere waarneming naar aanleiding van figuur 4.4 is de stijgende lijn. De stijgende lijn geeft de indicatie dat een hogere verwachte responswaarde een grotere absolute waarde van het deviance residu genereert. Echter deze indicatie wordt al weerlegd in figuur 4.3. Doordat de data die gebruikt is in figuur 4.3 betrouwbaarder is (doordat de steekproef groter is), kan er geen waarde gehecht worden aan de indicatie van figuur 4.4.

4.3.3. Pearson residuen

Met behulp van de Pearson residuen kan de aanname dat het aantal poststukken per week Poisson verdeeld is, gecontroleerd worden. Indien namelijk de Pearson residuen standaard normaal verdeeld zijn, dan kan de kwadratische som van de Pearson residuen (de Pearson chikwadraat toetsingsgrootheid) vergeleken worden met de kwantilen van de chikwadraat-verdeling met $(n-p-1)$ vrijheidsgraden. Naar aanleiding van deze vergelijking kan er geconcludeerd worden of de nulhypothese dat de Poisson-verdeling de echte verdelingsfunctie is van de responswaarde Y_i wel of niet verworpen moet worden.

De Pearson residuen zijn als volgt gedefinieerd:

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

Hierbij is

$V(\hat{\mu}_i)$ gelijk aan de verwachte variantie van het aantal poststukken per week.

De waarden van de Pearson residuen zijn weergegeven in bijlage IV.C(deel III en deel IV). Om te analyseren of de Pearson residuen standaard normaal verdeeld zijn, zijn in figuur 4.5 de residuen tegen de kwantilen van de standaard normale verdeling uitgezet. Indien de Pearson residuen standaard normaal verdeeld zijn, zal

- het bereik van de Pearson residuen overeenkomen met het bereik van de standaard normale verdeling
- er een rechte lijn waarneembaar zijn.

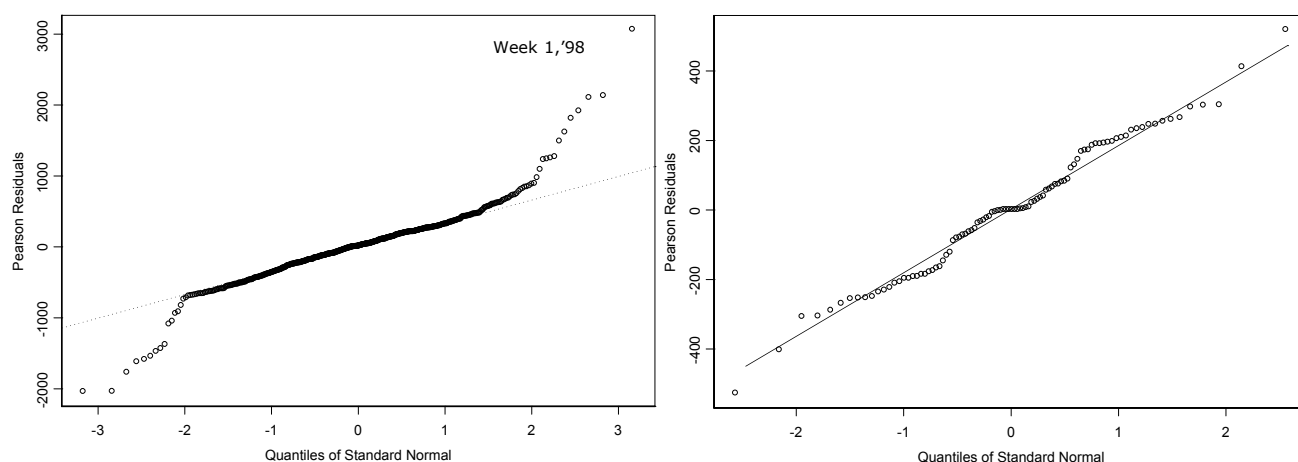
Log-lineair model

De aanname dat de Pearson residuen standaard normaal verdeeld zijn, wordt door figuur 4.5 weerlegd. Het bereik van de Pearson residuen van beide log-lineaire modellen is namelijk veel groter dan bij een standaard normale verdeling verwacht. Tevens is er geen sprake van een rechte lijn. Dat de Pearson

²⁸ Zie bijlage IV.H

²⁹ verwachting bepaald met als data de jaren 2001 en 2002. Hierbij is er rekening gehouden met de voor die jaren beschikbare hoeveelheden post die door voorgemelde grote partijen gegenereerd werden.

residuen niet standaard normaal verdeeld zijn, geeft al de indicatie dat de Poisson-aanname niet juist is. Met de Pearson-chikwadraat toets kan bekeken worden of deze indicatie bevestigd wordt door voorlopig toch aan te nemen dat de Pearson residuen standaard normaal verdeeld zijn.



Figuur 4.5: de Pearson residuen uitgezet tegen de kwantilen van de standaard normale verdeling (in de linker figuur zijn de Pearson residuen van het log-lineair model 4.1 weergegeven; in de rechter figuur zijn de Pearson residuen van het log-lineair model 4.2 weergegeven)

De Pearson chikwadraat toetsingsgrootheid, oftewel de kwadratische som van de Pearson residuen van het model gebruikt voor de linkerfiguur respectievelijk het model gebruikt voor de rechterfiguur, is gelijk aan:

$$\chi^2_{Pearson} = \sum_{i=1}^{n-p-1} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} = 121,2 \text{ respectievelijk } 1,3 \text{ miljoen}$$

Hierbij is

p gelijk aan het aantal parameters.
(Bij het log-lineair model 4.1 is p gelijk aan 86³⁰.
Bij het log-lineair model 4.2 is p gelijk aan 54.)

Onder de voorlopige aanname dat de Pearson residuen bij benadering standaard normaal verdeeld, is de Pearson chikwadraat toetsingsgrootheid χ^2_{n-p-1} verdeeld. De nullhypothese dat de Poisson-verdeling de echte verdelingsfunctie is van de responswaarde Y_i wordt verworpen als:

$$\chi^2_{Pearson} \geq \chi^2_{n-p-1, \alpha}$$

Hierbij is

α het significantie-niveau welke gelijkgesteld is aan 0.05.

Daar 121.1 respectievelijk 1,3 miljoen groter is dan $\chi^2_{647-86-1, 0.05} = 506.1$ respectievelijk

$\chi^2_{96-59-1, 0.05} = 23.3$ wordt de nullhypothese voor beide log-lineaire modellen verworpen en kan er aangenomen worden dat de responswaarde Y_i geen Poisson-verdeling bezit.

Log-quasi-likelihood model

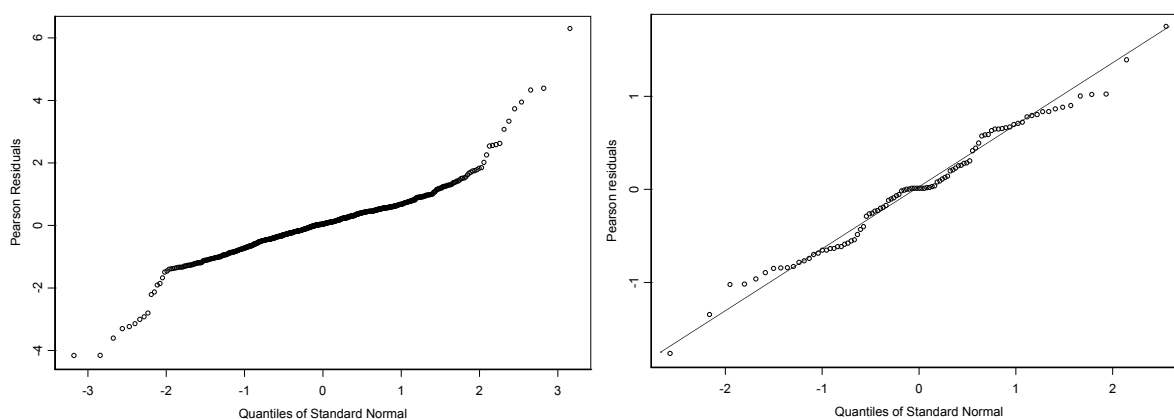
Indien de dispersion-parameter ongelijk is aan 1 dan wordt de aanname, dat de responswaarde Y_i geen Poisson-verdeling bezit, bevestigd. De dispersion-parameter is namelijk gelijk aan de geschatte variantie van Y_i gedeeld door de verwachting van Y_i . Doordat de verwachting van Y_i bij een Poisson-verdeling gelijk is aan de variantie van Y_i , wordt er bij een Poisson verdeling verwacht dat de

³⁰ Zie paragraaf 5.2 'Beschrijving model' (ii). Het jaar 2003 wordt alleen bij het maken van voorspellingen voor het jaar 2003 als regressie waarneming meegenomen in het model.

dispersion-parameter gelijk is aan 1. Indien Y_i dus niet Poisson verdeeld is, dan wordt er een dispersion-parameter verwacht die ongelijk is aan een.

De dispersion-parameter bij model 4.1 respectievelijk model 4.2 is gelijk aan 238.065 respectievelijk 88.622. Dit bevestigt de indicatie die gemaakt is naar aanleiding van de analyse van de Pearson residuen, namelijk de indicatie dat de responsvariabele Y_i bij beide modellen niet Poisson verdeeld is. Omdat bij het quasi-likelihood model geen verdeling verondersteld wordt, is het een mogelijkheid om bij een incorrecte aanname van bijvoorbeeld de Poisson-verdeling uit te wijken naar het quasi-likelihood model. In plaats van het veronderstellen van een verdeling, geeft een log-quasi-likelihood model aan via de dispersion-parameter hoe de variantie met de verwachting verandert. Bij de analyse van de Pearson residuen zal er daarom niet meer gekeken worden of de aanname van de verdeling goed is, maar zal er gekeken worden of de verwachting en de variantie van Y_i bij dit model een goede verhouding hebben. Hierbij is de variantie van Y_i gelijk aan de dispersion parameter vermenigvuldigd met de verwachting van Y_i , oftewel

$$\text{var}(Y_i) = \phi E(Y_i) = \phi \mu_i$$



Figuur 4.6: de Pearson residuen uitgezet tegen de kwantilen van de standaard normale verdeling (in de linker figuur zijn de Pearson residuen van het log-quasi-likelihood model 4.1 weergegeven; in de rechter figuur zijn de Pearson residuen van het log-quasi-likelihood model 4.2 weergegeven)

De waarden van de Pearson residuen zijn weergegeven in bijlage IV.I(deel I en deel II). Om te analyseren of de Pearson residuen standaard normaal verdeeld zijn, zijn in bovenstaand figuur de residuen tegen de kwantilen van de standaard normale verdeling uitgezet. De aanname dat de Pearson residuen standaard normaal verdeeld zijn, wordt door figuur 4.6 aan de ene kant bevestigd doordat het bereik van de Pearson residuen het bereik van de standaard normale verdeling benadert. Aan de andere kant wordt de aanname ontkracht, doordat er enkele punten zijn die niet op de rechte lijn zitten terwijl de omvang van de steekproef redelijk groot is. Om deze reden wordt het onderzoek of de variantie en de verwachting van de Y_i goed zijn gecontinueerd met een voorlopige aanname dat de Pearson residuen bij benadering standaard normaal verdeeld zijn.

De Pearson chikwadraat toetsingsgrootheid, oftewel de kwadratische som van de Pearson residuen van het model gebruikt voor de linkerfiguur respectievelijk het model gebruikt voor de rechterfiguur, is gelijk aan:

$$\chi^2_{\text{Pearson}} = \sum_{i=1}^{n-p-1} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} = 509,0 \text{ respectievelijk } 14,5$$

Hierbij is

p gelijk aan het aantal parameters.
(Bij het log-quasi-likelihood model 4.1 is p gelijk aan 86³¹.
Bij het log-quasi-likelihood model 4.2 is p gelijk aan 54.)

³¹ Zie paragraaf 5.2 'Beschrijving model' (ii). Het jaar 2003 wordt alleen bij het maken van voorspellingen voor het jaar 2003 als regressie waarneming meegenomen in het model.

Onder de aanname dat de Pearson residuen bij benadering standaard normaal verdeeld, kan er aangenomen worden dat de Pearson chikwadraat toetsingsgroottheid χ^2_{n-p-1} verdeeld is. De nulhypothese dat de variantie en de verwachting van de responswaarde Y_i van de log-quasi-likelihood model correct zijn wordt verworpen als:

$$\chi^2_{Pearson} \geq \chi^2_{n-p-1, \alpha}$$

Hierbij is

α het significantie-niveau welke gelijkgesteld is aan 0.05.

Daar 509,0 groter is dan $\chi^2_{647-86-1, 0.05} = 506.1$ wordt de nulhypothese voor het log-quasi-likelihood model 4.1 verworpen. Als de uitbijter week 1 van het jaar 1998 echter uit de data wordt verwijderd, dan wordt de nulhypothese niet verworpen. Om deze reden wordt er wel aangenomen dat de variantie en de verwachting van de responswaarde Y_i van het quasi-likelihood model correct is. De nulhypothese wordt niet verworpen bij het log-quasi-likelihood model 4.2. De Pearson chikwadraat toetsingsgroottheid is hier namelijk gelijk aan 14,5 en niet groter dan $\chi^2_{96-59-1, 0.05} = 23.3$.

Voor beide modellen wordt er dus aangenomen dat de variantie en de verwachting van Y_i goed zijn.

Klassiek lineair regressiemodel

Bij het klassiek lineair regressiemodel is er aangenomen dat Y_i normaal ($EY_i = \mu_i$, $\text{var}(Y_i) = \sigma_i^2$) verdeeld is. Indien deze aanname klopt, dan is de verwachting van Y_i onafhankelijk van de variantie van Y_i . Om deze reden is ook dit model onderzocht. Dit onderzoek is beschreven in bijlage IV.G. Omdat hieruit blijkt dat de normale verdelingsfunctie niet de verdelingsfunctie is van de responsvariabelen Y_i , zal dit model niet gebruikt worden voor het maken van voorspellingen.

4.3.4. Conclusie n.a.v. analyse fitting

Naar aanleiding van de analyse van de verschillende residuen, kan er geconcludeerd worden dat het log-lineaire model niet van toepassing is bij het modelleren van de wekelijkse hoeveelheid post. Bij de analyse van de Pearson residuen blijkt namelijk dat de Poisson-aanname incorrect is. Door deze incorrectheid worden tevens de fittings van het klassiek lineair regressiemodel en het log-quasi-likelihood moel geanalyseerd. Doordat de normale verdelingsfunctie niet de verdelingsfunctie is van de responsvariabelen Y_i zal het klassiek lineair regressiemodel niet gebruikt worden, maar wordt het log-quasi-likelihood model gebruikt bij het maken van voorspellingen.

4.4. Voorspellingen van log-quasi-likelihood modellen

In deze paragraaf worden de voorspellingen voor perioden van het jaar 2003 van beide log-quasi-likelihood modellen vergeleken met de WAP-voorspellingen en de gerealiseerde hoeveelheden post. De voorspellingen van de log-quasi-likelihood modellen worden bepaald met behulp van de geschatte regressie coëfficiënt $\hat{\beta}$, welke geschat is volgens de in paragraaf 4.2 beschreven schattingsmethode.

Bij de analyse van de voorspellingen wordt er, naast een vergelijkend onderzoek met de WAP-voorspellingen en de gerealiseerde hoeveelheden post, ook gekeken naar het verschil in de voorspellings-resultaten indien er voor:

- d) een periode van 48 weken voorspeld wordt, oftewel de weken 1 tot en met 48 van 2003.
- e) een periode van 8 weken voorspeld wordt, en wel voor de perioden:
 - week 17 tot en met 24 omdat daar grote fluctuatie is in de wekelijkse hoeveelheden post door pasen, pinksteren, etc.
 - week 41 tot en met 48 omdat daar weinig fluctuatie is in de wekelijkse hoeveelheden post
- c) elke week opnieuw voorspeld wordt. Daar uit eerdere analyse³² blijkt dat de resultaten van deze voorspellingsmethode vergelijkbaar zijn met de resultaten van de voorspellingsmethode waarbij er per 8 weken geprognosticeerd wordt behalve voor de eerste 8 weken, zal deze voorspellingsmethode voor de weken 1 tot en met 8 van 2003 geanalyseerd.

Per voorspellingsperiode zal in onderstaande paragrafen de voorspellingen geanalyseerd worden, om daarna in paragraaf 4.4.4 conclusies te trekken over de voorspellingen en over het verschil tussen de verschillende voorspellingsperioden.

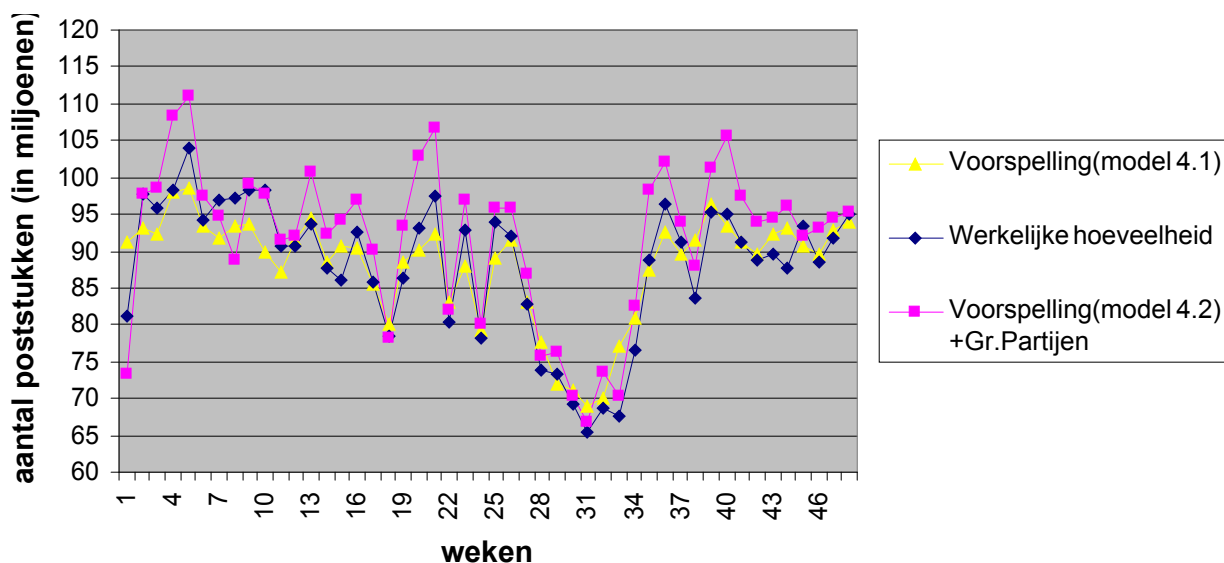
4.4.1. Voorspellingsperiode van 48 weken

Schatting regressiecoëfficiënt $\hat{\beta}$ behorende bij het jaar 2003

Om de hoeveelheden post in de weken 1 tot en met 48 van 2003 te voorspellen, is er eerst een schatting van de regressie coëfficiënt $\hat{\beta}$ voor het jaar 2003 nodig. Deze is namelijk nog niet geschat, doordat de regressie waarnemingen x_i van de data van de jaren 1990 tot en met 2002 niet het jaar 2003 bevatten.

De methode die gebruikt is om de $\hat{\beta}$ voor het jaar 2003 te schatten is de kleinste kwadraten methode, waarbij in dit geval een polynoom bepaald wordt. De kleinste kwadraten schatter wordt in Splus met de functie 'lsfit.poly' bepaald. Deze functie is ook gebruikt om de hoeveelheid post die in het jaar 2003 door grote partijen gegenereerd wordt te voorspellen. Hoe deze functie werkt, en wat de resultaten zijn, is beschreven in bijlage IV.D.

³² zie bijlage IV.B



Figuur 4.6: Vergelijking van de geprognosticeerde met de werkelijke hoeveelheid post in week 1 t/m 48 2003

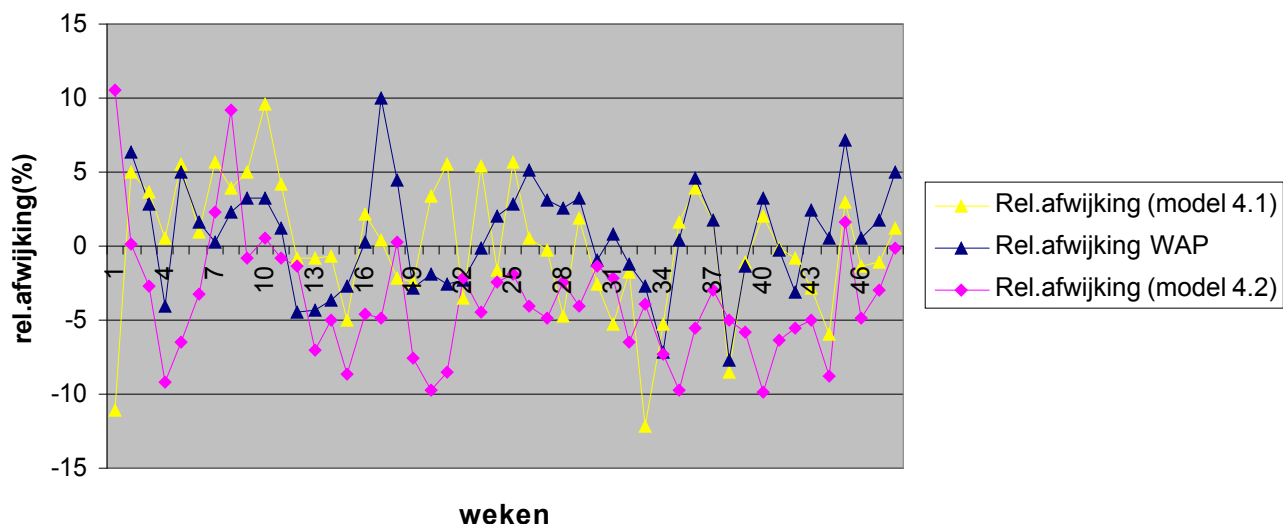
Prognoses t.o.v. werkelijke hoeveelheid

In figuur 4.6 zijn de wekelijkse hoeveelheid post en de geschatte wekelijkse hoeveelheden post weergegeven. Uit dit figuur blijkt dat de geschatte wekelijkse hoeveelheden post de werkelijke hoeveelheid post redelijk goed benaderen. Echter in de weken 5,7,8,9,10 en in de weken 20 21 is een duidelijke te lage voorspelling gegeven. In de weken 1,33 en 38 daar en tegen is er een te hoge schatting gegeven.

Van week 1 is in de cijfers van de fitting die weergegeven zijn in bijlage IV.D(deel I) te zien dat de fluctuatie van de relatieve afwijking erg groot is, waardoor de conclusie getrokken kan worden dat het voorspellen van week 1 moeilijk is. Met het log-quasi-likelihood model wordt in ieder geval niet het gewenste resultaat behaald.

Het verschil in de weken 5, 20 en 21 kan veroorzaakt worden door het grote aantal grote partijen in die weken. Verder kan de te lage voorspelling in de weken 7,8,9 en 10 een indicatie geven dat de hoeveelheden post in die weken structureel toenemen. Hetzelfde, alleen dan andersom, geldt voor de te hoge voorspellingen in de weken 33 en 38.

In figuur 4.6 is het duidelijk dat de prognoses met de voorgemelde grote partijen samen een grotere afwijking hebben t.o.v. de werkelijke hoeveelheid post dan in figuur 4.6. Ook hier is er een grote afwijking bij week 1, echter nu is er een te hoge voorspelling gemaakt in week 1. Week 2 en 3, en tevens in de weken 9 tot en met 13 is de afwijking niet zo groot. In de overige weken zijn er grote afwijkingen waarneembaar. Hierbij zijn de grote afwijkingen veroorzaakt door te hoge voorspellingen. Het aantal grote partijen was tenslotte hoger geschat dan het aantal grote partijen in werkelijkheid is (zie bijlage IV.D). Daar de totale hoeveelheid post goed geschat zou zijn, was de verwachting dat de voorspellingen te laag zouden zijn. Het tegendeel is echter waar. Uitzondering hierop zijn de weken 7 en 8. Bij de weken 7 en 8 is er een te lage voorspelling gegeven. De grote afwijkingen worden veroorzaakt door de kleine hoeveelheid data over voorgemelde grote partijen. Door te kijken naar de prognoses met de voorgemelde grote partijen van bijvoorbeeld de weken 17 tot en met 24 of de weken 40 tot en met 48, kan er geanalyseerd worden of er met een aantal datapunten extra de afwijking tussen de werkelijke hoeveelheid post en de prognoses samen met de voorgemelde grote partijen verkleint. Deze analyse wordt beschreven in paragraaf 4.4.2.



Figuur 4.7: Vergelijking van de relatieve afwijking van de realisatie t.o.v. de voorspelling (respectievelijk volgens WAP en de log-quasi-likelihood model len)

Prognoses t.o.v. prognoses WAP

In figuur 4.7 en in bijlage IV.F(deel I en deel IV) zijn de relatieve afwijkingen van de WAP-prognoses en de prognoses van de log-quasi-likelihood modellen weergegeven. Het gemiddelde van de absolute waarden van de relatieve afwijkingen geeft tevens een indicatie over de geschiktheid van een model voor voorspellingen. Het gemiddelde van de absolute waarden van de relatieve afwijkingen is gedefinieerd als het gemiddelde van:

$$\left| \frac{(y - \hat{\mu})}{\hat{\mu}} \right| * 100\% \text{ bij model 4.1}$$

$$\left| \frac{(y - \hat{\mu})}{\hat{\mu} + \text{Gr.Partijen}} \right| * 100\% \text{ bij model 4.2}$$

en wordt aangegeven als Gem*.

De relatieve afwijkingen van bijlage IV.E(deel I en deel IV) en gem* zijn in tabel 4.3 kort samengevat.

Interval	model 4.1		model 4.2		WAP	
	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.
[-2%,2%]	20	42.6%	10	21.3%	17	36.2%
[-5%,5%]	36	76.6%	30	63.8%	41	87.2%
[-10%,10%]	46	97.9%	47	100%	47	100%
[-15%,15%]	47	100%	47	100%	47	100%

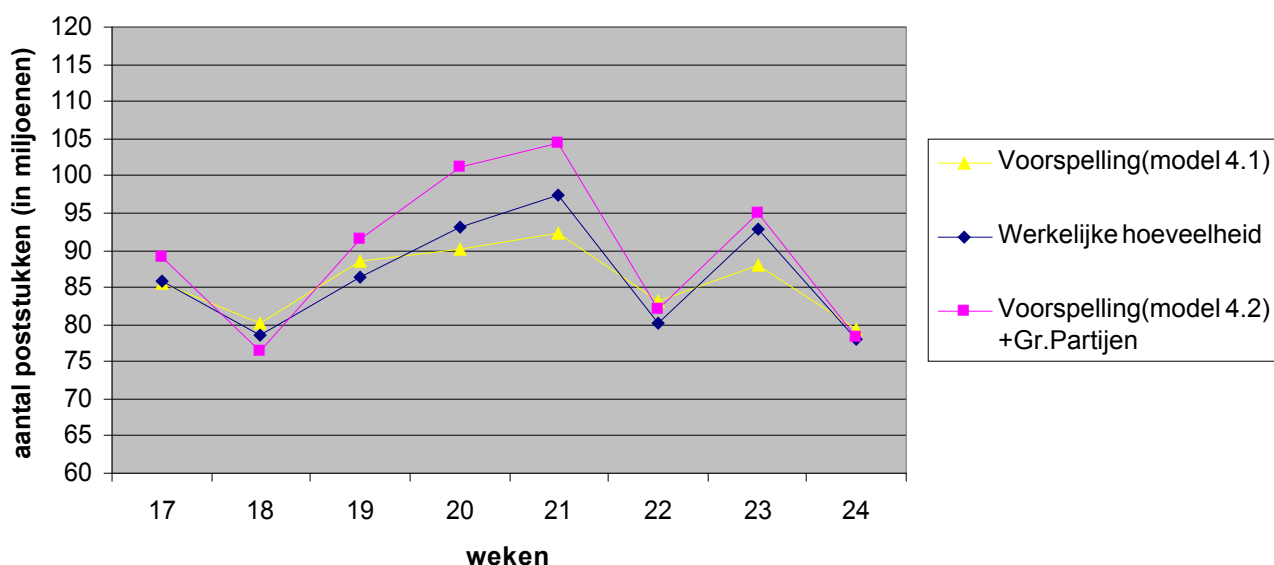
Wk	model 4.1			model 4.2			WAP		
	2-48	17-24	41-48	2-48	17-24	41-48	2-48	17-24	41-48
Gem*	3.3%	3.0%	2.0%	4.6%	5.0%	4.4%	3.0%	3.3%	2.6%

Tabel 4.3: Analyse van de relatieve afwijkingen

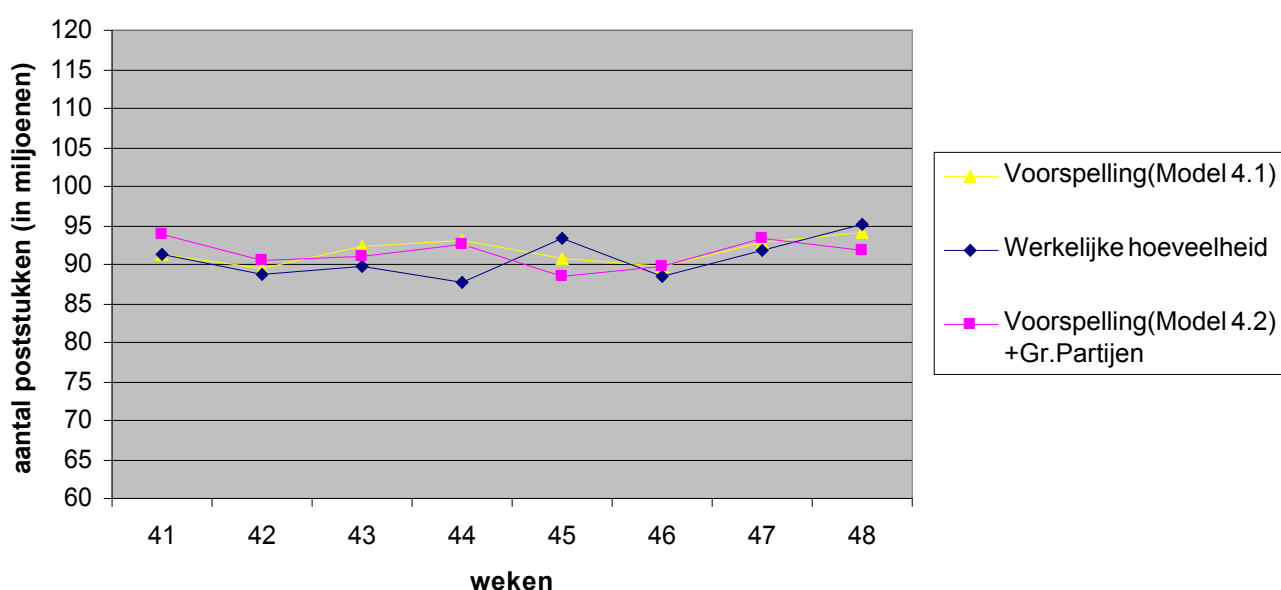
Doordat de WAP-groep de landelijke relatieve afwijkingen sinds week 2 van het jaar 2003 bijhoudt, wordt er bij de vergelijking van de verschillende prognoses gekeken naar de weken 2 tot en met 48. Uit tabel 4.3 en de figuur 4.7 blijkt dat het log-quasi-likelihood model 4.1 meer voorspellingen met een maximale afwijking van 2% berekent dan de WAP-groep. De spreiding van de overige voorspellingen is echter groter bij het log-quasi-likelihood model, dan bij de overige voorspellingen van WAP. Het log-quasi-likelihood model 4.2 heeft qua spreiding en qua interval [-2%,2%] minder goede resultaten dan de WAP-groep. Zoals al eerder vermeld is, wordt dit veroorzaakt door de kleine hoeveelheid

beschikbare data over grote partijen en is het afhankelijk van het onderzoek in paragraaf 4.4.2 of er geconcludeerd wordt dat er nog weinig of veel data nodig is om met dit model goede voorspellingen te berekenen.

4.4.2. Voorspellingsperiode van 8 weken



Figuur 4.8: Vergelijking van de geprognosticeerde met de werkelijke hoeveelheid post in week 17 t/m 24 2003



Figuur 4.9: Vergelijking van de geprognosticeerde met de werkelijke hoeveelheid post in week 41 t/m 48 2003

Nu de werkelijke hoeveelheden post van de weken 1 tot en met 16 respectievelijk de weken 1 tot en met 40 van 2003 inbegrepen zijn bij de data om te voorspellen, is het van belang om het voordeel van extra datapunten te analyseren. Hierbij wordt er geanalyseerd of een voorspelling per 8 weken in plaats van per jaar betere voorspellingen oplevert.

Bij de vergelijking van figuur 4.8 en figuur 4.9 met figuur 4.6 en bij vergelijking van de cijfers van deel I en II van bijlage IV.E, blijkt dat het toevoegen van de extra datapunten (de realisaties van week 1 tot en met 16) de voorspellingen van model 4.1 van week 17 tot en met 24 weinig verbetert. De voorspellingen van de weken 20 en 21 zijn nog steeds te laag, maar wel dichterbij de werkelijke

hoeveelheid post. Echter de afwijkingen in de weken 17, 19, 22 en 24 zijn weer groter geworden. Over de 8 weken totaal is er maar een kleine relatieve verbetering van de voorspellingen. Bij de weken 41 tot en met 48 is er daarentegen een kleine relatieve verslechtering. Voor model 4.1 is er de indicatie dat de omvang van de steekproef al zo groot is zodat extra datapunten weinig tot geen verbetering in de voorspellingen opleveren. Voor de stabielere periode week 41 tot en met 48 heeft het toevoegen van extra data zelfs een negatieve uitwerking.

Indien er gekeken wordt naar het voordeel van het toevoegen extra datapunten voor model 4.2 is aan de hand van figuur 4.8 en 4.9 duidelijk dat het verschil tussen de werkelijke hoeveelheid en de voorspellingen plus de grote partijen over het algemeen kleiner is geworden. De relatieve afwijkingen zijn dan ook met 1.09% respectievelijk 1.46% verbeterd voor de periode week 17 tot en met 24 respectievelijk de periode week 41 tot en met 48. Het grootste gedeelte van de voorspellingen zijn bij dit model echter nog steeds te hoog.

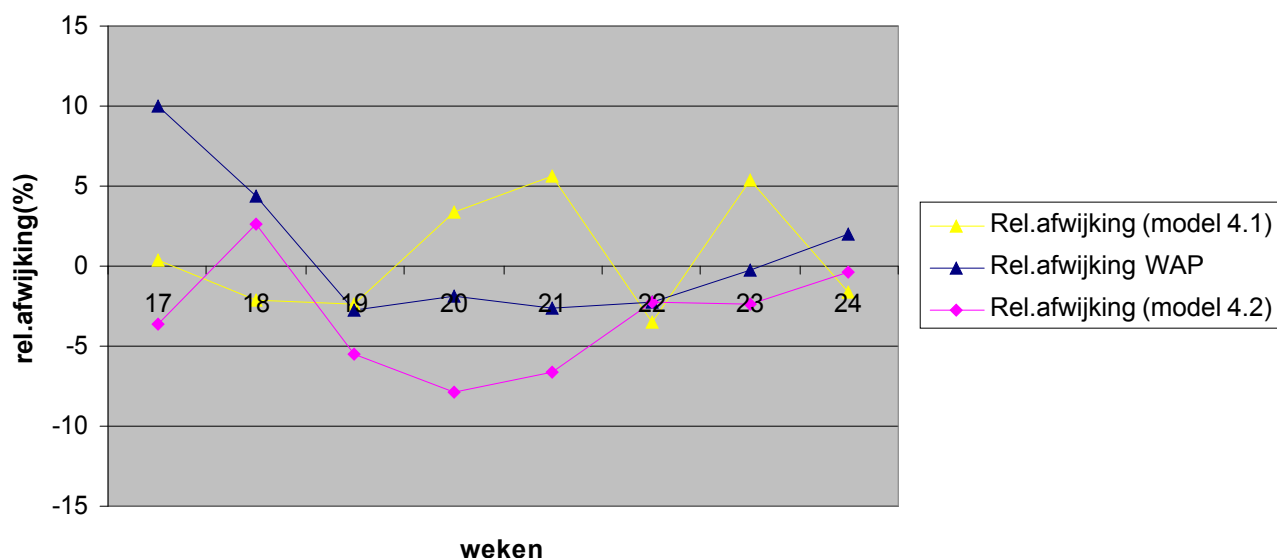
Prognoses t.o.v. prognoses WAP

Uit onderstaand tabel blijkt dat ondanks de verbeteringen van de voorspellingen van model 4.2 de weken 17 tot en met 24 nog steeds slechter voorspelt dan de WAP-groep. Voor de weken 41 tot en met 48 kan dat echter niet meer gezegd worden, daar er bij de WAP-prognose 3 i.p.v. 4 prognoses een absolute relatieve afwijking van maximaal 2% hebben. Daarnaast is de maximale absolute relatieve afwijking 7.2% bij WAP en bij het laatst genoemde model 5.5%. Dit levert de indicatie op dat model 4.2 voor het voorspellen van een stabielere periode beter wordt naarmate het aantal beschikbare datapunten toeneemt. Echter het model is qua resultaten nog steeds slechter dan model 4.1.

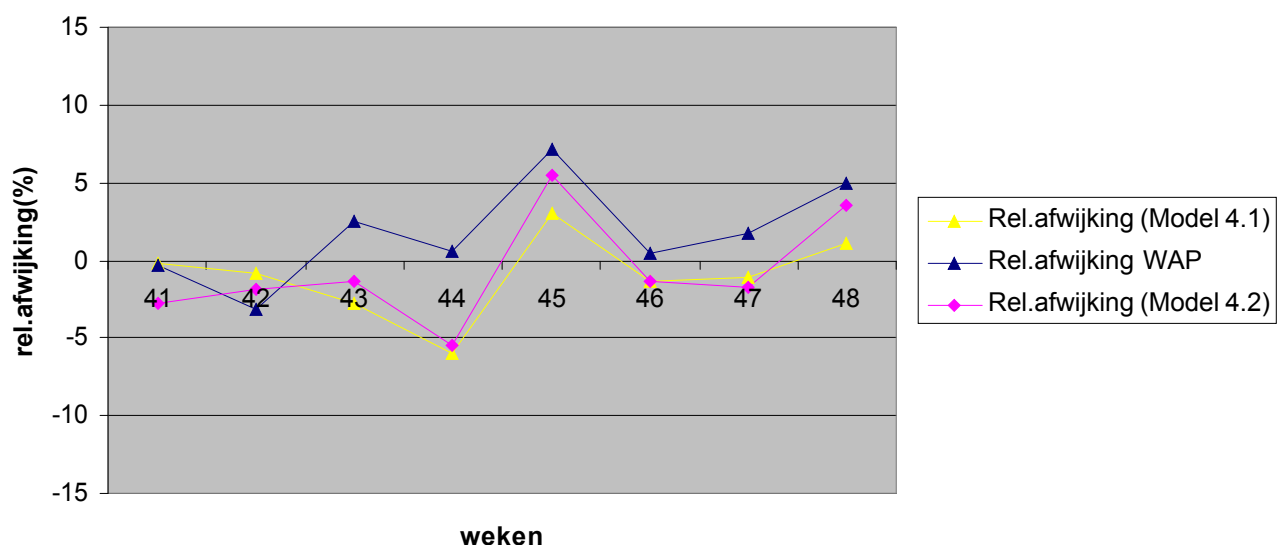
Interval	model 4.1				Model 4.2				WAP			
	Wk 17-24		Wk 41-48		Wk 17-24		Wk 41-48		Wk 17-24		Wk 41-48	
	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.	Aantal	Perc.
[-2%,2%]	3	13%	5	63%	1	12%	4	50%	3	38%	4	50%
[-5%,5%]	7	88%	7	88%	5	63%	6	75%	7	88%	7	88%
[-10%,10%]	8	100%	8	100%	8	100%	8	100%	8	100%	8	100%
[-15%,15%]	8	100%	8	100%	8	100%	8	100%	8	100%	8	100%

	model 4.1		Model 4.2		WAP	
	Wk 17-24	Wk 41-48	Wk 17-24	Wk 41-48	Wk 17-24	Wk 41-48
Gem*	2.9%	2.4%	3.9%	3.0%	3.3%	2.6%

Tabel 4.4: Analyse van de relatieve afwijkingen aan de hand van deel II en deel V van bijlage IV.E

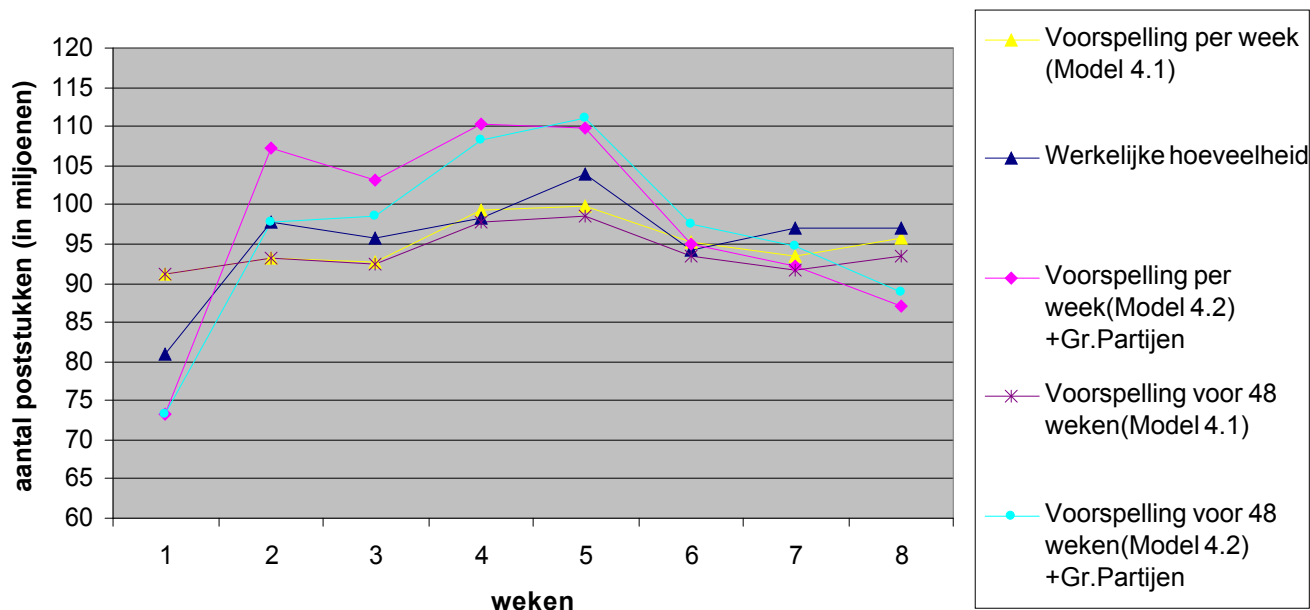


Figuur 4.10: Vergelijking van de relatieve afwijking van de realisatie t.o.v. de voorspelling voor de weken 17 tot en met 24 (respectievelijk volgens WAP en de log-quasi-likelihood model len)



Figuur 4.11: Vergelijking van de relatieve afwijking van de realisatie t.o.v. de voorspelling voor de weken 41 tot en met 48 (respectievelijk volgens WAP en de log-quasi-likelihood model len)

Uit de figuren 4.10 en 4.11 blijkt door een minder fluctuerende relatieve afwijking voor de weken 17 tot en met 24 dat de WAP-groep het beste grip heeft op deze periode die bekend staan om de fluctuaties in de hoeveelheid post. Figuur 4.10 geeft de indicatie dat dit veroorzaakt wordt door hun kennis over de voorgemelde grote partijen, daar model 4.2 een iets minder fluctuerende relatieve afwijking produceert dan model 4.1. Tevens houdt WAP in tegenstelling tot de log-quasi-likelihood modellen rekening met de extra invloed die bijvoorbeeld de week voor of na Pasen heeft.

4.4.3. Voorspellingsperiode van 1 week**Figuur 4.12:** Vergelijking van de geprognosticeerde met de werkelijke hoeveelheid post in week 1 t/m 8

In deze paragraaf wordt er geanalyseerd of een voorspelling week in plaats van per jaar of per 8 weken betere voorspellingen oplevert. Daar uit eerdere analyse³³ blijkt dat de resultaten van deze voorspellingsmethode voor de weken 9 tot en met 48 vergelijkbaar zijn met de resultaten van de voorspellingsmethode waarbij er per 8 weken geprognosticeerd wordt, zal deze voorspellingsmethode voor de weken 1 tot en met 8 van 2003 worden geanalyseerd.

Uit figuur 4.12 blijkt dat de voorspellingen per week van model 4.1 niet dichterbij de werkelijke hoeveelheid post zitten dan de voorspellingen die voor 48 weken worden gemaakt. Dit kan veroorzaakt worden door de grote invloed die week 1 op de voorspellingen per week heeft middels de invloed op het bepalen van de β van 2003. Bij het maken van de voorspellingen voor een periode voor 48 weken is er een β van 2003 bepaald naar aanleiding van de trend in de jaarhoeveelheden³⁴.

Figuur 4.12 geeft echter aan dat bij het model 4.2 de voorspellingen per week dichterbij de werkelijke hoeveelheden zitten dan de voorspellingen voor 48 weken. Dit levert wederom de indicatie op dat bij dit model extra datapunten de voorspelling verbeteren.

³³ zie bijlage IV.A

³⁴ zie bijlage IV.D

4.4.4. Voorspellingsintervallen

Voorspellingsintervallen geven de betrouwbaarheid van de gemaakte schatting $\hat{\mu}_i$ aan. De voorspellingsintervallen zijn bij het log-quasi-likelihood model als volgt gedefinieerd:

$$\hat{\mu}_i \pm \hat{\sigma}_i t_{n-p-1, (1-\alpha/2)}$$

Hierbij is

- $\hat{\mu}_i$ verwachte hoeveelheid post in week i
- $\hat{\sigma}_i$ de geschatte standaard-deviatie
- t de t-verdeling met de parameters (n-p-1) en (1- α /2)
- n het aantal waarnemingen die gebruikt worden om voorspellingen te maken
- p gelijk aan het aantal parameters.
 - Bij het log-lineair model 4.1 is p gelijk aan 86 of, als er een waarneming is die plaatsvindt in het jaar 2003, 87
 - Bij het log-lineair model 4.2 is p gelijk aan 54 of, als er een waarneming is die plaatsvindt in het jaar 2003, 55
- α significantie-niveau welke gelijk gesteld is aan 0,05

Doordat de standaard-deviatie per week en per model verschilt, zijn er in bijlage IV.E verschillende groottes van de voorspellingsintervallen waarneembaar. Door deze verschillende groottes en doordat model 4.1 betere voorspellingen maakt voor 2003 dan model 4.2, zijn er ook verschillen in het aantal realisaties dat in de voorspellingsintervallen vallen. Deze aantallen zijn weergegeven in tabel 4.5.

Periode	1-48	17-24		41-48		1-8	
Methode	Per 48 Aantal	Per 48 Aantal	Per 8 Aantal	Per 48 Aantal	Per 8 Aantal	Per 48 Aantal	Per wk Aantal
Model 4.1	28	6	6	7	6	2	7
Model 4.2	23	5	5	2	4	4	1

Tabel 4.5: Aantal realisaties die in het voorspellingsinterval (welke verschilt per model per methode) vallen

Uit tabel 4.5 blijkt dat model 4.1 betere resultaten heeft dan model 4.2, behalve in de periode week 1 tot en met 8. Daar heeft model 4.1 wel duidelijke betere resultaten qua voorspellingsintervallen als er per week voorspeld wordt. Deze worden verklaard door een grotere standaard-deviatie $\hat{\sigma}_i$ bij de methode 'per week'.

Als bij het gebruik van model 4.1 de eerste 8 weken met de methode 'per week' voorspeld wordt, en de overige voorspellingen met behulp van de methode 'per 48 weken' wordt voorspeld, verwacht ik dat meer dan 50% van de realisaties in de voorspellingsintervallen zullen vallen.

Bij het gebruik van model 4.2 zijn de cijfers nog te fluctueren om er verwachtingen over uit te spreken.

4.5. Conclusie

Het verwezenlijken van het doel, het ontwikkelen van een model dat het landelijke weekverkeer voorspelt, loopt wederom synchroon met het bepalen van de variabelen die van invloed zijn. Indien niet de goede variabelen bepaald zijn, zal er ook geen model ontwikkeld worden die goede voorspellingen bepaalt.

Per variabele kan er via de t-ratio en de deviance getest worden of de variabele wel of geen invloed heeft. Indien de variabele invloed heeft, dan moet de variabele meegenomen worden in het model. Indien alle variabelen die van invloed zijn op de hoeveelheid post bepaald zijn, worden de meest aannemelijke schatters voor de regressiecoëfficiënten β bepaald. Deze regressie coëfficiënten zijn direct een indicatie voor de grootte van de invloed van de verschillende variabelen. Door deze indicaties te vergelijken met de indicaties van de tijdreeks modellen en de Poisson modellen kunnen er conclusies getrokken worden over de invloed van de verschillende variabelen.

Bij het gebruik maken van een model moeten de aannames van het model wel getoetst worden. Indien de aannamen namelijk niet juist zijn, dan zal er gekeken moeten worden naar een ander model. Zodoende is er bij dit onderzoek gebleken dat de responsewaarden Y_i niet Poisson verdeeld zijn. Hierdoor was het log-quasi-likelihood model niet van toepassing. Het klassiek lineair model was tevens niet van toepassing, doordat de residuen niet normaal verdeeld zijn. Het model die wel van toepassing bleek, is het quasi-log-likelihood model.

Helaas hebben ook de voorspellingen van de log-quasi-likelihood model geen maximale relatieve afwijking van 2%. Niettemin zijn de voorspellingen voor 2003 voor de periode 17 tot en met 24, en de periode 41 tot en met 48 bij het model 4.1 in ieder geval beter dan de voorspellingen van de WAP-groep. Dit is geconcludeerd naar aanleiding van de spreiding van de relatieve afwijkingen en de gemiddelde relatieve afwijking. Vergeleken met de voorspellingen van de WAP-groep worden er in de periode week 1 tot en met 48 van het jaar 2003 met model 4.1 meer voorspellingen gemaakt met een maximale absolute relatieve afwijking van 2%, maar zijn er weer minder voorspellingen met een maximale absolute relatieve afwijking van 5%. Het gemiddelde absolute relatieve afwijking is bij model 4.1 echter maar 0,3% hoger.

Bij model 4.2 kan er geconcludeerd worden dat er een verbetering is in de voorspellingen als er meer datapunten toegevoegd worden in het model. Dit is een indicatie dat dit model in de toekomst de gewenste voorspellingen kan bepalen.

Naar aanleiding van de voorspellingsintervallen, kan er geconcludeerd worden dat model 4.1, waarbij de eerste 8 weken met de methode 'per week' en de overige weken 'per 48 weken' voorspeld wordt, het hoogste aantal realisaties in de voorspellingsintervallen zitten. Indien er dus gebruik gemaakt wordt van voorspellingsintervallen, wordt de bovenstaande combinatie van methodes geadviseerd.

5. Poisson model

5.1. Introductie

Het Poisson proces is een wiskundig model dat de momenten van toevallige incidenten beschrijft. In paragraaf 4.1 is er meer informatie over het Poisson proces.

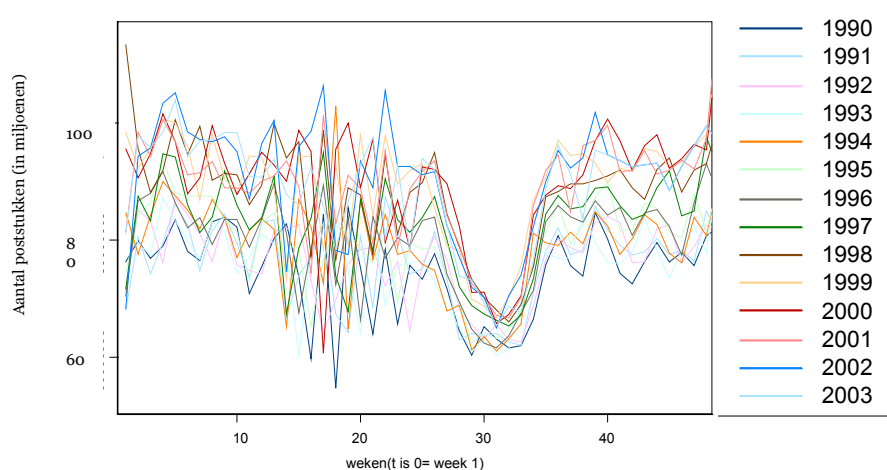
In dit hoofdstuk zullen vier verschillende modellen die gebruik maken van het Poisson proces geanalyseerd worden, namelijk

- (5.1) het model waarbij de responswaarde Y_{ij} gelijk is de hoeveelheid post in week i , jaar j ;
- (5.2) het model waarbij de responswaarde Y_{ij} gelijk is aan de hoeveelheid post zonder de hoeveelheid post die door grote partijen in week i , jaar j gegenereerd is;
- (5.3) het model waarbij de responswaarde Y_{ij} gelijk is aan de hoeveelheid post van week i jaar j waarbij de verwachte invloed van het weeknummer in de tijd gladgemaakt is;
- (5.4) het model waarbij de responswaarde Y_{ij} gelijk is aan de hoeveelheid post zonder de hoeveelheid post die door grote partijen in week i , jaar j gegenereerd is, waarbij de verwachte invloed van de het weeknummer in de tijd gladgemaakt is.

In paragraaf 5.2 zullen de modellen beschreven worden, en zal de schattingsmethode toegelicht worden. Bij het bepalen van de fitting wordt deze schattingsmethode gebruikt. De fitting van de 4 modellen zal in paragraaf 5.3 geanalyseerd worden. In paragraaf 5.4 zullen de voorspellingen per model worden geanalyseerd en vergeleken met de realisaties en de voorspellingen van de WAP-groep.

5.2. Beschrijving model

In het algemeen wordt verwacht dat de hoeveelheid post afhankelijk is van de totale post in het hele jaar en van het weeknummer. Dit blijkt ook uit onderstaande figuur. In deze figuur zijn de weken 1 tot en met 48 van jaar j vergeleken met de weken 1 tot en met 48 van de andere jaren. Hieruit blijkt dat bij een piek in week i in jaar j , er in andere jaren ook een piek is in week i . Ditzelfde geldt voor een dal in week i . Indien er een weekniveau bepaald wordt die onafhankelijk is van het jaar, zal een piek in week i die voorkomt bij meerdere jaren, ook een piek veroorzaken bij het bepaalde weekniveau van week i ten opzichte van de andere weekniveaus. Hieruit blijkt dat het weeknummer 'invloed heeft' op de wekelijkse hoeveelheid post. Dat de feestdagen, zoals Pasen, ook invloed hebben blijkt uit de toename van de fluctuatie in de weken 14 tot en met week 24.



Figuur 5.1: De hoeveelheid post gedurende de weken 1 tot en met 48 van de jaren 1990 tot en met 2003

De verwachting van een Poisson(λ_{ij}) verdeling is gelijk aan λ_{ij} . Omdat verwacht wordt dat de wekelijkse hoeveelheid post afhankelijk is van de totale post in het hele jaar, van het weeknummer en eventuele feestdagen is λ_{ij} voor model 5.1 als volgt gedefinieerd:

$$\lambda_{ij} = \left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right) \tau_j \quad (5.1)$$

Hierbij is

i	het weeknummer ($i=1,2,\dots,48$)
w_i	gelijk aan de invloed op de verwachte hoeveelheid post van week i .
α_k	gelijk aan het de invloed op de hoeveelheid post als feestdag k plaatsvindt in week i jaar j ,
τ_j	gelijk aan de verwachte hoeveelheid post die in de weken 1 tot en met 48 in jaar j gedistribueerd wordt
k	is $\begin{cases} 1 \text{ als er Pasen is in week } i \text{ jaar } j \\ 2 \text{ als er Hemelvaart is in week } i \text{ jaar } j \\ 3 \text{ als er Pinksteren is in week } i \text{ jaar } j \\ 4 \text{ als er koninginnedag is in week } i \text{ jaar } j \\ 5 \text{ als er bevrijdingsdag is in week } i \text{ jaar } j \end{cases}$

Voor model 5.2 wordt op dezelfde manier de λ_{ij} geformuleerd, alleen is hier de τ_j gelijk aan de verwachte hoeveelheid post die in de weken 1 tot en met 48 in jaar j gedistribueerd wordt, zonder de hoeveelheid post die in deze periode door grote partijen gegenereerd wordt.

Bij model 5.3 en 5.4 wordt het bijbehorende weekniveau als volgt gladgemaakt worden:

$$w_i^* = \begin{cases} (2/3)w_1 + (1/3)w_2 & \text{als } i = 1 \\ (1/3)w_{i-1} + (1/3)w_i + (1/3)w_{i+1} & \text{als } i = 2, \dots, 47 \\ (1/3)w_{47} + (2/3)w_{48} & \text{als } i = 48 \end{cases}$$

Deze w_i^* vervangt de waarde w_i in formule (5.1). Het verschil tussen model 5.3 en 5.4 is een verschillende definitie van τ_j . Bij model 5.3 is τ_j gelijk aan de verwachte hoeveelheid post die in de weken 1 tot en met 48 in jaar j gedistribueerd wordt. Bij model 5.4 is τ_j gelijk aan de verwachte hoeveelheid post die in de weken 1 tot en met 48 in jaar j gedistribueerd wordt, zonder de hoeveelheid post die in deze periode door grote partijen gegenereerd wordt.

Poisson

De responswaarde Y_{ij} is gelijk aan het aantal poststukken met of zonder grote partijen in week i jaar j . Veronderstel dat elk poststuk een kans p_{ij} heeft dat het in week i jaar j aangeboden wordt. Veronderstel daarnaast dat het aanbieden onafhankelijk van elkaar gebeurt. Dan heeft Y_{ij} een binomiale(τ_j, p_{ij}) verdeling. Doordat τ_j groot is, en p_{ij} klein, zal Y_{ij} Poisson($\tau_j p_{ij}$) verondersteld worden. Hierbij is p_{ij} gelijk aan:

$$p_{ij} = \left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right),$$

De nevenvoorwaarden hierbij zijn:

$$p_{ij} \geq 0$$

$$\sum_{i=1}^{48} p_{ij} = 1, j=1, 2, \dots, 13$$

Doordat de tijdsintervallen per week en per jaar disjunkt zijn van elkaar, kan er gesproken worden van onafhankelijke poisson processen per week per jaar, waarbij Y_{ij} Poisson verdeeld is met parameter ($\tau_j p_{ij}$).

Schattingsmethode

Om te bepalen wat de waarde is van p_{ij} is, wordt in dit model de meest aannemelijke schatter bepaald. De meest aannemelijke schatter wordt verkregen via het oplossen van de vergelijking:

$$\delta l(\theta, Y) / \delta l(\theta) = 0$$

Waarbij θ gelijk is aan (w, α, τ)

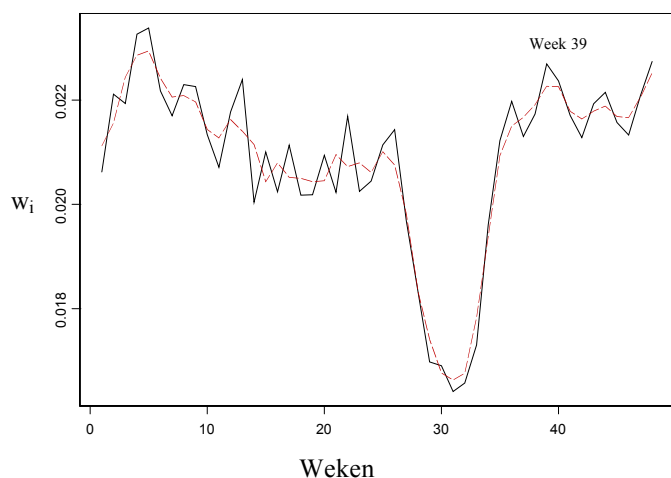
De log likelihood(= $l(\theta, Y)$) is bij dit model gelijk aan

$$l(\theta, Y) = \sum_{j=1}^{12} \sum_{i=1}^{48} \left(y_{ij} \log(\tau_j) + y_{ij} \log \left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right) - \log(y_{ij}!) - \left[\left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right) \tau_j \right] \right)$$

Bij het oplossen van de vergelijking is gebruik gemaakt van de Newton-methode. De oplossing van de vergelijking is beschreven in bijlage V.A.

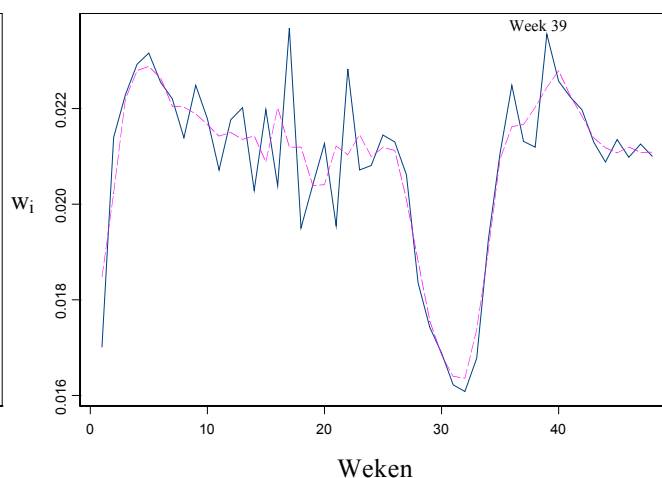
5.3. Fitting model

Voordat de fitting geanalyseerd wordt, zullen eerst de geschatte waarden voor het weekniveau en de invloed van de verschillende feestdagen geanalyseerd worden.



Figuur 5.2: Het geschatte seizoensinvloed per week van model 5.1 en 5.2

— Het geschatte weekniveau w_i per week van model 5.1
- - Het geschatte weekniveau w_i per week van model 5.3



Figuur 5.3: Het geschatte seizoensinvloed per week van model 5.3 en 5.4

— Het geschatte weekniveau w_i per week van model 5.2
- - Het geschatte weekniveau w_i per week van model 5.4

Feestdag	Model 5.1	Model 5.2	Model 5.3	Model 5.4
Pasen	-0.00003	-0.00003	0	0
Hemelvaart	-0.00002	-0.00002	0	0
Pinksteren	-0.00001	-0.00001	0	0
Koninginnedag	-0.00001	-0.00001	-	-
Bevrijdingsdag	-0.00001	-0.00001	-	-

Tabel 5.1: De geschatte invloeden van Pasen, Hemelvaart, Pinksteren, koninginnedag en bevrijdingsdag

Uit bovenstaande figuren en tabel blijkt dat

- er minder scherpe fluctuaties zijn in de modellen 5.2 en 5.4 doordat het weekniveau gladgemaakt is
- de weekniveaus in de weken 14 tot en met 24 bij model 5.3 scherper fluctueren dan bij model 5.1.
- de meest aannemelijke schatters van de invloeden van de verschillende feestdagen bij de modellen 5.1 en 5.2 erg klein zijn.
- de meest aannemelijke schatter van de invloed van Pasen, Hemelvaart en Pinksteren bij de modellen 5.3 en 5.4 0 is. De invloed van koninginnedag is niet geschat omdat deze feestdag in de jaren 2001 en 2002 beide keren valt in week 18. Het effect zit dus verwerkt in weekniveau 18. Bevrijdingsdag is geen officiële feestdag in de jaren 2001 en 2002. Om deze reden kan er geen effect van bevrijdingsdag bepaald worden.
- Week 39 in model 5.3 een relatieve hoge waarde heeft. Dit geeft de indicatie dat de hoeveelheid post in week 39 de laatste jaren is toegenomen.

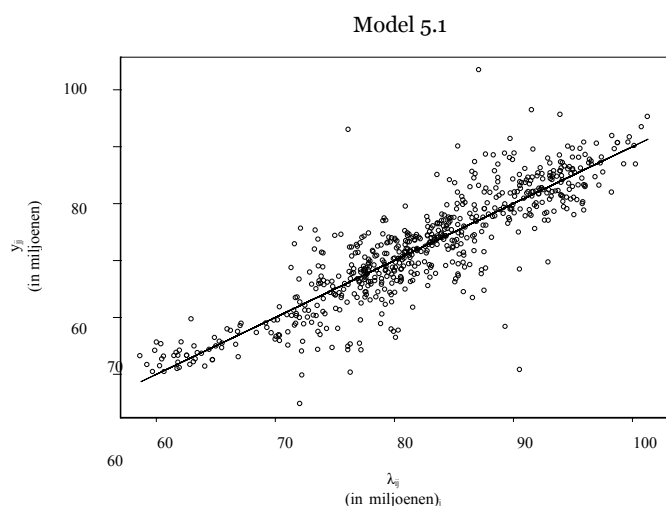
Fitting

De fitting wordt bij deze modellen geanalyseerd aan de hand van de response residuen.

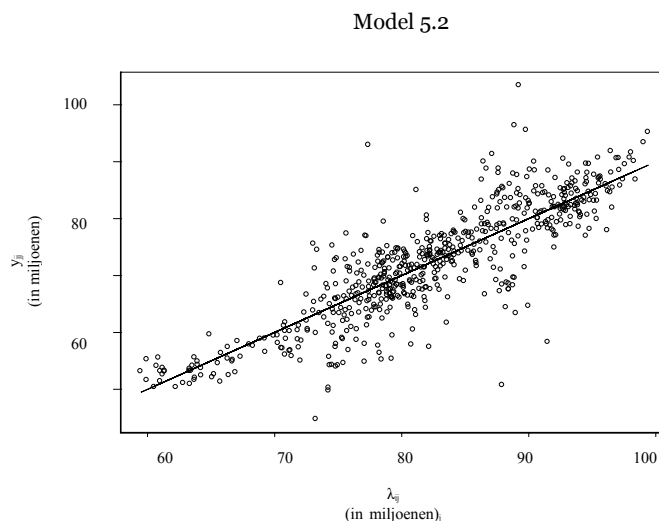
De response residuen zijn gelijk aan: $y_{ij} - \lambda_{ij}$, oftewel het verschil tussen de werkelijke hoeveelheid post in week i (in- of exclusief grote partijen) en de verwachte post in week i (in- of exclusief grote partijen).

Als de verschillen klein zijn, dan benaderen de λ_{ij} 's de y_{ij} 's sterk en zullen, indien de y_{ij} 's afgezet

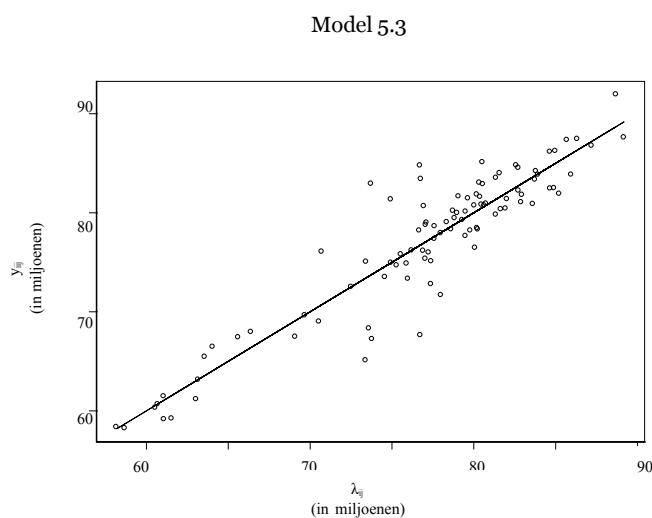
worden tegen de λ_{ij} 's, de datapunten (y_{ij}, λ_{ij}) sterk gecentreerd zijn rond de lijn $y=x$. Met behulp van de volgende figuren kan er geanalyseerd worden of de verschillen klein zijn.



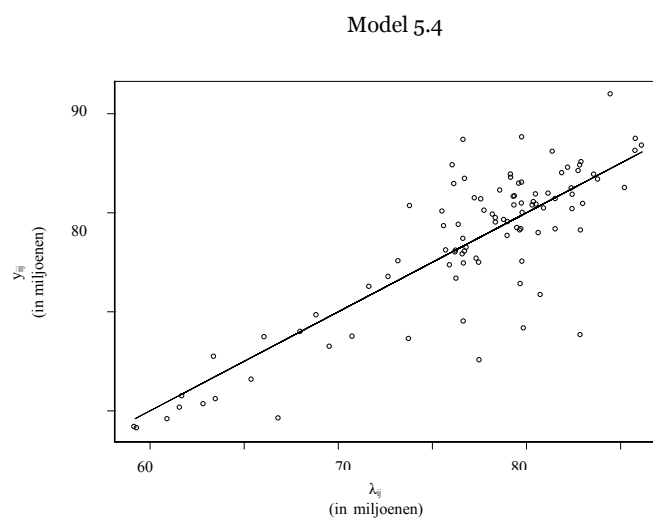
Figuur 5.4: λ_{ij} , de verwachte response Y_{ij} , o.v. de realisaties y_{ij} ven model 5.1



Figuur 5.5: λ_{ij} , de verwachte response Y_{ij} , t.o.v. de realisaties y_{ij} ven model 5.2



Figuur 5.6: λ_{ij} , de verwachte response Y_{ij} , t.o.v. de realisaties y_{ij} ven model 5.3



Figuur 5.7: λ_{ij} , de verwachte response Y_{ij} , t.o.v. de realisaties y_{ij} ven model 5.4

In de bovenstaande figuren is te zien dat het grootste deel van de datapunten niet gecentreerd zijn rond de lijn $y=x$, wat betekent dat er toch nog een redelijk verschil is tussen de werkelijke hoeveelheid wekelijkse post en de verwachte wekelijkse hoeveelheid post. In figuur 5.4 en 5.4 zijn er twee duidelijk uitbijters. Deze uitbijters, respectievelijk week 18 van het jaar 1992(koninginnedag) en week 19 van het jaar 2000(bevrijdingsdag), geven een indicatie dat de invloed van koninginnedag en/of bevrijdingsdag niet goed geschat is, of dat er een afwijkende hoeveelheid post is in die 2 weken. Omdat de bepaalde verwachtingen van de overige weken waarin koninginnedag, respectievelijk bevrijdingsdag vallen relatief de werkelijke waarde goed benadert, lijkt het er op dat er een afwijkende hoeveelheid post is in die 2 weken.

Door de relatieve afwijkingen te analyseren, kan er meer inzicht verkregen worden in het verschil $y_{ij} - \lambda_{ij}$. De relatieve afwijkingen zijn als volgt gedefinieerd:

$$\frac{y_{ij} - \lambda_{ij}}{\lambda_{ij}} * 100\% \text{ voor model 5.1 en 5.3}$$

$$\frac{y_{ij} - \lambda_{ij}}{\lambda_{ij} + G_{ij}} * 100\% \text{ voor model 5.2 en 5.4}$$

Hierbij is

G_{ij} gelijk aan de hoeveelheid post die door grote partijen is gegenereerd in week i jaar j

De relatieve afwijkingen zijn weergegeven in bijlage V.B en kort samengevat in onderstaand tabel,

Interval	Model 5.1		Model 5.2		Model 5.3		Model 5.4	
	Aantal v.d.624	Perc	Aantal v.d.96	Perc	Aantal v.d.624	Perc	Aantal v.d.96	Perc
[-2%,2%]	213	34.1%	48	50.0%	214	34.3%	43	44.8%
[-5%,5%]	448	71.8%	84	87.5%	432	69.2%	74	77.1%
[-10%,10%]	569	91.2%	92	95.8%	561	89.9%	88	91.7%
[-15%,15%]	604	96.8%	96	100%	600	96.2%	94	97.9%

Tabel 5.2: Overzicht van de response residuen

Het gemiddelde van de absolute waarden van de relatieve afwijkingen geeft tevens inzicht in de fitting van een model. Het gemiddelde van de absolute waarden van de relatieve afwijkingen is gedefinieerd als het gemiddelde van:

$$\left| \frac{y_{ij} - \lambda_{ij}}{\lambda_{ij}} \right| * 100\% \text{ voor model 5.1 en 5.3}$$

$$\left| \frac{y_{ij} - \lambda_{ij}}{\lambda_{ij} + G_{ij}} \right| * 100\% \text{ voor model 5.2 en 5.4}$$

en wordt aangegeven als Gem*.

	Model 5.1	Model 5.2	Model 5.3	Model 5.4
Gem*	4.1%	2.6%	4.5%	3.7%

Tabel 5.3: Overzicht van het gemiddelde van de absolute relatieve afwijkingen

Uit bovenstaande tabellen blijkt dat model 5.2, waarbij de responswaarde Y_{ij} gelijk is aan de hoeveelheid post zonder de hoeveelheid post die door grote partijen in week i, jaar j gegenereerd is, de beste fitting heeft. Model 5.4 heeft daarna de beste fitting. Modellen 5.2 en 5.4 hebben als overeenkomst dat met 53 parameters 96 waarnemingen gefit moeten worden. Bij de modellen 5.1 en 5.3 worden er met 68 parameters 624 waarnemingen gefit. Dit verschil verklaart het verschil in fitting en geeft tevens de indicatie dat er verwacht wordt dat de voorspellingen van model 5.1 en 5.3 beter zullen zijn.

Tevens blijkt uit bovenstaande tabellen dat het gladmaken op de beschreven manier geen verbetering in de fitting oplevert. Ik verwacht dat dit veroorzaakt wordt door het instellen van een te sterke afhankelijkheid tussen de weken,

5.4. Voorspellingen

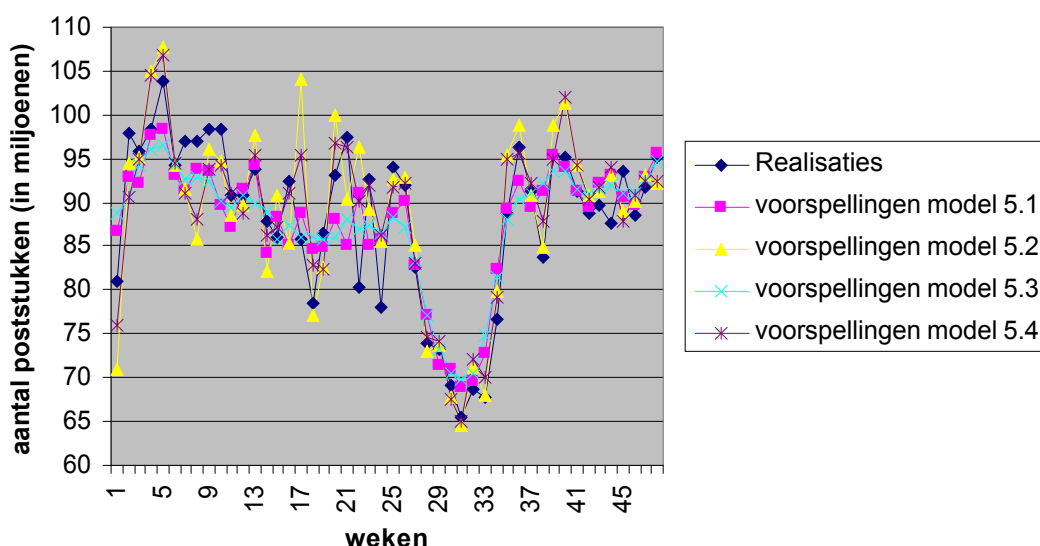
Nadat er een korte toelichting is beschreven over de voorspellingsmethode, zal de analyse van de voorspellingen voor 2003 beschreven worden.

Het maken van voorspellingen wordt gedaan door de geschatte weekniveaus en de geschatte waarden van de invloeden van de feestdagen in te vullen in onderstaande formule.

$$\lambda_{ij} = \left(\hat{w}_i + \sum_{k=1}^5 \hat{\alpha}_k 1_{k \in ij} \right) \hat{\tau}_j$$

De enige onbekende is de $\hat{\tau}_{14}$, oftewel de geschatte waarde van de totale hoeveelheid post in het jaar 2003. Het schatten van de totale hoeveelheid post in 2003(= τ_{14}) wordt gedaan volgens de kleinste kwadraten methode. Met deze methode wordt er een polynoom bepaald die de τ_j benadert. Een toelichting van deze schattingsmethode is beschreven in bijlage V.C.

Bij model 5.2 en 5.4 wordt na het bepalen van de λ_{ij} de G_{ij} , oftewel de hoeveelheid post die in week i jaar j door grote partijen is gegenereerd, er bij opgeteld.



Prognoses t.o.v. de werkelijke hoeveelheid

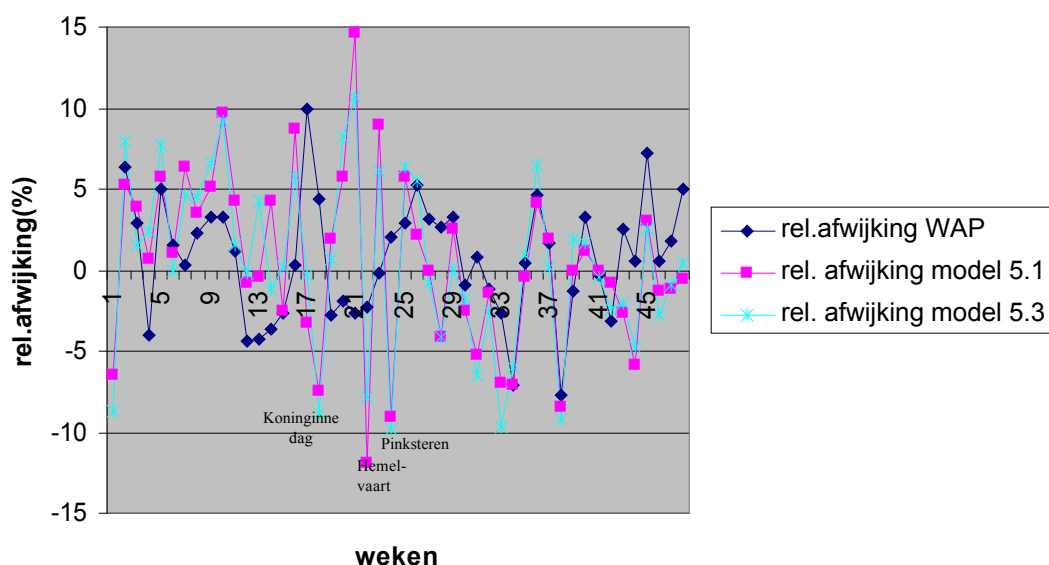
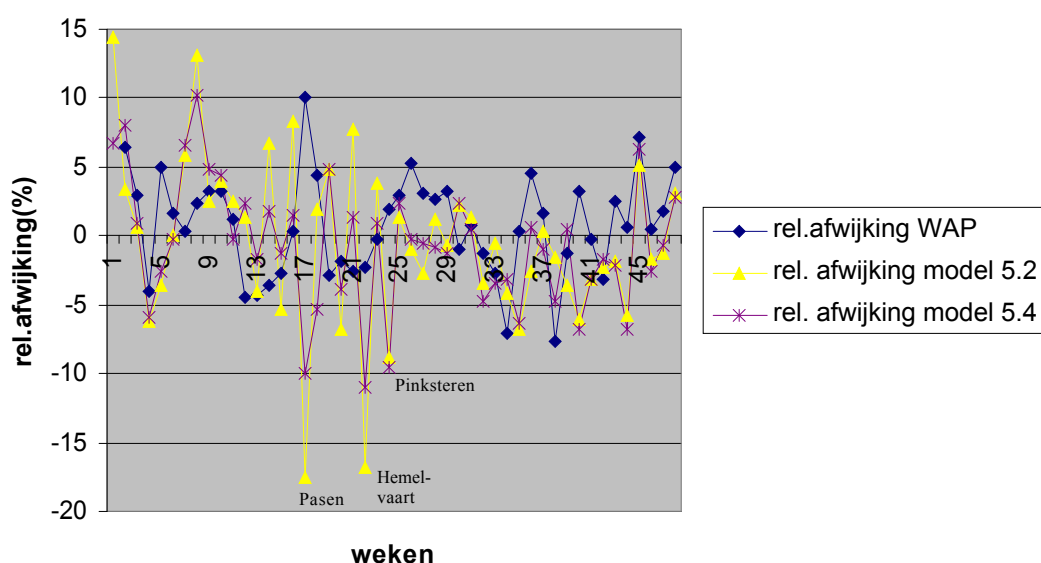
Figuur 5.7: Vergelijking van de geprognosticeerde met de werkelijke hoeveelheid post in week 1 t/m 48 2003

Figuur 5.7 geeft de indicatie dat voorspellingen van model 5.1 weinig afwijken van de voorspellingen van model 5.3. Ditzelfde blijkt na vergelijking van model 5.2 met model 5.4.

Verschillen tussen deze modellen zijn waar te nemen in de weken waarin de feestdagen vallen. Dit wordt verklaard door het verschil in de geschatte waarden van de invloeden van de feestdagen³⁵. Naar aanleidingen van deze waarnemingen kan er geconcludeerd worden dat er een indicatie is dat het gladmaken van de weekniveaus weinig invloed heeft. Deze indicatie kan worden bevestigd door de relatieve afwijkingen te analyseren.

Aan de hand van figuur 5.7 is er geen duidelijke indicatie welk model de werkelijke hoeveelheid post het beste benadert en of de geschatte waarden van de invloed van de feestdagen zorgen voor een betere benadering van de werkelijke hoeveelheid post. Verwacht wordt dat een analyse van de relatieve afwijkingen hierin meer inzicht geeft.

³⁵ Zie tabel 5.1

Prognoses t.o.v. prognoses van de WAP -groep**Figuur 5.8:** Vergelijking van de relatieve afwijking van WAP, model 5.1 en model 5.3**Figuur 5.9:** Vergelijking van de relatieve afwijking van WAP, model 5.2 en model 5.4

Na een analyse van de bovenstaande figuren blijkt dat de invloed voor Hemelvaart, Pinksteren en koninginnedag in de modellen 5.1 en 5.3 te klein geschat is. Bij model 5.2 en 5.4 is de invloed van Pasen, Hemelvaart en Pinksteren te klein geschat. Een aanpassing van deze schattingen is nog niet aan de orde, doordat er een afwijkende hoeveelheid post in het jaar 2003 kan zijn. Bij voorspellingen voor 2004 zal er echter wel rekening gehouden moeten worden met dit gegeven.

Uit de figuren lijkt het dat WAP betere prognoses maakt, dan de Poisson-modellen. Tabel 5.4, waarin een overzicht wordt gegeven van de spreiding van de relatieve afwijkingen, ondersteunt de waarneming dat WAP beter prognosticeert dan de Poisson-modellen. De gebruikte cijfers zijn weergegeven in bijlage V.D.

Interval	Model 5.1		Model 5.2		Model 5.3		Model 5.4		WAP	
	Aantal v.d.47	Perc	Aantal v.d.47	Perc	Aantal v.d.47	Perc	Aantal v.d.47	Perc	Aantal v.d.47	Perc
[-2%,2%]	16	34.0%	15	31.9%	17	36.2%	19	40.4%	17	36.2%
[-5%,5%]	29	61.7%	31	66.0%	28	59.6%	34	72.3%	41	87.2%
[-10%,10%]	46	97.9%	43	91.5%	46	97.9%	45	95.7%	47	100%
[-15%,15%]	47	100%	45	95.7%	47	100%	47	100%	47	100%

Tabel 5.4: Overzicht van de spreiding van de relatieve afwijkingen

	Model 5.1	Model 5.2	Model 5.3	Model 5.4	WAP
Gem*	4.1%	4.2%	4.0%	3.5%	3.0%

Tabel 5.5: Overzicht van de gemiddelde absolute relatieve afwijkingen

Tegen de verwachtingen in, blijkt model 5.4 beter te prognosticeren dan de andere Poisson-modellen. Dit kan verklaard worden doordat:

- er bij dit model gebruik wordt gemaakt van de aanwezige informatie omtrent grote partijen
- een extreme hoeveelheden post in een willekeurige week in de jaren 2001 en 2002 minder invloed heeft op de geschatte weekniveaus doordat deze weekniveaus gladgemaakt zijn
- de weekniveaus bepaald zijn naar aanleiding van 2001 en 2002. Dit geeft de indicatie dat een weekniveau die niet bepaald is met bijvoorbeeld data van meer dan 5 jaar geleden een betere voorspellingen oplevert.

Voorspellingsintervallen

De voorspellingsintervallen gebaseerd op de meest aannemelijke schatter van λ_{ij} zijn als volgt gedefinieerd:

$$\hat{\lambda}_{ij} \pm \sqrt{\hat{\lambda}_{ij} t_{n-p-1, (1-\alpha/2)}} \text{ voor model 5.1 en 5.3}$$

$$G_{ij} + \hat{\lambda}_{ij} \pm \sqrt{\hat{\lambda}_{ij} t_{n-p-1, (1-\alpha/2)}} \text{ voor model 5.2 en 5.4}$$

Hierbij is:

- $\hat{\lambda}_{ij}$ De voorspelde hoeveelheid post (in-/exclusief grote partijen) voor week i jaar j
- G_{ij} De hoeveelheid post die in week i jaar j door grote partijen is gegenereerd
- t De t-verdeling met de parameters (n-p-1) en (1- α /2)
- n De hoeveelheid weken die gebruikt is in de data om onder andere de weekniveaus te schatten
- α Gelijk aan het significantieniveau, dat gelijk gesteld is aan 0.05

Uit bijlage V.D blijkt dat de voorspellingsintervallen erg klein zijn. De realisaties vallen alleen in het voorspellingsinterval als de voorspellingen een relatieve afwijking hebben die kleiner is dan 0,025%. Hierdoor zijn er geen realisaties die in een voorspellingsinterval zitten. De voorspellingsintervallen geven dus geen goede indicatie voor de te verwachten hoeveelheid post.

5.5. Conclusie

Door het structuur van de jaren 1990 tot en met 2002 te analyseren, zijn er bij de Poisson-modellen drie respectievelijk 4 variabelen aangewezen die ervoor moesten zorgen dat de modellen een goede fitting en voorspellingen maakte. Dit zijn de variabelen:

- weeknummer
- invloed feestdagen
- totale hoeveelheid post in de weken 1 tot en met 48 per jaar
- en eventueel de hoeveelheid post die door grote partijen in week i jaar j gegenereerd is.

Na analyse van de fitting blijkt dat de informatie van grote partijen zorgt voor een betere fitting. Dit kan veroorzaakt worden doordat er bij deze modellen met bijna het zelfde aantal parameters minder datapunten worden gefit.

De modellen waarbij het weekniveau gladgemaakt is, scoren slechter qua fitting dan de modellen waar dit niet is gebeurd. Deze waarnemingen geven de indicatie dat de hoeveelheid post in week i een kleinere afhankelijkheid hebben met de week ervoor en erna dan bij model 5.2 en 5.4 gebruikt is.

Ook met de Poisson-modellen bleek het niet mogelijk om prognoses te maken die een maximale relatieve afwijking 2% hebben. Met meer data verwacht ik dat de voorspellingen van model 5.2 en 5.4 verbeteren. Dat model 5.4 nu betere prognoses maakt dan de andere Poisson-modellen, geeft de indicatie dat het gladmaken van de weekniveaus en het toevoegen van informatie over grote partijen de voorspellingen verbetert. Of met de modellen 5.2 en 5.4 over 2 jaar wel voorspellingen gemaakt kunnen worden die een maximale relatieve afwijking van 2% hebben, is nu nog geen zekerheid over te geven. Doordat dit model eenvoudiger is dan het quasi-log-likelihood model, adviseer ik TPG wel om de resultaten van deze modellen over 2 jaar te analyseren.

6. Conclusie

Dit onderzoek heeft als doel het ontwikkelen van een model dat het wekelijkse landverkeer voorspelt. Er zijn uiteindelijk 10 modellen ontwikkeld die gebaseerd zijn op vier algemene mathematische modellen. Conclusies over de kwaliteiten van de modellen met betrekking tot voorspellen worden in paragraaf 6.1 beschreven. In paragraaf 6.2 wordt er nagegaan of de variabelen die volgens de geïnterviewden van invloed zijn, ook volgens de modellen van invloed zijn. Aanbevelingen naar aanleiding van dit onderzoek worden in paragraaf 6.3 gedaan.

6.1 Conclusies wat betreft de voorspellingen

Naar aanleiding van de vergelijking van de voorspellingen voor 2003 met de werkelijke hoeveelheden post in het jaar 2003 en de voorspellingen die gemaakt zijn door de WAP-groep, kan er geconcludeerd worden dat er geen mathematisch model ontwikkeld is dat een maximale relatieve afwijking heeft van 2%. Er zijn echter wel modellen ontwikkeld die qua voorspellen ongeveer dezelfde resultaten hebben als WAP. Daarnaast zijn er modellen ontwikkeld waarvan ik verwacht dat hun voorspellingen zullen verbeteren als er meer datapunten beschikbaar zijn. Met deze verbetering verwacht ik dat deze modellen qua voorspellen ongeveer dezelfde resultaten zullen hebben als WAP, dan wel betere voorspellingen maken dan de WAP-groep. In deze paragraaf zal er per mathematisch model aangegeven worden hoe goed de voorspellingen zijn.

De tijdreeks modellen voorspellen, ondanks dat de fitting wel ongeveer overeenkomt met de log-quasi-likelihood modellen en de Poisson-modellen, duidelijk slechter dan de log-quasi-likelihood modellen en de Poisson-modellen. Dit wordt onder andere veroorzaakt doordat de daling van de hoeveelheid post niet in de geschatte trendlijn terug te vinden is. Door deze beperking verwacht ik niet dat de tijdreeks modellen betere voorspellingen zullen maken dan WAP en/of de overige mathematische modellen.

Naar aanleiding van de analyse van de voorspellingen van het log-quasi-likelihood model 4.1, blijkt dat er meer relatieve afwijkingen in het interval $[-2\%, 2\%]$ vallen bij model 4.1 dan bij WAP. Het gemiddelde van de absolute relatieve afwijkingen van de WAP-groep is echter 0,3% lager. De conclusie is dan ook dat de juistheid van de voorspellingen van het log-quasi-likelihood model 4.1 de juistheid van de voorspellingen van de WAP-groep benadert. Indien er met het log-quasi-likelihood model 4.1 voor de periode week 17 tot en met 24 van het jaar 2003 voorspeld wordt (waarbij er gebruik gemaakt wordt van de gerealiseerde data week 1 1990 tot en met week 16 2003), dan worden er betere voorspellingen gemaakt dan met WAP. Ditzelfde geldt voor de voorspellingen van het log-quasi-likelihood model 4.1 voor de periode week 41 tot en met 48 (waarbij er gebruik gemaakt wordt van de gerealiseerde data week 1 1990 tot en met week 40 2003). In deze periode zijn de voorspellingen die gemaakt werden door het log-quasi-likelihood model 4.1 voor een periode van 48 weken beter. Dit geeft de indicatie dat het toevoegen van gerealiseerde data van jaar j bij een model waar voldoende data beschikbaar is alleen een toegevoegde waarde heeft als er voorspellingen gemaakt moeten worden voor een periode waar in voorgaande jaren grote fluctuaties in de wekelijkse hoeveelheden post waren. Bij de periode week 2 tot en met 8 zijn de voorspellingen van de WAP-groep beter dan het log-quasi-likelihood model 4.1. Zelfs als er met het log-quasi-likelihood model 4.1 in die periode per week voorspeld wordt, blijken de WAP-voorspellingen beter te zijn. Omdat de verschillen niet zo groot zijn, kan het log-quasi-likelihood model 4.1 wel in de periode week 2 tot en met 8 gebruikt worden als ondersteunend mathematisch model voor het maken van voorspellingen. Daar het verschil tussen het per week voorspellen met het log-quasi-likelihood model 4.1 of per 48 weken minimaal is, maakt het weinig uit om in deze periode per week of per 48 weken te voorspellen.

Het log-quasi-likelihood model 4.2 genereert slechtere voorspellingen voor het jaar 2003 dan het log-quasi-likelihood model 4.1. Omdat de voorspellingen van het log-quasi-likelihood model 4.2 verbeteren naarmate er meer datapunten beschikbaar zijn en omdat met model 4.2 er een kleinere hoeveelheid post voorspeld moet worden³⁶, verwacht ik dat het log-quasi-likelihood model 4.2 in de toekomst op een gegeven moment betere voorspellingen zal maken dan het log-quasi-likelihood model 4.1.

Poisson-model 5.4, het model waarbij de informatie over grote partijen gebruikt wordt en waarbij het weekniveau gladgemaakt is, maakt de beste voorspellingen voor het jaar 2003 van de vier Poisson-

³⁶ Bij model 4.2 wordt de wekelijkse hoeveelheid post min de hoeveelheid post, dat door grote partijen gegenereerd wordt, voorspeld

modellen. Hierbij heeft Poisson-model 5.4 meer relatieve afwijkingen in het interval $[-2\%, 2\%]$ dan WAP, maar minder in het interval $[-5\%, 5\%]$. Het gemiddelde van de absolute waarden van de relatieve afwijkingen benadert de absolute waarden van WAP sterk. Met deze gegevens kan er geconcludeerd worden, dat Poisson-model 5.4 ongeveer net zo goed voorspelt als de WAP-groep. Van Poisson-model 5.4 verwacht ik dat de voorspellingen zullen verbeteren indien er meer datapunten beschikbaar zijn. Nu gebruikt het Poisson-model 5.4 namelijk alleen data uit de jaren 2001 en 2002 voor de voorspellingen van 2003.

6.2 Conclusies wat betreft variabelen die van invloed zijn

In deze paragraaf zullen er conclusies geformuleerd worden over de invloed van de variabelen die volgens de geïnterviewden van invloed zijn op de wekelijkse hoeveelheid post.

De variabelen die volgens de geïnterviewden van invloed zijn op de wekelijkse hoeveelheid post zijn:

- maandwisseling
- kwartaaleinde
- halfjaar-einde
- zomerdal(waaronder vakantiespreiding)
- feestdagen

Overige niet direct tijd gerelateerde variabelen zijn:

- opkomende concurrentie
- economische (negatieve) groei
- aantal zakelijke klanten
- aantal afgiftepunten

Deze niet tijd gerelateerde variabelen zijn indirect in de modellen opgenomen. Deze variabelen hebben namelijk invloed op respectievelijk de trend, op de bij jaar j behorende regressiecoëfficiënt β_j , en op de totale verwachte hoeveelheid post in een jaar τ_j . Daarnaast heeft het aantal zakelijke klanten invloed op de hoeveelheid post die door grote partijen gegenereerd wordt. De variabele 'hoeveelheid post die door grote partijen gegenereerd wordt' is opgenomen in het model bij elk van de drie algemene wiskundige modellen.

Bij het formuleren van conclusies over de invloed van de tijd gerelateerde variabelen is er meer waarde gehecht aan de indicaties van de log-quasi-likelihoodmodellen, omdat deze modellen in het algemeen een betere fitting hebben dan de overige modellen. Doordat bij de Poisson-modellen waar de weekniveaus gladgemaakt zijn de invloed van een week mede bepaald wordt door de omliggende weken, zijn deze modellen niet meegenomen voor het trekken van conclusies over de invloed van tijd gerelateerde variabelen. Tijdens het formuleren van conclusies over de invloeden van feestdagen is het Poisson model 5.3, die gebruik maakt van de informatie over voorgemelde grote partijen, niet meegenomen, omdat bij dit model de invloed van feestdagen geschat is op nul. Doordat in 9 van de 10 modellen alleen de weken 1 tot en met 48 van de verschillende jaren is geanalyseerd, is er de maandwisseling december-januari niet geanalyseerd. Hierdoor zijn er geen conclusies getrokken over de maandwisseling december-januari, het 4^e kwartaaleinde, en het 2^e halfjaareinde.

De invloed van de tijd gerelateerde variabelen is bepaald middels respectievelijk:

- de geschatte seizoensinvloed en de geschatte effecten van de feestdagen
- de geschatte waarden van de regressiecoëfficiënten β_j
- de geschatte waarden van de weekniveaus en de effecten van de feestdagen

Met deze schattingen zijn onderstaande conclusies getrokken over de tijd gerelateerde variabelen. Meer informatie over de invloed van de tijd gerelateerde variabelen is beschreven in bijlage VI.A.

De maandwisselingen zorgen over het algemeen niet voor een grotere hoeveelheid post. Alleen de maandwisselingen januari-februari, februari-maart, mei-juni en november-december lijken een grotere hoeveelheid post te veroorzaken. Dat de overige maandwisselingen niet voor een grotere hoeveelheid post zorgen, kan veroorzaakt worden door het gegeven dat de zomervakantie of feestdagen indirect invloed hebben op de geschatte waarde van een willekeurige maandwisseling. In bijlage VI.A is de invloed van de maandwisselingen per maandwisseling beschreven.

Van de eerste drie officiële kwartaaleinden, oftewel de kwartaaleinden gedurende de maandwisselingen maart-april, juni-juli en september-oktober, wordt er ook niet verwacht dat ze een grotere hoeveelheid post veroorzaken. Wordt er echter gekeken naar de weken die veel bedrijven aanhouden als kwartaaleinden (de weken 13, 26, 39 en 52) dan is er wel een positief effect op de hoeveelheid post waarneembaar. In bijlage VI.A is de invloed van de kwartaaleinden per kwartaaleinde beschreven.

Bij het eerste halfjaareinde is hetzelfde waarneembaar als bij het 2^e kwartaaleinde. Oftewel, de maandwisseling juni-juli geeft geen indicatie dat een halfjaar-einde voor een grotere hoeveelheid post zorgt. Echter als er alleen gekeken wordt naar week 26³⁷, dan kan er wel geconcludeerd worden dat het eerste halfjaareinde een grotere hoeveelheid post veroorzaakt.

Naar aanleiding van de analyse van de zomervakantie, kan er geconcludeerd worden dat de zomervakantie in de periode week 27 tot en met 35 een negatieve invloed heeft. Vooral in de weken 31 en 32 is de negatieve invloed groot.

Een onderzoek van de spreiding van de zomervakantie geeft aan dat Zuid-Nederland de grootste hoeveelheid post genereert. Noord-Nederland genereert de kleinste hoeveelheid post.

De feestdagen hebben allen een negatieve invloed op de hoeveelheid post. Met hoeveel poststukken de hoeveelheid post vermindert, is afhankelijk van de feestdag. Vooral koninginndag lijkt een kleine negatieve invloed te hebben, maar dit wordt veroorzaakt doordat de maandwisseling april-mei vaak gelijktijdig is. In bijlage VI.A is de invloed van de feestdagen per feestdag beschreven.

6.3 Aanbevelingen

In dit afstudeeronderzoek zijn de log-quasi-likelihood modellen en het Poisson model 5.4 naar voren gekomen als bruikbare modellen voor het maken van voorspellingen van het wekelijks landverkeer. Omdat de kans bestaat dat teveel data zorgen voor een kleinere flexibiliteit, is het wenselijk om per kwartaal een overzicht te maken van de resultaten van de modellen. Zo'n overzicht kan dan een indicatie geven dat een onderzoek naar de hoeveelheid data om de beste resultaten te verkrijgen benodigd is. Tevens wordt de kwaliteit van een model in de gaten gehouden.

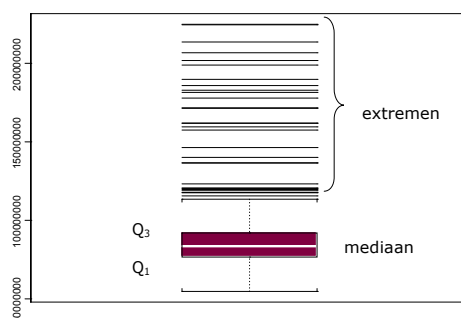
Vervolgonderzoek

- Bij de log-quasi-likelihood modellen is het mogelijk om variabelen in te voegen zoals 'week voor pasen', 'week na pasen', etcetera. Hierdoor is het mogelijk om te kijken of een feestdag bijvoorbeeld al invloed heeft op de hoeveelheid post een week voordat de feestdag plaatsvindt. Indien er daadwerkelijk sprake is van invloed, dan zullen deze variabelen misschien betere voorspellingen opleveren.
- Op dit moment worden alleen grote partijen die meer dan 200.000 poststukken bevatten in een week meegenomen in het model. Hiervoor is gekozen, omdat in het jaar 2001 er alleen maar gegevens beschikbaar zijn van grote partijen die meer dan 200.000 poststukken bevatten. Omdat er sinds 2002 ook de informatie omtrent andere grote partijen wordt geadministreerd, is het van belang om in bijvoorbeeld het jaar 2005 te analyseren of de resultaten van het log-quasi-likelihood model 4.2 en Poisson model 5.4 verbeteren als ook de informatie omtrent andere grote partijen wordt gebruikt.
- Omdat de vestigingen geïnteresseerd zijn in de hoeveelheid post die alleen zij moeten distribueren, is een onderzoek naar de toepassing van deze modellen per vestiging zeker interessant.
- Voor sorteercentra is het van belang of de te distribueren poststukken grote poststukken, kleine poststukken of poststukken zijn die niet door een machine kunnen. Een onderzoek naar de toepassing van de log-quasi-likelihood-modellen en de Poisson modellen is voor dit probleem ook zeker interessant.

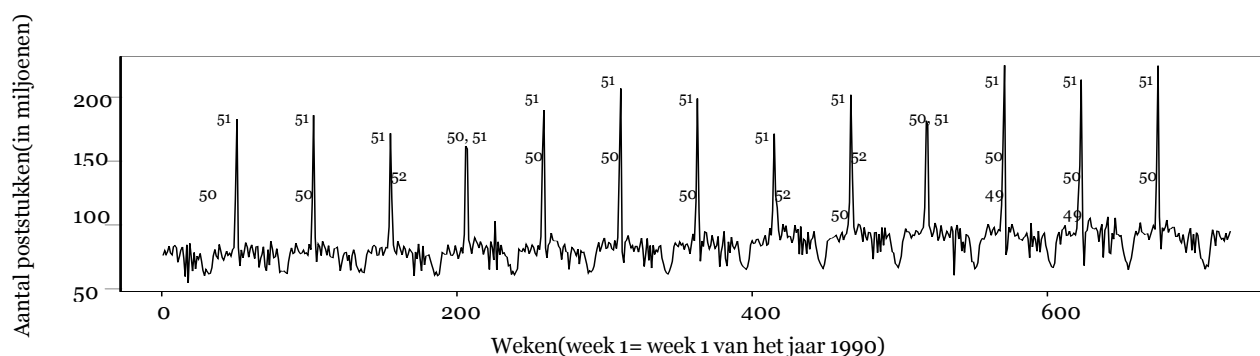
³⁷ de maandwisseling juni-juli is in de jaren 1990 tot en met 2002 of in week 26 of in week 27

Bijlage III.A: Extremen binnen de data

Met behulp van de boxplot kunnen de extremen bepaald worden. Deze punten bevinden zich dan namelijk buiten het gebied $[\text{mediaan} - 1,5(Q_3 - Q_1); \text{mediaan} + 1,5(Q_3 - Q_1)]$. Hierbij is Q_1 gelijk aan de hoeveelheid post op het eerste kwartiel, en Q_3 gelijk aan de hoeveelheid post op het derde kwartiel. De boxplot in figuur III.A.1 heeft de data gebruikt van de hoeveelheid briefpost aan huis inclusief de grote partijen. De locaties van de extremen zijn te zien in figuur III.A.2.



Figuur III.A.1: Boxplot met als data de hoeveelheid briefpost aan huis inclusief de grote partijen

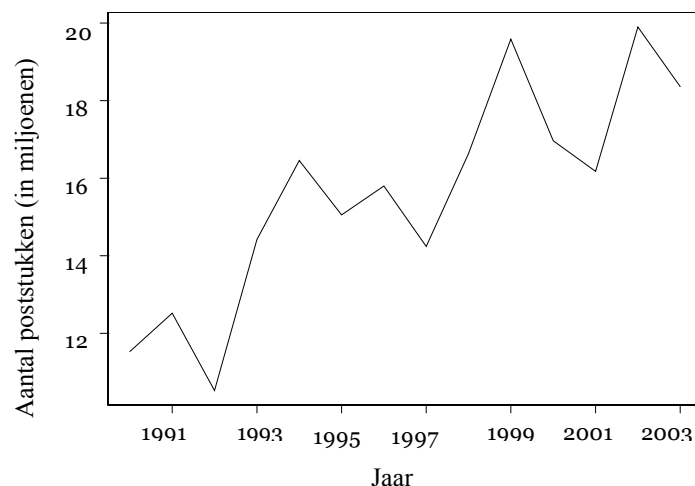


Figuur III.A.2: De hoeveelheid post gedurende de jaren 1990 tot en met 2003, waarbij de locaties van de extremen aangegeven zijn

Om fouten in de eindconclusie te voorkomen, is het beter om de extremen uit de data weg te laten.

Bijlage III.B: Toename zomerinvloed

Het verschil tussen de gemiddelde hoeveelheid post in het zomerdal met het gemiddelde hoeveelheid post in de weken waar geen zomerinvloed is, lijkt in figuur 4.3 met de jaren toe te nemen. In onderstaand figuur is het verschil tussen deze gemiddelden weergegeven.



Figuur III.B.1: Verschil tussen het gemiddelde van de weken 27 t/m 32 en de het gemiddelde van de weken 1 t/m 26 en 33 t/m 48

De cijfers behorende bij figuur III.B.1 zijn weergegeven in tabel III.B.1

	Gemiddelde hoeveelheid post in de weken 1 t/m 26 en de weken 33 t/m 48 (in miljoenen)	Gemiddelde hoeveelheid post in de weken 27 t/m 32 (in miljoenen)	Verschil (in miljoenen)
1990	75.9	64.4	11.5
1991	77.7	65.1	12.6
1992	77.6	67.1	10.5
1993	77.6	63.2	14.4
1994	80.8	64.3	16.5
1995	81.4	66.3	15.1
1996	81.8	66.0	15.8
1997	84.2	70.0	14.2
1998	90.1	73.5	16.6
1999	91.7	71.2	20.5
2000	91.5	74.6	16.9
2001	90.7	74.5	16.2
2002	92.8	72.9	19.9
2003	90.5	72.2	18.3

Tabel III.B.1: Het gemiddelde van de weken 1 t/m 26 en 33 t/m 48, het gemiddelde van de weken 27 t/m 32, en het verschil tussen deze gemiddelden.

Uit de cijfers en het figuur blijkt dat er aangenomen kan worden dat het verschil in het algemeen met de jaren toeneemt. Hierbij loopt de toename niet lineair maar erg fluctuerend. Deze fluctuaties zullen waarschijnlijk terug te vinden zijn in de residuen Y_t .

Bijlage III.D: State-space model & Kalman filter

De voorspellingen van het tijdreeks model worden in Splus berekend met het Kalmann filter. De Kalmann filter gebruikt variabelen van de waarnemings- en state-formule van het state-space model. De algemene waarnemings- en state-formule zijn als volgt gedefinieerd:

- waarnemings-formule: $Y_t = B_t C_t + W_t$
- state-formule: $C_{t+1} = F_t C_t + Z_t$

Onderstaand zijn de waarnemings-formule en de state-formule voor beide modellen gegeven³⁸.

De formule van het ARMA(1,2)-model voor model 3.1 is gelijk aan:

$$Y_t = 0.98Y_{t-1} + Z_t + 0.96Z_{t-1} - 0.05Z_{t-2}$$

De waarnemingsformule hierbij is gelijk aan:

$$Y_t = [-0.05, 0.96, 1] C_t$$

en de state-formule is:

$$C_t = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0.98 \end{bmatrix} C_t + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} Z_t$$

De formule van het ARMA(1,1)-model voor model 3.3 is gelijk aan:

$$Y_t = 0.98Y_{t-1} + Z_t + 0.93Z_{t-1}$$

De waarnemingsformule hierbij is gelijk aan:

$$Y_t = [0.93, 1] C_t$$

en de state-formule is:

$$C_t = \begin{bmatrix} 0 & 1 \\ 0 & 0.98 \end{bmatrix} C_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} Z_t$$

De kalman filter bepaalt de schatter van de $C_{t|t}$. Deze is gelijk aan $P_t(C_t)$. De bijbehorende covariantie-matrix is gedefinieerd als $\Omega_{t|t} = E[(Y_t - Y_{t|t})(Y_t - Y_{t|t})']$ en wordt bepaald door:

$$P_t C_t = P_{t-1} C_t + \Omega_{t|t} B_t' \Delta_t^{-1} (Y_t - B_t C_t)$$

Hierbij is

$$P_t(C) = (P_1(C_1), \dots, P_t(C_t)),$$

waarbij $P_t(C_i) = P(C_i | Y_0, Y_1, \dots, Y_t)$ de beste lineaire voorspeller is van C_i

$$\Omega_{t|t} = \Omega_t - \Omega_t B_t' \Delta_t^{-1} B_t \Omega_t$$

$$\Delta_t^{-1} = B_t \Omega_t B_t'$$

$$C_{t+1} = F_t C_t + \Theta_t \Delta_t^{-1} (Y_t - B_t C_t)$$

$$\Theta_t = F_t \Omega_t B_t'$$

³⁸ Voor een toelichting over de methode om deze modellen te herformuleren, verwijs ik u naar het boek "Introduction to Time Series, geschreven door Peter J. Brockwell en Richard A. Davis, © 1996 Springer Verlag New York

Bijlage III.D: cijfers fitting tijdreeks model(deel I)

Analyse van de relatieve afwijkingen van model 3.1

$$\frac{\left(X_t - \frac{X_t}{e^{R_t}}\right)}{\left(\frac{X_t}{e^{R_t}}\right)} * 100\%$$

Jaar week	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
1	-	-8,1	-5,9	10,4	11,2	-4,6	-12,0	-10,8	36,9	14,4	8,1	-8,7	-23,2
2	-2,9	0,9	4,0	-3,8	-5,5	0,2	0,7	0,6	6,2	-0,7	-5,3	2,3	-1,6
3	-4,3	-5,7	3,1	-5,7	3,0	4,1	3,8	-2,8	-2,8	3,1	-0,3	-0,6	2,1
4	-5,9	-4,6	-11,2	0,5	4,1	6,4	2,0	4,3	-4,5	2,6	1,2	0,1	4,1
5	0,1	3,0	1,0	-4,3	0,5	-0,9	-4,8	3,1	4,0	1,2	-4,1	-4,2	5,1
6	-1,0	4,3	3,5	-3,2	2,7	3,4	-4,6	-1,0	3,5	2,2	-8,2	-4,9	3,4
7	-0,6	-3,5	1,9	-4,7	1,1	6,0	-0,1	-4,4	11,1	-5,3	-2,8	-2,0	4,1
8	5,2	2,3	-6,8	0,7	4,2	2,7	-7,9	-4,0	-2,2	3,9	4,4	-2,4	0,7
9	5,9	5,9	2,2	-1,1	0,4	-2,4	-2,9	5,3	-1,4	-3,8	-2,2	-7,0	1,3
10	7,7	3,9	-3,2	-5,7	-3,9	-0,6	1,8	3,1	2,9	-2,4	-3,9	-2,7	3,0
11	-4,7	-0,8	-1,4	0,9	5,1	-7,9	-0,8	0,8	0,3	8,3	3,4	-0,3	-2,9
12	-3,8	6,0	-6,8	1,1	2,5	-0,4	-0,1	-1,9	-1,1	2,6	2,4	-2,9	2,2
13	0,0	3,5	-0,3	-3,8	-2,9	-0,8	4,2	3,4	7,6	-6,3	-2,9	-4,6	3,6
14	6,2	6,5	3,2	-3,1	-0,4	-4,8	-3,1	-1,3	4,1	-0,2	-3,2	1,0	-1,0
15	-5,5	-6,2	4,1	-5,9	6,1	-5,3	0,8	-8,3	6,7	2,6	6,3	-3,6	1,4
16	-2,8	-1,0	-6,7	0,1	1,8	0,3	-5,2	-0,8	8,0	4,4	4,1	2,3	5,7
17	7,1	4,0	-0,3	-9,2	-12,8	-8,2	5,2	10,0	7,0	-15,8	-19,2	8,4	10,9
18	-24,1	-7,6	-5,0	15,3	35,1	7,7	-6,7	-6,6	-9,6	9,0	10,9	-3,4	-11,2
19	11,5	-1,8	-3,1	-4,6	-4,6	0,0	1,2	-4,7	-1,7	4,6	9,9	-0,7	-2,1
20	-1,6	3,1	-0,6	-4,3	4,9	-4,5	-5,3	2,0	-3,4	7,4	-3,6	-3,9	-1,1
21	-3,0	-3,4	-1,9	3,0	2,1	7,8	-1,1	1,2	-0,2	1,2	2,9	0,5	3,6
22	-3,1	6,0	3,6	-7,8	-3,2	-6,3	-1,6	0,9	-0,8	1,3	-2,8	-1,8	6,1
23	-0,9	7,4	0,2	-1,7	-3,3	-2,6	-0,2	1,2	3,8	0,8	-3,2	-1,2	1,0
24	2,6	-1,9	-4,7	-0,5	-1,8	0,5	-1,5	-1,1	1,2	3,8	-1,2	-0,1	1,2
25	-1,5	-1,9	-1,1	-0,5	-5,5	-1,9	3,0	1,0	2,1	4,2	3,1	2,4	-1,3
26	2,8	5,1	4,0	-6,6	-7,9	-2,8	2,3	3,9	6,4	-5,8	1,2	3,0	-2,1
27	3,4	3,2	7,6	-4,5	-8,7	-0,8	-1,8	3,5	2,9	-5,3	7,2	0,7	-4,1
28	-0,2	-5,1	3,0	-2,1	0,0	0,3	-1,4	-0,2	3,0	1,1	5,8	1,7	-3,5
29	0,5	4,1	6,0	-1,1	-3,6	-2,6	-0,6	2,7	3,2	-2,2	-1,9	0,4	-2,6
30	9,0	4,1	2,3	3,0	0,3	0,5	-4,0	0,8	-1,0	-3,5	-1,8	-2,7	-4,4
31	8,5	7,5	8,9	1,3	-0,6	1,9	-2,4	2,0	-1,2	-3,3	-6,3	-4,1	-8,8
32	4,3	3,7	3,6	1,8	1,8	3,4	0,1	-0,6	-5,0	1,1	-4,9	-5,5	-1,7
33	0,2	-1,7	-1,9	10,6	1,4	3,5	1,9	-2,1	-4,7	4,4	-4,2	-4,5	-0,7
34	-5,2	-0,7	-4,8	10,1	10,6	2,5	-4,9	-5,4	-7,6	5,9	1,5	3,6	-2,3
35	-0,2	0,8	-1,8	-0,1	-0,2	0,8	2,0	0,5	-0,9	1,9	-2,3	2,4	-1,5
36	3,0	2,4	-0,7	-5,1	-4,6	3,1	2,1	1,0	-2,9	4,9	-4,0	1,6	0,9
37	-0,5	-0,3	0,6	-1,9	1,3	-2,4	2,7	1,4	1,8	5,2	-1,3	-6,1	0,8
38	-4,8	-1,3	-2,4	5,9	-2,9	-0,3	-0,5	-0,2	0,0	3,3	-0,4	4,3	0,7
39	4,6	3,6	-0,3	-4,6	-0,7	0,9	-0,8	-1,1	-3,7	-3,1	1,4	1,0	4,2
40	0,8	0,3	2,6	-7,5	-1,9	5,4	-2,1	0,7	-1,3	-5,1	6,9	5,0	-1,8
41	-3,9	-0,1	4,4	-4,6	-4,8	-1,5	2,5	0,1	3,1	0,7	6,1	-0,5	0,0
42	-4,4	-0,2	-2,4	0,3	0,5	0,6	-1,2	-0,7	1,7	2,3	2,2	1,5	0,9
43	-2,1	-2,5	-5,3	4,7	2,7	2,9	0,3	-2,8	-3,6	3,5	3,3	2,1	-1,8
44	1,2	-0,6	0,8	-3,3	-0,4	0,8	-0,1	1,5	0,9	1,8	4,1	-3,3	-2,5
45	-0,2	-1,2	4,7	-6,9	-3,8	-1,5	-0,9	7,8	6,2	1,7	0,6	0,0	-4,8
46	3,1	0,9	-2,4	0,2	-4,8	-1,7	-5,7	-0,2	0,7	3,6	3,4	3,2	0,9
47	-3,3	-4,8	-2,6	2,4	1,6	3,8	1,6	-3,1	1,0	2,8	2,3	-2,0	1,4
48	0,2	2,9	-0,9	-6,0	-4,8	-6,2	6,6	8,2	-0,7	1,9	-1,7	0,5	1,9

Bijlage III.D: cijfers fitting tijdreeks model(deel II)

Analyse van de relatieve afwijkingen van model 3.2

$$\frac{\left(X_t - \frac{X_t}{e^{Y_t}}\right)}{\left(\frac{X_t}{e^{Y_t}}\right)} * 100\%$$

Jaar week	'01	'02
1	3,9	-3,6
2	3,0	-2,8
3	-0,1	0,2
4	1,4	-1,2
5	-0,5	0,6
6	-2,8	3,1
7	-3,2	3,5
8	6,0	-5,6
9	-2,6	2,8
10	-0,8	0,9
11	0,0	0,2
12	-1,8	2,0
13	-2,2	2,4
14	0,1	0,2
15	-0,7	0,8
16	-2,0	0,1
17	-1,8	2,0
18	2,2	-2,1
19	0,1	-2,7
20	-4,6	4,9
21	1,0	8,4
22	-2,5	2,7
23	-9,4	0,1
24	0,9	-0,7
25	0,3	-0,1
26	-2,2	2,3
27	1,4	-1,3
28	-2,3	2,5
29	2,9	-2,7
30	3,1	-2,9
31	0,7	-0,6
32	-0,3	0,4
33	0,0	0,1
34	0,0	0,1
35	0,0	0,1
36	1,9	-1,7
37	-2,4	2,6
38	-2,0	2,1
39	3,9	-3,6
40	1,6	-1,4
41	-0,4	0,6
42	2,3	-2,1
43	2,2	-2,0
44	-0,3	0,5
45	1,6	-1,4
46	1,3	-1,2
47	0,9	-0,8
48	3,4	-3,2

Bijlage III.D: cijfers fitting tijdreeks model(deel III)

Analyse van de relatieve afwijkingen van model 3.3

$$\frac{\left(X_t - \frac{X_t}{e^{R_t}}\right)}{\left(\frac{X_t}{e^{R_t}}\right)} * 100\%$$

Jaar week	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
1	-	-8,3	-6,2	10,3	11,3	-4,6	-11,7	-11,1	36,5	14,3	8,1	-8,6	-23,2
2	-3,0	1,1	4,2	-4,3	-6,1	0,3	1,3	1,1	4,9	-1,3	-5,5	2,8	-0,5
3	-4,3	-5,9	2,8	-5,5	3,1	3,9	3,6	-2,9	-2,6	3,2	0,1	-0,7	2,0
4	-5,8	-4,5	-11,4	0,7	3,9	6,0	1,8	4,4	-4,0	2,5	1,3	0,1	3,8
5	0,3	2,9	1,5	-4,5	0,2	-1,2	-4,9	2,8	4,6	1,1	-4,1	-4,3	4,7
6	-1,1	3,9	3,2	-3,1	2,6	3,4	-4,5	-1,1	3,6	2,2	-8,0	-4,8	3,1
7	-0,6	-3,9	1,5	-4,7	1,0	5,7	0,0	-4,4	11,2	-5,3	-2,4	-1,9	3,9
8	5,1	2,2	-7,0	0,7	4,1	2,4	-8,0	-3,8	-2,3	4,2	4,4	-2,5	0,5
9	5,5	5,6	2,4	-1,4	0,2	-2,5	-2,7	5,4	-0,9	-3,9	-2,5	-7,0	1,3
10	7,4	3,5	-3,4	-5,9	-3,9	-0,5	1,7	2,8	3,3	-2,2	-3,8	-2,6	3,0
11	-5,1	-1,1	-1,4	1,0	5,3	-7,9	-1,1	0,7	0,5	8,4	3,4	-0,4	-2,9
12	-3,6	5,9	-6,9	0,9	2,3	-0,1	-0,3	-1,9	-0,9	2,2	2,1	-3,1	2,4
13	0,1	3,1	-0,2	-4,1	-2,9	-0,9	4,0	3,5	7,9	-6,3	-3,1	-4,7	3,5
14	7,6	4,0	4,3	-1,8	-2,0	-3,6	-2,1	-3,7	5,4	-3,6	-1,8	2,3	-5,0
15	-5,0	-5,4	4,7	-6,6	7,4	-4,2	-1,9	-7,1	7,8	3,8	7,4	-2,9	2,8
16	-3,7	-0,2	-6,7	0,6	2,1	-2,3	-4,8	0,2	2,2	4,9	4,0	-2,7	6,4
17	9,1	6,1	-1,2	-7,8	-11,1	-6,5	7,4	12,1	9,2	-14,2	-19,5	10,2	12,9
18	-24,6	-7,9	-5,3	15,3	35,9	7,8	-7,1	-7,1	-9,5	9,8	11,7	-3,9	-11,5
19	12,0	2,8	-3,7	-6,2	-2,0	-1,2	0,8	-1,3	-1,5	4,5	8,4	-1,2	-0,6
20	-2,2	3,1	-0,4	0,7	5,4	-4,3	-1,2	2,3	-2,7	7,4	-3,9	-3,6	-0,6
21	-1,7	-0,3	-5,4	-0,8	3,7	6,0	-4,5	2,5	-1,8	1,4	-0,5	-1,8	3,4
22	-3,5	5,3	6,1	-1,9	-3,7	-6,9	3,3	0,3	-0,8	0,9	-2,2	-2,1	5,7
23	5,1	4,7	-2,0	-3,7	-5,1	2,8	-2,4	-1,0	6,9	-1,1	-4,8	2,0	-1,0
24	1,3	-3,0	2,7	-1,6	-2,3	-0,3	-2,3	-1,8	0,5	3,3	3,4	-0,7	0,7
25	-2,2	-2,0	-1,4	-1,0	-5,4	-2,2	2,7	1,0	2,2	4,1	3,0	2,3	-1,1
26	2,3	4,9	3,5	-7,1	-7,8	-3,0	1,9	3,7	6,4	-5,8	0,9	2,8	-1,9
27	2,7	2,8	6,9	-4,7	-8,5	-0,9	-2,1	3,2	2,8	-4,9	6,9	0,5	-3,8
28	-0,8	-5,4	2,2	-2,4	0,2	0,1	-1,5	-0,4	3,0	1,4	5,3	1,6	-3,2
29	0,1	4,1	5,6	-1,5	-3,9	-2,9	-0,7	2,7	3,3	-2,2	-2,2	0,3	-2,4
30	8,5	3,7	1,8	2,5	0,2	0,3	-4,1	0,7	-1,0	-3,4	-1,7	-2,7	-4,3
31	7,7	7,1	8,5	0,7	-0,9	1,6	-2,4	2,0	-1,0	-3,1	-6,3	-4,0	-8,6
32	3,6	3,3	3,0	1,3	1,6	3,1	0,0	-0,6	-4,8	1,2	-4,7	-5,4	-1,4
33	-0,2	-1,9	-2,1	10,1	1,1	3,2	1,7	-2,0	-4,4	4,2	-4,1	-4,3	-0,7
34	-5,3	-0,6	-4,8	9,3	10,3	2,2	-5,1	-5,3	-7,4	5,7	1,5	3,7	-2,4
35	-0,2	0,8	-1,7	-0,7	-0,8	0,6	2,1	0,8	-0,6	1,6	-2,6	2,1	-1,5
36	2,8	2,3	-0,7	-5,2	-4,6	3,0	1,8	0,9	-2,9	4,9	-4,0	1,4	0,9
37	-0,8	-0,4	0,5	-1,8	1,4	-2,6	2,5	1,3	1,9	5,0	-1,3	-6,2	0,7
38	-4,9	-1,3	-2,5	5,8	-3,0	-0,2	-0,7	-0,3	-0,2	3,1	-0,6	4,6	0,6
39	4,6	3,6	-0,3	-4,9	-0,7	0,9	-0,8	-1,1	-3,8	-3,2	1,2	0,7	4,1
40	0,3	0,1	2,4	-7,4	-2,0	5,3	-2,1	0,7	-1,3	-4,9	6,6	4,9	-2,0
41	-4,1	-0,1	4,1	-4,3	-4,9	-1,8	2,5	0,1	3,0	1,0	5,6	-0,7	0,1
42	-4,4	-0,3	-2,7	0,3	0,6	0,7	-1,4	-0,7	1,4	2,3	1,8	1,6	0,9
43	-2,1	-2,5	-5,3	4,5	2,5	2,9	0,3	-2,7	-3,8	3,4	3,2	2,0	-1,9
44	1,1	-0,6	1,0	-3,6	-0,6	0,7	-0,1	1,6	1,0	1,7	4,0	-3,4	-2,5
45	-0,5	-1,2	4,5	-6,9	-3,9	-1,5	-0,9	7,7	6,0	1,7	0,4	0,2	-4,7
46	2,9	0,9	-2,7	0,4	-4,7	-1,6	-5,7	-0,6	0,3	3,6	3,4	3,2	1,1
47	-3,6	-4,9	-2,6	2,2	1,7	3,9	1,9	-3,1	0,9	2,7	2,2	-2,1	1,2
48	0,1	3,0	-0,9	-6,3	-5,0	-6,4	6,4	8,4	-0,8	1,8	-1,7	0,6	1,7

Bijlage III.D: cijfers fitting tijdreeks model(deel IV)

Analyse van de relatieve afwijkingen van model 3.4

$$\frac{\left(X_t - \frac{X_t}{e^{Y_t}}\right)}{\left(\frac{X_t}{e^{Y_t}}\right)} * 100\%$$

Jaar week	'01	'02
1	3,6	-3,5
2	2,7	-2,6
3	-0,4	0,4
4	1,1	-1,1
5	-0,8	0,8
6	-3,1	3,2
7	-3,5	3,6
8	5,7	-5,4
9	-2,9	2,9
10	-1,1	1,1
11	-0,3	0,3
12	-2,1	2,1
13	-2,5	2,5
14	1,2	-1,2
15	-0,9	1,0
16	-3,0	3,1
17	-2,1	2,1
18	2,0	-1,9
19	-2,1	2,2
20	-4,8	5,1
21	-5,4	5,7
22	-2,8	2,8
23	0,3	-0,3
24	0,6	-0,6
25	0,0	0,0
26	-2,4	2,5
27	1,1	-1,1
28	-2,6	2,7
29	2,6	-2,6
30	2,8	-2,8
31	0,4	-0,4
32	-0,6	0,6
33	-0,3	0,3
34	-0,2	0,2
35	-0,3	0,3
36	1,6	-1,6
37	-2,7	2,7
38	-2,2	2,3
39	3,6	-3,5
40	1,3	-1,3
41	-0,7	0,7
42	2,0	-2,0
43	1,9	-1,9
44	-0,6	0,6
45	1,3	-1,3
46	1,0	-1,0
47	0,7	-0,7
48	3,2	-3,1

Bijlage III.E: Cijfers voorspellingen(deel I)

Onderstaand de cijfers behorende bij de voorspellingen van model 3.1

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling	
						Model 3.1	WAP
Week 1	90,5	[90,2	, 90,8]	81,1	-10,4 %	
Week 2	97,9	[97,5	, 98,3]	97,8	-0,1 %	6,4 %
Week 3	97,0	[96,6	, 97,4]	95,8	-1,2 %	2,9 %
Week 4	102,7	[102,4	, 103,1]	98,4	-4,2 %	-4,0 %
Week 5	103,5	[103,1	, 103,8]	103,9	0,4 %	5,0 %
Week 6	98,2	[97,8	, 98,5]	94,2	-4,0 %	1,6 %
Week 7	95,9	[95,6	, 96,3]	97,0	1,1 %	0,3 %
Week 8	98,6	[98,3	, 99,0]	97,1	-1,6 %	2,3 %
Week 9	98,7	[98,3	, 99,0]	98,4	-0,3 %	3,3 %
Week 10	94,5	[94,1	, 94,8]	98,4	4,1 %	3,3 %
Week 11	91,6	[91,2	, 91,9]	90,8	-0,8 %	1,2 %
Week 12	96,4	[96,0	, 96,7]	90,8	-5,8 %	-4,4 %
Week 13	99,2	[98,8	, 99,6]	93,7	-5,5 %	-4,3 %
Week 14	96,5	[96,1	, 96,8]	87,8	-9,0 %	-3,6 %
Week 15	96,6	[96,2	, 96,9]	86,0	-10,9 %	-2,7 %
Week 16	75,9	[75,6	, 76,2]	92,5	21,8 %	0,3 %
Week 17	78,0	[77,7	, 78,3]	85,8	10,0 %	10,0 %
Week 18	89,5	[89,1	, 89,8]	78,5	-12,3 %	4,4 %
Week 19	96,6	[96,2	, 96,9]	86,4	-10,5 %	-2,8 %
Week 20	96,3	[95,9	, 96,6]	93,0	-3,4 %	-1,9 %
Week 21	82,1	[81,8	, 82,4]	97,5	18,8 %	-2,6 %
Week 22	84,8	[84,5	, 85,2]	80,2	-5,4 %	-2,3 %
Week 23	83,2	[82,9	, 83,5]	92,7	11,4 %	-0,2 %
Week 24	82,8	[82,5	, 83,2]	78,1	-5,8 %	2,0 %
Week 25	93,7	[93,3	, 94,0]	94,0	0,3 %	2,9 %
Week 26	95,0	[94,7	, 95,4]	92,0	-3,2 %	5,2 %
Week 27	87,3	[87,0	, 87,6]	82,7	-5,3 %	3,1 %
Week 28	81,2	[80,9	, 81,6]	73,9	-9,1 %	2,6 %
Week 29	75,3	[75,0	, 75,6]	73,2	-2,8 %	3,2 %
Week 30	75,1	[74,8	, 75,4]	69,2	-7,8 %	-1,0 %
Week 31	73,0	[72,7	, 73,2]	65,4	-10,4 %	0,8 %
Week	73,7	[73,4	, 74,0]	68,7	-6,8 %	-1,2 %

32 Week								
33 Week	76,9	[76,6	,	77,2]	67,7	-12,0 %
34 Week	87,0	[86,7	,	87,4]	76,6	-12,0 %
35 Week	94,3	[94,0	,	94,7]	88,8	-5,9 %
36 Week	97,7	[97,3	,	98,0]	96,2	-1,4 %
37 Week	94,7	[94,3	,	95,0]	91,2	-3,7 %
38 Week	96,5	[96,2	,	96,9]	83,7	-13,3 %
39 Week	100,9	[100,5	,	101,3]	95,4	-5,5 %
40 Week	99,4	[99,0	,	99,8]	95,1	-4,3 %
41 Week	96,4	[96,0	,	96,8]	91,2	-5,4 %
42 Week	94,6	[94,2	,	94,9]	88,8	-6,1 %
43 Week	97,5	[97,1	,	97,8]	89,7	-8,0 %
44 Week	98,5	[98,1	,	98,9]	87,6	-11,1 %
45 Week	95,9	[95,5	,	96,2]	93,5	-2,5 %
46 Week	94,8	[94,4	,	95,1]	88,5	-6,6 %
47 Week	98,2	[97,8	,	98,5]	91,8	-6,5 %
48 Week	101,1	[100,7	,	101,4]	95,1	-5,9 %

Bijlage III.E: Cijfers voorspellingen(deel II)

Onderstaand de cijfers behorende bij de voorspellingen van model 3.2

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling	
						Model 3.2	WAP
Week 1	48,3	[48,2	, 48,5]	81,1	67,7 %	
Week 2	66,3	[66,1	, 66,4]	97,8	47,6 %	6,4 %
Week 3	65,8	[65,6	, 66,0]	95,8	45,6 %	2,9 %
Week 4	74,7	[74,5	, 74,9]	98,4	31,8 %	-4,0 %
Week 5	77,1	[76,9	, 77,3]	103,9	34,8 %	5,0 %
Week 6	64,4	[64,2	, 64,6]	94,2	46,4 %	1,6 %
Week 7	62,2	[62,0	, 62,4]	97,0	55,8 %	0,3 %
Week 8	57,4	[57,2	, 57,6]	97,1	69,1 %	2,3 %
Week 9	66,3	[66,1	, 66,5]	98,4	48,5 %	3,3 %
Week 10	65,9	[65,7	, 66,1]	98,4	49,3 %	3,3 %
Week 11	61,2	[61,0	, 61,4]	90,8	48,3 %	1,2 %
Week 12	60,8	[60,6	, 61,0]	90,8	49,2 %	-4,4 %
Week 13	68,5	[68,3	, 68,7]	93,7	36,8 %	-4,3 %
Week 14	60,3	[60,1	, 60,5]	87,8	45,7 %	-3,6 %
Week 15	61,9	[61,7	, 62,1]	86,0	39,0 %	-2,7 %
Week 16	54,0	[53,9	, 54,2]	92,5	71,1 %	0,3 %
Week 17	62,5	[62,3	, 62,6]	85,8	37,4 %	10,0 %
Week 18	51,2	[51,1	, 51,4]	78,5	53,2 %	4,4 %
Week 19	59,3	[59,1	, 59,5]	86,4	45,7 %	-2,8 %
Week 20	71,8	[71,6	, 72,0]	93,0	29,6 %	-1,9 %
Week 21	61,2	[61,1	, 61,4]	97,5	59,3 %	-2,6 %
Week 22	59,6	[59,4	, 59,7]	80,2	34,7 %	-2,3 %
Week 23	62,6	[62,4	, 62,8]	92,7	48,2 %	-0,2 %
Week 24	55,2	[55,0	, 55,4]	78,1	41,4 %	2,0 %
Week 25	64,4	[64,2	, 64,6]	94,0	46,0 %	2,9 %
Week 26	64,7	[64,5	, 64,9]	92,0	42,2 %	5,2 %
Week 27	57,7	[57,5	, 57,9]	82,7	43,2 %	3,1 %
Week 28	48,7	[48,6	, 48,9]	73,9	51,6 %	2,6 %
Week 29	50,6	[50,4	, 50,7]	73,2	44,7 %	3,2 %
Week 30	45,4	[45,3	, 45,6]	69,2	52,5 %	-1,0 %
Week 31	43,1	[42,9	, 43,2]	65,4	51,9 %	0,8 %
Week	49,9	[49,7	, 50,0]	68,7	37,7 %	-1,2 %

32 Week								
33 Week	45,8	[45,7	,	46,0]	67,7	47,6 %
34 Week	54,4	[54,3	,	54,6]	76,6	40,8 %
35 Week	67,5	[67,3	,	67,7]	88,8	31,6 %
36 Week	69,0	[68,8	,	69,2]	96,2	39,5 %
37 Week	62,8	[62,6	,	63,0]	91,2	45,3 %
38 Week	57,0	[56,8	,	57,2]	83,7	46,8 %
39 Week	67,6	[67,4	,	67,8]	95,4	41,0 %
40 Week	71,4	[71,2	,	71,6]	95,1	33,2 %
41 week	64,7	[64,5	,	64,9]	91,2	40,9 %
42 Week	61,8	[61,6	,	62,0]	88,8	43,8 %
43 Week	63,2	[63,0	,	63,4]	89,7	42,0 %
44 Week	65,4	[65,2	,	65,6]	87,6	33,9 %
45 Week	60,7	[60,5	,	60,9]	93,5	54,0 %
46 Week	62,4	[62,2	,	62,6]	88,5	42,0 %
47 Week	64,9	[64,7	,	65,1]	91,8	41,5 %
48	64,4	[64,2	,	64,5]	95,1	47,7 %

Bijlage III.E: Cijfers voorspellingen(deel III)

Onderstaand de cijfers behorende bij de voorspellingen van model 3.3

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling	
						Model 3.3	WAP
Week 1	90,6	[90,2	, 90,9]	81,1	-10,5 %	
Week 2	97,9	[97,5	, 98,2]	97,8	-0,1 %	6,4 %
Week 3	97,0	[96,6	, 97,3]	95,8	-1,2 %	2,9 %
Week 4	102,7	[102,3	, 103,1]	98,4	-4,2 %	-4,0 %
Week 5	103,4	[103,0	, 103,8]	103,9	0,5 %	5,0 %
Week 6	98,1	[97,8	, 98,5]	94,2	-4,0 %	1,6 %
Week 7	95,9	[95,5	, 96,2]	97,0	1,1 %	0,3 %
Week 8	98,6	[98,2	, 98,9]	97,1	-1,5 %	2,3 %
Week 9	98,6	[98,2	, 99,0]	98,4	-0,2 %	3,3 %
Week 10	94,4	[94,1	, 94,8]	98,4	4,2 %	3,3 %
Week 11	91,5	[91,2	, 91,9]	90,8	-0,8 %	1,2 %
Week 12	96,3	[96,0	, 96,7]	90,8	-5,8 %	-4,4 %
Week 13	99,1	[98,8	, 99,5]	93,7	-5,4 %	-4,3 %
Week 14	95,1	[94,8	, 95,5]	87,8	-7,7 %	-3,6 %
Week 15	95,5	[95,1	, 95,9]	86,0	-9,9 %	-2,7 %
Week 16	94,6	[94,2	, 95,0]	92,5	-2,3 %	0,3 %
Week 17	81,8	[81,5	, 82,2]	85,8	4,9 %	10,0 %
Week 18	89,4	[89,1	, 89,7]	78,5	-12,2 %	4,4 %
Week 19	95,9	[95,5	, 96,2]	86,4	-9,8 %	-2,8 %
Week 20	95,9	[95,5	, 96,2]	93,0	-3,0 %	-1,9 %
Week 21	101,5	[101,1	, 101,9]	97,5	-3,9 %	-2,6 %
Week 22	85,6	[85,2	, 86,0]	80,2	-6,3 %	-2,3 %
Week 23	94,9	[94,6	, 95,3]	92,7	-2,3 %	-0,2 %
Week 24	79,3	[79,0	, 79,7]	78,1	-1,6 %	2,0 %
Week 25	93,6	[93,2	, 94,0]	94,0	0,4 %	2,9 %
Week 26	94,9	[94,6	, 95,3]	92,0	-3,1 %	5,2 %
Week 27	87,2	[86,9	, 87,5]	82,7	-5,2 %	3,1 %
Week 28	81,2	[80,9	, 81,5]	73,9	-9,0 %	2,6 %
Week 29	75,3	[75,0	, 75,5]	73,2	-2,8 %	3,2 %
Week 30	75,0	[74,8	, 75,3]	69,2	-7,7 %	-1,0 %
Week 31	72,9	[72,6	, 73,2]	65,4	-10,3 %	0,8 %
Week	73,6	[73,3	, 73,9]	68,7	-6,7 %	-1,2 %

32 Week								
33 Week	76,8	[76,5	,	77,1]	67,7	-11,9 %
34 Week	86,9	[86,6	,	87,3]	76,6	-11,9 %
35 Week	94,2	[93,9	,	94,6]	88,8	-5,8 %
36 Week	97,6	[97,2	,	97,9]	96,2	-1,3 %
37 Week	94,6	[94,2	,	94,9]	91,2	-3,6 %
38 Week	96,4	[96,1	,	96,8]	83,7	-13,2 %
39 Week	100,8	[100,4	,	101,2]	95,4	-5,4 %
40 Week	99,3	[98,9	,	99,6]	95,1	-4,2 %
41 Week	96,3	[95,9	,	96,7]	91,2	-5,3 %
42 Week	94,4	[94,1	,	94,8]	88,8	-6,0 %
43 Week	97,4	[97,0	,	97,7]	89,7	-7,9 %
44 Week	98,4	[98,0	,	98,8]	87,6	-11,0 %
45 Week	95,7	[95,4	,	96,1]	93,5	-2,4 %
46 Week	94,7	[94,3	,	95,0]	88,5	-6,5 %
47 Week	98,1	[97,7	,	98,4]	91,8	-6,4 %
48 Week	100,9	[100,5	,	101,3]	95,1	-5,8 %

Bijlage III.E: Cijfers voorspellingen(deel I)

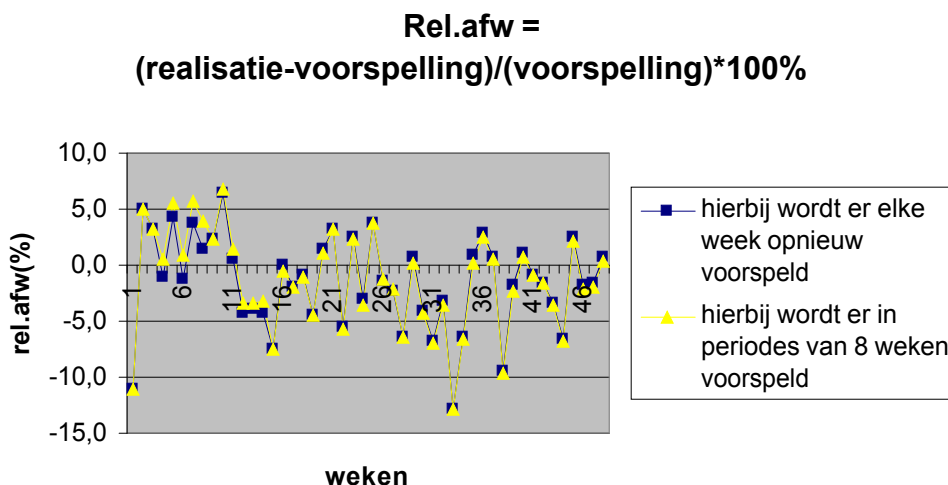
Onderstaand de cijfers behorende bij de voorspellingen van model 3.1

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling			
						Model 3.1		WAP	
Week 1	46,4	[46,3	, 46,6]	81,1	74,6	%		
Week 2	63,8	[63,7	, 64,0]	97,8	53,2	%	6,4	%
Week 3	63,3	[63,1	, 63,5]	95,8	51,4	%	2,9	%
Week 4	72,1	[71,9	, 72,3]	98,4	36,5	%	-4,0	%
Week 5	74,5	[74,3	, 74,7]	103,9	39,5	%	5,0	%
Week 6	61,8	[61,6	, 62,0]	94,2	52,4	%	1,6	%
Week 7	59,7	[59,6	, 59,9]	97,0	62,4	%	0,3	%
Week 8	55,0	[54,8	, 55,1]	97,1	76,5	%	2,3	%
Week 9	63,7	[63,5	, 63,9]	98,4	54,4	%	3,3	%
Week 10	63,4	[63,2	, 63,6]	98,4	55,1	%	3,3	%
Week 11	58,9	[58,7	, 59,1]	90,8	54,2	%	1,2	%
Week 12	58,4	[58,2	, 58,6]	90,8	55,5	%	-4,4	%
Week 13	66,0	[65,9	, 66,2]	93,7	42,0	%	-4,3	%
Week 14	57,1	[56,9	, 57,2]	87,8	53,9	%	-3,6	%
Week 15	59,4	[59,2	, 59,6]	86,0	44,8	%	-2,7	%
Week 16	60,1	[59,9	, 60,3]	92,5	53,8	%	0,3	%
Week 17	56,1	[55,9	, 56,3]	85,8	53,0	%	10,0	%
Week 18	49,0	[48,9	, 49,2]	78,5	60,0	%	4,4	%
Week 19	57,7	[57,6	, 57,9]	86,4	49,7	%	-2,8	%
Week 20	69,4	[69,2	, 69,5]	93,0	34,1	%	-1,9	%
Week 21	70,8	[70,6	, 71,0]	97,5	37,8	%	-2,6	%
Week 22	47,7	[47,5	, 47,9]	80,2	68,2	%	-2,3	%
Week 23	63,5	[63,3	, 63,7]	92,7	46,0	%	-0,2	%
Week 24	41,8	[41,7	, 42,0]	78,1	86,7	%	2,0	%
Week 25	62,0	[61,8	, 62,1]	94,0	51,7	%	2,9	%
Week 26	62,3	[62,2	, 62,5]	92,0	47,7	%	5,2	%
Week 27	55,4	[55,2	, 55,6]	82,7	49,2	%	3,1	%
Week 28	46,6	[46,5	, 46,8]	73,9	58,3	%	2,6	%
Week 29	48,6	[48,5	, 48,8]	73,2	50,6	%	3,2	%
Week 30	43,5	[43,4	, 43,6]	69,2	59,2	%	-1,0	%
Week 31	41,2	[41,1	, 41,4]	65,4	58,6	%	0,8	%
Week	48,1	[47,9	, 48,2]	68,7	42,9	%	-1,2	%

32 Week								
33 Week	43,9	[43,8	,	44,1]	67,7	54,0 % -2,7 %
34 Week	52,3	[52,1	,	52,4]	76,6	46,6 % -7,1 %
35 Week	65,1	[64,9	,	65,3]	88,8	36,4 % 0,4 %
36 Week	66,4	[66,3	,	66,6]	96,2	44,9 % 4,6 %
37 Week	60,4	[60,2	,	60,5]	91,2	51,1 % 1,7 %
38 Week	54,6	[54,4	,	54,8]	83,7	53,2 % -7,7 %
39 Week	65,0	[64,8	,	65,2]	95,4	46,8 % -1,3 %
40 Week	68,9	[68,7	,	69,1]	95,1	38,1 % 3,3 %
41 week	62,2	[62,0	,	62,4]	91,2	46,5 % -0,3 %
42 Week	59,3	[59,1	,	59,4]	88,8	49,8 % -3,1 %
43 Week	60,8	[60,6	,	60,9]	89,7	47,6 % 2,5 %
44 Week	63,0	[62,9	,	63,2]	87,6	39,0 % 0,6 %
45 Week	58,3	[58,1	,	58,5]	93,5	60,4 % 7,2 %
46 Week	60,0	[59,8	,	60,2]	88,5	47,6 % 0,5 %
47 Week	62,5	[62,3	,	62,7]	91,8	46,9 % 1,8 %
48	62,0	[61,8	,	62,2]	95,1	53,4 % 5,0 %

Bijlage IV.A: Analyse effect verschillende voorspelperiodes

In deze bijlage wordt er gekeken of er een groot verschil is, als er met het log-quasi-likelihood model elke week opnieuw wordt voorspeld of dat er elke 8 weken opnieuw wordt voorspeld. In onderstaand figuur is duidelijk te zien dat het verschil qua voorspellingen door een andere grootte van de voorspellingsperiode alleen in de weken 4 tot en met 8 waarneembaar is.



Figuur IV.A.1: De relatieve afwijkingen bij de voorspellingen van het log-lineair model voor het jaar 2003.

De cijfers van figuur IV.A.1 zijn weergegeven in bijlage IV.E (deel II en deel III). Uit deze cijfers blijkt dat het grootste verschil in de relatieve afwijkingen in week 8 plaatsvindt. Hierbij is het verschil van 2,5% ten voordele van de voorspelling waarbij er elke week opnieuw voorspeld wordt. In tabel IV.A.1 is een korte samenvatting van de cijfers gegeven.

	Week 1 t/m week 8	Week 9 t/m wk. 48	Week 1 t/m wk. 48
Gemiddeld verschil	0.82%	0.29%	0.37%
Gemiddelde van (KT _i -MT _i)	-0.64%	0.05%	-0.07%

KT_i = absolute rel.afwijking waarbij er per week opnieuw voorspeld wordt

MT_i = absolute rel.afwijking waarbij er elke 8 weken opnieuw wordt voorspeld

Tabel IV.A.1: Samenvatting van de cijfers van bijlage IV.E(deel II en deel III) waarbij de samenvatting gespecificeerd is voor een analyse van het verschil.

Met behulp van het gemiddelde van (KT_i – MT_i) kan er geconcludeerd worden dat het voorspellen per week duidelijk beter is in de periode van week 1 tot en met week 8. Gemiddeld heeft de voorspelling per week in de periode week 1 tot en met week 8 een betere relatieve afwijking van 0.64%.

In de periode van week 9 tot en met week 48 is het echter duidelijk dat voorspellingen per 8 weken de voorkeur hebben. De gemiddelde verbetering van de relatieve afwijking is in deze periode maar 0.05%. Het scheelt echter wel veel rekenwerk en daarom is er voor gekozen om in dit onderzoek de periode week 1 tot en met week 8 per week te voorspellen, en de periode week 9 tot en met 48 per 8 weken te voorspellen.

Bijlage IV.B: De regressie-waarnemingen x_i

Toelichting

De waarnemingen x_i zijn te splitsen in dichotome nominale variabelen, waarbij de variabele 2 categorieën heeft, en polytome nominale variabelen. Polytome nominale variabelen beschikken over meer dan 2 categorieën. Door hierbij dummy variabelen te gebruiken die een 1 voor een bepaalde categorie hebben en een 0 voor de overige categorieën, kan de waarde van de invloed van een categorie-onderdeel bepalen zonder al van tevoren een bepaalde waarde aan dit onderdeel gegeven te hebben door het gebruik van 2, 3, 4, etc.

Variabelen model 4.2

De dichotome nominale variabelen in dit model zijn de variabelen:

- Pasen
 - 0= geen officiële feestdag dankzij Pasen,
 - 1= wel een officiële feestdag dankzij Pasen
- Hemelvaartsdag
 - 0= geen officiële feestdag dankzij Hemelvaartsdag,
 - 1= wel een officiële feestdag dankzij Hemelvaartsdag
- Pinksteren
 - 0= geen officiële feestdag dankzij Pinksteren,
 - 1= wel een officiële feestdag dankzij Pinksteren

De polytomous nominale variabelen in model 4.2 zijn de variabelen:

- Jaar: 0=2001, 1=2002, 2=2003³⁹
- Week: 0=week 1,...,47=week 48
- Vakantiespreiding⁴⁰
 - 0= er is geen zomervakantie bij het basisonderwijs
 - 1= alleen het basisonderwijs in regio noord is op zomervakantie
 - 2= alleen het basisonderwijs in de regio's noord en midden zijn op zomervak.
- Maandwisseling⁴¹
 - 0= geen maandwisseling
 - 1= einde september

Variabelen model 4.1

Naast de variabelen die in model 4.2 gebruikt worden, zijn de volgende variabelen toegevoegd bij model 4.1. Dit is mogelijk doordat er bij model 4.1 meer datapunten zijn.

- Valentijnsdag
 - 0= geen Valentijnsdag gedurende maandag tot en met zaterdag,
 - 1= wel Valentijnsdag gedurende maandag tot en met zaterdag
- Koninginnedag
 - 0= geen officiële feestdag dankzij Koninginnedag,
 - 1= wel een officiële feestdag dankzij Koninginnedag
- Bevrijdingsdag
 - 0= geen officiële feestdag dankzij Bevrijdingsdag,
 - 1= wel een officiële feestdag dankzij Bevrijdingsdag
- Mailmomenten
 - 0= geen toestemming in die week voor een 3^e ongeadresseerde partij,
 - 1= wel toestemming in die week voor een 3^e partij.
- Week: 48=week 49, 49= week 50, 50=week 52⁴²
- Kerst
 - 0= 0 officiële feestdagen gedurende maandag t/m zaterdag dankzij Kerst,

³⁹ Wordt niet gebruikt bij het analyseren van de fitting. Omdat deze variabele wel gebruikt wordt voor voorspellingen, is deze variabele wel opgenomen in het model.

⁴⁰ De invloed van de overige 'vakantiespreidings-vormen' kunnen niet met model 4.2 bepaald worden door het tekort aan datapunten.

⁴¹ Van de overige maanden is het nummer van de week waarin de maandwisseling voorkomt in de jaren 2001 en 2002 hetzelfde en daarom is de invloed van een van deze maandwisselingen al m.b.v. de invloed van dat weeknummer bepaald.

⁴² Daar in elk jaar een extreem in week 51 voorkomt, is de weekvariabele voor deze week weggelaten.

- 1= 1 officiële feestdag gedurende maandag t/m zaterdag dankzij Kerst,
 2= 2 officiële feestdagen gedurende maandag t/m zaterdag

De polytomous nominale variabelen 'Jaar', 'vakantiespreiding' en 'maandwisseling' zijn door de volgende polytomous nominale variabelen vervangen.

Jaar: 0=1990,...,13= 1993

- Vakantiespreiding⁴³

- 0= er is geen zomervakantie bij het basisonderwijs
 1= alleen het basisonderwijs in regio noord is op zomervakantie
 2= alleen het basisonderwijs in regio midden is op zomervakantie
 3= alleen het basisonderwijs in regio zuid is op zomervakantie
 4= alleen het basisonderwijs in de regio's noord en midden zijn op zomervak.
 5= alleen het basisonderwijs in de regio's noord en zuid zijn op zomervak.
 6= alleen het basisonderwijs in de regio's midden en zuid zijn op zomervak.

- Maandwisseling⁴⁴

- | | |
|------------------------|--------------------|
| 0= geen maandwisseling | 5= einde juli |
| 1= einde januari | 6= einde augustus |
| 2= einde maart | 7= einde september |
| 3= einde april | 8= einde november |
| 4= einde juni | 9= einde december |

Niet-aanwezige variabelen

Interessante waarnemingen zoals de economische groei in een bepaald jaar en/of de invloed van een kwartaaleinde, worden op indirecte wijze meegenomen in het model via respectievelijk de variabelen 'jaar', 'week' en/of 'maandwisseling'.

De interactie tussen een willekeurig jaar en een willekeurig week is niet meegenomen in het model doordat dit de voorspellingen verslechtert. Het voordeel van het meenemen van de jaar-week interactie is een betere fitting⁴⁵. Beide gevolgen is logischerwijs te verklaren door het volgende: indien het model een waarde moet schatten voor bijvoorbeeld de interactie tussen het jaar 1998 en week 23 dan wijst het model een waarde aan deze variabele toe, zodat het verschil tussen de schatter voor week 23 van het jaar 1998 en de realisatie dan zo goed als nul is. Doordat een goede fitting al grotendeels bereikt wordt via de interactie-variabele, neemt de waarde van de schatting van onder andere de invloed van het jaar $\hat{\beta}_{jaar1998}$ en $\hat{\beta}_{week23}$ af. Deze minder goede schattingen worden echter wel direct dan wel indirect gebruikt voor het maken van voorspellingen, wat er dus voor zorgt dat de voorspellingen verslechteren. Daar het toevoegen van de jaar-week interacties niet overeenkomstig is met het doel⁴⁶ van dit onderzoek, worden deze variabelen niet in het model opgenomen.

⁴³ Doordat bijvoorbeeld de periode dat alle regio's op zomervakantie zijn, elk jaar gedurende de weken 30, 31 en 32 is, is de invloed van 'alle regio's zijn op zomervakantie' meegenomen in het model via de dummy-variabelen week 30, 31 en 32 van de week-variabele.

⁴⁴ Van de overige maanden is het nummer van de week waarin de maandwisseling voorkomt elk jaar hetzelfde en daarom is de invloed van een van deze maandwisselingen al m.b.v. de invloed van dat weeknummer bepaald.

⁴⁵ Het woord fitting wordt in de wiskunde gebruikt om aan te geven hoe goed de geschatte waarden lijken op de gerealiseerde waarden.

⁴⁶ Doel van dit onderzoek is het bepalen van de variabelen die van invloed zijn op het landelijke weekverkeer en het ontwikkelen van een model om het landelijke weekverkeer te voorspellen.

Voorbeeld database

Van een gedeelte van de uiteindelijke data-invoer is in tabel A.1 een korte weergave gegeven.

hoeveelPost	hulpJaar1991	hulpJaar1992	hulpJaar1993	hulpJaar1994	hulpJaar1995	hulpJaar1996	hulpJaar1997	hulpJaar1998	hulpJaar1999	hulpJaar2000	hulpJaar2001	hulpJaar2002	hulpJaar2003	hulp/wk2	hulp/wk3
31174707	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
31713739	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
82471945	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
77071219	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
71343880	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
hulp/wk24	hulp/wk25	hulp/wk26	hulp/wk27	hulp/wk28	hulp/wk29	hulp/wk30	hulp/wk31	hulp/wk32	hulp/wk33	hulp/wk34	hulp/wk35	hulp/wk36	---	Pasen	Pinksteren
0	1	0	0	0	0	0	0	0	0	0	0	0	---	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	---	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	---	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	---	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	---	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	---	0	0
vakN	vakM	vakZ	vakNM	vakNZ	vakMZ	m/wis1	m/wis3	m/wis4	m/wis6	m/wis7	m/wis8	m/wis9	m/wis11	m/wis12	naalMomenten
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Tabel IV.A.1: Data van de weken 25 tot en met 29 in het jaar 2002

Validatie

Met behulp van de t-ratio of de deviance wordt er voor de dichotome en/ of polytome nominale variabelen per variabele bepaald of deze een significante invloed heeft op het wekelijkse landverkeer. Indien dit het geval is, dan dient deze variabele meegenomen te worden in het model. Uit dit onderzoek blijkt dat de genoemde variabelen per model allemaal een significante invloed hebben en dat dus alle variabelen in het model horen.

De t-ratio is gelijk aan

$$\frac{\hat{\beta}_i}{\sqrt{\text{geschatte Var}(\hat{\beta}_i)}}.$$

Indien de absolute waarde van de t-ratio voor β_i groter is dan $t_{n-p-1;1-(\alpha/2)}$, dan is de regressie waarneming x_i significant bij een niveau α .

Hierbij is

n het aantal waarnemingen
p het aantal parameters

In tabel IV.A.2 is per regressie waarneming waar de absolute waarde van de t-ratio niet groter is dan de bijbehorende $t_{n-p-1;1-(\alpha/2)}$ het vakje grijs getint.

Hieruit blijkt dat:

- koninginndag volgens de t-ratio geen significante invloed heeft. Dit kan verklaard worden door het feit dat koninginndag bijna altijd in week 18 valt.
- Het jaar 1992 en 1993 bij model 4.1 volgens de t-ratio geen significante invloed hebben.
- Veel weken volgens de t-ratio geen significante invloed hebben. Bij model 4.2 zijn er meer weken met een significante invloed dan bij model 4.1.
- Het specificeren van de spreiding van vakanties volgens de t-ratio geen significante invloed heeft
- Veel maandwisselingen volgens de t-ratio geen significante invloed hebben.

Met de deviance kan er tevens per regressiewaarneming bepaald worden of er sprake is van een significante invloed. Tevens kan er met de deviance bepaald worden of een groep variabelen een significante invloed.

De deviance meet het verschil tussen het model 4.1 respectievelijk het model 4.2 en het model waarbij het aantal variabelen gelijk is aan het aantal waarnemingen. Bij dit laatste model zal de fitting logischerwijs perfect zijn.

De deviance D is gelijk aan

$$\frac{D}{\phi} = 2[l(\tilde{\beta}, Y) - l(\hat{\beta}, Y)]$$

Hierbij is

ϕ	gelijk aan de dispersion-parameter
$l(\tilde{\beta}, Y)$	gelijk aan de meest aannemelijke functie van het model waarbij het aantal variabelen gelijk is aan het aantal waarnemingen
$l(\hat{\beta}, Y)$	gelijk aan de meest aannemelijke functie van model 4.1 respectievelijk model 4.2

De deviance is bij benadering chikwadraat verdeeld met $n-p-1$ vrijheidsgraden. Hierdoor kan er getoetst worden of een kleiner model beter is.

Indien

$$\sum_{i=j}^{j+m} Deviance_j \leq \chi_{m,1-\alpha}$$

dan wordt de nulhypothese dat het kleinere model beter is niet verworpen.

Hierbij is

j	een variabele die de waarde 1,...,p kan aannemen
m	het verschil in aantal variabelen tussen het gewone model en het kleinere model
α	het significantie-niveau welke gelijk gesteld is aan 0.05

Door de deviances die zijn weergegeven in tabel IV.A.2 te vergelijken met $\chi_{1,0.95} = 3.84$ wordt er per variabele gekeken of deze eventueel weggelaten kan worden.

Omdat de nulhypothese verworpen wordt, kan er geconcludeerd worden dat het weglaten van 1 variabele geen verbetering in de fitting oplevert.

Omdat de nulhypothese voor elke variabele verworpen wordt indien er per variabele geanalyseerd wordt, zal de nulhypothese ook verworpen worden indien er per groep geanalyseerd wordt. Een vermeerdering van de m met q variabelen zal namelijk een $\chi_{m,1-\alpha}$ opleveren die kleiner is dan $q^* \chi_{1,0.95}$.

Omdat alle variabelen volgens de deviance van invloed zijn op de hoeveelheid post, zullen de aangegeven variabelen meegenomen worden in de modellen.

	Geschatte β :		t-ratio		Deviance	
			t _{647-86-1,0,975=1,96}	t _{96-58-1,0,975=1,96}		
Intercept	18,1337	17,9747	469,91	677,41	843.487.649	75.989.686
Nieuwjaarsdag	0,0279		0,61		11.461	
Valentijnsdag	0,0307		1,09		2.320.883	
Hemelvaartsdag	-0,1304	-0,2215	-6,69	-5,60	9.468.843	2.320.495
Pasen	-0,1580	-0,2341	-8,33	-6,83	24.809.106	3.230.100
Pinksteren	-0,1015	-0,1215	-5,49	-3,10	6.730.910	328.551
koninginnedag	-0,0332	-0,0455	-1,03	-0,92	6.399.096	965.562
Bevrijdingsdag	-0,0969		-2,43		409.756	
1 ^e kerstdag	-0,0187		-0,25		89.172	
2 ^e kerstdag	-0,0285		-0,43		5.382.724	
Jaar 1991	0,0219		1,94		25.552.640	
Jaar 1992	0,0231		2,04		25.550.333	
Jaar 1993	0,0207		1,84		34.862.445	
Jaar 1994	0,0558		4,96		19.518.218	
Jaar 1995	0,0658		5,88		19.423.059	
Jaar 1996	0,0729		6,59		17.691.295	
Jaar 1997	0,1029		9,28		6.448.426	
Jaar 1998	0,1687		15,37		1.517.866	
Jaar 1999	0,1839		16,80		8.806.965	
Jaar 2000	0,1795		16,09		8.258.107	
Jaar 2001	0,1698		15,47		12.423.955	
Jaar 2002	0,1923	-0,0383	17,75	-5,42	76.023.460	2.625.998
Week 2	0,0496	0,2288	1,23	6,49	2.514.444	55.891
Week 3	0,0408	0,2696	1,01	7,72	1.751.634	593.446
Week 4	0,0989	0,2977	2,45	8,57	11.536.748	1.347.302
Week 5	0,0989	0,3078	1,90	8,88	13.394.003	1.757.045
Week 6	0,0510	0,2810	1,26	8,06	3.575.560	1.004.530
Week 7	0,0032	0,2656	0,07	7,60	102.639	681.899
Week 8	0,0514	0,2281	1,27	6,47	3.479.245	130.880
Week 9	0,0546	0,2782	1,35	7,98	4.466.790	1.046.811
Week 10	0,0122	0,2465	0,30	7,02	525.761	407.042
Week 11	-0,0178	0,1959	-0,44	5,52	62.856	113
Week 12	0,0314	0,2454	0,78	6,99	2.023.158	412.209
Week 13	0,0633	0,2570	1,34	7,34	5.954.440	661.001
Week 14	0,0001	0,2839	0,00	7,35	133.056	467.260
Week 15	0,0214	0,2560	0,53	7,31	1.355.981	704.118
Week 16	0,0199	0,2930	0,48	7,54	1.433.515	1.161.064
Week 17	0,1216	0,3308	2,93	9,59	3.055.120	3.625.685
Week 18	0,1146	0,1585	2,39	3,67	1.872.436	38.369
Week 19	-0,0015	0,2837	-0,04	7,17	13.934	494.357
Week 20	0,0153	0,2233	0,38	6,33	835.473	303.785
Week 21	0,0418	0,3100	1,00	6,26	1.526.639	433.600
Week 22	0,0653	0,2940	1,61	8,46	7.313.559	2.279.254
Week 23	0,0007	0,2569	0,02	6,39	451.681	662.457
Week 24	-0,0119	0,2015	-0,29	5,68	136.783	146.174
Week 25	0,0028	0,2307	0,07	6,55	786.681	598.028
Week 26	0,0319	0,2239	0,71	6,34	1.935.259	516.010
Week 27	-0,0986	0,2052	-1,43	4,82	1.882.576	109.071
Week 28	-0,0829	0,0747	-1,08	2,04	13.566.105	1.151.309
Week 29	-0,1599	0,0229	-2,08	0,62	37.867.229	2.960.177
Week 30	-0,2219	-0,0087	-5,35	-0,23	43.189.362	4.657.097
Week 31	-0,2529	-0,0489	-5,11	-1,30	60.092.679	7.320.779
Week 32	-0,2371	-0,0575	-5,70	-1,52	61.959.994	8.650.171
Week 33	-0,1408	-0,0159	-1,84	-0,34	49.800.045	6.995.660
Week 34	-0,1062	0,1231	-1,52	3,41	13.108.239	1.462.636
Week 35	-0,0048	0,2126	-0,06	6,01	1.558.545	29.979
Week 36	0,0437	0,2780	1,08	7,97	19.070	458.462
Week 37	0,0105	0,2246	0,26	6,36	1.501.998	449
Week 38	0,0306	0,2188	0,76	6,19	409.356	2.724
Week 39	0,0830	0,3125	1,99	8,17	695.307	1.907.611
Week 40	0,0671	0,2703	1,62	7,06	160.018	874.365
Week 41	0,0293	0,2672	0,72	7,64	390.317	713.971
Week 42	0,0094	0,2549	0,23	7,27	1.973.241	626.833
Week 43	0,0396	0,2237	0,98	6,34	292.972	236.387
Week 44	0,0495	0,2040	1,23	5,75	75.531	93.022
Week 45	0,0230	0,2263	0,57	6,41	1.727.754	473.733
Week 46	0,0117	0,2087	0,29	5,89	4.409.257	405.005
Week 47	0,0458	0,2218	1,13	6,28	2.289.684	1.237.160

Week 48	0,0581	0,2097	1,11	5,92	1.329.244	3.066.683
Week 49	0,1130		2,76		214.447	
Week 50	0,2710		5,33		6.644.897	
Week 52	0,1407		2,68		1.255.086	
Noord heeft zomervakantie	0,0535	-0,0292	0,91	-0,60	477.424	31.524
Midden heeft zomervakantie	0,0325		0,55		85.203	
Zuid heeft zomervakantie	0,0145		0,27		13.165	
Noord en Midden hebben zomervakantie	-0,0513	0,0015	-0,82	0,03	3.738	68
Noord en Zuid hebben zomervakantie	-0,0562		-0,84		4.803	
Midden en Zuid hebben zomervakantie	-0,0635		-0,94		199.158	
Eind januari	0,0059		0,16		6.346	
Eind maart	-0,0032		-0,11		471	
Eind april	-0,2152		-9,26		20.408.943	
Eind juni	-0,0216		-0,84		161.038	
Eind juli	0,0019		0,06		821	
Eind augustus	-0,0241		-0,63		90.878	
Eind september	-0,0167	0,0235	-0,82	0,74	159.584	46.988
Eind november	0,0185		0,50		63.431	
Eind december	-0,0688		-2,61		1.616.550	
Mailmoment	0,0052		0,30		21.942	

Tabel IV.B.1: Analyse van de t-ratio's en de devoances

Bijlage IV.C: cijfers fitting log-lineair modell(deel I)

Analyse van de respons residuen van model 4.1. In onderstaand tabel de resultaten van

$$\frac{(y - \hat{\mu})}{(\hat{\mu})} * 100\%$$

Jaar week	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
1	-1,2	-7,8	-3,5	14,5	11,1	-7,2	-15,2	-10,3	33,2	13,6	6,4	-11,0	-21,8
2	1,3	-0,1	5,4	-1,4	-7,0	0,0	2,2	0,0	2,6	-2,6	-4,0	5,2	-1,4
3	-1,7	-7,1	4,1	-3,0	1,8	3,5	4,6	-3,9	-4,6	1,7	1,3	1,7	1,0
4	-4,7	-6,1	-10,3	3,1	2,7	5,9	2,9	3,1	-6,2	0,5	2,4	2,4	2,9
5	0,1	1,1	2,2	-2,1	-0,6	-1,0	-3,7	2,0	2,1	0,5	-2,9	-2,0	4,1
6	-1,2	2,3	4,2	-0,9	1,9	3,7	-3,4	-1,7	1,4	1,1	-7,0	-2,6	2,8
7	-1,5	-5,8	2,0	-3,3	2,9	5,6	0,3	-5,6	8,4	-6,9	0,5	-0,5	3,2
8	5,2	1,1	-5,6	3,0	0,9	3,6	-6,7	-4,2	-3,3	3,2	2,1	-0,2	1,1
9	5,9	4,2	3,3	0,6	0,1	-1,5	-2,0	4,5	-2,4	-5,0	-1,8	-5,3	1,6
10	8,2	2,3	-2,6	-4,1	-4,1	0,3	2,2	2,1	1,5	-3,6	-3,4	-1,3	3,3
11	-3,9	-2,2	-0,9	2,4	4,7	-7,3	-0,7	0,0	-1,2	6,5	3,5	0,7	-2,7
12	-2,6	4,9	-6,3	2,5	2,2	0,3	0,3	-2,4	-2,4	0,5	2,5	-1,9	2,2
13	0,7	2,1	0,0	-2,4	-2,9	-0,5	4,6	2,9	5,8	-7,4	-2,7	-3,7	4,0
14	10,3	2,3	6,6	1,5	-17,9	-1,4	0,5	-5,2	6,4	-4,8	0,3	4,6	-4,0
15	-3,8	-5,7	5,3	-23,1	25,6	-4,0	-4,0	-7,5	6,8	2,2	7,1	-1,9	3,3
16	-8,6	-1,5	-6,7	18,2	1,5	-6,5	-4,9	-1,4	-0,1	2,5	3,8	-4,0	6,2
17	-0,4	-4,3	7,4	5,1	0,4	5,2	-2,0	1,1	-1,4	-1,4	-13,1	0,8	3,5
18	-8,2	0,4	-16,2	2,5	19,5	0,9	2,5	1,4	-1,2	-8,3	4,5	2,9	-1,7
19	14,1	-4,5	-1,1	-2,1	-18,3	1,5	4,1	-7,0	0,4	-0,5	11,5	2,5	-3,3
20	-1,2	3,5	0,5	-17,5	22,7	-3,4	-7,8	2,8	-2,6	6,9	-2,4	-1,7	1,3
21	-6,9	-5,2	-1,6	19,3	-7,3	3,7	0,0	-0,4	-4,5	-1,1	3,9	-2,8	3,6
22	-2,1	6,6	0,2	-15,9	10,0	-5,1	-1,4	1,9	-0,1	1,6	-5,9	0,4	8,7
23	-2,7	6,8	0,2	10,1	-1,5	-3,6	0,4	1,2	-7,9	0,1	-2,6	-1,5	2,6
24	2,1	-2,1	-6,1	0,5	-0,2	0,7	-0,9	-1,0	0,6	3,1	-2,3	1,0	3,0
25	-2,6	-2,4	-1,6	-0,3	-4,7	-2,5	2,9	0,5	1,0	2,8	2,7	2,8	-0,1
26	2,4	2,8	3,9	-5,8	-6,7	-2,7	2,9	1,8	3,7	-6,3	0,0	4,1	-0,2
27	0,0	2,3	6,5	-4,5	-7,0	0,2	-3,9	2,9	4,0	-6,5	8,5	2,9	-5,2
28	-1,2	-6,2	1,8	-1,6	0,5	-0,5	-1,2	-0,5	1,7	0,3	6,2	2,3	-2,6
29	-0,2	3,1	5,4	-0,7	-3,5	-3,3	-0,4	2,7	2,1	-3,1	-1,0	1,2	-1,8
30	8,5	3,7	2,7	2,9	-0,2	0,0	-3,5	1,1	-1,2	-5,1	-1,2	-1,5	-3,6
31	8,1	7,3	9,6	1,2	-0,9	1,3	-1,9	2,5	-1,3	-4,7	-5,9	-3,0	-8,2
32	4,0	3,4	4,0	1,5	0,9	2,5	-0,1	-0,5	-5,6	-1,2	0,0	-5,0	-1,8
33	1,3	-0,8	-0,8	10,6	0,2	2,5	2,7	-0,9	-5,3	2,0	-4,9	-4,5	-0,2
34	-4,6	1,5	-1,8	8,1	7,8	2,3	-4,8	-3,3	-6,8	1,6	-1,1	4,0	-2,6
35	0,6	3,2	1,4	-1,0	-2,2	1,2	2,6	3,0	-2,4	-1,8	-4,7	2,9	-1,4
36	3,0	2,4	0,0	-4,0	-4,6	3,1	1,8	0,8	-2,4	3,0	-4,8	1,7	0,3
37	-0,3	0,0	1,4	-0,6	1,4	-2,1	2,9	1,5	0,0	3,6	-2,2	-5,6	0,4
38	-4,6	-1,1	-1,8	6,9	-3,0	0,0	-0,2	-0,1	-2,0	1,8	-1,6	4,8	0,2
39	3,9	2,9	1,2	-2,8	0,1	2,0	-1,1	-1,7	-6,4	-3,4	1,1	2,1	3,0
40	1,6	1,3	2,3	-7,1	-2,9	4,8	-0,8	1,8	-2,5	-6,9	4,8	4,7	-1,2
41	-3,8	0,1	4,9	-3,7	-5,1	-1,2	2,9	0,2	0,7	-0,7	4,8	0,0	-0,1
42	-4,3	0,0	-1,8	0,9	0,1	1,2	-0,8	-0,5	-0,7	0,7	1,4	2,4	0,7
43	-2,3	-2,3	-4,5	5,1	2,1	3,5	0,8	-2,5	-5,8	1,9	2,9	2,9	-1,9
44	0,9	-0,4	1,5	-2,8	-0,9	1,5	0,5	1,8	-1,2	0,4	3,8	-2,4	-2,6
45	-0,7	-1,2	5,0	-6,3	-4,2	-0,7	-0,4	7,9	3,6	0,3	0,4	1,0	-5,0
46	2,6	0,8	-2,0	0,7	-5,2	-0,9	-5,2	-0,1	-1,7	2,2	3,3	4,1	0,6
47	-3,7	-4,9	-1,9	2,6	1,0	4,6	2,2	-2,5	-1,0	1,6	2,4	-1,5	0,4
48	-0,4	2,5	-0,5	-5,9	-5,7	-5,7	6,7	8,6	-1,0	0,2	-2,1	1,3	1,3
49	-2,3	-6,0	-4,7	7,8	2,2	0,4	-3,2	-1,3	-0,9	12,3	Extr.	Extr.	-5,2
50	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	2,0	Extr.	Extr.	Extr.	Extr.	Extr.
51	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.
52	-2,0	5,1	-2,1	2,0	-2,0	-5,6	22,1	Extr.	Extr.	0,0	-18,3	-5,5	4,6

Bijlage IV.C: cijfers fitting log-lineair modell(deel II)

Analyse van de deviance residuen van model 4.1. In onderstaand tabel de resultaten van

$$r_i^D = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i}$$

Jaar week	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
1	-103	-677	-303	1219	954	-655	-1418	-937	2911	1239	605	-1074	-2124
2	115	-12	481	-126	-645	-3	203	4	248	-255	-389	502	-132
3	-148	-641	362	-268	168	321	421	-366	-445	165	123	164	101
4	-430	-563	-961	281	251	546	271	298	-617	46	240	239	290
5	12	104	205	-198	-59	-95	-350	187	207	46	-295	-196	409
6	-104	209	376	-77	174	338	-313	-161	137	106	-690	-257	277
7	-131	-518	178	-296	258	505	32	-522	791	-673	51	-52	314
8	456	97	-506	271	83	329	-629	-396	-325	308	205	-24	103
9	518	378	294	55	5	-138	-189	417	-235	-490	-179	-522	157
10	706	200	-232	-367	-373	27	196	190	144	-349	-330	-121	312
11	-334	-188	-79	206	412	-658	-62	0	-114	608	325	64	-252
12	-232	432	-569	223	195	27	26	-221	-229	50	238	-182	217
13	63	185	0	-222	-271	-46	422	275	556	-734	-264	-365	387
14	875	188	568	131	-1645	-126	46	-439	591	-427	24	432	-354
15	-335	-505	467	-2132	2044	-361	-341	-700	642	207	670	-180	315
16	-708	-134	-596	1445	138	-547	-447	-126	-8	240	361	-355	590
17	-35	-399	566	424	36	440	-188	102	-143	-128	-1119	83	352
18	-641	30	-1523	224	1752	86	207	122	-108	-845	426	261	-154
19	1197	-372	-92	-186	-1679	137	363	-609	41	-41	1067	239	-295
20	-107	305	44	-1596	1845	-305	-675	257	-248	655	-233	-161	120
21	-578	-445	-145	1565	-671	315	2	-35	-407	-106	374	-251	334
22	-188	594	18	-1481	865	-477	-125	178	-14	158	-549	36	844
23	-219	587	19	821	-131	-308	38	105	-752	11	-243	-136	244
24	180	-183	-507	39	-19	61	-79	-94	56	290	-207	95	285
25	-225	-214	-144	-24	-422	-222	261	44	91	265	256	261	-7
26	208	247	340	-518	-606	-242	258	162	351	-603	0	385	-23
27	-3	194	547	-385	-603	15	-341	251	354	-604	764	258	-485
28	-98	-514	149	-130	38	-44	-101	-44	145	29	540	203	-236
29	-19	244	421	-53	-276	-267	-29	219	177	-266	-89	104	-158
30	646	287	211	224	-19	3	-286	90	-101	-438	-99	-129	-310
31	611	556	731	93	-73	105	-155	197	-111	-395	-497	-251	-698
32	304	260	312	114	69	198	-5	-37	-475	-100	0	-418	-149
33	100	-61	-63	827	18	203	216	-75	-453	174	-429	-384	-17
34	-388	129	-154	685	670	199	-416	-290	-617	143	-103	358	-239
35	52	279	123	-84	-200	106	231	268	-230	-174	-458	276	-138
36	263	213	-3	-357	-422	278	169	74	-229	288	-473	165	27
37	-25	-3	122	-54	128	-193	262	137	-1	341	-209	-536	36
38	-405	-97	-158	611	-275	2	-21	-8	-192	176	-156	459	23
39	349	264	112	-252	5	185	-107	-161	-636	-339	103	201	294
40	139	121	206	-648	-269	443	-69	167	-240	-683	466	453	-115
41	-338	12	430	-326	-467	-110	263	21	70	-65	462	4	-11
42	-377	-1	-160	75	10	105	-75	-48	-62	62	133	224	72
43	-204	-207	-405	455	190	317	71	-237	-558	180	276	280	-186
44	79	-39	138	-248	-79	134	43	163	-117	34	371	-229	-251
45	-64	-110	442	-564	-385	-68	-34	718	341	28	37	95	-483
46	224	67	-179	62	-472	-81	-477	-5	-161	211	315	383	57
47	-331	-441	-172	236	90	422	203	-234	-93	153	236	-149	39
48	-32	231	-48	-541	-535	-539	622	807	-95	15	-209	132	127
49	-211	-560	-440	716	209	35	-306	-123	-92	1210	Extr.	Extr.	-533
50	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	204	Extr.	Extr.	Extr.	Extr.	Extr.
51	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.
52	-179	468	-214	183	-187	-518	2029	Extr.	Extr.	0	-1832	-555	458

Bijlage IV.C: cijfers fitting log-lineair modell(deel III)

Analyse van de pearson residuen van model 4.1. In onderstaand tabel de resultaten van

$$r_i^P = \frac{(y_i - \hat{\mu}_i)}{\sqrt{V(\hat{\mu}_i)}}$$

Jaar week	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
1	-103	-668	-301	1247	971	-647	-1381	-920	3060	1266	611	-1054	-2042
2	115	-12	485	-125	-637	-3	204	4	249	-254	-386	506	-132
3	-148	-633	364	-267	168	323	424	-364	-442	165	123	164	101
4	-427	-558	-944	282	252	552	272	299	-610	46	241	240	292
5	12	104	206	-198	-59	-95	-348	188	208	46	-294	-195	411
6	-104	209	378	-77	175	340	-311	-161	137	106	-682	-256	279
7	-131	-513	179	-295	259	509	32	-517	802	-665	51	-52	315
8	460	98	-501	273	83	330	-622	-393	-324	309	206	-24	103
9	523	381	295	55	5	-138	-189	420	-235	-486	-178	-518	158
10	716	201	-231	-365	-371	27	196	190	144	-347	-328	-121	314
11	-332	-188	-79	207	415	-650	-61	0	-114	615	327	64	-251
12	-231	436	-562	224	195	27	26	-220	-229	50	239	-182	218
13	64	186	0	-221	-269	-46	426	276	562	-725	-263	-362	389
14	890	189	574	132	-1594	-125	46	-435	597	-423	24	435	-351
15	-333	-501	471	-2045	2127	-359	-339	-691	649	208	678	-180	317
16	-697	-134	-590	1487	139	-541	-444	-126	-8	241	363	-353	596
17	-35	-396	573	428	36	444	-187	102	-142	-128	-1094	83	354
18	-632	30	-1480	225	1806	86	208	122	-108	-833	430	262	-153
19	1224	-370	-92	-186	-1625	138	366	-602	42	-41	1087	240	-294
20	-107	307	44	-1547	1911	-303	-666	258	-247	663	-232	-161	120
21	-571	-441	-145	1613	-663	317	2	-35	-404	-106	377	-250	336
22	-188	601	18	-1440	879	-473	-124	178	-14	158	-544	36	856
23	-218	593	19	835	-131	-306	38	105	-742	11	-242	-135	245
24	180	-182	-502	39	-19	61	-79	-94	56	291	-206	96	287
25	-224	-213	-144	-24	-418	-221	262	44	91	266	257	263	-7
26	209	248	343	-512	-599	-240	260	163	353	-597	0	388	-23
27	-3	195	553	-382	-595	15	-339	253	357	-597	774	260	-481
28	-98	-508	150	-129	38	-44	-100	-44	145	29	546	204	-235
29	-19	246	424	-53	-275	-266	-29	220	178	-265	-89	104	-158
30	655	289	212	225	-19	3	-284	90	-101	-434	-98	-128	-308
31	619	563	742	94	-73	105	-154	198	-111	-391	-492	-249	-688
32	306	261	314	115	69	199	-5	-37	-470	-100	0	-415	-149
33	100	-61	-63	841	18	204	217	-75	-449	175	-426	-382	-17
34	-385	129	-154	694	679	200	-413	-288	-610	144	-103	361	-238
35	52	281	123	-84	-200	106	232	269	-229	-174	-455	277	-137
36	264	214	-3	-354	-419	280	170	74	-228	290	-469	166	27
37	-25	-3	122	-54	128	-192	263	137	-1	343	-209	-531	37
38	-402	-97	-158	618	-274	2	-21	-8	-191	177	-155	463	23
39	351	265	112	-250	5	186	-107	-161	-629	-337	103	201	296
40	139	121	207	-640	-268	446	-69	167	-239	-675	469	457	-115
41	-336	12	434	-324	-463	-110	264	21	70	-65	466	4	-11
42	-374	-1	-160	75	10	106	-75	-48	-62	62	133	224	72
43	-203	-206	-402	459	191	319	72	-236	-553	181	277	282	-185
44	79	-39	138	-247	-79	134	43	164	-117	35	374	-229	-250
45	-64	-109	445	-558	-383	-68	-34	727	344	28	37	95	-479
46	225	68	-179	62	-468	-81	-473	-5	-161	212	317	386	57
47	-329	-437	-171	237	90	425	203	-233	-93	153	237	-149	39
48	-32	232	-48	-536	-530	-533	628	818	-95	15	-208	132	128
49	-210	-554	-436	725	210	35	-304	-122	-92	1234	Extr.	Extr.	-528
50	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	205	Extr.	Extr.	Extr.	Extr.	Extr.
51	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.
52	-179	472	-213	184	-187	-513	2099	Extr.	Extr.	0	-1773	-550	462

Bijlage IV.C: cijfers fitting log-lineair modell(deel IV)

Analyse van de residuen van model 4.2.

Respons resid. $\frac{(y - \hat{\mu})}{(\hat{\mu} + gr.part)} * 100\%$			Deviance resid. $r_i^D = sign(y_i - \hat{\mu}_i) \sqrt{d_i}$			Pearson resid. $r_i^P = \frac{(y_i - \hat{\mu}_i)}{\sqrt{V(\hat{\mu}_i)}}$		
Jaar week	'01	'02	Jaar week	'01	'02	Jaar week	'01	'02
1	3,1	-3,4	1	300	-310	1	302	-308
2	2,5	-2,5	2	263	-271	2	265	-270
3	0,0	0,0	3	-2	2	3	-2	2
4	1,2	-1,1	4	129	-132	4	129	-131
5	-0,4	0,3	5	-38	39	5	-38	39
6	-2,5	2,4	6	-256	258	6	-254	259
7	-2,8	2,8	7	-291	294	7	-290	295
8	5,2	-4,6	8	513	-533	8	518	-528
9	-2,3	2,2	9	-233	235	9	-232	236
10	-0,6	0,6	10	-64	65	10	-64	65
11	0,0	0,0	11	3	-3	11	3	-3
12	-0,7	0,7	12	-73	73	12	-73	74
13	-1,9	1,8	13	-193	196	13	-193	196
14	-0,7	0,9	14	-71	82	14	-71	82
15	-0,6	0,5	15	-55	56	15	-55	56
16	-0,9	0,7	16	-79	72	16	-79	72
17	-1,5	1,5	17	-165	167	17	-164	167
18	0,0	0,0	18	0	0	18	0	0
19	-1,8	2,2	19	-180	203	19	-179	204
20	-3,9	4,0	20	-406	408	20	-403	411
21	-2,0	1,9	21	-198	191	21	-197	191
22	-2,1	2,0	22	-225	228	22	-224	228
23	-2,1	1,8	23	-194	184	23	-193	185
24	0,8	-0,8	24	80	-82	24	80	-81
25	0,3	-0,3	25	30	-30	25	30	-30
26	-1,7	1,8	26	-187	189	26	-186	190
27	0,0	0,0	27	0	0	27	0	0
28	-1,9	2,0	28	-187	189	28	-186	190
29	2,7	-2,5	29	231	-238	29	233	-237
30	2,9	-2,7	30	244	-251	30	245	-250
31	0,7	-0,7	31	59	-61	31	59	-61
32	-0,2	0,2	32	-20	20	32	-20	20
33	0,0	0,0	33	0	0	33	0	0
34	0,1	-0,1	34	8	-8	34	8	-8
35	0,1	-0,1	35	6	-6	35	6	-6
36	1,7	-1,6	36	171	-176	36	172	-175
37	-2,1	2,1	37	-208	211	37	-207	211
38	-1,5	1,6	38	-169	171	38	-168	171
39	2,5	-2,2	39	245	-255	39	246	-254
40	2,4	-2,4	40	253	-257	40	254	-256
41	-0,3	0,3	41	-34	35	41	-34	35
42	2,1	-2,0	42	207	-213	42	208	-212
43	1,9	-1,8	43	193	-198	43	194	-197
44	-0,2	0,2	44	-23	24	44	-23	24
45	1,4	-1,4	45	145	-148	45	145	-148
46	1,2	-1,1	46	120	-123	46	120	-122
47	0,9	-0,8	47	88	-90	47	88	-89
48	2,8	-2,6	48	299	-308	48	301	-306

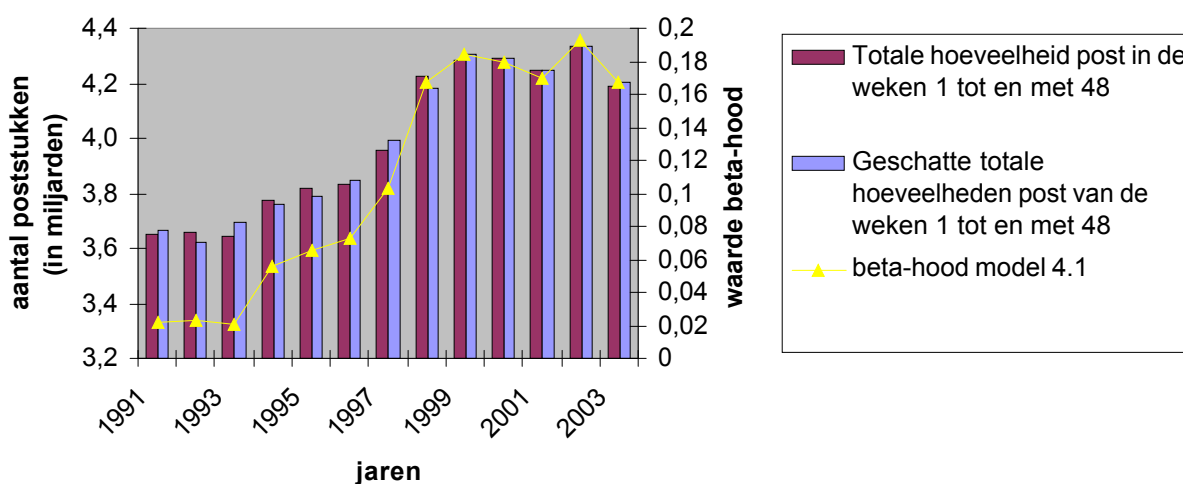
Bijlage IV.D: Schatten van de β voor 2003

Om de hoeveelheden post in de weken 1 tot en met 48 van 2003 te voorspellen, is er eerst een schatting van de regressie coëfficiënt $\hat{\beta}$ voor het jaar 2003 nodig. Deze is namelijk nog niet geschat, doordat de regressie waarnemingen x_i van de data van de jaren 1990 tot en met 2002 niet het jaar 2003 bevatten.

De methode die gebruikt is om de $\hat{\beta}$ voor het jaar 2003 te schatten is de kleinste kwadraten methode, waarbij in dit geval een polynoom bepaald wordt. In Splus wordt met de functie 'lsfit.poly' deze methode uitgevoerd.

Model 4.1

De $\hat{\beta}$ voor het jaar 2003 wordt bepaald door aan deze functie een orde voor de te bepalen polynoom mee te geven en de per jaar totale gerealiseerde hoeveelheden post van de weken 1 tot en met 48 als data mee te geven. De functie bepaalt bij het log-quasi-likelihood model waarbij er geen rekening wordt gehouden met voorgemelde grote partijen een 8-orde polynoom waarbij het totale kwadratische verschil tussen de geschatte polynoom-waarden en de gerealiseerde waarden minimaal is. Met de polynoom kan dan ook de totale hoeveelheid post van de weken 1 tot en met 48 voor het jaar 2003 geschat worden. Door de geschatte waarden te vergelijken met de $\hat{\beta}$ van de jaren 1990 tot en met 2002, is de $\hat{\beta}$ voor het jaar 2003 geschat.



Figuur IV.D.1: vergelijking hoeveelheden post met de geschatte $\hat{\beta}$

Doordat in de figuur IV.D.1 bij de jaren 1998 tot en met 2002 de geschatte totale hoeveelheden post van de weken 1 tot en met 48 samenvallen met de $\hat{\beta}$, is er voor gekozen om de $\hat{\beta}$ van het jaar 2003 te bepalen naar aanleiding van bovenstaand figuur. In onderstaand tabel zijn de bij de figuur behorende cijfers weergegeven.

	Totale hoeveelheid post in de weken 1 tot en met 48 (in miljarden)	Geschatte totale hoeveelheden post van de weken 1 tot en met 48 (in miljarden)	$\hat{\beta}$ van model 4.1
1990	zit bij intercept	zit bij intercept	zit bij intercept
1991	3,65	3,67	0,0219
1992	3,66	3,62	0,0231
1993	3,64	3,69	0,0207
1994	3,78	3,76	0,0558

1995	3,82	3,79	0,0657
1996	3,83	3,85	0,0730
1997	3,96	3,99	0,1029
1998	4,23	4,18	0,1668
1999	4,29	4,30	0,1840
2000	4,29	4,29	0,1795
2001	4,25	4,25	0,1698
2002	4,33	4,34	0,1923
2003	4,19	4,20	0,1667

Tabel IV.D.1: Cijfers behorende bij figuur E.1*Voorgemelde grote partijen*

Bij model 4.2 is voor de voorspellingen van 2003 alleen data van de jaren 2001 en 2002 beschikbaar. Dit vergroot de kans van een foutieve schatting. Daar het aandeel van de voorgemelde grote partijen ongeveer 15% in de totale post is, is er voor gekozen om de grote partijen te schatten met de data van 2001 en 2002. Door deze schatting van de in tabel IV.D.1 weergegeven schatting voor de totale hoeveelheid post van het jaar 2003 af te trekken, is er een schatting gemaakt van de totale hoeveelheid post min de voorgemelde grote partijen van het jaar 2003.

De resultaten van de functie `lsfit.poly` met behulp van de gegevens over voorgemelde grote partijen zijn weergegeven in tabel IV.D.2.

	Totale hoeveelheid voorgemelde gr.partijen in de weken 1 tot en met 48 (in miljoenen)	Geschatte totale hoeveelheid voorgemelde gr.partijen in de weken 1 tot en met 48 (in miljoenen)	$\hat{\beta}$ van model 4.2
2001	482,6	482,6	Zit bij intercept
2002	715,2	715,2	-0.0383
2003	776,8	947,8	-0.0345

Tabel IV.D.2: Cijfers van de grote partijen en de $\hat{\beta}$ van de jaar 2002 en 2003

De $\hat{\beta}$ van het jaar 2003 is bepaald door de geschatte waarden voor de grote partijen van de geschatte totale hoeveelheid af te trekken wat resulteert in een geschatte hoeveelheid van 3,3 miljard. In het jaar 2002 was dit nog gelijk aan 3,6 miljard. Door de verhouding tussen 2002 en 2003 door te berekenen bij de $\hat{\beta}$ van 2002, resulteert dit in een geschatte $\hat{\beta}$ van -0.0345 voor 2003.

Bijlage IV.E: Cijfers voorspellingen(deel I)

Onderstaand de cijfers behorende bij de voorspellingen van model 4.1 voor een periode van 48 weken

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling	
						Model 4.1	WAP
Week 1	91,2	[86,7	, 95,9]	81,1	-11,1 %	
Week 2	93,2	[90,2	, 96,2]	97,8	5,0 %	6,4 %
Week 3	92,4	[89,4	, 95,4]	95,8	3,7 %	2,9 %
Week 4	97,9	[94,8	, 101,1]	98,4	0,5 %	-4,0 %
Week 5	98,5	[95,3	, 101,7]	103,9	5,5 %	5,0 %
Week 6	93,3	[90,4	, 96,4]	94,2	1,0 %	1,6 %
Week 7	91,7	[88,7	, 94,9]	97,0	5,7 %	0,3 %
Week 8	93,3	[90,3	, 96,5]	97,1	4,0 %	2,3 %
Week 9	93,6	[90,7	, 96,7]	98,4	5,1 %	3,3 %
Week 10	89,8	[86,9	, 92,7]	98,4	9,6 %	3,3 %
Week 11	87,1	[84,3	, 90,0]	90,8	4,2 %	1,2 %
Week 12	91,5	[88,5	, 94,6]	90,8	-0,8 %	-4,4 %
Week 13	94,5	[89,1	, 100,1]	93,7	-0,8 %	-4,3 %
Week 14	88,4	[83,2	, 93,9]	87,8	-0,7 %	-3,6 %
Week 15	91,1	[87,0	, 95,3]	86,0	-5,5 %	-2,7 %
Week 16	90,4	[87,3	, 93,7]	92,5	2,2 %	0,3 %
Week 17	85,5	[81,3	, 90,0]	85,8	0,4 %	10,0 %
Week 18	77,6	[74,3	, 80,9]	78,5	1,2 %	4,4 %
Week 19	88,5	[85,5	, 91,7]	86,4	-2,4 %	-2,8 %
Week 20	90,0	[87,1	, 93,1]	93,0	3,3 %	-1,9 %
Week 21	92,5	[89,0	, 96,1]	97,5	5,5 %	-2,6 %
Week 22	83,5	[79,4	, 87,8]	80,2	-3,9 %	-2,3 %
Week 23	88,7	[85,6	, 92,0]	92,7	4,5 %	-0,2 %
Week 24	79,2	[75,6	, 82,8]	78,1	-1,4 %	2,0 %
Week 25	88,9	[86,0	, 91,9]	94,0	5,7 %	2,9 %
Week 26	89,6	[86,4	, 92,9]	92,0	2,7 %	5,2 %
Week 27	84,8	[81,2	, 88,5]	82,7	-2,5 %	3,1 %
Week 28	77,5	[73,9	, 81,4]	73,9	-4,7 %	2,6 %
Week 29	71,8	[68,3	, 75,4]	73,2	1,9 %	3,2 %
Week 30	71,0	[68,4	, 73,7]	69,2	-2,5 %	-1,0 %
Week 31	69,0	[66,4	, 71,6]	65,4	-5,2 %	0,8 %
Week 32	69,9	[67,4	, 72,6]	68,7	-1,8 %	-1,2 %

32 Week								
33 Week	72,3	[69,2	,	75,5]	67,7	-6,4 %
34 Week	80,9	[77,4	,	84,5]	76,6	-5,3 %
35 Week	87,4	[83,6	,	91,4]	88,8	1,6 %
36 Week	92,6	[89,6	,	95,7]	96,2	3,9 %
37 Week	89,6	[86,7	,	92,6]	91,2	1,8 %
38 Week	91,4	[88,5	,	94,4]	83,7	-8,5 %
39 Week	96,3	[92,7	,	100,1]	95,4	-1,0 %
40 Week	93,2	[89,7	,	96,9]	95,1	2,0 %
41 Week	91,3	[88,4	,	94,3]	91,2	-0,1 %
42 Week	89,5	[86,6	,	92,5]	88,8	-0,8 %
43 Week	92,2	[89,3	,	95,3]	89,7	-2,8 %
44 Week	93,2	[90,2	,	96,2]	87,6	-6,0 %
45 Week	90,7	[87,8	,	93,7]	93,5	3,0 %
46 Week	89,7	[86,8	,	92,7]	88,5	-1,3 %
47 Week	92,8	[89,8	,	95,9]	91,8	-1,1 %
48 Week	95,7	[92,6	,	98,9]	95,1	-0,7 %

Bijlage IV.E: Cijfers voorspellingen(deel II)

Onderstaand de cijfers behorende bij de voorspellingen van model 4.1 voor een periode van 8 weken

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling	
						Model 4.1	WAP
Week 1	91,2	[86,7	, 95,9]	81,1	-11,1 %	
Week 2	93,2	[90,2	, 96,2]	97,8	5,0 %	6,4 %
Week 3	92,4	[89,4	, 95,4]	95,8	3,7 %	2,9 %
Week 4	97,9	[94,8	, 101,1]	98,4	0,5 %	-4,0 %
Week 5	98,5	[95,3	, 101,7]	103,9	5,5 %	5,0 %
Week 6	93,3	[90,4	, 96,4]	94,2	1,0 %	1,6 %
Week 7	91,7	[88,7	, 94,9]	97,0	5,7 %	0,3 %
Week 8	93,3	[90,3	, 96,5]	97,1	4,0 %	2,3 %
Week 9	95,6	[91,3	, 100,1]	98,4	3,0 %	3,3 %
Week 10	91,6	[87,4	, 95,9]	98,4	7,4 %	3,3 %
Week 11	88,9	[84,8	, 93,1]	90,8	2,2 %	1,2 %
Week 12	93,4	[89,1	, 97,8]	90,8	-2,8 %	-4,4 %
Week 13	96,4	[90,2	, 103,0]	93,7	-2,7 %	-4,3 %
Week 14	90,2	[84,2	, 96,6]	87,8	-2,7 %	-3,6 %
Week 15	92,9	[87,9	, 98,3]	86,0	-7,4 %	-2,7 %
Week 16	92,3	[87,9	, 96,9]	92,5	0,2 %	0,3 %
Week 17	87,2	[82,5	, 92,2]	85,8	-1,6 %	10,0 %
Week 18	79,0	[75,4	, 82,8]	78,5	-0,7 %	4,4 %
Week 19	90,2	[86,6	, 94,0]	86,4	-4,2 %	-2,8 %
Week 20	91,7	[88,2	, 95,4]	93,0	1,4 %	-1,9 %
Week 21	94,2	[90,1	, 98,4]	97,5	3,6 %	-2,6 %
Week 22	84,8	[80,4	, 89,4]	80,2	-5,4 %	-2,3 %
Week 23	90,4	[86,7	, 94,2]	92,7	2,6 %	-0,2 %
Week 24	80,7	[76,7	, 84,8]	78,1	-3,2 %	2,0 %
Week 25	90,4	[87,2	, 93,7]	94,0	4,0 %	2,9 %
Week 26	91,0	[87,6	, 94,6]	92,0	1,1 %	5,2 %
Week 27	86,1	[82,3	, 90,1]	82,7	-4,0 %	3,1 %
Week 28	78,8	[75,0	, 82,8]	73,9	-6,3 %	2,6 %
Week 29	73,0	[69,3	, 76,8]	73,2	0,3 %	3,2 %
Week 30	72,2	[69,3	, 75,1]	69,2	-4,1 %	-1,0 %
Week 31	70,1	[67,3	, 73,0]	65,4	-6,7 %	0,8 %
Week	71,1	[68,3	, 74,0]	68,7	-3,4 %	-1,2 %

32 Week								
33 Week	73,2	[70,0	,	76,6]	67,7	-7,6 %
34 Week	81,9	[78,3	,	85,7]	76,6	-6,5 %
35 Week	88,5	[84,6	,	92,6]	88,8	0,3 %
36 Week	93,7	[90,5	,	97,0]	96,2	2,7 %
37 Week	90,6	[87,6	,	93,8]	91,2	0,6 %
38 Week	92,5	[89,4	,	95,7]	83,7	-9,5 %
39 Week	97,4	[93,7	,	101,4]	95,4	-2,1 %
40 Week	94,3	[90,6	,	98,2]	95,1	0,9 %
41 Week	91,9	[88,9	,	95,0]	91,2	-0,8 %
42 Week	90,1	[87,1	,	93,1]	88,8	-1,5 %
43 Week	92,8	[89,8	,	95,9]	89,7	-3,4 %
44 Week	93,8	[90,8	,	96,9]	87,6	-6,6 %
45 Week	91,3	[88,4	,	94,4]	93,5	2,3 %
46 Week	90,3	[87,4	,	93,3]	88,5	-1,9 %
47 Week	93,5	[90,5	,	96,7]	91,8	-1,8 %
48 Week	96,5	[93,3	,	99,7]	95,1	-1,4 %

Bijlage IV.E: Cijfers voorspellingen(deel III)

Onderstaand de cijfers behorende bij de voorspellingen van model 4.1 voor een periode van 1 week

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling			
						Model 4.1		WAP	
Week 1	91,2	[86,7	, 95,9]	81,1	-11,1	%		
Week 2	93,2	[82,6	, 105,1]	97,8	5,0	%	6,4	%
Week 3	90,1	[83,0	, 97,8]	95,8	6,3	%	2,9	%
Week 4	97,6	[91,2	, 104,4]	98,4	0,8	%	-4,0	%
Week 5	98,4	[92,7	, 104,5]	103,9	5,6	%	5,0	%
Week 6	94,4	[89,4	, 99,6]	94,2	-0,2	%	1,6	%
Week 7	92,7	[88,1	, 97,7]	97,0	4,6	%	0,3	%
Week 8	95,0	[90,4	, 99,7]	97,1	2,2	%	2,3	%
Week 9	95,6	[91,3	, 100,1]	98,4	3,0	%	3,3	%
Week 10	91,9	[87,9	, 96,1]	98,4	7,0	%	3,3	%
Week 11	89,8	[86,0	, 93,8]	90,8	1,1	%	1,2	%
Week 12	94,4	[90,5	, 98,5]	90,8	-3,9	%	-4,4	%
Week 13	97,2	[91,2	, 103,5]	93,7	-3,5	%	-4,3	%
Week 14	91,3	[85,6	, 97,3]	87,8	-3,8	%	-3,6	%
Week 15	93,3	[88,6	, 98,1]	86,0	-7,8	%	-2,7	%
Week 16	92,1	[88,3	, 96,0]	92,5	0,4	%	0,3	%
Week 17	87,2	[82,5	, 92,2]	85,8	-1,6	%	10,0	%
Week 18	79,0	[75,4	, 82,8]	78,5	-0,6	%	4,4	%
Week 19	90,1	[86,6	, 93,8]	86,4	-4,1	%	-2,8	%
Week 20	91,4	[88,0	, 94,9]	93,0	1,8	%	-1,9	%
Week 21	93,9	[90,1	, 97,9]	97,5	3,9	%	-2,6	%
Week 22	84,7	[80,4	, 89,2]	80,2	-5,3	%	-2,3	%
Week 23	90,2	[86,8	, 93,8]	92,7	2,8	%	-0,2	%
Week 24	80,4	[76,6	, 84,3]	78,1	-2,8	%	2,0	%
Week 25	90,4	[87,2	, 93,7]	94,0	4,0	%	2,9	%
Week 26	91,2	[87,7	, 94,8]	92,0	0,9	%	5,2	%
Week 27	86,3	[82,5	, 90,3]	82,7	-4,2	%	3,1	%
Week 28	78,8	[75,0	, 82,8]	73,9	-6,3	%	2,6	%
Week 29	72,5	[69,1	, 76,1]	73,2	0,9	%	3,2	%
Week 30	72,1	[69,3	, 75,0]	69,2	-4,0	%	-1,0	%
Week 31	70,0	[67,3	, 72,8]	65,4	-6,5	%	0,8	%
Week	70,8	[68,1	, 73,6]	68,7	-3,0	%	-1,2	%

32 Week								
33 Week	73,2	[70,0	,	76,6]	67,7	-7,6 %
34 Week	81,8	[78,2	,	85,5]	76,6	-6,3 %
35 Week	87,9	[84,1	,	91,8]	88,8	1,1 %
36 Week	93,4	[90,2	,	96,6]	96,2	3,1 %
37 Week	90,4	[87,4	,	93,5]	91,2	0,9 %
38 Week	92,3	[89,2	,	95,4]	83,7	-9,3 %
39 Week	97,0	[93,3	,	100,8]	95,4	-1,7 %
40 Week	93,9	[90,3	,	97,6]	95,1	1,3 %
41 Week	91,9	[88,9	,	95,0]	91,2	-0,8 %
42 Week	90,1	[87,1	,	93,1]	88,8	-1,4 %
43 Week	92,8	[89,8	,	95,9]	89,7	-3,3 %
44 Week	93,6	[90,7	,	96,7]	87,6	-6,5 %
45 Week	91,1	[88,1	,	94,1]	93,5	2,6 %
46 Week	90,1	[87,2	,	93,1]	88,5	-1,7 %
47 Week	93,3	[90,3	,	96,4]	91,8	-1,6 %
48 Week	96,1	[93,1	,	99,3]	95,1	-1,1 %

Bijlage IV.E: Cijfers voorspellingen(deel IV)

Onderstaand de cijfers behorende bij de voorspellingen van model 4.2 voor een periode van 48 weken

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling	
						Model 4.2	WAP
Week 1	73,4	[70,2	, 76,7]	81,1	10,5 %	
Week 2	97,7	[94,1	, 101,4]	97,8	0,2 %	6,4 %
Week 3	98,5	[94,9	, 102,3]	95,8	-2,7 %	2,9 %
Week 4	108,3	[104,6	, 112,1]	98,4	-9,1 %	-4,0 %
Week 5	111,1	[107,4	, 114,9]	103,9	-6,4 %	5,0 %
Week 6	97,4	[93,8	, 101,2]	94,2	-3,3 %	1,6 %
Week 7	94,8	[91,2	, 98,6]	97,0	2,3 %	0,3 %
Week 8	88,9	[85,3	, 92,6]	97,1	9,2 %	2,3 %
Week 9	99,2	[95,6	, 103,0]	98,4	-0,8 %	3,3 %
Week 10	97,8	[94,2	, 101,5]	98,4	0,6 %	3,3 %
Week 11	91,6	[88,1	, 95,2]	90,8	-0,9 %	1,2 %
Week 12	92,7	[89,2	, 96,5]	90,8	-2,1 %	-4,4 %
Week 13	100,8	[97,2	, 104,5]	93,7	-7,0 %	-4,3 %
Week 14	93,6	[89,2	, 98,4]	87,8	-6,2 %	-3,6 %
Week 15	94,1	[90,5	, 97,9]	86,0	-8,6 %	-2,7 %
Week 16	97,2	[92,6	, 102,0]	92,5	-4,9 %	0,3 %
Week 17	89,5	[84,3	, 95,3]	85,8	-4,2 %	10,0 %
Week 18	78,2	[73,5	, 83,3]	78,5	0,3 %	4,4 %
Week 19	93,5	[88,8	, 98,4]	86,4	-7,6 %	-2,8 %
Week 20	103,0	[99,5	, 106,7]	93,0	-9,7 %	-1,9 %
Week 21	106,6	[99,9	, 113,9]	97,5	-8,5 %	-2,6 %
Week 22	83,3	[77,6	, 89,5]	80,2	-3,7 %	-2,3 %
Week 23	97,0	[92,3	, 102,0]	92,7	-4,4 %	-0,2 %
Week 24	80,1	[74,3	, 86,4]	78,1	-2,5 %	2,0 %
Week 25	95,8	[92,3	, 99,5]	94,0	-1,9 %	2,9 %
Week 26	95,9	[92,4	, 99,6]	92,0	-4,1 %	5,2 %
Week 27	86,8	[81,9	, 92,1]	82,7	-4,8 %	3,1 %
Week 28	75,7	[68,4	, 83,9]	73,9	-2,4 %	2,6 %
Week 29	76,2	[69,3	, 84,1]	73,2	-4,0 %	3,2 %
Week 30	70,2	[67,1	, 73,5]	69,2	-1,3 %	-1,0 %
Week 31	66,8	[63,8	, 70,1]	65,4	-2,1 %	0,8 %
Week 32	73,4	[70,4	, 76,7]	68,7	-6,5 %	-1,2 %

32 Week								
33 Week	70,4	[65,9	,	75,2]	67,7	-3,9 %
34 Week	82,7	[79,3	,	86,2]	76,6	-7,3 %
35 Week	98,4	[94,9	,	102,0]	88,8	-9,7 %
36 Week	102,0	[98,3	,	105,8]	96,2	-5,6 %
37 Week	94,0	[90,5	,	97,7]	91,2	-3,0 %
38 Week	88,0	[84,5	,	91,7]	83,7	-5,0 %
39 Week	101,2	[96,7	,	106,0]	95,4	-5,8 %
40 Week	105,5	[101,0	,	110,2]	95,1	-9,8 %
41 week	97,3	[93,7	,	101,1]	91,2	-6,3 %
42 Week	94,0	[90,4	,	97,7]	88,8	-5,5 %
43 Week	94,4	[90,9	,	98,1]	89,7	-5,0 %
44 Week	96,0	[92,5	,	99,7]	87,6	-8,8 %
45 Week	92,0	[88,5	,	95,7]	93,5	1,6 %
46 Week	93,1	[89,6	,	96,8]	88,5	-4,9 %
47 Week	96,1	[92,5	,	99,8]	91,8	-4,4 %
48	95,2	[91,7	,	98,9]	95,1	-0,1 %

Bijlage IV.E: Cijfers voorspellingen(deel V)

Onderstaand de cijfers behorende bij de voorspellingen van model 4.2 voor een periode van 8 weken

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)					Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling			
							Model 4.2		WAP	
Week 17	89,2	[83,9	,	94,8		-3,7	%	10,0	%
Week 18	76,5	[71,9	,	81,5		2,6	%	4,4	%
Week 19	91,4	[86,9	,	96,2		-5,5	%	-2,8	%
Week 20	101,0	[97,6	,	104,6		-7,9	%	-1,9	%
Week 21	104,5	[98,0	,	111,6		-6,7	%	-2,6	%
Week 22	81,6	[76,0	,	87,7		-1,7	%	-2,3	%
Week 23	95,0	[90,5	,	99,8		-2,4	%	-0,2	%
Week 24	78,4	[72,7	,	84,6		-0,4	%	2,0	%
Week 41	93,8	[91,7	,	95,9		-2,7	%	-0,3	%
Week 42	90,5	[88,4	,	92,6		-1,9	%	-3,1	%
Week 43	91,0	[88,9	,	93,1		-1,4	%	2,5	%
Week 44	92,6	[90,6	,	94,7		-5,4	%	0,6	%
Week 45	88,6	[86,5	,	90,7		5,5	%	7,2	%
Week 46	89,8	[87,7	,	91,8		-1,4	%	0,5	%
Week 47	92,6	[90,6	,	94,7		-0,9	%	1,8	%
Week 48	91,8	[89,8	,	93,9		3,5	%	5,0	%

Bijlage IV.E: Cijfers voorspellingen(deel VI)

Onderstaand de cijfers behorende bij de voorspellingen van model 4.2 voor een periode van 1 week

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling	
						Model 4.2	WAP
Week 1	73,4	[71,7	, 75,0]	81,1	10,5 %	
Week 2	107,4	[103,1	, 111,8]	97,8	-8,9 %	6,4 %
Week 3	103,0	[99,8	, 106,4]	95,8	-7,0 %	2,9 %
Week 4	110,3	[107,3	, 113,3]	98,4	-10,7 %	-4,0 %
Week 5	109,7	[106,9	, 112,7]	103,9	-5,3 %	5,0 %
Week 6	94,9	[92,2	, 97,7]	94,2	-0,7 %	1,6 %
Week 7	92,2	[89,7	, 94,8]	97,0	5,2 %	0,3 %
Week 8	87,1	[84,6	, 89,6]	97,1	11,5 %	2,3 %

Bijlage IV.F: Toets voor uitbijters

Aan de hand van de deviance en de response residuen is er een indicatie dat week 1 van het jaar 1998 een uitbijter is. Voordat dit getoetst kan worden, wordt het model geherdefinieerd. Stel dat $\delta_1 > 0$ geldt:

- i. $Y_i \sim \text{Poisson}(\mu_i)$, Y_i is hier het aantal poststukken in week i .
- ii. $\eta_i = x_i^T \beta$
 $= \beta_0 + \beta_1 \mathbf{1}\{\text{Valentijnsdag} \in i\} + \beta_2 \mathbf{1}\{\text{Pasen} \in i\} + \beta_3 \mathbf{1}\{\text{Hemelvaartsdag} \in i\} + \beta_4 \mathbf{1}\{\text{Pinksteren} \in i\}$
 $+ \beta_5 \mathbf{1}\{\text{Koninginnedag} \in i\} + \beta_6 \mathbf{1}\{\text{Bevrijdingsdag} \in i\} + \beta_7 \mathbf{1}\{\text{Mailmoment} \in i\}$
 $+ \beta_{8 \dots 20} \mathbf{1}\{\text{Jaar 1991} \dots \text{Jaar 2002} \in i\} + \beta_{21 \dots 70} \mathbf{1}\{\text{Week 2} \dots \text{Week 50, Week 52} \in i\} + \beta_{71, 72} \mathbf{1}\{\text{Kerst 1, Kerst 2} \in i\}$
 $+ \beta_{73 \dots 78} \mathbf{1}\{\text{Vakantie 1} \dots \text{Vakantie 6} \in i\}$
 $+ \beta_{79 \dots 87} \mathbf{1}\{\text{Eind jan, eind maart, eind april, eind juni} \dots \text{eind aug, eind nov, eind dec} \in i\} + \delta_1 \mathbf{1}\{\text{Week 1 '98} \in i\}$
- iii. $\eta_i = g(\mu_i) = \log(\mu_i)$

De toets voor uitbijters toets de hypothesen:

$$H_0: \delta_1 = 0, \beta \text{ willekeurig}$$

$$H_1: \delta_1 \neq 0, \beta \text{ willekeurig}$$

De nulhypothese wordt verworpen als de t-ratio groter is dan $t_{n-p-2; 1-(\alpha/2)}$

Hierbij is

- n het aantal waarnemingen ($n=647$)
- p het aantal variabelen ($p=87$)

Doordat de t-ratio van δ_1 gelijk is aan 3271,3 en de $t_{n-p-2; 1-(\alpha/2)} = t_{647-87-2; 0.975} = 1.9642$ wordt de nulhypothese verworpen. Hierdoor kan er geconcludeerd worden dat week 1 jaar 1998 een uitbijter is.

Bijlage IV.G Klassiek lineair regressiemodel

Doordat de hypothese dat $Y_i \sim \text{Poisson}(\mu)$ verdeeld is verworpen wordt door de Pearson chkwadraat-toets, heb ik tevens het klassieke lineaire regressiemodel onderzocht. De kans dat de Pearson chkwadraat-toets niet bij dit model verworpen wordt, is groter doordat bij een normale verdeling de variantie van de responsvariabele Y_i niet gelijk hoeft te zijn aan de verwachting van de verwachting van Y_i . Er is echter wel een andere eis bij het klassieke lineaire regressiemodel, en wel de eis dat de response residuen normaal verdeeld moeten zijn. In deze bijlage zal eerst het model beschreven worden, om vervolgens kort de resultaten te beschrijven. Hierbij zal eerst bekeken worden of de residuen normaal verdeeld zijn, om onnodig verder onderzoek te voorkomen.

Het klassieke lineaire regressiemodel heeft de volgende eigenschappen:

- de responsvariabele Y_i is Normaal(μ, σ^2) verdeeld
- $Y_i = x_i^T \beta + \varepsilon$
- schaal respons is gelijk aan (- ,)
- de verdeling van de response residuen is onafhankelijk en gelijk normaal($0, \sigma^2$) verdeeld.

Het verschil met het log-lineair model, is een verschil in de aanname van de verdelingsfunctie van de responsvariabele Y_i . Tevens wordt er bij een log-lineair model een linkfunctie gespecificeerd, zodat er van tevoren bepaald wordt dat de verwachting van de responsvariabele Y_i niet negatief is. Doordat de responsvariabele Y_i bij dit model gelijk is aan het aantal poststukken per week, heeft een linkfunctie geen toegevoegde waarde. Het aantal poststukken per week varieerde in de jaren 1990 tot en met 2002 tussen de 54,7 en de 113,5 miljoen. Door deze grote getallen is de kans dat er een negatieve verwachting van de responsvariabele Y_i bepaald wordt erg klein.

Beschrijving model

In onderstaand model is het onderzochte model beschreven:

- iv. $Y_i \sim \text{Normaal}(\mu, \sigma^2)$,
 Y_i het aantal poststukken in week i .
 μ_i verwachting van het aantal poststukken in week i .
 σ^2 gelijk aan de variantie van Y_i
 n het totaal aantal weken in de jaren 1990 tot en met 2002, zonder de weken waarin een extreme hoeveelheid post gegenereerd is ($n=647$)
 m (het aantal weken waarvoor een voorspelling gegenereerd wordt)- n

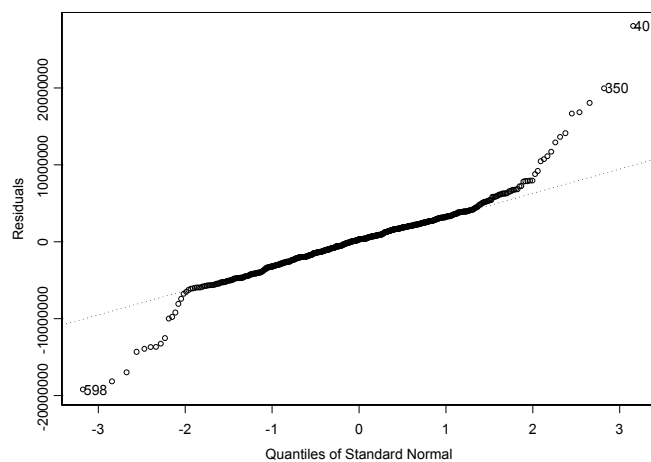
v. $\eta_i = x_i^T \beta$
 $= \beta_0 + \beta_1 \mathbf{1}\{\text{Valentijnsdag} \in i\} + \beta_2 \mathbf{1}\{\text{Pasen} \in i\} + \beta_3 \mathbf{1}\{\text{Hemelvaartsdag} \in i\} + \beta_4 \mathbf{1}\{\text{Pinksteren} \in i\}$
 $+ \beta_5 \mathbf{1}\{\text{Koninginnedag} \in i\} + \beta_6 \mathbf{1}\{\text{Bevrijdingsdag} \in i\} + \beta_7 \mathbf{1}\{\text{Mailmoment} \in i\}$
 $+ \beta_8 \dots 20 \mathbf{1}\{\text{Jaar } 1991 \dots \text{Jaar } 2003^{47} \in i\} + \beta_{21} \dots 70 \mathbf{1}\{\text{Week } 2 \dots \text{Week } 50, \text{Week } 52 \in i\} + \beta_{71,72} \mathbf{1}\{\text{Kerst } 1, \text{Kerst } 2 \in i\}$
 $+ \beta_{73} \dots 78 \mathbf{1}\{\text{Vakantie } 1 \dots \text{Vakantie } 6 \in i\}$
 $+ \beta_{79} \dots 87 \mathbf{1}\{\text{Eind jan, eind maart, eind april, eind juni} \dots \text{eind sept, eind nov, eind dec} \in i\}$

vi. $\eta_i = g(\mu_i) = \mu_i$

Om te bepalen wat de verwachting van Y_i is, wordt ook in dit model de meest aannemelijke schatter bepaald van de regressie coëfficiënten β . Dit gaat op dezelfde manier als bij het log-lineair model en het log-quasi-likelihood model.

⁴⁷ het jaar 2003 wordt alleen bij het maken van voorspellingen voor het jaar 2003 als regressie waarneming meegenomen in het model.

Resultaten



Figuur IV.G.1: de residuen uitgezet tegen de kwantielen van de standaard normale verdeling

Uit dit figuur blijkt dat de residuen van het klassieke lineaire regressie model niet normaal verdeeld zijn. Om deze reden is er geen verder onderzoek verricht naar het klassiek lineair regressiemodel.

Bijlage IV.H Kernschatter

De kernschatter wordt bepaald door de ‘smooth-functie’ $s(x)$ te definiëren. Het definiëren van $s(x)$ vindt plaats door het voltooien van de volgende stappenplan:

Stap i)	Neem een deviance residu $\sqrt{ r_0^D }$. Vind de k dichtstbijzijnde buren van dit punt, en noem deze $N(\sqrt{ r_0^D })$
Stap ii)	Bereken het grootste verschil tussen $\sqrt{ r_0^D }$ en een andere punt die in $N(\sqrt{ r_0^D })$ zit, oftewel $\Delta(\sqrt{ r_0^D }) = \max_{N(\sqrt{ r_0^D })} \sqrt{ r_0^D } - \sqrt{ r_1^D } $
Stap iii)	Wijs gewichten toe aan elke punt in $N(\sqrt{ r_0^D })$ waarbij er gebruik gemaakt wordt van de driehoeks gewogen functie: $W\left(\frac{ \sqrt{ r_0^D } - \sqrt{ r_1^D } }{\Delta(\sqrt{ r_0^D })}\right)$ waarbij $W(u) = \begin{cases} (1-u^3)^3 & \text{als } 0 \leq u < 1 \\ 0 & \text{anders} \end{cases}$
Stap iv)	Bereken de gewogen kleinste kwadraten schatter van y op $N(\sqrt{ r_0^D })$. Noem de geschatte waarde $\hat{y}_0 = s(x_0)$
Stap v)	Herhaal dit voor elke deviance residu

Bijlage IV.C: cijfers fitting log-quasi-likelihood modell(deel I)

Analyse van de pearson residuen van model 4.1. In onderstaand tabel de resultaten van

$$r_i^P = \frac{(y_i - \hat{\mu}_i)}{\sqrt{V(\hat{\mu}_i)}}$$

Jaar week	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
1	-0,07	-0,43	-0,19	0,81	0,63	-0,42	-0,90	-0,60	1,98	0,82	0,40	-0,68	-1,32
2	0,07	-0,01	0,31	-0,08	-0,41	0,00	0,13	0,00	0,16	-0,16	-0,25	0,33	-0,09
3	-0,10	-0,41	0,24	-0,17	0,11	0,21	0,27	-0,24	-0,29	0,11	0,08	0,11	0,07
4	-0,28	-0,36	-0,61	0,18	0,16	0,36	0,18	0,19	-0,40	0,03	0,16	0,16	0,19
5	0,01	0,07	0,13	-0,13	-0,04	-0,06	-0,23	0,12	0,13	0,03	-0,19	-0,13	0,27
6	-0,07	0,14	0,25	-0,05	0,11	0,22	-0,20	-0,10	0,09	0,07	-0,44	-0,17	0,18
7	-0,08	-0,33	0,12	-0,19	0,17	0,33	0,02	-0,34	0,52	-0,43	0,03	-0,03	0,20
8	0,30	0,06	-0,32	0,18	0,05	0,21	-0,40	-0,25	-0,21	0,20	0,13	-0,02	0,07
9	0,34	0,25	0,19	0,04	0,00	-0,09	-0,12	0,27	-0,15	-0,32	-0,12	-0,34	0,10
10	0,46	0,13	-0,15	-0,24	-0,24	0,02	0,13	0,12	0,09	-0,23	-0,21	-0,08	0,20
11	-0,22	-0,12	-0,05	0,13	0,27	-0,42	-0,04	0,00	-0,07	0,40	0,21	0,04	-0,16
12	-0,15	0,28	-0,36	0,15	0,13	0,02	0,02	-0,14	-0,15	0,03	0,15	-0,12	0,14
13	0,04	0,12	0,00	-0,14	-0,17	-0,03	0,28	0,18	0,36	-0,47	-0,17	-0,23	0,25
14	0,58	0,12	0,37	0,09	-1,03	-0,08	0,03	-0,28	0,39	-0,27	0,02	0,28	-0,23
15	-0,22	-0,32	0,31	-1,33	1,38	-0,23	-0,22	-0,45	0,42	0,13	0,44	-0,12	0,21
16	-0,45	-0,09	-0,38	0,96	0,09	-0,35	-0,29	-0,08	-0,01	0,16	0,24	-0,23	0,39
17	-0,02	-0,26	0,37	0,28	0,02	0,29	-0,12	0,07	-0,09	-0,08	-0,71	0,05	0,23
18	-0,41	0,02	-0,96	0,15	1,17	0,06	0,13	0,08	-0,07	-0,54	0,28	0,17	-0,10
19	0,79	-0,24	-0,06	-0,12	-1,05	0,09	0,24	-0,39	0,03	-0,03	0,70	0,16	-0,19
20	-0,07	0,20	0,03	-1,00	1,24	-0,20	-0,43	0,17	-0,16	0,43	-0,15	-0,10	0,08
21	-0,37	-0,29	-0,09	1,05	-0,43	0,21	0,00	-0,02	-0,26	-0,07	0,24	-0,16	0,22
22	-0,12	0,39	0,01	-0,93	0,57	-0,31	-0,08	0,12	-0,01	0,10	-0,35	0,02	0,55
23	-0,14	0,38	0,01	0,54	-0,08	-0,20	0,02	0,07	-0,48	0,01	-0,16	-0,09	0,16
24	0,12	-0,12	-0,33	0,03	-0,01	0,04	-0,05	-0,06	0,04	0,19	-0,13	0,06	0,19
25	-0,15	-0,14	-0,09	-0,02	-0,27	-0,14	0,17	0,03	0,06	0,17	0,17	0,17	0,00
26	0,14	0,16	0,22	-0,33	-0,39	-0,16	0,17	0,11	0,23	-0,39	0,00	0,25	-0,01
27	0,00	0,13	0,36	-0,25	-0,39	0,01	-0,22	0,16	0,23	-0,39	0,50	0,17	-0,31
28	-0,06	-0,33	0,10	-0,08	0,02	-0,03	-0,07	-0,03	0,09	0,02	0,35	0,13	-0,15
29	-0,01	0,16	0,28	-0,03	-0,18	-0,17	-0,02	0,14	0,12	-0,17	-0,06	0,07	-0,10
30	0,42	0,19	0,14	0,15	-0,01	0,00	-0,18	0,06	-0,07	-0,28	-0,06	-0,08	-0,20
31	0,40	0,36	0,48	0,06	-0,05	0,07	-0,10	0,13	-0,07	-0,25	-0,32	-0,16	-0,45
32	0,20	0,17	0,20	0,07	0,04	0,13	0,00	-0,02	-0,30	-0,06	0,00	-0,27	-0,10
33	0,06	-0,04	-0,04	0,55	0,01	0,13	0,14	-0,05	-0,29	0,11	-0,28	-0,25	-0,01
34	-0,25	0,08	-0,10	0,45	0,44	0,13	-0,27	-0,19	-0,40	0,09	-0,07	0,23	-0,15
35	0,03	0,18	0,08	-0,05	-0,13	0,07	0,15	0,17	-0,15	-0,11	-0,29	0,18	-0,09
36	0,17	0,14	0,00	-0,23	-0,27	0,18	0,11	0,05	-0,15	0,19	-0,30	0,11	0,02
37	-0,02	0,00	0,08	-0,04	0,08	-0,12	0,17	0,09	0,00	0,22	-0,14	-0,34	0,02
38	-0,26	-0,06	-0,10	0,40	-0,18	0,00	-0,01	-0,01	-0,12	0,11	-0,10	0,30	0,02
39	0,23	0,17	0,07	-0,16	0,00	0,12	-0,07	-0,10	-0,41	-0,22	0,07	0,13	0,19
40	0,09	0,08	0,13	-0,41	-0,17	0,29	-0,04	0,11	-0,15	-0,44	0,30	0,30	-0,07
41	-0,22	0,01	0,28	-0,21	-0,30	-0,07	0,17	0,01	0,05	-0,04	0,30	0,00	-0,01
42	-0,24	0,00	-0,10	0,05	0,01	0,07	-0,05	-0,03	-0,04	0,04	0,09	0,15	0,05
43	-0,13	-0,13	-0,26	0,30	0,12	0,21	0,05	-0,15	-0,36	0,12	0,18	0,18	-0,12
44	0,05	-0,03	0,09	-0,16	-0,05	0,09	0,03	0,11	-0,08	0,02	0,24	-0,15	-0,16
45	-0,04	-0,07	0,29	-0,36	-0,25	-0,04	-0,02	0,47	0,22	0,02	0,02	0,06	-0,31
46	0,15	0,04	-0,12	0,04	-0,30	-0,05	-0,31	0,00	-0,10	0,14	0,21	0,25	0,04
47	-0,21	-0,28	-0,11	0,15	0,06	0,28	0,13	-0,15	-0,06	0,10	0,15	-0,10	0,02
48	-0,02	0,15	-0,03	-0,35	-0,34	-0,35	0,41	0,53	-0,06	0,01	-0,13	0,09	0,08
49	-0,14	-0,36	-0,28	0,47	0,14	0,02	-0,20	-0,08	-0,06	0,80	Extr.	Extr.	-0,34
50	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	0,13	Extr.	Extr.	Extr.	Extr.	Extr.
51	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.	Extr.
52	-0,12	0,31	-0,14	0,12	-0,12	-0,33	1,36	Extr.	Extr.	0,00	-1,15	-0,36	0,30

Bijlage IV.I: cijfers fitting log-quasi likelihood modell(deel II)

Analyse van de Pearson residuen van model 4.2.

$$r_i^P = \frac{(y_i - \hat{\mu}_i)}{\sqrt{V(\hat{\mu}_i)}}$$

Jaar week	'01	'02
1	1,01	-1,03
2	0,89	-0,91
3	-0,01	0,01
4	0,43	-0,44
5	-0,13	0,13
6	-0,85	0,87
7	-0,97	0,99
8	1,74	-1,77
9	-0,78	0,79
10	-0,21	0,22
11	0,01	-0,01
12	-0,24	0,25
13	-0,65	0,66
14	-0,24	0,27
15	-0,18	0,19
16	-0,27	0,24
17	-0,55	0,56
18	0,00	0,00
19	-0,60	0,68
20	-1,35	1,38
21	-0,66	0,64
22	-0,75	0,77
23	-0,65	0,62
24	0,27	-0,27
25	0,10	-0,10
26	-0,63	0,64
27	0,00	0,00
28	-0,63	0,64
29	0,78	-0,80
30	0,82	-0,84
31	0,20	-0,20
32	-0,07	0,07
33	0,00	0,00
34	0,03	-0,03
35	0,02	-0,02
36	0,58	-0,59
37	-0,70	0,71
38	-0,56	0,58
39	0,83	-0,85
40	0,85	-0,86
41	-0,11	0,12
42	0,70	-0,71
43	0,65	-0,66
44	-0,08	0,08
45	0,49	-0,50
46	0,40	-0,41
47	0,29	-0,30
48	1,01	-1,03

Bijlage V.A: Meest aannemelijke schatter

In deze bijlage wordt een toelichting van de Newton procedure gegeven. Tijdens deze procedure worden er meest aannemelijke schatters bepaald. Deze procedure wordt vaak toegepast als er geen expliciete oplossing is tijdens het bepalen van het maximum van de aannemelijkheidfunctie l. Om aan te geven dat er geen expliciete oplossing is, zal eerst de afgeleiden van de aannemelijkheidfunctie beschreven worden. Na deze beschrijving volgt de toelichting van de Newton procedure.

$$\begin{aligned}
 Y_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
 \lambda_{ij} &= \left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right) \tau_j \\
 l(\mathbf{w}, \mathbf{\alpha}, \mathbf{\tau}) &= \sum_{j=1}^{13} \sum_{i=1}^{48} \log P_0(Y_{ij} = y_{ij}) \\
 &= \sum_{j=1}^{13} \sum_{i=1}^{48} \log \left(\frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \right) e^{-\lambda_{ij}} \\
 &= \sum_{j=1}^{13} \sum_{i=1}^{48} (y_{ij} \log(\lambda_{ij}) - \log(y_{ij}!) - \lambda_{ij}) \\
 &= \sum_{j=1}^{13} \sum_{i=1}^{48} \left(y_{ij} \log \left[\left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right) \tau_j \right] - \log(y_{ij}!) - \left[\left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right) \tau_j \right] \right) \\
 &= \sum_{j=1}^{13} \sum_{i=1}^{48} \left(y_{ij} \log(\tau_j) + y_{ij} \log \left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right) - \log(y_{ij}!) - \left[\left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right) \tau_j \right] \right) \\
 \frac{\partial}{\partial \tau_j} l(\mathbf{w}, \mathbf{\alpha}, \mathbf{\tau}) &= \sum_{i=1}^{48} \left(\frac{y_{ij}}{\tau_j} - \left[w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right] \right) = 0 \\
 &\Leftrightarrow \frac{1}{\tau_j} \sum_{i=1}^{48} y_{ij} = 1 \\
 &\Leftrightarrow \hat{\tau}_j = \sum_{i=1}^{48} y_{ij} \quad \text{Hierbij is}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial}{\partial \alpha_k} l(\mathbf{w}, \mathbf{\alpha}, \mathbf{\tau}) &= \sum_{j=1}^{13} \frac{y_{k(j)j}}{w_{k(j)} + \sum_{l=1}^5 \alpha_l 1_{l \in k(j)j}} - \sum_{j=1}^{13} \sum_{i=1}^{48} \hat{\tau}_j 1_{k \in ij} = 0 \\
 \frac{\partial}{\partial w_i} l(\mathbf{w}, \mathbf{\alpha}, \mathbf{\tau}) &= \sum_{j=1}^{13} \frac{y_{ij}}{w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij}} - \sum_{j=1}^{13} \hat{\tau}_j = 0
 \end{aligned}$$

$k(j)$ gelijk aan het weeknummer waarbij in jaar j feestdag k valt.

Indien er in geen enkel jaar een feestdag is in week i dan volgt uit bovenstaande dat w_i gelijk is aan

$$w_i = \frac{\sum_{j=1}^{13} y_{ij}}{\sum_{j=1}^{13} \sum_{i=1}^{48} y_{ij}}$$

Omdat er geen expliciete oplossing is voor $\frac{\partial}{\partial \alpha_k} l(\mathbf{w}, \mathbf{\alpha}, \mathbf{\tau}) = 0$ en voor $\frac{\partial}{\partial w_i} l(\mathbf{w}, \mathbf{\alpha}, \mathbf{\tau}) = 0$, wordt de

Newton-procedure gebruikt voor het bepalen van de meest aannemelijke schatter. Deze Newton-procedure werkt als volgt:

Stap i) Neem startwaarden voor \dot{w} en $\dot{\alpha}$ en voeg deze samen in vector v_1 , $\dot{v}_1 = (\dot{w}, \dot{\alpha})$.

De startwaarde voor \dot{w} wordt gegeven door $w_i = \frac{\sum_{j=1}^{13} y_{ij}}{\sum_{j=1}^{13} \sum_{i=1}^{48} y_{ij}}$, de startwaarden voor

$\dot{\alpha}$ zijn gelijk aan nul. Ga naar stap ii)

Stap ii) Bepaal $l_1(\dot{w}, \dot{\alpha}, \tau)$ waarbij $(\dot{w}, \dot{\alpha}) = \dot{v}_1$ en ga naar stap iii)

Stap iii) Bepaal het minimum van

$$f(\dot{v}) \cong f(\dot{v}_1) + (\dot{v} - \dot{v}_1)^T \nabla f(\dot{v}_1) + \frac{1}{2} (\dot{v} - \dot{v}_1)^T \nabla^2 f(\dot{v}_1) (\dot{v} - \dot{v}_1).$$

zodat de onbekende parameter $\dot{v} = (\dot{w}_2, \dot{\alpha}_2)$ bepaald wordt en ga naar stap iv)

Hierbij is

$$f(\dot{v}_1) = l_1(\dot{w}, \dot{\alpha}, \tau)$$

$\dot{v}_2 = (\dot{w}_2, \dot{\alpha}_2)$ een vector met nog onbekende parameters \dot{w}_2 en $\dot{\alpha}_2$.

$$\nabla f(\dot{v}_1) = \begin{pmatrix} \frac{\partial}{\partial w_1} f(\dot{v}_1) \\ \dots \\ \frac{\partial}{\partial w_{48}} f(\dot{v}_1) \\ \frac{\partial}{\partial \alpha_1} f(\dot{v}_1) \\ \dots \\ \frac{\partial}{\partial \alpha_5} f(\dot{v}_1) \end{pmatrix}$$

$$\nabla^2 f(\dot{v}_1) = \begin{pmatrix} \frac{\partial}{\partial w_1^2} f(\dot{v}_1) & \dots & \dots & \frac{\partial}{\partial w_1 w_{48}} f(\dot{v}_1) & \frac{\partial}{\partial w_1 \alpha_1} f(\dot{v}_1) & \dots & \frac{\partial}{\partial w_1 \alpha_5} f(\dot{v}_1) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial}{\partial w_{48} w_1} f(\dot{v}_1) & \dots & \dots & \frac{\partial}{\partial w_{48}^2} f(\dot{v}_1) & \frac{\partial}{\partial w_{48} \alpha_1} f(\dot{v}_1) & \dots & \frac{\partial}{\partial w_{48} \alpha_5} f(\dot{v}_1) \\ \frac{\partial}{\partial \alpha_1 w_1} f(\dot{v}_1) & \dots & \dots & \frac{\partial}{\partial \alpha_1 w_{48}} f(\dot{v}_1) & \frac{\partial}{\partial \alpha_1^2} f(\dot{v}_1) & \dots & \frac{\partial}{\partial \alpha_1 \alpha_5} f(\dot{v}_1) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial}{\partial \alpha_5 w_1} f(\dot{v}_1) & \dots & \dots & \frac{\partial}{\partial \alpha_5 w_{48}} f(\dot{v}_1) & \frac{\partial}{\partial \alpha_5 \alpha_1} f(\dot{v}_1) & \dots & \frac{\partial}{\partial \alpha_5^2} f(\dot{v}_1) \end{pmatrix}$$

$$\frac{\partial}{\partial w_i w_h} f(\dot{v}_1) = \begin{cases} \sum_{j=1}^{13} \left[- \frac{y_{ij}}{\left(w_i + \sum_{k=1}^5 \alpha_k 1_{k \in ij} \right)^2} \right] & \text{als } i = h \\ 0 & \text{anders} \end{cases}$$

$$\begin{aligned}
\frac{\partial}{\partial \alpha_k w_i} f(\mathbf{r}_{\mathbf{v}_1}) &= \left\{ \sum_{j=1}^{13} \left[- \frac{y_{k(j)j}}{\left(w_{k(j)} + \sum_{l=1}^5 \alpha_l 1_{l \in k(j)} \right)^2} \right] \right\} \\
&= \frac{\partial}{\partial w_i \alpha_k} f(\mathbf{r}_{\mathbf{v}_1}) = \left\{ \sum_{j=1}^{13} \left[- \frac{y_{k(j)j}}{\left(w_{k(j)} + \sum_{l=1}^5 \alpha_l 1_{l \in k(j)} \right)^2} \right] \right\} \quad \begin{array}{l} \text{als } k(j) = i \\ \text{anders} \end{array} \\
\frac{\partial}{\partial w_i \alpha_k} f(\mathbf{r}_{\mathbf{v}_1}) &= \left\{ \sum_{j=1}^{13} \left[- \frac{y_{k(j)j}}{\left(w_{k(j)} + \sum_{l=1}^5 \alpha_l 1_{l \in k(j)} \right)^2} \right] \right\} \quad \begin{array}{l} \text{als} \\ \text{als} \end{array} \\
\frac{\partial}{\partial \alpha_k \alpha_m} f(\mathbf{r}_{\mathbf{v}_1}) &= \left\{ \sum_{j=1}^{13} \left[- \frac{y_{m(j)j} 1_{\{k(j)=m(j)\}}}{\left(w_{m(j)} + \sum_{l=1}^5 \alpha_l 1_{l \in k(j), l \in m(j)} \right)^2} \right] \right\} \quad \begin{array}{l} \text{als} \\ \text{als} \end{array}
\end{aligned}$$

Stap iv) Bepaal $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$

Indien deze groter is dan $l_1(\mathbf{w}, \hat{\alpha}, \hat{\tau})$, vervang dan de vector $\mathbf{v}_1 = (\mathbf{w}, \hat{\alpha})$ door $\mathbf{v} = (\mathbf{w}_2, \hat{\alpha}_2)$, oftewel \mathbf{v}_1 wordt $(\mathbf{w}_2, \hat{\alpha}_2)$ en ga naar stap ii)

Indien deze $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$ kleiner is dan $l_1(\mathbf{w}, \hat{\alpha}, \hat{\tau})$, ga dan naar stap v)

Indien deze $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$ gelijk is, dan is \mathbf{w}_2 gelijk aan de meest aannemelijke schatter voor w , en $\hat{\alpha}_2$ gelijk aan de meest aannemelijke schatter voor α .

Stap v) Bepaal het gemiddelde van de vectoren $\mathbf{v}_1 = (\mathbf{w}, \hat{\alpha})$ en $\mathbf{v} = (\mathbf{w}_2, \hat{\alpha}_2)$ en noem deze \mathbf{v}_{gem} . Vervang $\mathbf{v} = (\mathbf{w}_2, \hat{\alpha}_2)$ door \mathbf{v}_{gem} , oftewel \mathbf{v} wordt \mathbf{v}_{gem} . Ga naar stap vi)

Stap vi) Bepaal $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$

Indien deze $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$ kleiner is dan $l_1(\mathbf{w}, \hat{\alpha}, \hat{\tau})$, en er wordt nog wel een stijging van de $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$ verwacht ga dan naar stap v)

Indien deze $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$ kleiner is dan $l_1(\mathbf{w}, \hat{\alpha}, \hat{\tau})$, en er wordt geen stijging van de $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$ verwacht, dan is \mathbf{w}_1 gelijk aan de meest aannemelijke schatter voor w , en $\hat{\alpha}_1$ gelijk aan de meest aannemelijke schatter voor α .

Indien deze $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$ gelijk is, dan is \mathbf{w}_2 gelijk aan de meest aannemelijke schatter voor w , en $\hat{\alpha}_2$ gelijk aan de meest aannemelijke schatter voor α .

Indien deze $l(\mathbf{w}_2, \hat{\alpha}_2, \hat{\tau})$ groter is dan $l_1(\mathbf{w}, \hat{\alpha}, \hat{\tau})$, vervang dan de vector $\mathbf{v}_1 = (\mathbf{w}, \hat{\alpha})$ door $\mathbf{v} = (\mathbf{w}_2, \hat{\alpha}_2)$, oftewel \mathbf{v}_1 wordt $(\mathbf{w}_2, \hat{\alpha}_2)$ en ga naar stap ii)

Minimalisatie van $f(\mathbf{v})$

Het minimaliseren van $f(\mathbf{v}) \cong f(\mathbf{v}_1) + (\mathbf{v} - \mathbf{v}_1)^T \nabla f(\mathbf{v}_1) + \frac{1}{2}(\mathbf{v} - \mathbf{v}_1)^T \nabla^2 f(\mathbf{v}_1)(\mathbf{v} - \mathbf{v}_1)$, wordt gedaan door $\nabla^2 f(\mathbf{v}_1) * \mathbf{v}$ gelijk te stellen aan $\nabla^2 f(\mathbf{v}_1) * \mathbf{v}_1 - \nabla f(\mathbf{v}_1)$, omdat:

$$\begin{aligned} f(\mathbf{v}) &\cong f(\mathbf{v}_1) + (\mathbf{v} - \mathbf{v}_1)^T \nabla f(\mathbf{v}_1) + \frac{1}{2}(\mathbf{v} - \mathbf{v}_1)^T \nabla^2 f(\mathbf{v}_1)(\mathbf{v} - \mathbf{v}_1) \\ &= f(\mathbf{v}_1) + \frac{1}{2}(\mathbf{v} - \mathbf{v}_1 + (\nabla^2 f(\mathbf{v}_1)^{-1} \nabla f(\mathbf{v}_1)) \nabla^2 f(\mathbf{v}_1)(\mathbf{v} - \mathbf{v}_1 + (\nabla^2 f(\mathbf{v}_1)^{-1} \nabla f(\mathbf{v}_1)) \\ &\quad - \frac{1}{2} \nabla f(\mathbf{v}_1)^T (\nabla^2 f(\mathbf{v}_1)^{-1} \nabla f(\mathbf{v}_1)) \end{aligned}$$

Het minimum van $f(\mathbf{v})$ wordt dan bepaald door de vergelijking $(\mathbf{v} - \mathbf{v}_1 + (\nabla^2 f(\mathbf{v}_1)^{-1} \nabla f(\mathbf{v}_1)) = 0$, omdat de overige gedeeltes van de formule onafhankelijk zijn van \mathbf{v} . Door

$(\mathbf{v} - \mathbf{v}_1 + (\nabla^2 f(\mathbf{v}_1)^{-1} \nabla f(\mathbf{v}_1))$ gelijk te stellen aan nul, zal het minimum van $f(\mathbf{v})$ bepaald worden. Hieruit volgt dat

$$\nabla^2 f(\mathbf{v}_1) * \mathbf{v} = \nabla^2 f(\mathbf{v}_1) * \mathbf{v}_1 - \nabla f(\mathbf{v}_1)$$

Bijlage V.B: cijfers fitting Poisson model(deel I)

Analyse van de response residuen. In onderstaand tabel de resultaten van:

$$\frac{(y - \hat{\lambda})}{\hat{\lambda}} * 100\% \text{ van model 5.1}$$

Jaar week	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
1	3,4	-9,8	-5,7	12,3	8,7	-5,5	-10,9	-12,2	30,1	11,3	8,0	-7,1	-23,6
2	1,0	-0,3	5,1	-1,4	-7,1	-0,2	2,4	-0,1	2,3	-2,6	-4,5	4,7	-1,6
3	-2,0	-7,3	3,7	-3,0	1,6	3,3	4,7	-4,1	-4,9	1,7	0,7	1,1	0,7
4	-5,1	-6,3	-10,6	3,0	2,4	5,5	2,9	2,9	-6,6	1,0	1,8	1,8	2,5
5	-0,2	0,9	1,9	-2,2	-0,8	-1,2	-3,6	1,8	1,8	-0,1	-3,5	-2,5	3,8
6	-1,6	2,0	3,8	-1,0	1,6	3,4	-3,4	-1,9	1,0	1,0	-7,6	-3,3	2,5
7	-1,4	-5,6	2,1	-3,0	0,0	5,8	0,8	-5,4	8,4	-6,5	-2,7	-0,7	3,3
8	4,2	0,3	-6,4	2,4	3,2	2,8	-7,2	-4,9	-4,2	2,6	4,1	-1,4	0,2
9	5,4	3,9	2,8	0,5	-0,3	-1,8	-2,1	4,2	-2,8	-5,1	-2,5	-5,9	1,2
10	7,8	1,9	-3,1	-4,3	-4,4	0,0	2,1	1,8	1,1	-3,8	-4,1	-1,9	2,9
11	-4,3	-2,5	-1,4	2,2	4,3	-7,6	-0,7	-0,3	-1,7	6,4	2,7	0,0	-3,1
12	-3,1	4,5	-6,8	2,3	1,7	-0,1	0,2	-2,7	-2,9	0,8	1,7	-2,6	2,3
13	0,2	2,0	-0,5	-2,6	-3,3	-0,9	4,5	2,6	5,6	-7,6	-3,4	-4,4	3,5
14	15,5	-9,5	11,6	6,6	-14,6	3,4	5,7	-15,8	11,0	-15,4	4,7	9,8	-14,8
15	-1,8	-3,6	7,5	-21,2	9,6	-1,9	-15,9	-5,4	9,0	4,6	9,6	-0,1	5,5
16	-17,4	4,3	-1,3	7,1	7,5	-15,3	1,0	4,5	-9,7	8,8	9,5	-13,4	12,3
17	11,7	7,5	-17,0	-4,7	-9,1	-4,8	10,4	13,5	10,5	-13,5	-33,0	12,8	16,1
18	-24,1	-8,5	-3,1	16,0	35,0	7,0	-6,3	-7,5	-10,1	7,4	10,3	-3,4	-10,5
19	18,5	-12,8	2,7	1,9	-15,0	5,5	8,5	-15,0	4,3	-8,4	17,1	6,2	-11,3
20	0,5	5,4	2,3	-15,8	9,8	-1,6	-17,2	4,8	-0,9	9,1	-1,0	-0,2	3,1
21	-11,5	-7,1	6,3	13,5	0,4	-1,4	8,5	-2,3	-9,3	-2,9	12,0	-7,4	1,5
22	1,1	10,2	-9,0	-12,9	3,0	-1,9	-7,4	5,4	3,1	5,2	-14,4	3,4	12,3
23	-9,3	9,9	3,0	2,8	1,4	-10,1	3,6	4,2	-5,3	3,2	-0,1	-8,5	5,5
24	3,5	-0,6	-13,7	2,2	1,3	2,2	0,9	0,5	2,0	4,8	-10,5	2,2	4,5
25	-3,0	-2,8	-2,0	-0,4	-5,0	-2,8	2,9	0,2	0,5	2,7	2,0	2,1	-0,5
26	1,3	4,0	2,8	-6,5	-7,6	-3,6	2,2	3,0	4,8	-7,0	0,1	2,8	-1,3
27	1,9	2,1	6,3	-4,4	-8,7	-1,7	-1,7	2,7	1,5	-6,4	6,1	0,6	-3,3
28	-1,6	-6,0	1,9	-2,4	-0,6	-0,8	-1,2	-0,7	1,8	-0,5	4,7	1,7	-3,0
29	-0,6	3,3	5,5	-1,5	-4,5	-3,6	-0,4	2,4	2,2	-3,9	-2,4	0,6	-2,2
30	7,9	3,2	2,1	2,6	-0,5	-0,4	-3,7	0,7	-1,8	-5,2	-2,0	-2,3	-4,1
31	7,6	6,9	9,0	1,0	-1,5	0,9	-2,1	2,1	-1,8	-5,0	-6,6	-3,7	-8,6
32	3,9	3,3	3,9	1,7	0,9	2,5	0,2	-0,4	-5,7	-1,0	-5,4	-5,3	-1,8
33	0,1	-1,8	-1,2	10,5	0,5	2,7	1,9	-1,9	-5,6	1,9	-5,1	-4,5	-1,3
34	-5,2	-0,7	-4,1	10,1	9,6	1,8	-4,9	-5,4	-9,0	3,4	0,1	3,1	-3,1
35	-0,1	0,7	-1,1	0,7	-0,8	0,5	2,2	0,6	-2,6	-0,2	-3,7	1,9	-2,1
36	2,8	2,3	-0,3	-3,9	-4,7	2,9	2,0	0,7	-4,9	3,1	-5,3	1,3	0,1
37	-0,7	-0,4	0,9	-0,7	1,1	-2,5	2,9	1,2	-0,5	3,5	-2,9	-6,2	0,0
38	-5,0	-1,4	-2,2	6,8	-3,3	-0,3	-0,3	-0,3	-2,4	1,7	-2,3	4,1	-0,2
39	4,4	3,5	0,0	-3,6	-1,0	0,9	-0,3	-1,0	-6,0	-4,3	-0,4	0,6	3,5
40	0,2	0,1	2,6	-6,5	-2,5	5,3	-1,7	0,6	-3,8	-6,3	4,9	4,8	-2,4
41	-4,2	-0,2	4,4	-3,8	-5,4	-1,5	2,9	0,0	0,3	-0,8	4,1	-0,6	-0,5
42	-4,7	-0,4	-2,3	0,7	-0,2	0,8	-0,9	-0,8	-1,1	0,5	0,7	1,7	0,3
43	-2,7	-2,6	-4,9	5,0	1,8	3,2	0,7	-2,8	-6,2	1,7	2,2	2,3	-2,3
44	0,5	-0,8	1,1	-2,9	-1,2	1,1	0,4	1,5	-1,6	0,2	3,1	-3,0	-2,9
45	-1,1	-1,6	4,6	-6,4	-4,6	-1,1	-0,4	7,6	3,1	0,2	-0,3	0,3	-5,3
46	2,1	0,4	-2,5	0,6	-5,5	-1,2	-5,3	-0,3	-2,1	2,1	2,6	3,4	0,2
47	-4,2	-5,3	-2,4	2,4	0,6	4,2	2,1	-2,8	-1,5	1,4	1,7	-1,8	0,4
48	-0,7	2,3	-0,9	-5,9	-6,0	-6,0	6,8	8,4	-3,2	0,6	-2,2	0,8	1,0

Bijlage V.B: cijfers fitting Poisson model(deel II)

Analyse van de response residuen. In onderstaand tabel de resultaten van:

$$\frac{(y - \hat{\lambda})}{\hat{\lambda} + gr.part} * 100\% \text{ van model 5.2}$$

Jaar week	'01	'02
1	8,7	-5,5
2	-7,1	-0,2
3	1,6	3,3
4	2,4	5,5
5	-0,8	-1,2
6	1,6	3,4
7	0,0	5,8
8	3,2	2,8
9	-0,3	-1,8
10	-4,4	0,0
11	4,3	-7,6
12	1,7	-0,1
13	-3,3	-0,9
14	-14,6	3,4
15	9,6	-1,9
16	7,5	-15,3
17	-9,1	-4,8
18	35,0	7,0
19	-15,0	5,5
20	9,8	-1,6
21	0,4	-1,4
22	3,0	-1,9
23	1,4	-10,1
24	1,3	2,2
25	-5,0	-2,8
26	-7,6	-3,6
27	-8,7	-1,7
28	-0,6	-0,8
29	-4,5	-3,6
30	-0,5	-0,4
31	-1,5	0,9
32	0,9	2,5
33	0,5	2,7
34	9,6	1,8
35	-0,8	0,5
36	-4,7	2,9
37	1,1	-2,5
38	-3,3	-0,3
39	-1,0	0,9
40	-2,5	5,3
41	-5,4	-1,5
42	-0,2	0,8
43	1,8	3,2
44	-1,2	1,1
45	-4,6	-1,1
46	-5,5	-1,2
47	0,6	4,2
48	-6,0	-6,0

Bijlage V.B: cijfers fitting Poisson model(deel III)

Analyse van de response residuen. In onderstaand tabel de resultaten van:

$$\frac{(y - \hat{\lambda})}{\hat{\lambda}} * 100\% \text{ van model 5.3}$$

Jaar week	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
1	1,0	-11,9	-7,9	9,6	6,1	-7,7	-13,0	-14,3	27,0	8,7	5,5	-9,3	-25,4
2	3,7	2,3	7,8	1,2	-4,7	2,4	5,0	2,5	4,9	0,0	-2,0	7,4	1,0
3	-4,2	-9,4	1,4	-5,2	-0,6	1,0	2,3	-6,2	-7,1	-0,6	-1,6	-1,1	-1,5
4	-3,4	-4,7	-9,0	4,8	4,2	7,4	4,7	4,7	-4,9	2,7	3,5	3,6	4,3
5	1,7	2,9	3,8	-0,3	1,1	0,7	-1,7	3,8	3,7	1,8	-1,6	-0,6	5,8
6	-2,6	0,9	2,7	-2,0	0,5	2,3	-4,4	-3,0	-0,1	-0,1	-8,6	-4,3	1,4
7	-3,0	-7,1	0,4	-4,5	-1,7	4,1	-0,8	-6,9	6,7	-8,1	-4,3	-2,3	1,7
8	5,2	1,2	-5,5	3,4	4,2	3,7	-6,3	-4,0	-3,3	3,5	5,0	-0,4	1,1
9	6,9	5,3	4,2	1,9	1,1	-0,5	-0,8	5,6	-1,5	-3,8	-1,2	-4,7	2,6
10	7,3	1,5	-3,5	-4,7	-4,9	-0,5	1,7	1,3	0,6	-4,2	-4,5	-2,4	2,4
11	-6,8	-5,1	-4,0	-0,5	1,6	-10,1	-3,4	-2,9	-4,3	3,6	0,0	-2,6	-5,6
12	-2,5	5,2	-6,2	3,0	2,4	0,5	0,8	-2,1	-2,3	1,5	2,4	-2,0	2,9
13	4,9	6,7	4,1	1,9	1,2	3,7	9,3	7,4	10,5	-3,3	1,1	0,0	8,3
14	9,5	-11,9	5,7	1,1	-16,9	-2,0	0,2	-18,0	5,2	-17,6	-0,7	4,1	-17,1
15	1,0	-0,9	10,5	-18,9	12,6	0,9	-13,5	-2,7	12,1	7,6	12,6	2,8	8,4
16	-19,6	1,5	-3,9	4,2	4,7	-17,6	-1,7	1,7	-12,1	5,9	6,6	-15,7	9,4
17	15,1	10,7	-14,5	-1,8	-6,4	-1,9	13,7	16,9	13,9	-10,9	-31,0	16,2	19,6
18	-25,3	-10,0	-4,7	14,2	32,9	5,3	-7,8	-8,9	-11,5	5,7	8,6	-4,9	-11,9
19	17,1	-13,8	1,5	0,7	-16,0	4,2	7,1	-16,1	3,0	-9,5	15,6	4,9	-12,3
20	2,9	7,9	4,7	-13,7	12,4	0,8	-15,3	7,3	1,5	11,7	1,4	2,2	5,5
21	-14,6	-10,3	2,6	9,6	-3,1	-4,8	4,8	-5,7	-12,4	-6,3	8,1	-10,6	-2,0
22	5,8	15,4	-4,8	-8,8	7,8	2,6	-3,1	10,3	7,9	10,1	-10,4	8,2	17,5
23	-11,7	7,0	0,3	0,1	-1,3	-12,5	0,9	1,4	-7,8	0,5	-2,7	-10,9	2,7
24	2,7	-1,4	-14,4	1,3	0,5	1,4	0,1	-0,3	1,2	4,0	-11,2	1,4	3,7
25	-2,4	-2,1	-1,4	0,2	-4,4	-2,2	3,5	0,8	1,1	3,3	2,7	2,7	0,1
26	4,6	7,4	6,1	-3,5	-4,6	-0,5	5,5	6,4	8,2	-3,9	3,4	6,1	2,0
27	1,2	1,4	5,6	-5,1	-9,3	-2,4	-2,3	2,1	0,8	-7,0	5,4	-0,1	-4,0
28	-1,6	-6,0	1,9	-2,4	-0,6	-0,8	-1,2	-0,7	1,8	-0,5	4,7	1,7	-3,0
29	-3,1	0,8	2,9	-3,9	-6,8	-6,0	-2,8	-0,1	-0,3	-6,3	-4,8	-1,9	-4,6
30	8,8	4,1	3,0	3,5	0,3	0,4	-2,9	1,5	-0,9	-4,4	-1,1	-1,5	-3,3
31	6,1	5,4	7,6	-0,3	-2,8	-0,4	-3,3	0,8	-3,1	-6,3	-7,8	-5,0	-9,8
32	2,7	2,2	2,7	0,5	-0,3	1,3	-0,9	-1,5	-6,8	-2,1	-6,4	-6,3	-2,9
33	-2,8	-4,7	-4,1	7,2	-2,5	-0,3	-1,1	-4,8	-8,4	-1,1	-7,9	-7,4	-4,2
34	-4,0	0,5	-3,0	11,4	10,9	3,0	-3,8	-4,3	-7,9	4,6	1,3	4,4	-2,0
35	1,2	2,1	0,2	2,1	0,6	1,9	3,6	1,9	-1,2	1,2	-2,4	3,3	-0,8
36	5,0	4,5	1,9	-1,8	-2,6	5,2	4,3	3,0	-2,8	5,4	-3,2	3,5	2,3
37	-2,4	-2,1	-0,8	-2,4	-0,6	-4,1	1,1	-0,5	-2,1	1,7	-4,5	-7,8	-1,7
38	-5,7	-2,2	-3,0	6,0	-4,1	-1,1	-1,1	-1,1	-3,2	0,9	-3,1	3,3	-1,0
39	6,4	5,5	2,0	-1,8	0,9	2,9	1,7	0,9	-4,2	-2,4	1,5	2,6	5,5
40	0,7	0,6	3,1	-6,0	-2,0	5,8	-1,2	1,1	-3,3	-5,8	5,4	5,3	-2,0
41	-4,6	-0,6	4,0	-4,1	-5,8	-1,9	2,5	-0,4	-0,1	-1,2	3,7	-1,0	-0,9
42	-6,3	-2,0	-3,9	-1,0	-1,9	-0,8	-2,5	-2,4	-2,7	-1,1	-1,0	0,0	-1,3
43	-2,1	-2,0	-4,3	5,7	2,4	3,8	1,4	-2,1	-5,6	2,4	2,8	2,9	-1,6
44	1,7	0,4	2,3	-1,7	0,0	2,4	1,7	2,7	-0,5	1,4	4,4	-1,8	-1,8
45	-1,7	-2,1	4,0	-6,9	-5,0	-1,6	-0,9	7,0	2,6	-0,4	-0,8	-0,2	-5,8
46	0,6	-1,1	-4,0	-1,0	-7,0	-2,8	-6,7	-1,9	-3,6	0,5	1,0	1,8	-1,4
47	-4,0	-5,1	-2,3	2,6	0,7	4,4	2,2	-2,7	-1,3	1,5	1,8	-1,6	0,6
48	0,3	3,3	0,1	-5,0	-5,0	-5,1	7,8	9,5	-2,2	1,6	-1,3	1,7	1,9

Bijlage V.B: cijfers fitting Poisson model(deel IV)

Analyse van de response residuen. In onderstaand tabel de resultaten van:

$$\frac{(y - \hat{\lambda})}{\hat{\lambda} + gr.part} * 100\% \text{ van model 5.4}$$

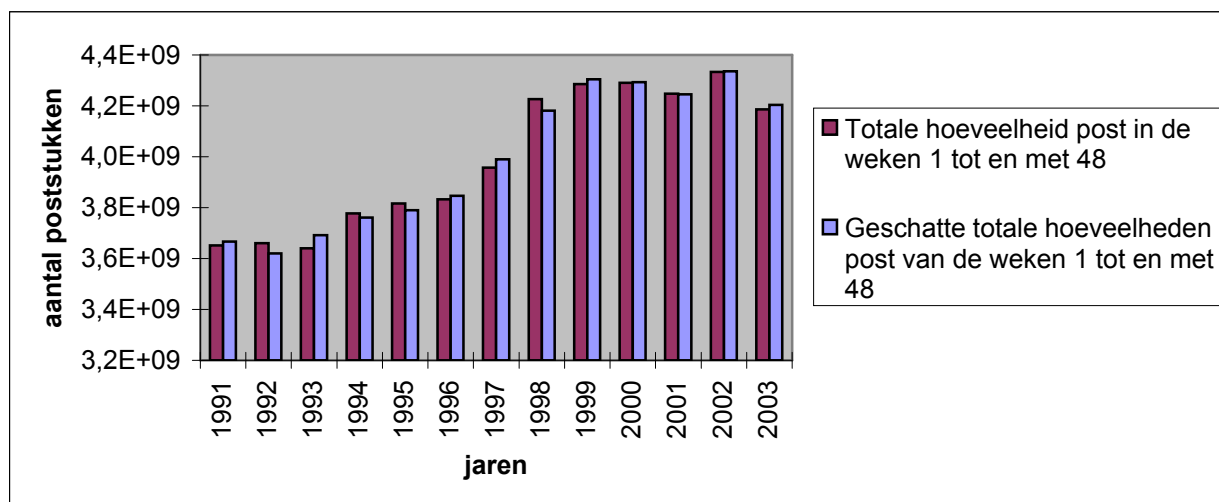
Jaar week	'01	'02
1	-3,7	-10,1
2	7,3	2,0
3	0,2	0,4
4	1,6	-0,7
5	0,6	1,3
6	-3,0	2,1
7	-2,3	3,4
8	2,3	-6,7
9	0,0	4,6
10	-0,3	1,1
11	-3,0	-2,9
12	-0,6	2,5
13	0,7	4,3
14	2,0	-14,3
15	4,2	4,9
16	-17,1	3,4
17	8,3	11,1
18	-5,5	-9,0
19	7,8	-7,8
20	-0,5	7,8
21	-12,7	-0,8
22	5,1	8,9
23	-10,4	4,2
24	0,0	-1,4
25	1,2	0,7
26	-1,2	2,6
27	3,6	1,0
28	-4,1	-0,1
29	1,8	-3,2
30	2,9	-2,5
31	-0,5	-1,7
32	-1,9	-1,2
33	-3,2	-2,9
34	0,9	0,9
35	0,4	0,4
36	5,2	1,6
37	-3,7	0,6
38	-4,7	-1,6
39	8,3	0,7
40	0,4	-2,2
41	-0,6	0,2
42	2,5	-1,5
43	1,4	-2,2
44	-1,6	-0,9
45	2,5	-0,3
46	0,1	-2,0
47	1,5	-0,2
48	2,3	-2,9

Bijlage V.C: Schatten van de hoeveelheid post voor 2003

Om de hoeveelheden post in de weken 1 tot en met 48 van 2003 te voorspellen, is er eerst een schatting van de regressie coëfficiënt τ_{2003} nodig.

De methode die gebruikt is om de $\hat{\tau}$ voor het jaar 2003 te schatten is de kleinste kwadraten methode, waarbij in dit geval een polynoom bepaald wordt. In Splus wordt met de functie 'lsfit.poly' deze methode uitgevoerd.

De τ_{2003} wordt bepaald door aan deze functie een orde voor de te bepalen polynoom mee te geven en de per jaar totale gerealiseerde hoeveelheden post van de weken 1 tot en met 48 als data mee te geven. De functie bepaalt bij het Poisson model waarbij er geen rekening wordt gehouden met voorgemelde grote partijen een 8^e-orde polynoom waarbij het totale kwadratische verschil tussen de geschatte polynoomwaarden en de gerealiseerde waarden minimaal is. Met de polynoom kan dan ook de totale hoeveelheid post van de weken 1 tot en met 48 voor het jaar 2003 geschat worden.



Figuur V.C.1: vergelijking hoeveelheden post met de geschatte $\hat{\tau}_j$

In onderstaand tabel zijn de bij de figuur behorende cijfers weergegeven.

	Totale hoeveelheid post in de weken 1 tot en met 48 (in miljarden)	Geschatte totale hoeveelheden post van de weken 1 tot en met 48 (in miljarden)
1990	zit bij intercept	zit bij intercept
1991	3,65	3,67
1992	3,66	3,62
1993	3,64	3,69
1994	3,78	3,76
1995	3,82	3,79
1996	3,83	3,85
1997	3,96	3,99
1998	4,23	4,18
1999	4,29	4,30
2000	4,29	4,29
2001	4,25	4,25
2002	4,33	4,34
2003	4,19	4,20

Tabel V.C.1: Cijfers behorende bij figuur C.1

Voorgemelde grote partijen

Bij het model waarbij de responswaarde Y_{ij} gelijk is aan het aantal poststukken min de voorgemelde grote partijen in deze week in week i , is voor de voorspellingen van 2003 alleen data van de jaren 2001

en 2002 beschikbaar. Dit vergroot de kans van een foutieve schatting. Daar het aandeel van de voorgemelde grote partijen ongeveer 15% in de totale post is, is er voor gekozen om de grote partijen te schatten met de data van 2001 en 2002. Door deze schatting van de in tabel V.C.2 weergegeven schatting voor de totale hoeveelheid post van het jaar 2003 af te trekken, is er een schatting gemaakt van de totale hoeveelheid post min de voorgemelde grote partijen van het jaar 2003.

De resultaten van de functie lsfit.poly met behulp van de gegevens over voorgemelde grote partijen zijn weergegeven in tabel V.C.2.

	Totale hoeveelheid voorgemelde gr.partijen in de weken 1 tot en met 48 (in miljoenen)	Geschatte totale hoeveelheid voorgemelde gr.partijen in de weken 1 tot en met 48 (in miljoenen)
2001	482,6	482,6
2002	715,2	715,2
2003	776,8	947,8

Tabel V.C.2: Cijfers van de grote partijen en de geschatte hoeveelheden van het jaar 2002 en 2003

Echter, naar aanleiding van de kennis van de WAP-groep wordt voor de geschatte hoeveelheid voorgemelde grote partijen in de weken 1 tot en met 48 van het jaar 2003 het gerealiseerde hoeveelheid voorgemelde grote partijen van het jaar 2002 genomen. Volgens de WAP-groep is de hoeveelheid voorgemelde grote partijen in het jaar 2001 gelijk aan de hoeveelheid voorgemelde grote partijen in het jaar 2002, maar is er op een andere manier geadministreerd.

Bijlage V.D: Cijfers voorspellingen(deel I)

Onderstaand de cijfers behorende bij de voorspellingen van model 5.1

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling	
						Model 5.1	WAP
Week 1	86,7	[86,7	, 86,7]	81,1	-6,5 %	
Week 2	93,0	[92,9	, 93,0]	97,8	5,2 %	6,4 %
Week 3	92,2	[92,2	, 92,2]	95,8	3,9 %	2,9 %
Week 4	97,8	[97,8	, 97,8]	98,4	0,6 %	-4,0 %
Week 5	98,3	[98,3	, 98,3]	103,9	5,7 %	5,0 %
Week 6	93,2	[93,2	, 93,2]	94,2	1,1 %	1,6 %
Week 7	91,2	[91,2	, 91,2]	97,0	6,3 %	0,3 %
Week 8	93,7	[93,7	, 93,7]	97,1	3,6 %	2,3 %
Week 9	93,6	[93,5	, 93,6]	98,4	5,1 %	3,3 %
Week 10	89,7	[89,7	, 89,7]	98,4	9,7 %	3,3 %
Week 11	87,1	[87,0	, 87,1]	90,8	4,3 %	1,2 %
Week 12	91,5	[91,5	, 91,5]	90,8	-0,8 %	-4,4 %
Week 13	94,1	[94,1	, 94,2]	93,7	-0,4 %	-4,3 %
Week 14	84,2	[84,2	, 84,3]	87,8	4,2 %	-3,6 %
Week 15	88,3	[88,3	, 88,3]	86,0	-2,6 %	-2,7 %
Week 16	85,1	[85,1	, 85,1]	92,5	8,7 %	0,3 %
Week 17	88,7	[88,7	, 88,7]	85,8	-3,3 %	10,0 %
Week 18	84,8	[84,8	, 84,8]	78,5	-7,4 %	4,4 %
Week 19	84,8	[84,8	, 84,9]	86,4	1,9 %	-2,8 %
Week 20	88,0	[88,0	, 88,0]	93,0	5,7 %	-1,9 %
Week 21	85,0	[85,0	, 85,1]	97,5	14,7 %	-2,6 %
Week 22	91,1	[91,1	, 91,1]	80,2	-11,9 %	-2,3 %
Week 23	85,1	[85,1	, 85,1]	92,7	8,9 %	-0,2 %
Week 24	85,9	[85,9	, 85,9]	78,1	-9,1 %	2,0 %
Week 25	88,9	[88,8	, 88,9]	94,0	5,7 %	2,9 %
Week 26	90,1	[90,1	, 90,1]	92,0	2,1 %	5,2 %
Week 27	82,8	[82,7	, 82,8]	82,7	-0,1 %	3,1 %
Week 28	77,1	[77,0	, 77,1]	73,9	-4,1 %	2,6 %
Week 29	71,4	[71,4	, 71,4]	73,2	2,5 %	3,2 %
Week 30	71,1	[71,0	, 71,1]	69,2	-2,6 %	-1,0 %
Week 31	69,0	[69,0	, 69,0]	65,4	-5,2 %	0,8 %
Week 32	69,7	[69,7	, 69,7]	68,7	-1,4 %	-1,2 %

Week 33	72,7	[72,7	,	72,7]	67,7	-7,0	%	-2,7	%
Week 34	82,4	[82,4	,	82,4]	76,6	-7,1	%	-7,1	%
Week 35	89,2	[89,2	,	89,2]	88,8	-0,4	%	0,4	%
Week 36	92,4	[92,4	,	92,4]	96,2	4,2	%	4,6	%
Week 37	89,5	[89,5	,	89,6]	91,2	1,9	%	1,7	%
Week 38	91,3	[91,3	,	91,4]	83,7	-8,4	%	-7,7	%
Week 39	95,4	[95,4	,	95,4]	95,4	0,0	%	-1,3	%
Week 40	94,0	[94,0	,	94,0]	95,1	1,2	%	3,3	%
Week 41	91,2	[91,2	,	91,3]	91,2	-0,1	%	-0,3	%
week 42	89,4	[89,4	,	89,5]	88,8	-0,8	%	-3,1	%
Week 43	92,2	[92,2	,	92,2]	89,7	-2,7	%	2,5	%
Week 44	93,1	[93,1	,	93,1]	87,6	-5,9	%	0,6	%
Week 45	90,7	[90,7	,	90,7]	93,5	3,1	%	7,2	%
Week 46	89,7	[89,6	,	89,7]	88,5	-1,3	%	0,5	%
Week 47	92,8	[92,8	,	92,9]	91,8	-1,1	%	1,8	%
Week 48	95,6	[95,6	,	95,6]	95,1	-0,5	%	5,0	%

Bijlage V.D: Cijfers voorspellingen(deel II)

Onderstaand de cijfers behorende bij de voorspellingen van model 5.2

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling			
						Model 5.2		WAP	
Week 1	70,9	[70,8	, 70,9]	81,1	14,4	%		
Week 2	94,6	[94,5	, 94,6]	97,8	3,4	%	6,4	%
Week 3	95,2	[95,2	, 95,3]	95,8	0,6	%	2,9	%
Week 4	105,0	[105,0	, 105,0]	98,4	-6,3	%	-4,0	%
Week 5	107,7	[107,7	, 107,7]	103,9	-3,5	%	5,0	%
Week 6	94,2	[94,1	, 94,2]	94,2	0,1	%	1,6	%
Week 7	91,6	[91,6	, 91,6]	97,0	5,9	%	0,3	%
Week 8	85,8	[85,8	, 85,8]	97,1	13,1	%	2,3	%
Week 9	96,0	[96,0	, 96,0]	98,4	2,5	%	3,3	%
Week 10	94,7	[94,7	, 94,7]	98,4	3,9	%	3,3	%
Week 11	88,6	[88,6	, 88,6]	90,8	2,5	%	1,2	%
Week 12	89,6	[89,6	, 89,6]	90,8	1,3	%	-4,4	%
Week 13	97,6	[97,6	, 97,6]	93,7	-4,0	%	-4,3	%
Week 14	82,3	[82,2	, 82,3]	87,8	6,7	%	-3,6	%
Week 15	90,9	[90,9	, 90,9]	86,0	-5,3	%	-2,7	%
Week 16	85,4	[85,4	, 85,4]	92,5	8,2	%	0,3	%
Week 17	104,0	[104,0	, 104,0]	85,8	-17,5	%	10,0	%
Week 18	77,0	[77,0	, 77,0]	78,5	1,9	%	4,4	%
Week 19	82,5	[82,5	, 82,5]	86,4	4,8	%	-2,8	%
Week 20	99,9	[99,8	, 99,9]	93,0	-6,8	%	-1,9	%
Week 21	90,5	[90,5	, 90,5]	97,5	7,8	%	-2,6	%
Week 22	96,4	[96,4	, 96,4]	80,2	-16,8	%	-2,3	%
Week 23	89,3	[89,3	, 89,3]	92,7	3,8	%	-0,2	%
Week 24	85,6	[85,6	, 85,7]	78,1	-8,9	%	2,0	%
Week 25	92,7	[92,7	, 92,7]	94,0	1,3	%	2,9	%
Week 26	92,9	[92,9	, 92,9]	92,0	-0,9	%	5,2	%
Week 27	85,0	[85,0	, 85,0]	82,7	-2,7	%	3,1	%
Week 28	73,0	[73,0	, 73,0]	73,9	1,2	%	2,6	%
Week 29	73,7	[73,6	, 73,7]	73,2	-0,6	%	3,2	%
Week 30	67,8	[67,8	, 67,8]	69,2	2,2	%	-1,0	%
Week 31	64,5	[64,5	, 64,5]	65,4	1,4	%	0,8	%
Week	71,1	[71,1	, 71,2]	68,7	-3,5	%	-1,2	%

32 Week								
33 Week	68,0	[68,0	,	68,0]	67,7	-0,5 %
34 Week	79,9	[79,9	,	79,9]	76,6	-4,1 %
35 Week	95,3	[95,3	,	95,4]	88,8	-6,8 %
36 Week	98,7	[98,7	,	98,7]	96,2	-2,5 %
37 Week	90,9	[90,9	,	90,9]	91,2	0,3 %
38 Week	85,0	[85,0	,	85,0]	83,7	-1,6 %
39 Week	98,8	[98,8	,	98,9]	95,4	-3,5 %
40 Week	101,3	[101,3	,	101,3]	95,1	-6,1 %
41 week	94,1	[94,1	,	94,2]	91,2	-3,1 %
42 Week	90,8	[90,8	,	90,9]	88,8	-2,3 %
43 Week	91,3	[91,3	,	91,4]	89,7	-1,8 %
44 Week	93,0	[93,0	,	93,0]	87,6	-5,8 %
45 Week	88,9	[88,9	,	89,0]	93,5	5,1 %
46 Week	90,1	[90,1	,	90,1]	88,5	-1,7 %
47 Week	93,0	[93,0	,	93,0]	91,8	-1,3 %
48	92,2	[92,2	,	92,2]	95,1	3,1 %

Bijlage V.D: Cijfers voorspellingen(deel III)

Onderstaand de cijfers behorende bij de voorspellingen van model 5.3

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)					Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling				
							Model 5.3		WAP		
Week 1	88,8	[88,8	,	88,8]	81,1	-6,5	%		
Week 2	90,6	[90,6	,	90,6]	97,8	5,2	%	6,4	%
Week 3	94,3	[94,3	,	94,3]	95,8	3,9	%	2,9	%
Week 4	96,1	[96,1	,	96,1]	98,4	0,6	%	-4,0	%
Week 5	96,4	[96,4	,	96,4]	103,9	5,7	%	5,0	%
Week 6	94,2	[94,2	,	94,3]	94,2	1,1	%	1,6	%
Week 7	92,7	[92,7	,	92,7]	97,0	6,3	%	0,3	%
Week 8	92,8	[92,8	,	92,9]	97,1	3,6	%	2,3	%
Week 9	92,3	[92,3	,	92,3]	98,4	5,1	%	3,3	%
Week 10	90,1	[90,1	,	90,1]	98,4	9,7	%	3,3	%
Week 11	89,4	[89,4	,	89,4]	90,8	4,3	%	1,2	%
Week 12	90,9	[90,9	,	90,9]	90,8	-0,8	%	-4,4	%
Week 13	90,0	[89,9	,	90,0]	93,7	-0,4	%	-4,3	%
Week 14	88,9	[88,9	,	88,9]	87,8	4,2	%	-3,6	%
Week 15	85,9	[85,9	,	85,9]	86,0	-2,6	%	-2,7	%
Week 16	87,4	[87,4	,	87,4]	92,5	8,7	%	0,3	%
Week 17	86,1	[86,1	,	86,2]	85,8	-3,3	%	10,0	%
Week 18	86,1	[86,1	,	86,1]	78,5	-7,4	%	4,4	%
Week 19	85,9	[85,9	,	85,9]	86,4	1,9	%	-2,8	%
Week 20	86,0	[86,0	,	86,0]	93,0	5,7	%	-1,9	%
Week 21	88,1	[88,1	,	88,1]	97,5	14,7	%	-2,6	%
Week 22	87,0	[87,0	,	87,0]	80,2	-11,9	%	-2,3	%
Week 23	87,4	[87,4	,	87,4]	92,7	8,9	%	-0,2	%
Week 24	86,6	[86,6	,	86,6]	78,1	-9,1	%	2,0	%
Week 25	88,3	[88,3	,	88,3]	94,0	5,7	%	2,9	%
Week 26	87,2	[87,2	,	87,3]	92,0	2,1	%	5,2	%
Week 27	83,3	[83,3	,	83,3]	82,7	-0,1	%	3,1	%
Week 28	77,1	[77,0	,	77,1]	73,9	-4,1	%	2,6	%
Week 29	73,2	[73,2	,	73,2]	73,2	2,5	%	3,2	%
Week 30	70,5	[70,5	,	70,5]	69,2	-2,6	%	-1,0	%
Week 31	69,9	[69,9	,	69,9]	65,4	-5,2	%	0,8	%
Week	70,5	[70,4	,	70,5]	68,7	-1,4	%	-1,2	%

32 Week								
33 Week	74,9	[74,9	,	75,0]	67,7	-7,0 %
34 Week	81,5	[81,4	,	81,5]	76,6	-7,1 %
35 Week	88,0	[88,0	,	88,0]	88,8	-0,4 %
36 Week	90,4	[90,4	,	90,4]	96,2	4,2 %
37 Week	91,1	[91,1	,	91,1]	91,2	1,9 %
38 Week	92,1	[92,1	,	92,1]	83,7	-8,4 %
39 Week	93,6	[93,6	,	93,6]	95,4	0,0 %
40 Week	93,5	[93,5	,	93,6]	95,1	1,2 %
41 week	91,6	[91,5	,	91,6]	91,2	-0,1 %
42 week	91,0	[90,9	,	91,0]	88,8	-0,8 %
43 Week	91,6	[91,6	,	91,6]	89,7	-2,7 %
44 Week	92,0	[92,0	,	92,0]	87,6	-5,9 %
45 Week	91,1	[91,1	,	91,2]	93,5	3,1 %
46 Week	91,1	[91,0	,	91,1]	88,5	-1,3 %
47 Week	92,7	[92,7	,	92,7]	91,8	-1,1 %
48 Week	94,7	[94,6	,	94,7]	95,1	-0,5 %

Bijlage V.D: Cijfers voorspellingen(deel IV)

Onderstaand de cijfers behorende bij de voorspellingen van model 5.4

	Voorspelling van het aantal poststukken per week plus betrouwbaarheidsinterval (in miljoenen)				Werkelijke hoeveelheid poststukken per week (in miljoenen)	Relatieve afwijking: (Werkelijk-voorspelling) voorspelling	
						Model 5.4	WAP
Week 1	76,0	[75,9	, 76,0]	81,1	6,7 %	
Week 2	90,5	[90,5	, 90,5]	97,8	8,1 %	6,4 %
Week 3	94,9	[94,9	, 95,0]	95,8	0,9 %	2,9 %
Week 4	104,5	[104,5	, 104,5]	98,4	-5,8 %	-4,0 %
Week 5	106,7	[106,7	, 106,7]	103,9	-2,6 %	5,0 %
Week 6	94,5	[94,5	, 94,5]	94,2	-0,3 %	1,6 %
Week 7	91,0	[91,0	, 91,1]	97,0	6,5 %	0,3 %
Week 8	88,0	[88,0	, 88,0]	97,1	10,3 %	2,3 %
Week 9	93,9	[93,9	, 93,9]	98,4	4,8 %	3,3 %
Week 10	94,2	[94,2	, 94,2]	98,4	4,4 %	3,3 %
Week 11	91,1	[91,1	, 91,1]	90,8	-0,3 %	1,2 %
Week 12	88,7	[88,7	, 88,7]	90,8	2,4 %	-4,4 %
Week 13	95,3	[95,3	, 95,3]	93,7	-1,6 %	-4,3 %
Week 14	86,2	[86,2	, 86,3]	87,8	1,8 %	-3,6 %
Week 15	87,1	[87,1	, 87,1]	86,0	-1,2 %	-2,7 %
Week 16	91,1	[91,1	, 91,1]	92,5	1,5 %	0,3 %
Week 17	95,3	[95,3	, 95,3]	85,8	-10,0 %	10,0 %
Week 18	82,9	[82,9	, 82,9]	78,5	-5,3 %	4,4 %
Week 19	82,5	[82,5	, 82,5]	86,4	4,8 %	-2,8 %
Week 20	96,8	[96,8	, 96,9]	93,0	-3,9 %	-1,9 %
Week 21	96,3	[96,3	, 96,3]	97,5	1,3 %	-2,6 %
Week 22	90,1	[90,1	, 90,1]	80,2	-11,0 %	-2,3 %
Week 23	91,9	[91,9	, 91,9]	92,7	0,9 %	-0,2 %
Week 24	86,3	[86,3	, 86,3]	78,1	-9,5 %	2,0 %
Week 25	91,8	[91,8	, 91,8]	94,0	2,3 %	2,9 %
Week 26	92,3	[92,2	, 92,3]	92,0	-0,2 %	5,2 %
Week 27	83,2	[83,2	, 83,2]	82,7	-0,6 %	3,1 %
Week 28	74,5	[74,5	, 74,5]	73,9	-0,9 %	2,6 %
Week 29	74,1	[74,1	, 74,1]	73,2	-1,2 %	3,2 %
Week 30	67,6	[67,6	, 67,6]	69,2	2,4 %	-1,0 %
Week 31	65,1	[65,1	, 65,1]	65,4	0,4 %	0,8 %
Week	72,1	[72,1	, 72,1]	68,7	-4,8 %	-1,2 %

32 Week								
33 Week	70,1	[70,1	,	70,1]	67,7	-3,5 %
34 Week	79,1	[79,1	,	79,1]	76,6	-3,1 %
35 Week	94,9	[94,9	,	94,9]	88,8	-6,4 %
36 Week	95,7	[95,7	,	95,7]	96,2	0,6 %
37 Week	92,1	[92,1	,	92,2]	91,2	-1,0 %
38 Week	87,9	[87,9	,	87,9]	83,7	-4,8 %
39 Week	94,9	[94,9	,	95,0]	95,4	0,4 %
40 Week	102,0	[102,0	,	102,1]	95,1	-6,8 %
41 week	94,2	[94,2	,	94,2]	91,2	-3,2 %
42 week	90,4	[90,3	,	90,4]	88,8	-1,8 %
43 Week	91,6	[91,6	,	91,7]	89,7	-2,1 %
44 Week	94,0	[94,0	,	94,0]	87,6	-6,8 %
45 Week	88,0	[87,9	,	88,0]	93,5	6,3 %
46 Week	90,9	[90,9	,	90,9]	88,5	-2,6 %
47 Week	92,4	[92,4	,	92,4]	91,8	-0,6 %
48 Week	92,5	[92,5	,	92,5]	95,1	2,8 %

Bijlage VI.A Informatie over de invloed van tijd gerelateerde variabelen

In deze bijlage is de invloed van een maandwisseling, kwartaalwisseling, halfjaarwisseling en feestdag beschreven. Tevens is er een indicatie gegeven van de grootte van de invloed van deze tijd gerelateerde variabelen. Aan het eind van het document is de analyse van de invloed van de zomervakantie beschreven.

Maandwisseling januari -februari

Alleen het log-quasi likelihood model 4.1 houdt direct rekening met de invloed van een maandwisseling januari-februari via de variabele 'eind januari'. Omdat in de jaren 2001 en 2002 de maandwisselingen januari-februari in week 5 vallen, is bij de modellen waar er gebruik gemaakt wordt van informatie omtrent voorgemelde grote partijen, de invloed van deze maandwisseling indirect in de invloed van week 5 verwerkt. Bij de overige modellen wordt er indirect rekening gehouden met de invloed van de maandwisseling januari-februari via de berekening van de seizoensinvloed en de weekniveaus voor week 4 en week 5. Omdat de maandwisseling januari-februari in de jaren 1990 tot en met 2002 vaker voorkomt in week 5, is er bij deze modellen een piek te zien in week 5. De grootte van de invloed van de maandwisseling januari-februari is direct bepaald met het log-quasi-likelihood model 4.1 en wordt geschat op een vermenigvuldiging van $e^{\hat{\beta}_{79}} = e^{0,0059} = 1,0059$ met de verwachte hoeveelheid post in week i jaar j indien in week i de maand januari eindigt. In week 5 van het jaar 2003 resulteerde de maandwisseling januari-februari in een verwachte hoeveelheid post van 98,5 miljoen poststukken in plaats van een verwachte hoeveelheid post van 97,9 miljoen poststukken als er geen maandwisseling was geweest in week 5 2003. Het verschil in week 5 2003 is dus maar 0,6 miljoen poststukken. Omdat de werkelijke hoeveelheid post 103,9 miljoen poststukken bedroeg in week 5 2003, geeft dit de indicatie dat er een grotere invloed van een maandwisseling januari-februari op de hoeveelheid post is. Het zou kunnen zijn dat de grootte van de invloed in het begin van de periode 1990 tot en met 2002 een stuk lager was, waardoor de geschatte invloed van de maandwisseling januari-februari van model 4.1 te laag geschat is. Bij het maken van voorspellingen met het log-quasi-likelihood model 4.1 voor een week waarin de maandwisseling januari-februari plaatsvindt, zal er rekening gehouden moeten worden met een eventuele te lage schatting van de grootte van de invloed van de maandwisseling.

Maandwisseling februari -maart

De grootte van de invloed van de maandwisseling februari-maart is bij de drie mathematische modellen indirect verwerkt in de geschatte seizoensinvloed, regressiecoëfficiënt of het weekniveau van week 9, doordat de maandwisseling februari-maart in de jaren 1990 tot 2002 elk jaar in week 9 is. Alleen bij de log-quasi-likelihood modellen en de modellen waar de hoeveelheid post is verminderd met de hoeveelheid post die door grote partijen is gegenereerd, is een duidelijk maximum waarneembaar in week 9. Dit maximum kan veroorzaakt worden doordat er een maximum is door de maandwisseling februari-maart in de jaren 2001 en 2002 er of door een relatief grote hoeveelheid post die door grote partijen gegenereerd is in week 9 van de jaren 2001 en 2002. Omdat er in deze jaren in week 9 er geen relatief grote hoeveelheid post wordt gegenereerd door grote partijen, kan er geconcludeerd worden dat in de jaren 2001 en 2002 een duidelijk maximum is door de maandwisseling februari-maart. Dat dit maximum niet zichtbaar is in de tijdreeksmodellen en de Poisson-modellen waarbij er geen rekening is gehouden met grote partijen, zal dan veroorzaakt worden doordat de invloed van de maandwisseling februari-maart in het begin van de periode 1990 tot en met 2002 waarschijnlijk lager was. Hierdoor zal er bij het maken van voorspellingen met de tijdreeksmodellen en de Poisson-modellen (waarbij er geen rekening is gehouden met grote partijen) voor een week waarin de maandwisseling februari-maart plaatsvindt, rekening gehouden moeten worden met een te lage schatting. Omdat de maandwisseling februari-maart in de jaren 1990 tot en met 2002 elk jaar in week 9 is, kan het verschil met een week 9 zonder de maandwisseling februari-maart niet geanalyseerd worden. Hierdoor is de invloed van de maandwisseling februari-maart niet direct schatbaar.

Maandwisseling maart -april, 1^e kwartaaleinde

Het einde van het eerste kwartaal en tegelijkertijd de maandwisseling maart-april valt in de jaren 1990 tot 2002 in de weken 13 of 14. Omdat in sommige jaren ook Pasen valt in week 14, wordt de geschatte seizoensinvloed, regressiecoëfficiënt en weekniveau van week 14 mede beïnvloed door de invloed van Pasen.

Alleen het log-quasi likelihood model 4.1 houdt direct rekening met de invloed van een maandwisseling maart-april via de variabele 'eind maart' en de invloed van Pasen. Omdat in de jaren

2001 en 2002 de maandwisselingen maart-april in week 13 vallen, is bij de modellen waar er gebruik gemaakt wordt van informatie omtrent voorgemelde grote partijen, de invloed van deze maandwisseling indirect in de invloed van week 13 verwerkt. Bij deze modellen is dan ook een piek zichtbaar in week 13. Bij de overige modellen wordt er indirect rekening gehouden met de invloed van de maandwisseling maart-april via de berekening van de seizoensinvloed en de weekniveaus voor week 13 en week 14. Omdat er ondanks dat de maandwisseling maart-april in de jaren 1990 tot en met 2002 vaker voorkomt in week 14 een piek zichtbaar is in week 13 bij deze modellen, kan er geconcludeerd worden dat de invloed van asen groter is dan de maandwisseling maart-april. De grootte van de invloed van de maandwisseling maart-april is direct bepaald met het log-quasi-likelihood model 4.1 en wordt geschat op een vermenigvuldiging van $e^{\hat{\beta}_{80}} = e^{-0,0032} = 0,9968$ met de verwachte hoeveelheid post in week i jaar j indien in week i de maand maart eindigt. In het jaar 2003 resulteerde de maandwisseling januari-februari in een verwachte hoeveelheid post van 88,4 miljoen poststukken in week 14 (de week waarbij in het jaar 2003 de maandwisseling maart-april plaatsvond) in plaats van een verwachte hoeveelheid post van 88,7 miljoen poststukken als er geen maandwisseling was geweest in week 14 2003. Het verschil in week 14 2003 is dus maar 0,3 miljoen poststukken. Omdat de werkelijke hoeveelheid post 87,8 miljoen poststukken bedroeg in week 14 2003, geeft dit de indicatie dat de grootte van de invloed van de maandwisseling maart-april goed geschat is.

Maandwisseling april 1-mei

Alleen het log-quasi likelihood model 4.1 houdt direct rekening met de invloed van een maandwisseling april mei via de variabele 'eind april'. Omdat in de jaren 2001 en 2002 de maandwisselingen januari-februari in week 18 vallen, is bij de modellen waar er gebruik gemaakt wordt van informatie omtrent voorgemelde grote partijen, de invloed van deze maandwisseling indirect in de invloed van week 18 verwerkt. Bij de overige modellen wordt er indirect rekening gehouden met de invloed van de maandwisseling april-mei via de berekening van de seizoensinvloed en de weekniveaus voor week 17 en week 18. Ondanks dat de maandwisseling april-mei in de jaren 1990 tot en met 2002 maar 1 keer voorkomt in week 17, is er bij deze modellen geen piek te zien in week 18. Dit wordt veroorzaakt door de invloed van koninginndag. Koninginndag valt namelijk 11 keer op een maandag tot en met zaterdag in dezelfde week als de maandwisseling april-mei. Deze officiële feestdag zorgt ervoor dat er minder post gegenereerd in de week waarin koninginndag valt en zorgt er daardoor ook voor dat er geen piek zichtbaar is in week 18.

De grootte van de invloed van de maandwisseling april-mei is direct bepaald met het log-quasi-likelihood model 4.1 en wordt geschat op een vermenigvuldiging van $e^{\hat{\beta}_{81}} = e^{-0,2152} = 0,8064$ met de verwachte hoeveelheid post in week i jaar j indien in week i de maand april eindigt. In week 18 van het jaar 2003 resulteerde de maandwisseling april-mei in een verwachte hoeveelheid post van 77,6 miljoen poststukken in plaats van een verwachte hoeveelheid post van 96,2 miljoen poststukken als er geen maandwisseling was geweest in week 18 2003. Het verschil in week 18 2003 is dus 18,6 miljoen poststukken. Omdat de werkelijke hoeveelheid post 78,5 miljoen poststukken bedroeg in week 18 2003, geeft dit de indicatie dat de grootte van de invloed goed geschat is.

Maandwisseling mei -juni

De grootte van de invloed van de maandwisseling mei-juni is bij de drie mathematische modellen indirect verwerkt in de geschatte seizoensinvloed, regressiecoëfficiënt of het weekniveau van week 22, doordat de maandwisseling mei-juni in de jaren 1990 tot 2002 elk jaar in week 22 is. Alleen bij het log-quasi-likelihood model 4.2 en de tijdreeksmodel 4.4 is er geen piek waarneembaar in week 22. Dit wordt veroorzaakt door het feit dat in het jaar 2001 en in het jaar 2002 er een officiële feestdag valt in week 21, waardoor er bij het log-quasi-likelihood model 4.2 en de tijdreeksmodel 4.4 een piek is in week 21. Dit geeft de indicatie dat er met deze modellen voor een week 21 een te hoge schatting gegeven zal worden indien er geen feestdag is in week 21. Bij tijdreeks model 3.2 en Poisson model 5.2 is er wel een piek waarneembaar in week 22, omdat:

- bij het tijdreeks model 3.2 de invloed van Hemelvaart en Pinksteren is bepaald zonder gebruik te maken van de informatie van week 21 van de jaren 2001 en 2002.
- bij het Poisson model 5.2 de grootte van de invloed van een willekeurige feestdag op nul geschat wordt

Omdat de maandwisseling mei-juni in de jaren 1990 tot en met 2002 elk jaar in week 22 is, kan het verschil met een week 22 zonder de maandwisseling mei-juni niet geanalyseerd worden. Hierdoor is de invloed van de maandwisseling mei-juni niet direct schatbaar.

Maandwisseling juni -juli, 2^e kwartaal einde, 1^e halfjaareinde

Het einde van het eerste halfjaar, het tweede kwartaal en tegelijkertijd de maandwisseling juni-juli valt in de jaren 1990 tot 2002 in de weken 26 of 27. Ondanks dat de maandwisseling juni-juli 1 keer vaker voorkomt in week 27 in de jaren 1990 tot en met 2002, is de geschatte seizoensinvloed, regressiecoëfficiënt of het weekniveau van week 26 hoger dan die van week 27. Dit wordt veroorzaakt doordat in week 27 de basisscholen van 1 van de drie gebieden Noord-, Midden- of Zuid-Nederland al vakantie hebben in week 27 en doordat veel bedrijven hun kwartalen sluiten in respectievelijk de weken 13, 26, 39, en 52.

De grootte van de invloed van de maandwisseling juni-juli is direct bepaald met het log-quasi-

likelihood model 4.1 en wordt geschat op een vermenigvuldiging van $e^{\hat{\beta}_{82}} = e^{-0,0216} = 0,9786$ met de verwachte hoeveelheid post in week i jaar j indien in week i de maand juni eindigt. In week 27 van het jaar 2003 resulteerde de maandwisseling juni-juli in een verwachte hoeveelheid post van 84,8 miljoen poststukken in plaats van een verwachte hoeveelheid post van 86,7 miljoen poststukken als er geen maandwisseling was geweest in week 27 2003. Het verschil in week 27 2003 is dus 1,9 miljoen poststukken. Omdat de werkelijke hoeveelheid post 82,7 miljoen poststukken bedroeg in week 27 2003, geeft dit de indicatie dat de grootte van de invloed goed geschat is.

Maandwisseling juli -augustus

De maandwisseling juli-augustus valt in de jaren 1990 tot 2002 twaalf keer in de week 31 en een keer in week 32, oftewel midden in de bouwvak. Deze bouwvak is dan ook waarschijnlijk de oorzaak dat er ondanks het vaker voorkomen van een maandwisseling in week 31 er een minimum in week 31 is bij de modellen die gebruik maken van de data uit de jaren 1990 tot en met 2002. Bij de modellen die gebruik maken van de informatie over voorgemelde grote partijen en daardoor alleen gebruik maken van data uit de jaren 2001 en 2002, is er een minimum in week 32. Dit geeft de indicatie dat de laatste jaren de laagste hoeveelheid post in de bouwvak verschoven is van week 31 naar week 32. Dit kan veroorzaakt worden doordat de laatste jaren de maandwisseling juli-augustus misschien meer effect heeft.

De grootte van de invloed van de maandwisseling juli-augustus is direct bepaald met het log-quasi-

likelihood model 4.1 en wordt geschat op een vermenigvuldiging van $e^{\hat{\beta}_{83}} = e^{-0,0019} = 0,9981$ met de verwachte hoeveelheid post in week i jaar j indien in week i de maand juli eindigt. In week 31 van het jaar 2003 resulteerde de maandwisseling juli-augustus in een verwachte hoeveelheid post van 73,0 miljoen poststukken in plaats van een verwachte hoeveelheid post van 73,1 miljoen poststukken als er geen maandwisseling was geweest in week 31 2003. Het verschil in week 31 2003 is dus 0,1 miljoen poststukken. Omdat de werkelijke hoeveelheid post 65,4 miljoen poststukken bedroeg in week 31 2003, geeft dit de indicatie dat de grootte van de invloed van de maandwisseling kleiner moet zijn dan -0,0019, of dat de regressiecoëfficiënt die bij week 31 hoort te groot is. Omdat er verwacht wordt dat een maandwisseling een positief effect heeft, neem ik aan dat de regressiecoëfficiënt te groot is.

Maandwisseling augustus -september

De maandwisseling augustus-september valt in de jaren 1990 tot 2002 zeven keer in week 35 en zes keer in week 36.

Ondanks dat de maandwisseling vaker voorkomt in week 35, is er in alle modellen een piek waarneembaar in week 36. Dit wordt veroorzaakt door het feit dat er in week 35 nog steeds invloed van de zomervakantie is op de hoeveelheid post.

De grootte van de invloed van de maandwisseling augustus-september is direct bepaald met het log-

quasi-likelihood model 4.1 en wordt geschat op een vermenigvuldiging van $e^{\hat{\beta}_{84}} = e^{-0,0241} = 0,9762$ met de verwachte hoeveelheid post in week i jaar j indien in week i de maand augustus eindigt. In week 35 van het jaar 2003 resulteerde de maandwisseling augustus-september in een verwachte hoeveelheid post van 94,3 miljoen poststukken in plaats van een verwachte hoeveelheid post van 96,6 miljoen poststukken als er geen maandwisseling was geweest in week 35 2003. Het verschil in week 35 2003 is dus 2,3 miljoen poststukken. Omdat de werkelijke hoeveelheid post 88,8 miljoen poststukken bedroeg in week 35 2003, geeft dit de indicatie dat de grootte van de invloed van de maandwisseling kleiner moet zijn dan -0,0241, of dat de regressiecoëfficiënt die bij week 35 hoort te groot is. Omdat er verwacht wordt dat een maandwisseling een positief effect heeft, neem ik aan dat de regressiecoëfficiënt te groot is.

Maandwisseling september -oktober, 3^e kwartaaleinde

De maandwisseling september-oktober valt in de jaren 1990 tot 2002 twee keer in week 39 en elf keer in week 4036.

Ondanks dat de maandwisseling vaker voorkomt in week 40, is er in alle modellen een piek waarneembaar in week 39. Dit wordt veroorzaakt doordat veel bedrijven hun kwartalen sluiten in de weken 13, 26, 39 en 52.

De grootte van de invloed van de maandwisseling september-oktober is direct bepaald met het log-quasi-likelihood model 4.1 en met het log-quasi-likelihood model 4.2.

Het log-quasi-likelihood model 4.1 schat de grootte van de invloed van de maandwisseling op een vermenigvuldiging van $e^{\hat{\beta}_{85}} = e^{-0,0167} = 0,9834$ met de verwachte hoeveelheid post in week i jaar j indien in week i de maand september eindigt. In week 40 van het jaar 2003 resulteerde de maandwisseling augustus-september in een verwachte hoeveelheid post van 99,4 miljoen poststukken in plaats van een verwachte hoeveelheid post van 101,1 miljoen poststukken als er geen maandwisseling was geweest in week 40 2003. Het verschil in week 40 2003 is dus 1,7 miljoen poststukken. Omdat de werkelijke hoeveelheid post 95,1 miljoen poststukken bedroeg in week 40 2003, geeft dit de indicatie dat de grootte van de invloed van de maandwisseling kleiner moet zijn dan -0,0167, of dat de regressiecoëfficiënt die bij week 40 hoort te groot is. Omdat er verwacht wordt dat een maandwisseling een positief effect heeft, neem ik aan dat de regressiecoëfficiënt bij model 4.1 te groot is.

Het log-quasi-likelihood model 4.2 schat de grootte van de invloed van de maandwisseling op een vermenigvuldiging van $e^{\hat{\beta}_{55}} = e^{0,0235} = 1,0238$ met de verwachte hoeveelheid post in week i jaar j indien in week i de maand september eindigt. In week 40 van het jaar 2003 resulteerde de maandwisseling augustus-september in een verwachte hoeveelheid post van 105,5 miljoen poststukken in plaats van een verwachte hoeveelheid post van 103,6 miljoen poststukken als er geen maandwisseling was geweest in week 40 2003. Het verschil in week 40 2003 is dus 1,9 miljoen poststukken. Omdat de werkelijke hoeveelheid post 95,1 miljoen poststukken bedroeg in week 40 2003, geeft dit de indicatie dat de grootte van de invloed van de maandwisseling kleiner moet zijn dan 0,0235, of dat de regressiecoëfficiënt die bij week 40 hoort te groot is.

Maandwisseling oktober -november

De grootte van de invloed van de maandwisseling oktober-november is bij de drie mathematische modellen indirect verwerkt in de geschatte seizoensinvloed, regressiecoëfficiënt of het weekniveau van week 44, doordat de maandwisseling mei-juni in de jaren 1990 tot 2002 elk jaar in week 44 valt. Bij de modellen die gebruik maken van de informatie over voorgemelde grote partijen en daardoor alleen gebruik maken van data uit de jaren 2001 en 2002, is er geen piek waarneembaar in week 44. Dit kan veroorzaakt worden door een grote hoeveelheid grote partijen in week 44. In de jaren 2001 en 2002 blijkt echter dat er in week 44 geen sprake is van een relatief grote hoeveelheid post die door grote partijen gegenereerd is. Dat er dus geen piek waarneembaar is in week 44 bij de modellen die gebruik maken van de informatie over voorgemelde grote partijen, kan daardoor alleen verklaard worden dat de maandwisseling oktober-november geen positief effect heeft op de hoeveelheid post in de jaren 2001 en 2002.

Maandwisseling november -december

De maandwisseling november-december valt in de jaren 1990 tot 2002 zeven keer in week 48 en zes keer in week 49.

Alleen het log-quasi-likelihood model gebruikt tevens de weken 49 tot en met 52 (zonder de weken met een extreme hoeveelheid post) tijdens het maken van schattingen en is daardoor ook het enige model waarmee een conclusie gemaakt kan worden over de maandwisseling november-december. De grootte van de invloed van de maandwisseling november-december is direct bepaald met het log-quasi-likelihood model 4.1 en wordt geschat op een vermenigvuldiging van $e^{\hat{\beta}_{86}} = e^{0,0185} = 1,01867$ met de verwachte hoeveelheid post in week i jaar j indien in week i de maand november eindigt. In week 48 van het jaar 2003 resulteerde de maandwisseling november-december in een verwachte hoeveelheid post van 101,1 miljoen poststukken in plaats van een verwachte hoeveelheid post van 99,2 miljoen poststukken als er geen maandwisseling was geweest in week 48 2003. Het verschil in week 48 2003 is dus 1,9 miljoen poststukken. Omdat de werkelijke hoeveelheid post 95,1 miljoen poststukken bedroeg in week 48 2003, geeft dit de indicatie dat de grootte van de invloed van de maandwisseling kleiner moet zijn dan 0,0185, of dat de regressiecoëfficiënt die bij week 48 hoort te groot is.

Pasen

Pasen valt in de jaren 1990 tot 2002 in de weken 14, 15, 16 of 17. In het jaar 2003 valt Pasen in week 17. Omdat in sommige jaren ook de maandwisseling maart-april valt in week 14, wordt de geschatte grootte van de invloed van Pasen mede beïnvloed door de maandwisseling maart-april.

De grootte van de invloed van Pasen is bij elk mathematisch model geschat. In tabel 6.1 is de invloed van Pasen per model weergegeven.

	Tijdreeks model 3.1	Tijdreeks model 3.2	Log-quasi- likeliheid model 4.1	Log-quasi- likeliheid model 4.2	Poisson model 5.1
(1) Geschatte invloed Pasen	-0.225	-0.224	-0.1580	-0.2341	-0.00003
(2) Verwachte hoeveelheid post week 17 2003 incl. invloed Pasen	78,0 miljoen	62,5 miljoen	85,5 miljoen	89,5 miljoen	88,7 miljoen
(3) Verwachte hoeveelheid post week 17 2003 excl. invloed Pasen	97,8 miljoen	72,8 miljoen	100,1 miljoen	107,5 miljoen	88,8 miljoen
(4) Verschil tussen (2) en (3)	19,8 miljoen	10,3 miljoen	14,6 miljoen	18,0 miljoen	0,1 miljoen

Tabel 6.1: De geschatte invloed van Pasen

Omdat de fitting van de log-quasi-likeliheid modellen beter zijn dan de fitting van de overige modellen, wordt er meer waarde gehecht aan de door deze modellen geschatte grootte van de invloed van Pasen. De door het log-quasi-likeliheid model 4.1 verwachte hoeveelheid post van week 17 2003 inclusief invloed Pasen zit het dichtst bij de gerealiseerde hoeveelheid post (85,7 miljoen). Dit geeft de indicatie dat het log-quasi-likeliheid model 4.1 de grootte van de invloed van Pasen beter schat dan het log-quasi-likeliheid model 4.2.

Hemelvaart

Hemelvaart valt in de jaren 1990 tot 2002 in de weken 19, 20, 21 of 22. In het jaar 2003 valt Hemelvaart in week 17. Omdat in sommige jaren ook de maandwisseling mei-juni valt in week 22, wordt de geschatte grootte van de invloed van Hemelvaart mede beïnvloed door de maandwisseling mei-juni.

De grootte van de invloed van Hemelvaart is bij elk mathematisch model geschat. In tabel 6.1 is de invloed van Hemelvaart per model weergegeven.

	Tijdreeks model 3.1	Tijdreeks model 3.2	Log-quasi- likeliheid model 4.1	Log-quasi- likeliheid model 4.2	Poisson model 5.1
(1) Geschatte invloed Hemelvaart	-0.178	-0.146	-0.1304	-0.2215	-0.00002
(2) Verwachte hoeveelheid post week 22 2003 incl. invloed Hemelvaart	84,8 miljoen	59,6 miljoen	83,5 miljoen	83,3 miljoen	91,1 miljoen
(3) Verwachte hoeveelheid post week 22 2003 excl. invloed Hemelvaart	101,4 miljoen	66,3 miljoen	95,1 miljoen	98,4 miljoen	91,2 miljoen
(4) Verschil tussen (2) en (3)	15,3 miljoen	6,7 miljoen	11,6 miljoen	15,1 miljoen	0,1 miljoen

Tabel 6.2: De geschatte invloed van Hemelvaart

Omdat de fitting van de log-quasi-likelihood modellen beter zijn dan de fitting van de overige modellen, wordt er meer waarde gehecht aan de door deze modellen geschatte grootte van de invloed van Pasen. De verwachte hoeveelheid post voor week 22 2003 van beide log-quasi-likelihood modellen verschillen weinig van elkaar. Hierdoor is het verschil met de gerealiseerde hoeveelheid post (80,2 miljoen) ongeveer even groot. Omdat de schattingen bij log-quasi-likelihood model 4.1 naar aanleiding van meet data zijn bepaald, verwacht ik dat de schatting van de grootte van de invloed van Hemelvaart van model 4.1 beter is. Omdat de schatting van de grootte van de invloed van Hemelvaart van model 4.2 gemaakt is naar aanleiding van gegevens van de jaren 2001 en 2002, geeft deze schatting de indicatie dat de invloed van Hemelvaart groter is geworden in de laatste jaren.

Pinksteren

Pinksteren valt in de jaren 1990 tot 2002 in de weken 21, 22, 23 of 24. In het jaar 2003 valt Pinksteren in week 24. Omdat in sommige jaren ook de maandwisseling mei-juni valt in week 22, wordt de geschatte grootte van de invloed van Pinksteren mede beïnvloed door de maandwisseling mei-juni. De grootte van de invloed van Pinksteren is bij elk mathematisch model geschat. In tabel 6.1 is de invloed van Pinksteren per model weergegeven.

	Tijdreeks model 3.1	Tijdreeks model 3.2	Log-quasi- likelihood model 4.1	Log-quasi- likelihood model 4.2	Poisson model 5.1
(1) Geschatte invloed Hemelvaart	-0.114	-0.068	-0.1015	-0.1215	-0.00001
(2) Verwachte hoeveelheid post week 24 2003 incl. invloed Hemelvaart	82,8 miljoen	55,2 miljoen	79,2 miljoen	80,1 miljoen	85,90 miljoen
(3) Verwachte hoeveelheid post week 24 2003 excl. invloed Hemelvaart	92,9 miljoen	58,2 miljoen	87,7 miljoen	88,8 miljoen	85,95 miljoen
(4) Verschil tussen (2) en (3)	10,1 miljoen	3,0 miljoen	7,5 miljoen	8,7 miljoen	0,05 miljoen

Tabel 6.3: De geschatte invloeden van Pinksteren

Omdat de fitting van de log-quasi-likelihood modellen beter zijn dan de fitting van de overige modellen, wordt er meer waarde gehecht aan de door deze modellen geschatte grootte van de invloed van Pinksteren. De door het log-quasi-likelihood model 4.1 verwachte hoeveelheid post van week 24 2003 inclusief invloed Pinksteren zit het dichtst bij de gerealiseerde hoeveelheid post (78,1 miljoen). Dit geeft de indicatie dat het log-quasi-likelihood model 4.1 de grootte van de invloed van Pinksteren beter schat dan het log-quasi-likelihood model 4.2.

Koninginnedag

Omdat in de jaren 1995 en 2000 koninginnedag valt op een zondag, wordt er geen invloed verwacht van koninginnedag op de hoeveelheid post. Gedurende de overige jaren in de periode 1990 tot en met 2002 valt koninginnedag tegelijkertijd met de maandwisseling april-mei in week 18. Dit is ook het geval in het jaar 2003. In het jaar 1990 valt in week 18 naast koninginnedag en de maandwisseling april-mei, ook nog eens bevrijdingsdag. De geschatte grootte van de invloed van koninginnedag wordt daarom mede beïnvloed door de maandwisseling en bevrijdingsdag.

Bij de tijdreeks modellen is er aangenomen dat koninginnedag valt in week 18, waardoor het effect van koninginnedag in de seizoensinvloed van week 18 is meegenomen. Aangezien in de jaren 1995 en 2000 er geen invloed verwacht wordt van koninginnedag, wordt er door tijdreeks model 3.1 de invloed van een ontbrekende koninginnedag in week 18 geschat.

Bij de log-quasi-likelihood modellen 4.1 en 4.2 en het Poisson model 5.1 is de invloed van een koninginnedag in week 18 geschat. In tabel 6.4 is de invloed van wel of niet koninginnedag in week 18 per model weergegeven.

	Tijdreeks model 3.1	Log-quasi- likeliheid model 4.1	Log-quasi- likeliheid model 4.2	Poisson model 5.1
(1) Geschatte invloed koninginndag	-	-0.0332	-0.0455	-0.00001
(2) Geschatte invloed geen koninginndag	0,155	-	-	-
(3) Verwachte hoeveelheid post week 18 2003 incl. invloed koninginndag	89,5 miljoen	77,6 miljoen	78,2 miljoen	84,77 miljoen
(4) Verwachte hoeveelheid post week 18 2003 excl. invloed koninginndag	104,5 miljoen	80,2 miljoen	81,4 miljoen	84,81 miljoen
(5) Verschil tussen (3) en (4)	15,0 miljoen	2,6 miljoen	3,2 miljoen	0,04 miljoen

Tabel 6.4: De geschatte invloed van koninginndag

Omdat de fitting van de log-quasi-likeliheid modellen beter zijn dan de fitting van de overige modellen, wordt er meer waarde gehecht aan de door deze modellen geschatte grootte van de invloed van Koninginndag. De door het log-quasi-likeliheid model 4.2 verwachte hoeveelheid post van week 18 2003 inclusief invloed koninginndag zit het dichtst bij de gerealiseerde hoeveelheid post (78,5 miljoen). Dit geeft de indicatie dat het log-quasi-likeliheid model 4.2 de grootte van de invloed van Pinksteren beter schat dan het log-quasi-likeliheid model 4.1.

Bevrijdingsdag

Bevrijdingsdag is een officiële feestdag in de jaren 1990, 1995 en 2000. Omdat tijdreeks model 3.2 en log-quasi-likeliheid model 4.2 alleen gebruik maken van data van de jaren 2001 en 2002, kan de grootte van de invloed van deze officiële feestdag niet met deze modellen geschat worden. De tijdreeks model 3.1, die gebruik maakt van de data van de jaren 1990 en 2002, kan geen schatting geven over de grootte van de invloed, omdat er met de drie weken waarin bevrijdingsdag een officiële feestdag is verschillende indicaties gegeven worden over de grootte van de invloed van bevrijdingsdag. Bij de log-quasi-likeliheid modellen 4.1 en het Poisson model 5.1 is de invloed van bevrijdingsdag wel geschat. In tabel 6.5 is de invloed van bevrijdingsdag voor deze modellen weergegeven.

	Log-quasi- likeliheid model 4.1	Poisson model 5.1
(1) Geschatte invloed koninginndag	-0.0969	-0.0001
(2) Verwachte hoeveelheid post week 19 2003 incl. invloed bevrijdingsdag	88,5 miljoen	84,81 miljoen
(3) Verwachte hoeveelheid post week 19 2003 excl. invloed bevrijdingsdag	97,5 miljoen	84,84 miljoen
(4) Verschil tussen (3) en (4)	9,0 miljoen	0,03 miljoen

--	--	--

Tabel 6.5: Invloed van bevrijdingsdag

Omdat de fitting van de log-quasi-likelihood modellen beter zijn dan de fitting van de overige modellen, wordt er meer waarde gehecht aan de door deze modellen geschatte grootte van de invloed van bevrijdingsdag.

Zomervakantie

In de jaren 1990 tot en met 2002 is er in de weken 27 tot en met 35 sprake van een negatieve invloed op de hoeveelheid post omdat er dan veel mensen op vakantie zijn. Tijdens de bouwvak welke in de weken 30, 31 en 32 elk jaar plaatsvindt, wordt de minimale hoeveelheid post per jaar gegenereerd. De grootte van de invloed van de zomervakantie is bij de drie mathematische modellen indirect verwerkt in de geschatte seizoensinvloed, regressiecoëfficiënt of het weekniveau van week 27 tot en met 35. De log-quasi-likelihood modellen hebben voor zover mogelijk ook de grootte van de invloed van de vakantiespreiding geschat.

Het log-quasi-likelihood model 4.2 heeft alleen de invloed van ‘alleen de basisscholen in Noord-Nederland hebben vakantie’ en de invloed van ‘Noord- en Midden-Nederland hebben vakantie’ geschat.

Indien in week i jaar j alleen de basisscholen in Noord-Nederland vakantie hebben, dan verwacht het log-quasi-likelihood model 4.2 dat de verwachte hoeveelheid post in week i jaar j met 2,88% vermindert indien in week i jaar j alleen Noord-Nederland vakantie heeft.

Indien in week i jaar j de basisscholen in Noord- en Midden-Nederland vakantie hebben, dan verwacht het log-quasi-likelihood model 4.2 dat de verwachte hoeveelheid post in week i jaar j met 0,15% toeneemt indien in week i jaar j Noord- en Midden-Nederland vakantie hebben.

Het log-quasi-likelihood model 4.1 heeft met behulp van de data van de jaren 1990 tot en met 2002 de invloeden van:

- ‘alleen de basisscholen in Noord-Nederland hebben vakantie’
- ‘alleen de basisscholen in Midden-Nederland hebben vakantie’
- ‘alleen de basisscholen in Zuid-Nederland hebben vakantie’
- ‘de basisscholen in Noord- en Midden-Nederland hebben vakantie’
- ‘de basisscholen in Noord- en Zuid-Nederland hebben vakantie’

kunnen schatten. Het log-quasi-likelihood model verwacht de volgende effecten:

Op vakantie:	Alleen Noord-Nl.	Alleen Midden-Nl.	Alleen Zuid-Nl.	Noord- en Midden-Nl.	Noord- en Zuid-Nl.	Midden- en Zuid-Nl.
Geschatte invloed	+5,50%	+3,30%	+1,46%	-5,00%	-5,47%	-6,15%

Tabel 6.6: Overzicht geschatte invloeden vakantie

Uit tabel 6.6 blijkt dat Zuid-Nederland de grootste hoeveelheid post creëert. Noord-Nederland creëert de kleinste hoeveelheid post.

In de regressiecoëfficiënten behorende bij week 30, 31 en 32 is de invloed ‘alle basisscholen hebben vakantie’ meegenomen.

Referenties

- [1] "S-Plus 6 for Windows - Guide to Statistics", Insightfull Coporation, July 2001; © 1988-2001
- [2] P.J. Brockwell & R.A.Davis, "Introduction to Time Series and Forecasting", © 1996; Springer-Verlag New york Inc.
- [3] C.Chatfield, "The analysis of time series" 4^e editie 1989 © 1975 Chapman and Hall
- [4] J. DiNardo & J. Johnston, "Econometric Methods", © 1997 The McGraw-Hill Companies, Inc.
- [5] A.J. Dobson, "An introduction to Generalized Linear Models", © 1990, Chapman and Hall
- [6] M.C.M. de Gunst, "Statistical models", © 2001 Vrije Universteit
- [7] M.C.M. de Gunst & Prof. Dr. A.W. van der Vaart "Statistische Data Analyse", © 1995 Vrije Universiteit
- [8] K. van Harn,& P.J. Holewijn, "Inleiding in de waarschijnlijkheidsrekening", © 1993 Vrije Universiteit
- [9] P.K. Kitandis & R.W. Lane, "Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method", © 1984 University of Iowa
- [10] J.F.C. Kingman,"Poisson processes", © 1993 Clarendon Press
- [11] P.McCullagh, J.A.Nelder, "Generalized Linear Models", © 1983;2^e druk 1984, Chapman and Hall
- [12] T.C. Mills, "Time series techniques for economists"; © 1889 Cambridge University Press
- [13] J. Oosterhoff & A.W. van der Vaart, "Algemene Statistiek",2^e druk 1999 Vrije Universiteit
- [14] A.B. Owen,"Empirical likelihood", © 2001 Chapman and Hall