

# Lokale Activiteiten Monitor

Is het iets waard, of kan het in de openhaard?





Afstudeerscriptie van Bart den Nijs in het kader van de studie Bedrijfskunde en Informatica, aan de Vrije Universiteit te Amsterdam.

Utrecht, September 2006

Begeleiders:

Aart de Vos

Harry van Zanten

Drs. F.J. Oevering

Eerste begeleider, Vrije Universiteit

Tweede begeleider, Vrije Universiteit

Rabobank Nederland



## Voorwoord

Voor u ligt de scriptie van mijn afstudeerstage voor de studie Bedrijfskunde en Informatica. Ik heb mijn stage volbracht bij Rabobank Nederland te Utrecht op de afdeling Kennis en Economisch Onderzoek (KEO) / Regionaal onderzoek. Ik heb onderzoek gedaan naar de regionaal economische-conjunctuurontwikkelingen. De data geanalyseerd en vervolgens voorspellingen gemaakt. Hierbij heb ik gebruik gemaakt van statistische data analyse en voorspellingstechnieken uit de tijdreeksanalyse.

Ik wil hierbij nog even een aantal mensen persoonlijk bedanken:

Rogier Aalders	:	Voor de gezelligheid en het halen van de thee.
Arjan Bakkeren	:	Voor hetzelfde commentaar als bij Rogier.
Gilbert Bal	:	Voor de leuke gesprekken over alle meiden en de gezellige vrijdagen waarbij we tijdens het werk muziek zaten te luisteren.
Sergej Bulterman	:	Voor de vriendelijke en uitbundige glimlach.
Frits Oevering	:	Voor het vertrouwen en de motivatie. Hij wist me telkens weer te stimuleren en door hem voelde ik me erg op mijn gemak.
Willem van der Velden	:	Voor het geduld en eveneens het vertrouwen.
Enrico Versteegh	:	Voor de onbekende muziek.
Aart de Vos	:	Voor het sturen in de goede richting.
Lennart de Waard	:	Voor de grappige verhalen en het lachen.

De vriendelijke vrouwen van het secretariaat voor alle fijne gesprekken en de vrolijke gezichten. In het bijzonder Nel Vinke voor alle administratieve rompslomp wat voornamelijk aan het einde van mijn stageduur nodig was.

De rest van KEO en iedereen van KIRA waarmee ik een prima verstandhouding had.

Tenslotte wil ik nog mijn allerliefste moeder bedanken voor al haar goede zorg en geweldige opvoeding. Tot mijn grote spijt maakt ze dit niet meer mee, maar ik weet zeker dat ze enorm trots op me is. Mam ik hou van u.

Voor eventuele vragen en opmerkingen naar aanleiding van deze scriptie, kunt u mailen naar:

[bdnijs@few.vu.nl](mailto:bdnijs@few.vu.nl)

Utrecht, September 2006

Bart den Nijs

<b>Inhoudsopgave</b>		<b>Blz.</b>
	<a href="#">Inleiding</a>	<a href="#">7</a>
<b>H.1</b>	<a href="#">Samenvatting en conclusies</a>	
1.1	<a href="#">Samenvatting</a>	<a href="#">8</a>
1.2	<a href="#">Conclusies</a>	<a href="#">11</a>
1.3	<a href="#">Aanbevelingen voor vervolgonderzoek</a>	<a href="#">12</a>
<b>H.2</b>	<a href="#">Probleemstelling stageopdracht</a>	
2.1	<a href="#">Omschrijving probleem</a>	<a href="#">13</a>
2.2	<a href="#">Aanpak probleemstelling</a>	<a href="#">15</a>
<b>H.3</b>	<a href="#">Resultaten dataonderzoek</a>	
3.1	<a href="#">Onderbouwen keuze van Monitor reeksen</a>	<a href="#">17</a>
3.2	<a href="#">Externe reeksen (CBS, EIM)</a>	<a href="#">19</a>
3.3	<a href="#">Uitwerking probleemstelling</a>	<a href="#">20</a>
3.3.1	<a href="#">Vergelijken Monitor reeksen met externe reeksen</a>	<a href="#">21</a>
3.3.2	<a href="#">Decomposities van Monitor reeksen</a>	<a href="#">26</a>
3.3.3	<a href="#">Voorspellingen van Monitor reeksen</a>	<a href="#">29</a>
<b>H.4</b>	<a href="#">Bijlagen</a>	
4.1	<a href="#">Structurele tijdreeksmodellen</a>	<a href="#">31</a>
4.2	<a href="#">State Space Form</a>	<a href="#">33</a>
4.3	<a href="#">Kalman filter</a>	<a href="#">37</a>
4.4	<a href="#">Bruikbare programmatuur</a>	<a href="#">43</a>
4.5	<a href="#">Achtergrondinformatie Rabobank Nederland</a>	<a href="#">45</a>
4.6	<a href="#">Beschrijving Lokale Activiteiten Monitor</a>	<a href="#">49</a>
4.7	<a href="#">Uitleg bij de gebruikte reeksen</a>	<a href="#">70</a>
4.8	<a href="#">Indeling van bedrijfstakken</a>	<a href="#">79</a>
4.9	<a href="#">Uitdraai STAMP</a>	<a href="#">81</a>
4.10	<a href="#">Tabellen en figuren</a>	<a href="#">89</a>
4.11	<a href="#">Woordenlijst</a>	<a href="#">93</a>
4.12	<a href="#">Literatuurlijst</a>	<a href="#">95</a>

## Inleiding



Regionaal-economische data zijn schaars. Vaak zijn de data laat vrijgegeven met als gevolg dat ze minder bruikbaar zijn voor het maken van prognoses. Je kunt de meest betrouwbare prognoses maken als je beschikt over de meest recente gegevens. Rabobank Nederland zou dit graag willen doen. Om deze reden hebben ze in 1998 een eigen instrument ontwikkeld. Dit instrument is van groot belang op de afdeling regionaal onderzoek, waar ik mijn stage heb gevolgd. In dit verslag is te lezen hoe het allemaal in zijn werking gaat en hoe met behulp van statistische technieken voorspellingen gemaakt kunnen worden die de Rabobank kan gebruiken als toegevoegde waarde bij hun prognoses.

Het verslag bestaat uit 4 hoofdstukken, te beginnen met de samenvatting en conclusies. In dit hoofdstuk worden ook aanbevelingen gegeven voor vervolgonderzoek. Het tweede hoofdstuk geeft een beschrijving van de probleemstelling en de aanpak van het onderzoek. Het derde hoofdstuk bevat de uitwerkingen van de vraagstukken die gesteld zijn in de probleemstelling bij hoofdstuk 2. Ten slotte bevindt zich in het laatste hoofdstuk de bijlagen. Alle achtergrondinformatie en achterliggende theorie zijn hier te vinden.

---

## H.1 Samenvatting en conclusies

De eerste paragraaf van dit hoofdstuk bevat de samenvatting (1.1). Kort zal hier de belangrijkste inhoud van het stageverslag te vinden zijn. Dit moet al een duidelijk beeld geven van de werkwijze en aanpak. In de tweede paragraaf staan de conclusies (1.2). Dit hoofdstuk zal afgesloten worden met aanbevelingen voor vervolgonderzoek (1.3).

### 1.1 Samenvatting

Voor financiële instellingen, zoals de Rabobank, zijn economische data van groot belang. Voor de afdeling Regionaal Onderzoek is natuurlijk regionaal-economische data zeer kostbaar. Helaas zijn deze erg schaars. Als het er al is, dan is het over het algemeen niet recent. Als je voorspellingen wil maken aan de hand van historische gegevens, dan is het van groot belang dat de data zo actueel mogelijk is. Rabobank Nederland zou veel waarde hebben aan data die tijdig worden vrijgegeven. Om deze reden is de Rabobank begonnen met het maken van een instrument, de Lokale Activiteiten Monitor (LAM). Rabobank Nederland krijgt maandelijks de transacties van hun zakelijke klanten door de rekenkamer aangeleverd in een groot database bestand. De gegevens zijn binnen 2 maanden aanwezig, dus behoorlijk recent. Ze bevatten de credit en debet saldo van de bedrijven als ook hun postcodes. Het is een handig instrument om de juiste data op te zoeken uit een grote database van gegevens. Vanwege de toevoeging van de postcodes kan je de gewenste gegevens per regio opvragen. Als je dit van elke maand doet en deze vervolgens achter elkaar plaatst, ontstaan er regionaal-economische tijdreeksen. In het vervolg zullen dit de LAM reeksen genoemd worden.

Rabobank Nederland heeft een drietal vragen opgesteld met betrekking tot de LAM en de reeksen.

De probleemstelling luidt:

1. Zeggingskracht: is het wat?
2. Zit er een regionale conjunctuurmeter in?
3. Zit er een voorspeller in?

#### **Ad. 1. Zeggingskracht: is het wat?**

Voor de Rabobank is het handig om prognoses te maken en te onderbouwen aan de hand van LAM reeksen. Dit is ook wat ze in de voorbije jaren gedaan hebben. Nu is de vraag of de reeksen wel een zeggingskracht hebben. Kunnen ze naar buiten treden met deze gegevens. Is het mogelijk voor de Rabobank om een persbericht in de krant te plaatsen waarbij staat dat de prognoses gebaseerd zijn op reeksen van de LAM? De LAM zal om deze reden getoetst moeten worden. Dit toetsen gebeurt door de reeksen te vergelijken met externe reeksen die door de media als betrouwbaar worden beschouwd.



Het Centraal Bureau voor de Statistiek (CBS) is daar een goed voorbeeld van. Ze verzamelen al sinds 1899 statistieke gegevens en publiceren dit op het internet.

Om een breder en betrouwbaarder oordeel over de LAM data te kunnen geven, is er ook nog gekeken naar reeksen van het Economisch Instituut voor Midden- en kleinbedrijf (EIM). Er is gekozen voor het EIM, omdat die voornamelijk gegevens verzamelen van het midden- en kleinbedrijf. Dit komt het meest overeen met de zakelijke klanten van de Rabobank.

## Ad. 2. Zit er een regionale conjunctuurmeter in?

Een conjunctuurmeter is een instrument die aangeeft hoe de situatie van de conjunctuur is. Regionaal is dit nog niet op de markt, dus zou het een mooie toevoeging zijn. De Rabobank heeft dan een streepje voor op andere ondernemingen. De data moet eerst wel een bepaalde waarde van zeggingskracht hebben anders heeft het geen nut. Vraag 1 van de probleemstelling moet daarom met een bevredigend antwoord naar voren komen.

Als je de reeksen zou kunnen opsplitsen in losse componenten, dan is het mogelijk om hieruit een conjunctuurmeter te halen. Er bestaat een methode die dit kan. Deze methode heet het Kalman filter en is gespecialiseerd in het maken van decomposities van structurele tijdreeksen in de componenten trend, seizoen, cyclus en irregulier [Kalman 1960]. Een structurele tijdreeks kan er in formule vorm als volgt uit zien:

$$\begin{array}{llll}
 1) & y_t & = \mu_t + \psi_t + \gamma_t + \varepsilon_t & \text{voor } t = 1, 2, \dots, T \text{ met } \varepsilon \sim \text{NID}(0, \sigma_\varepsilon^2) \\
 2) & \mu_t & = \mu_{t-1} + \beta_{t-1} + \eta_t & \eta \sim \text{NID}(0, \sigma_\eta^2) \\
 3) & \psi_t & = \text{cyclus } (\kappa_t \text{ storing}) & \\
 4) & \beta_t & = \beta_{t-1} + \zeta_t & \zeta \sim \text{NID}(0, \sigma_\zeta^2)
 \end{array}$$

Om het Kalman filter hier te kunnen gebruiken, moet deze formule omgezet worden naar een toestandsruimte (Engelse term is State Space Form, kortweg SSF). De algemene vorm van een SSF is:

$$\begin{array}{llll}
 5) & \alpha_{t+1} & = & T_t \alpha_t + H_t \varepsilon_t, \quad \text{voor } t = 1, \dots, n \quad \text{met } \alpha_1 \sim \text{N}(\bar{a}, P), \\
 6) & y_t & = & Z_t \alpha_t + G_t \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, I),
 \end{array}$$

waar  $\alpha_t$  de toestands-vector (state-vector) is [Koopman en Ooms 2001].

Voor maandelijkse reeksen is de toestands-vector gelijk aan  $(\mu_t, \beta_t, \gamma_t, \gamma_{t-1}, \dots, \gamma_{t-10}, \gamma_{t-11})'$ . De matrices  $T_t$ ,  $Z_t$ ,  $H_t$  en  $G_t$  heten systeem matrices en deze zijn bekend. Ze staan vast door middel van de keuzes die je maakt van het model.

Aan de hand van de Maximum Likelihood Methode schat het Kalman filter de onbekende parameters. Dit wordt gedaan via de voorspellingsfout decompositie [Harvey 1989]:

$$l = \log p(y_1, \dots, y_n; \varphi) = \sum_{t=1}^n \log p(y_t | y_1, \dots, y_{t-1}; \varphi) = -\frac{n-d}{2} \log(2\pi) - \frac{1}{2} \sum_{t=d+1}^n (\log |F_t| + v_t' F_t^{-1} v_t)$$

De uiteindelijke state-vector is de meest optimale schatting, omdat dan de residuen (afwijking van het geschatte waarde ten opzichte van de werkelijke waarde) geminimaliseerd zijn. De schatting komt dan het beste overeen met de werkelijke waarden.

Als je de structuur van een tijdreeks kan schatten, kan je met behulp van deze schatting iets zeggen over de conjunctuur. Aangezien de LAM reeksen op regionaal niveau zijn, kun je hier spreken van een regionale conjunctuurmeter.

### Ad. 3. Zit er een voorspeller in?

Iedereen is wel bekend met de spreuk "Regeren is vooruitzien". Als je vooruit kan zien (voorspellen wat er gaat gebeuren), dan kan je daar op inspringen. De Rabobank kan er veel profijt van hebben. Ze hebben dan letterlijk een streepje voor op de rest.

Deze vraag sluit aan op de 2<sup>e</sup>. Daar is namelijk met gebruik van het Kalman filter de tijdreeks in componenten gemodelleerd. Het geeft de 'Final State Vector'. Dit is de laatste toestand van de tijdreeks in losse componenten. Zou je hier de volgende stap mee willen berekenen, dan maak je wederom weer gebruik van het Kalman filter.

Achteraf, als de werkelijke waarde bekend geworden is, weet je pas of je voorspelling goed is geweest. Om te testen of je model goede voorspellingen maakt, kun je voorspellingen maken binnen de reeks (in de literatuur beter bekend als 'predictive testing', [Koopman, Harvey, Doornik, en Shephard (2000)]). Stel je hebt 90 waarnemingen. Laat de laatste 12 waarnemingen achterwege en voorspel deze met de overige 78 waarnemingen. Vervolgens kun je de gegevens naast elkaar leggen en je conclusies trekken.

---

## 1.2 Conclusies

De probleemstelling van dit onderzoek bevat 3 vraagstukken. In paragraaf 1.1 zijn ze al even kort samengevat. In hoofdstuk 2 staan ze wat uitgebreider beschreven. In deze paragraaf zullen de vragen kort aan bod komen met de gevonden conclusies.

- Zeggingskracht: is het wat?

Het antwoord hierop is ja. Het is een grote database aan bruikbare financiële gegevens. Als je het vergelijkt met reeksen van het CBS, vind je goede correlaties op bijna alle bedrijfstakken. Meer dan 91% van de reeksen op COROP niveau vertonen een goede correlatie. CBS gebruikt de data om prognoses te doen en om conclusies te trekken uit historische gebeurtenissen. De Rabobank zou vanwege de goede correlatie hetzelfde kunnen doen. De LAM data heeft om deze reden wel zeggingskracht. Op een reeks mag je weinig concluderen, daarom is er ook gekeken naar reeksen van het EIM. Deze vertonen zelfs nog hogere correlaties dan het CBS.

- Zit er een regionale conjunctuurmeter in?

Een regionale conjunctuurmeter houdt in, dat het instrument de situatie van de conjunctuur kan aangeven voor de opgegeven regio. Door de reeksen op te delen in losse componenten kan je dit doen. Via gebruik van het Kalman filter is dit mogelijk. Van elke reeks kan een state-vector berekend worden. Als de state-vector bekend is, dan kan het weer als een structurele tijdreeks worden weergegeven. Vul de geschatte parameters in en maak je conclusies. Zo kan je dit van elk jaar doen. Vergelijk de jaren met elkaar en je weet hoe de situatie van de conjunctuur er voor staat.

- Zit er een voorspeller in?

Nu de onbekende parameters van het model geschat zijn door het Kalman filter, is het mogelijk om voorspellingen te maken. Je hebt nu immers de 'Final State Vector', want de parameters zijn bekend. Voorspellingen binnen de reeks bleken erg goed te zijn (zie voorbeelden in figuur 24, 25 en 26 van bijlagen 4.10), dus mag je er van uit gaan dat ze dat ook zijn voor buiten de reeks. Je kunt dus inderdaad spreken van een voorspeller.

---

### 1.3 Aanbevelingen voor vervolgonderzoek

Een aantal punten die nog verder onderzocht zouden kunnen worden:

- De LAM zou nog verder doorontwikkeld moeten worden. Totale periode in een keer opgeven, maar dat het dan wel gebruik blijft maken van een 2-jarig panel.
- Meer onderzoek naar verklarende variabelen die in het model kunnen worden meegegeven.
- Standaard Excel bladen om gemakkelijk lange reeksen te produceren.
- Onderzoeken of het mogelijk is om CBS reeksen nauwkeurig te schatten aan de hand van de LAM reeksen. Het is dus nodig om een bepaalde factor te vinden die het verband legt tussen de reeksen.
- Vergelijken met de huidige voorspellingstechnieken die gebruikt worden. Is het ook toe te passen op andere gebieden, zoals bij Macro Economisch Onderzoek (MEO).
- Onderzoeken of een bepaalde regio of bedrijfstak gebruikt kan worden als een indicator voor de conjunctuur.

## H.2 Probleemstelling stageopdracht

De faculteit 'Exacte Wetenschappen' van de Vrije Universiteit beschikt over een stagebureau waar studenten terecht kunnen voor stageopdrachten die op dat moment beschikbaar zijn. De stageopdracht voor Rabobank Nederland heeft ook daar in het bestand gestaan. De opdracht bestaat uit 3 vraagstukken die in de eerste paragraaf van dit hoofdstuk beschreven zullen worden (2.1). In de tweede paragraaf van dit hoofdstuk staat de aanpak van deze probleemstelling beschreven (2.2).

### 2.1 Omschrijving probleem

Het is lastig om bruikbare regionaal-economische data te vinden. De data die daarvoor toegankelijk is, is erg schaars en over het algemeen zwaar verouderd. Om deze reden heeft de Rabobank de LAM ontwikkeld om als conjunctuurwijzer voor de locale (d.w.z., onderverdeelt naar geografische locatie) economie te gebruiken. De Rabobank verzamelt de maandelijkse transacties van hun zakelijke klanten sinds januari 1998. Dit zijn transacties op de debet- en op de creditzijde van de rekeningen. Met behulp van deze data en het instrument willen ze de regionaal-economische ontwikkelingen, dat wil zeggen, de conjunctuurontwikkelingen zoals deze worden gerapporteerd door het CBS, van Nederland in kaart brengen en bij voorkeur ook kunnen voorspellen.

Om met de Monitor naar buiten te kunnen treden, moet het instrument worden geijkt aan andere economische indicatoren, zoals de CBS-omzetindices en het verloop van de conjunctuur zoals gerapporteerd door het CBS. Het CBS treedt met soortgelijke informatie naar buiten voor andere branches dan de dienstverlening waar de Rabobank zich op richt. Echter, er is een positieve correlatie tussen deze branches ondervonden, zodat de prestatie van de Monitor geijkt en gemeten kan worden.

De vragen die ze graag beantwoord willen zien, zijn de volgende:

1. Zeggingskracht: is het wat?

Is de data die de Rabobank maandelijks verzamelt en bewerkt in de LAM voldoende genoeg om bepaalde uitspraken te kunnen doen over de huidige situatie (conjunctuur/economische ontwikkeling) op regionaal niveau? Als het antwoord hierop ja is, kan de Rabobank het toepassen en ermee naar buiten treden. Dit geeft ze dan een leading position, omdat de data eerder aanwezig is dan die van het CBS.

2. Zit er een regionale conjunctuurmeter in?

Om te onderzoeken of er een conjunctuurmeter in zit, is het noodzakelijk dat de data te structureren is. De basisstructuur waar het om gaat zijn de volgende vier ([Harvey 1989] en [Koopman, Shephard en Doornik 1998]):

- Trend

Is er een tendentie aanwezig van een reeks om op de lange termijn te stijgen of te dalen?

- Seizoen

Zijn er invloeden aanwezig die betrekking hebben op regelmatige fluctuaties die binnen het tijdsverloop van een jaar optreden (zoals maandelijks of per kwartaal)?

- Cyclus

Gedrag hiervan is gelijk aan die van seizoen, alleen dan betrekking op schommelingen met een periode langer dan een jaar. Bekendst voorbeeld is natuurlijk de economische conjunctuur.

- Irregulier

Is er verder een bepaalde structuur te herkennen in de uitgaven die niet te classificeren valt in de voorgaande 3 categorieën?

Het voordeel hiervan is namelijk, dat je dan uitleg kan geven waarom sommige waarden hoog/laag zijn en tevens helpt dit om op korte termijn voorspellingen te kunnen maken. Met gebruik van dit kan je de tijdreeksen modelleren.

3. Zit er een voorspeller in?

Als je decomposities van reeksen kan maken, is het niet ondenkbaar dat je ook een redelijke voorspeller eruit kan ontwikkelen. Dit is uiteraard erg gewenst om je leading position te kunnen handhaven. Dit heeft grote voordelen, omdat je dan veel eerder uitspraken kan doen over de toekomstige ontwikkelingen.

## 2.2 Aanpak probleemstelling



### Ad. 1

Op de site van het CBS is een heleboel data opgeslagen en vrij te verkrijgen. Deze data zit opgeslagen in een systeem wat 'Statline' heet. Daar is de mogelijkheid om te zoeken op kernwoord en periode, via het selecteren door middel van themabomen of een kaart (dit is uiteraard alleen mogelijk voor regionale cijfers).

Zoeken via de themabomen is over het algemeen een betere optie. Dan zie je in totaal 10 groepen voor je staan. Er zijn een groot aantal groepen doorlopen, maar in principe is de data uit maar een groep gehaald. Dit was de groep 'Macro-economie, financiële instellingen'. Deze is dan weer opgedeeld in 4 subgroepen: Nationale rekeningen, Conjunctuurgegevens, Financiële instellingen en markten, Prijzen. Uit elk van deze subgroepen zijn een aantal data reeksen gehaald waarvan gedacht wordt dat ze wel eens relevant zouden kunnen zijn.

Er zijn ook nog andere instanties die data reeksen bijhouden. Een voorbeeld daarvan is EIM, CPB, NWO, DNB en misschien wel andere banken. Aangezien het CBS door meerdere instellingen wordt gezien als de grootste en meest betrouwbare data, zal voornamelijk daarmee vergeleken worden. Er is nog wel een andere instantie waar naar gekeken is, omdat die betere cijfers heeft voor de agrarische sector en tevens ook onderscheid kan maken tussen de grote van bedrijven (klein- midden- en grootbedrijf). Dit is het Economisch Instituut voor Midden- en klein bedrijf (EIM).

Sommige reeksen moeten aangepast worden. Het komt soms voor dat er reeksen bestaan tot een bepaald jaar en dan gaan de reeksen verder vanaf een ander jaar waarbij er een nieuw indexjaar is opgenomen. Door deze aan elkaar te schakelen, krijg je langere reeksen. Dit is niet erg ingewikkeld, alleen vergt het een klein beetje tijd en inzicht.

Je werkt uiteraard het liefst met maandcijfers, want dan heb je meerdere waarnemingen. Het is altijd beter om met meerdere waarnemingen te werken, want dit zal veel nauwkeuriger resultaten geven. Helaas zijn er niet altijd data te vinden die op maandniveau zijn opgeslagen (zoals bij EIM, daar hebben ze alleen jaarcijfers) en dan zit er helaas niks anders op. Je kunt dan wel correlaties berekenen alleen moet je erg voorzichtig zijn met je redeneringen.

### Werkwijze

Er zijn een heleboel mogelijkheden om te onderzoeken. Hiermee wordt bedoeld dat er enorm veel gebieden zijn opgeslagen waarnaar getoetst kan worden. Je kunt op postcodes, provincies of COROP's (Coördinatie Commissie Regionaal Onderzoekprogramma) onderzoek doen. Het is verstandig om met COROP's te werken, omdat dat een veelgebruikte indeling van Nederland is bij een heleboel instanties. In totaal zijn er 40 COROPs aanwezig. Deze reeksen ga je vervolgens correleren tegen de reeksen van

het CBS. Indien de waarde van de correlatie voldoende hoog wordt geacht, is er sprake van samenhang tussen de reeksen wat natuurlijk graag geconstateerd wordt.

### **Ad. 2**

Hiervoor zal gebruik gemaakt worden van het Kalman filter. Een methode die door dhr. Kalman in 1960 is bedacht en vooral in de systeemtheorie veel gebruikt is. Dit recursieve algoritme kan decomposities maken van tijdreeksen. De traditionele decompositie van een tijdreeks is gegeven door observatie = trend + seizoen + storing (in formulevorm  $y_t = \mu_t + \gamma_t + \varepsilon_t$ ) [Koopman 2000].

Met behulp van State Space modellen (zie 4.2 voor nadere uitleg) kan dit berekend worden. Er zal wel een computerprogramma bij gebruikt moeten worden, want er komt nogal veel rekenwerk bij kijken. Matrix vermenigvuldigingen die Excel niet aankan. Het programma dat hiervoor gebruikt is, is speciaal ontworpen voor dit soort vraagstukken. Het heet STAMP (Structural Time Series Analyser, Modeller and Predictor) en zal in 4.4 nader behandeld worden. Enige wat nog nodig is, is het correct modelleren en interpreteren van de conclusies.

### **Ad. 3**

Eerst checken of het mogelijk is om voorspellingen te doen binnen de huidige reeksen. Onderzoeken of de gemaakte voorspellingen overeenkomen met de werkelijke waarden. Vervolgens zou dit dan mogelijk moeten zijn voor toekomstige waarden. Het Kalman filter kan hierbij ook toegepast worden, want die update telkens als er nieuwe waarden vrij komen. Dit gebeurt eveneens allemaal in het programma STAMP.



---

## H.3 Resultaten dataonderzoek

In dit hoofdstuk staan de resultaten van dit onderzoek beschreven. Tevens vind je hier de uitwerkingen en antwoorden op de vraagstukken. In de eerste paragraaf zijn de keuzes van de te gebruiken reeksen te lezen (3.1). De volgende paragraaf geeft uitleg over de externe reeksen die gekozen zijn en waarom (3.2). De laatste paragraaf geeft de uitwerkingen van de 3 vraagstukken uit de probleemstelling die beschreven zijn in hoofdstuk 2 (3.3). Hier zijn voorbeelden gebruikt van LAM data en een uitdraai van resultaten afkomstig uit het computerprogramma STAMP.

### 3.1 Onderbouwen keuze van Monitor reeksen

Er zijn een heleboel mogelijkheden om reeksen te vormen en die vervolgens te toetsen. Vanwege de grote verscheidenheid in keuzes, moet je bepaalde grenzen stellen anders is het niet overzichtelijk meer en te veel tijdrovend. In dit hoofdstuk zal nader ingegaan worden op de keuzes die genomen zijn om dit te bereiken.

Allereerst is er gekozen om te werken met acht 'hoofd'-bedrijfstukken. De LAM maakt gebruik van gegevens van 21 verschillende bedrijfstukken maar levert ook direct de acht 'hoofd' bedrijfstukken (Zie bijlagen 4.6). Dit heeft een aantal voordelen, waaronder ook het voordeel dat deze indeling met grote regelmaat terug te vinden is bij reeksen van externe bronnen. Dit is ook te zien in de bijlagen van hoofdstuk 4 (4.8).

Met de nieuwe LAM heb ik alle mogelijke keuzes berekend voor het samenstellen van een reeks. Dit zijn in totaal 18 mogelijkheden (Ruim/Strikt/Strikst, Alle bedrijven/Dec.bedrijven/Bedrijven periode bestaan en Met schaling/zonder schaling). Hiervan zijn de correlaties berekend ten opzichte van de betreffende CBS reeksen en deze in een Excel sheet geplaatst. Dit om na te gaan met welke instellingen je de beste resultaten zou behalen. Als extra is er ook nog gekeken naar een 1-jarig of 2-jarig bedrijven panel. Een 1-jarig bedrijven panel houdt in dat de gekozen bedrijven ten minste voor een volledig jaar moeten bestaan. 2-jarig spreekt nu natuurlijk voor zich. Zie tabel 1 en 2 in de bijlagen voor het resultaat. De resultaten van Alle en dec bedrijven bleken identiek te zijn, dus zijn de mogelijkheden gereduceerd tot 12.

---

Tabel 1 laat de correlaties zien per bedrijfstak samen met het CBS en het EIM voor een 1-jarig bedrijfspanel. De correlaties zijn geordend per externe bron en aflopend gesorteerd, waarbij de bovenste de hoogste sommatie van correlaties heeft (laatste kolom). Dit is ook voor een 2-jarig bedrijfspanel gedaan (tabel 2). Uit deze bevindingen zijn een aantal dingen te concluderen. Zo is het overduidelijk dat het beter is om geen gebruik te maken van de schaling (zie 4.6 voor uitleg van schaling). De correlaties zijn 'altijd' slechter met schaling dan zonder schaling. Ook is te zien dat je bij bedrijvenselectie het beste kan kiezen dat ze de gehele periode bestaan. Bij alle vier de varianten bestaat de top 3 uit Strikt per, Strikst per en Ruim per. Als je dan nog dieper hierin kijkt, kan je concluderen dat de beste keuze Strikt is. Bij 3 van de 4 staat deze variant op nummer 1 en Strikt heeft van alle mogelijkheden ook de meeste correlaties met de hoogste waarden uit de betreffende kolom (met gele achtergrond). Van nu af aan zullen de reeksen geproduceerd worden met de volgende instellingen: Omzet Credit, Geen schaling, Verkorte bedrijfstakken, 2-jarig bedrijven panel en gehele periode bestaan.

Vervolgens ga je denken over een keuze van de indeling naar regio. Dit kan op een aantal manieren. Je kunt op provincie reeksen maken. Je zou zelfs nog specifieker kunnen kijken, dus naar COROP (Coördinatie Commissie Regionaal Onderzoekprogramma) of naar gemeente. Vanwege de omvang van het aantal gemeenten (467) lijkt me dit geen optie, omdat je dan te veel informatie hebt (indien gewenst wel mogelijk natuurlijk). Reeksen op basis van provincies is ook geen optie, omdat je dan in totaal maar met 12 verschillende waarden heb. Uiteindelijk kom je tot het besluit dat je beter met COROP gebieden kan werken, omdat je dan een grotere hoeveelheid reeksen kan produceren en vergelijken. Nederland is opgedeeld in 40 COROP gebieden, maar de grotere steden zijn vaak nog opgedeeld in meerdere COROP's. Dit kan soms wel nuttig zijn om gebruik van te maken.

---

## 3.2 Externe reeksen (CBS, EIM)



Rabobank Nederland wil natuurlijk weten of de gegevens van de Rabobank klanten wel iets waard zijn. Het zou ideaal zijn als blijkt dat ze aan de hand van deze gegevens prognoses en voorspellingen kunnen doen die identiek zijn aan die van het CBS. Het CBS bestaat al sinds 1899 en staat bekend als een nationaal, onpartijdig en betrouwbaar statistisch instituut.

Naast het CBS zijn er natuurlijk ook andere reeksen mogelijk waarmee je zou kunnen vergelijken. Voorbeelden hiervan zijn EIM en DNB (De Nederlandsche Bank). Groot nadeel van de twee laatstgenoemde instellingen is, dat ze geen gegevens op maandniveau publiceren, maar alleen op jaarniveau. Dit is natuurlijk minder nauwkeurig, want dan heb je maar 7 waarnemingen per reeks. Bij het CBS zijn er een flink aantal reeksen te vinden die wel op maandniveau gepubliceerd zijn en tevens ook meer regionale cijfers.

Uit dit alles blijkt dat de meeste waarde gehecht wordt aan gegevens van het CBS. Zoeken naar een samenhang tussen de LAM reeksen en de CBS reeksen is de voornaamste vereiste. De methode en toepassing hiervoor staan beschreven in de volgende paragraaf.

---

### 3.3 Uitwerking probleemstelling



Er zijn een aantal mogelijkheden om reeksen te vergelijken met elkaar. Het meest gebruikelijke en de bekendste methode is die van de pearson-correlatie [Bain L.J. en Engelhardt M. 1992]. Deze correlatiecoëfficiënt is een maat voor de sterkte van de lineaire samenhang tussen de verschillende reeksen. De waarde kan variëren van -1 tot +1. Een correlatiecoëfficiënt van +1 (-1) betekent dat er een positief (negatief) lineair verband is tussen de beide reeksen.

Een hoge correlatie tussen twee variabelen zegt niet altijd dat de samenhang daadwerkelijk zo goed is. Soms kan het zelfs voorkomen dat de betreffende variabelen (hier reeksen) totaal niets met elkaar te maken hebben. Dat zal hier dan niet het geval zijn, omdat goed gelet is op de definities van de reeksen bij de verschillende instellingen. Alleen de reeksen worden met elkaar vergeleken waarvan je zeker weet dat dezelfde definities bedoeld worden. De beschrijving van deze definities zijn te vinden in de bijlagen (4.7) waar alle gebruikte reeksen uitgelegd zijn. Nu mag je bij een correlatie van ongeveer 0,70 wel spreken van een hoge correlatie, dus een duidelijke samenhang. In de volgende paragraaf zijn de berekeningen en de resultaten van de correlaties te zien.

### 3.3.1 Vergelijken Monitor reeksen met externe reeksen



Zoals al eerder in hoofdstuk 2 is vermeld, zal eerst begonnen worden met het vergelijken van reeksen op jaarcijfers. Vervolgens kan er later gekeken worden naar maandcijfers.

Aangezien het CBS de belangrijkste indicator is om mee te vergelijken, zal daar mee begonnen worden. In het vervolg zal er nog correlaties berekeningen gemaakt worden tussen EIM en de Rabobank.

Van geheel Nederland is er een reeks gemaakt vanaf januari 1998 t/m augustus 2005. In het vervolg zal van dezelfde reeks gebruik gemaakt worden om mee te werken. De reeks van Nederland is vergeleken met soortgelijke reeksen van het CBS en EIM. Van CBS zijn helaas niet alle bedrijfstakken te vinden, maar bij EIM gelukkig wel. Enig nadeel is dat EIM met jaarcijfers werkt in plaats van maandcijfers. Dit heeft tot gevolg dat er maar 6 waarden vergeleken kunnen worden. Dit neemt niet weg dat de resultaten best behoorlijk zijn te noemen.

Er is een reeks van de omzetindices van geheel Nederland per bedrijfstak (de 8 in totaal) geproduceerd die deels eerder is berekend door de LAM. De laatste 2 jaar dat nog niet gedaan waren, zijn toegevoegd. Nu is er dus een reeks van de Credit omzet per bedrijfstak over geheel Nederland van de periode januari 1998 t/m augustus 2005.

Op Statline van CBS is er gezocht naar jaarindices waarmee ze vergeleken kunnen worden. Eveneens de datasets die te vinden zijn op de kennisbank van EIM. Met beide zijn er correlaties berekend om te zien of deze overeenkomen. Tabel 3 laat de correlatie zien van Rabobank tegen CBS op jaarbasis:

Agrarisch	0,90
Industrie	0,46
Bouw	0,71
Handel	0,72
Horeca en recreatie	0,99
Transport	0,93
Commerciële dienstverlening	-
Overige Dienstverlening	-

Hiervan zijn er op CBS maar gegevens tot en met 2003

Iets vergelijkbaars op CBS was alleen van 2001 tot en met 2004

Geen vergelijkbare reeksen kunnen vinden op CBS

Idem als bij commerciële dienstverlening

Bron: CBS; Bewerking: Rabobank Tabel 3.

Het is opvallend dat industrie behoorlijk lager ligt dan de andere bedrijfstakken die bekend zijn. Eens kijken wat de correlatie zal zijn als de data van de Rabobank uitgezet word tegen de data van EIM (allen gegevens vanaf 1998 tot en met 2004), zie hiervoor tabel 4:

Agrarisch	0,81
Industrie	0,40
Bouw	0,92
Handel	0,91
Horeca en recreatie	0,98
Transport	0,95
Commerciële dienstverlening	0,92
Overige Dienstverlening	0,77

Bron: EIM; Bewerking: Rabobank Tabel 4.

Ook hier scoort de correlatie van industrie veel lager dan de andere branches. De rest is wel beduidend goed te noemen.

De reeksen die hier beschreven staan, zijn opgehoogd zoals beschreven in hoofdstuk 4. In het vorige hoofdstuk bleek al dat zonder ophoging beter correleerde dan met. Even ter bevestiging zal ik daarom ook hier beide varianten gebruiken. De komende tabellen bevatten dezelfde reeksen als bij tabel 3 en 4 alleen dan zonder de ophogingfactor er in verwerkt (zie tabel 5 en 6). De correlaties in het groen zijn een verbetering of van gelijke waarden aan de eerder berekende correlaties:

Agrarisch	0,90
Industrie	0,61
Bouw	0,75
Handel	0,77
Horeca en recreatie	0,96
Transport	0,94
Commerciële dienstverlening	-
Overige Dienstverlening	-

Hiervan zijn er op CBS maar gegevens tot en met 2003

Iets vergelijkbaars op CBS was alleen van 2001 tot en met 2004

Geen vergelijkbare reeksen kunnen vinden op CBS

Idem als bij commerciële dienstverlening

Bron: CBS; Bewerking: Rabobank Tabel 6.

Dit is helemaal niet slecht. Zonder ophoging correleren de LAM data beter met de data van het CBS dan met. Enige bedrijfstak die gedaald is in zijn correlatie, is die van Horeca en recreatie. Daling is maar minimaal (3,01%) en de stijging van sommige bedrijfstakken is aanzienlijk. Zo is de correlatie van de bedrijfstak industrie gestegen met meer dan 33%, van bouw met bijna 6%, handel meer dan 7% en transport met bijna 1,5%.

Eens kijken wat er verandert als hetzelfde wordt gedaan met de jaarcijfers van EIM. Resultaat van de nieuw berekende correlaties, waarbij de data van de Rabobank zonder ophoging vergeleken wordt met jaarcijfers van het EIM, is te vinden in tabel 7.

Agrarisch	0,81
Industrie	0,61
Bouw	0,96
Handel	0,91
Horeca en recreatie	0,99
Transport	0,96
Commerciële dienstverlening	0,87
Overige Dienstverlening	0,85

Bron: EIM; Bewerking: Rabobank Tabel 7.

Er is wederom een flinke stijging in de bedrijfstak industrie (meer dan 52%), bijna 4% in de bouw, handel een lichte daling van nog geen 1 %, Horeca en transport beide een lichte stijging, commerciële dienstverlening een daling van bijna 5% en als laatste heeft de overige dienstverlening een stijging van bijna 10%. Het merendeel vertoont dus daadwerkelijk een verbetering van de correlatie. Je vraagt je nu zeker af of het wel nodig is om de gegevens van de Rabobank op te hogen aan de gegevens van de KvK. Misschien is de data op zichzelf toch al voldoende om mee te werken. Dit is natuurlijk erg interessant en iets wat de bank graag hoort.

Voor de bedrijfstakken Agrarisch, Industrie, Bouw, Handel en Horeca was het ook mogelijk correlaties te kunnen berekenen op maandindices. Dit is vergeleken met de cijfers die van Statline CBS te halen zijn. Agrarisch is als enige tot eind 2003 en de rest is tot halverwege 2005.

De resultaten hiervan staan in tabel 8:

Agrarisch	0,60	Wederom maar gegevens tot en met 2003
Industrie	0,50	
Bouw	0,75	(0,87 op kwartaal niveau)
Handel	0,68	
Horeca en recreatie	0,70	
Transport	-	
Commerciële dienstverlening	-	
Overige Dienstverlening	-	

Bron: CBS; Bewerking: Rabobank Tabel 8.

En weer is het opvallend dat industrie beduidend minder correlatie vertoond dan de andere bedrijfstakken.

Ook hier even kijken wat de correlaties zullen zijn als er niet gebruik gemaakt zou worden van een ophogingfactor. Dit resultaat is in tabel 9 te zien waar wederom de stijgingen in het groen genoteerd zijn:

Agrarisch	0,60	Wederom maar gegevens tot en met 2003
Industrie	0,55	
Bouw	0,65	(0,87 op kwartaal niveau)
Handel	0,65	
Horeca en recreatie	0,67	
Transport	-	
Commerciële dienstverlening	-	
Overige Dienstverlening	-	

Bron: CBS; Bewerking: Rabobank Tabel 9.

Jammer genoeg zijn de mooie resultaten nu iets minder mooi als op jaarniveau. Agrarisch is uiteraard gelijk gebleven. Industrie wederom behoorlijk gestegen alleen ditmaal is de bedrijfstak Bouw flink gedaald (meer dan 13%) en eveneens zijn de overige 2 bedrijfstakken ook lichtjes gedaald.

Het is erg opvallend dat de correlatie in de agrarische sector redelijk laag ligt (0,60 op maandbasis terwijl dit op jaarbasis gelijk was aan 0,90). Hier ligt de oorzaak waarschijnlijk in het verschil in definities van agrarisch. De gegevens die van het CBS gebruikt zijn, zijn afkomstig van LEI (Landbouw Economisch Instituut). Dit zijn omzetindices van de landbouw en visserij prijzen. De definitie van Rabobank is uitgebreider, wat resulteert in verschillen door andere seizoensinvloeden. Als je dan op jaarindices kijkt, zullen deze seizoensinvloeden niet nadrukkelijk aanwezig zijn met als gevolg dat de correlaties hoger zullen uitvallen.

Na het zien van al deze berekeningen, valt er toch nog duidelijk iets op. Alle keren scoort de bedrijfstak industrie lager dan de overige bedrijfstakken. Wat zou hier de oorzaak van kunnen zijn?

Mogelijke antwoorden kunnen zijn:

- Dat de Rabobank niet groot vertegenwoordigd is in die branche. Misschien werken ze in de industrie veel met meerdere rekeningen. Het zou zelfs zo kunnen zijn dat ze bij elke bank een rekening courant hebben. Dat houdt dan in dat er behoorlijk wat vuiligheid aanwezig is wat een grote invloed kan hebben op de berekening van de correlatie.
- Dat Rabobank niet de grote namen (zoals Shell, Hoogovens, DSM etc.) als klant heeft. Grote bedrijven hebben natuurlijk de meeste invloed, omdat die veel grotere bedragen spenderen en meerdere transacties hebben dan de kleinere bedrijven.
- Dat zowel het CBS als de EIM andere definities hanteren dan de Rabobank. Wanneer behoort iets tot de bedrijfstak industrie?

Dit alles bij elkaar kan natuurlijk leiden tot vuiligheid en onnauwkeurigheid in de data.

Aangezien het bekend is dat de Rabobank niet nadrukkelijk vertegenwoordigd is bij de grote bedrijven in de industrie, zijn er opnieuw correlaties berekend. Bij de kennissite van EIM kun je kiezen voor de grootte van het bedrijf (helaas bij CBS niet kunnen vinden). Keuzes zijn: kleinbedrijf [0-9 werkn], middenbedrijf [10-99 werkn], grootbedrijf [100 en meer werkn] en alle bedrijven. Als je deze alle 4 afzonderlijk bekijkt, vind je de correlaties in tabel 10.

kleinbedrijf [0-9 werkn]	0,52
middenbedrijf [10-99 werkn]	0,52
midden- en kleinbedrijf	0,52
grootbedrijf [100 en meer werkn]	0,38
alle bedrijven	0,40

Bron:EIM;Bewerking:Rabobank Tabel 10.

En nogmaals zonder ophoging:

kleinbedrijf [0-9 werkn]	0,71
middenbedrijf [10-99 werkn]	0,71
Midden- en kleinbedrijf	0,59
grootbedrijf [100 en meer werkn]	0,71
alle bedrijven	0,61

Bron:EIM;Bewerking:Rabobank Tabel 11.

Uit beide tabellen kun je opmaken dat de correlatie van Rabobank met CBS gegevens van alle bedrijven inderdaad negatief beïnvloed wordt. Het is daarom beter om bij industrie te werken met middenbedrijf, want dit geeft de hoogste correlatie.

Uiteindelijk wil de Rabobank weten of de LAM ingezet kan worden om te gebruiken als een conjunctuurmeter op regionaal niveau, dus zal er nu correlaties berekend worden van COROP's. Deze zijn voor alle bedrijfstakken te vinden op Statline, alleen helaas geen maandcijfers. Het enige wat ze publiceren is op jaarniveau. In tabel 12 zijn 43 COROP's (Utrecht is opgedeeld in 4) te vinden. Om snel



een overzicht te krijgen, zijn er gradaties aangegeven in de hoogte van de correlatie met behulp van kleuren.

In totaal zijn er 344 cellen waarvan er 47 een zwakke correlatie hebben. Dit komt overeen met 13,66%, wat helemaal geen slechte score is. De zwakke correlaties zijn voornamelijk in de agrarische sector te vinden.

Dit kan 2 dingen betekenen:

- CBS houdt weinig data bij op het agrarische gebied
- Verkeerd gekozen gegevens

Waarschijnlijk is hier beide gaande. Het CBS heeft namelijk weinig gegevens op het gebied van de agrarische sector met als gevolg dat je snel de verkeerde gegevens gebruikt om mee te vergelijken. Voor de overige bedrijfstakken was dit niet ter sprake. Als je opnieuw het totale aantal cellen telt, maar dan zonder de kolom agrarisch, is het gelijk aan 301. Het aantal rode cellen (die met een zwakke correlatie) dat dan over blijft is 27. In procenten is dit 8,97 (bijna 5% lager).

### 3.3.2 Decomposities van Monitor reeksen



Het is meestal verstandig om met een simpel model (het Basic Structural Model) te beginnen. Van hieruit zijn een aantal stappen te volgen om tot een beter model te komen.

Een BSM ziet er als volgt uit:

- 1)  $y_t = \mu_t + \gamma_t + \varepsilon_t$  voor  $t = 1, 2, \dots, T$  met  $\varepsilon \sim \text{NID}(0, \sigma_\varepsilon^2)$
- 2)  $\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$   $\eta \sim \text{NID}(0, \sigma_\eta^2)$
- 3)  $\beta_t = \beta_{t-1} + \zeta_t$   $\zeta \sim \text{NID}(0, \sigma_\zeta^2)$

$y_t$  staat voor de geobserveerde tijdreeks.

$\mu_t$  de lineaire trend die bestaat uit een stochastisch niveau (2) en een stochastische helling (3).

$\gamma_t$  de seizoenscomponent. Deze kan op verschillende manieren worden gespecificeerd.

$\varepsilon_t$  de irregular (storingsterm).

Door gebruik te maken van de autocorrelatiefunctie (ACF), kun je nagaan of er nog een trend en een cyclus aanwezig zijn in je model (zie figuur 23 in bijlagen 4.10 voor een voorbeeld van een ACF). Een ACF bestaat uit een reeks van correlaties  $\rho_1, \rho_2, \dots, \rho_m$ . Hierbij is  $\rho_1$  de correlatiecoëfficiënt tussen de data die precies een tijdstap verwijderd zijn:  $(y_1, y_2), (y_2, y_3), \dots, (y_{n-1}, y_n)$ .  $\rho_2$  staat dan voor de correlatie tussen data die twee tijdstappen verschillen:  $(y_1, y_3), (y_2, y_4), \dots, (y_{n-2}, y_n)$ . Enzovoorts tot en met  $\rho_m$ . De schattingen voor  $\rho_1, \rho_2, \dots, \rho_m$  worden aangeduid met de notatie  $R_1, R_2, \dots, R_m$ . Als er in de residureeks nog een trendmatig signaal aanwezig is, dan zal de reeks  $R_1, R_2, R_3, \dots$  langzaam uitdempen. Als er in de data nog een cyclus aanwezig is, dan zullen er alternerende negatieve en positieve correlaties zichtbaar zijn. Bijvoorbeeld een cyclus met periode 12, heeft alternerende correlaties bij  $R_6, R_{12}, R_{18}, R_{24}, \dots$  enzovoort [Visser 2002].

Er zijn meerdere trendmodellen mogelijk. In paragraaf 1 van de bijlagen worden deze uitgelegd (4.1). Een keuze tussen deze trendmodellen kan gemaakt worden door gebruik van de ACF van elk model na stationair maken door discreet differentiëren ( $y_t, \Delta y_t = y_t - y_{t-1}$  en  $\Delta^2 y_t = \Delta(\Delta y_t)$ ). Als namelijk blijkt dat de reeks  $y_t$  stationair is ( $R_1, R_2, R_3, \dots$  doven snel uit), dan is geen trendmodel nodig. Als de reeks op de ACF zeer snel stationair is ( $-0,5 \leq R_1 \leq 0,0$  en  $R_2 \approx R_3 \approx \dots \approx 0,0$ ), dan is trendmodel (ii) de beste keuze. Als de ACF van  $\Delta y_t$  langzaam uitdooft, dan is het beter om te kijken naar de reeks  $\Delta^2 y_t$ . Als voor deze reeks geldt dat  $R_1/R_2 \approx -4,0$  en  $-0,67 \leq R_1 \leq 0,0$  en  $0,0 \leq R_2 \leq 0,17$ , dan maak je gebruik van model (iii). Geldt er alleen maar  $-0,67 \leq R_1 \leq 0,0$  en  $0,0 \leq R_2 \leq 0,17$ , dan model (i).

Om na te gaan of het model correct geschat is, wordt er gekeken naar de residuen. Residuen zijn afwijkingen van de geschatte waarde op tijdstip  $t$  met de werkelijke waarde op hetzelfde tijdstip  $t$ . Er

zijn een aantal toetsen mogelijk om na te gaan of de reeks goed geschat is (4.3). In de bijlagen staan deze beschreven en aan de hand van een uitdraai (uit STAMP) zal nu kort een voorbeeld volgen. Het betreft hier data van geheel Nederland afkomstig uit de LAM (volledige uitdraai is te vinden in 4.9 van de bijlage).

NL\_index = Trend + Trigo seasonal + Irregular.

Summary statistics

	NL_index	
Std.Error	11.419	Deze wil je zo laag mogelijk hebben.
Normality	4.5336	Deze heeft een Chi-kwadraat(2) verdeling met als kritieke waarden 5.991 bij 5% significantie niveau -> $H_0$ (dat het een Normale verdeling betreft) verwerpen.
H( 23)	7.1624	Goldfeld-Quandt toetsingsgrootte voor heteroskedasticiteit, deze is gelijk aan de ratio laatste m tegen de eerste m sommatie gekwadrateerde residuen $H(m) = \frac{\sum_{t=m+1}^n e_t^2}{\sum_{t=1}^m e_t^2}$ , een $F_{m,m}$ -verdeling. Kritieke waarde voor $F_{23,23} = 2,014424183$ . F-waarde is groter, dus $H_0$ (geen Heteroskedasticiteit) niet verwerpen.
DW	2.1502	Durbin-Watson toetsingsgrootte, deze heeft een $N(2,4/T)$ -verdeling. Dicht bij 2 duidt op geen seriële correlatie, ver beneden 2 op een positieve en ver erboven op een negatieve correlatie.
Q(18, 6)	13.117	Dit is de toetsingsgrootte van Box-Ljung voor test van residuen seriële correlatie gebaseerd op de eerste P residuen autocorrelaties en q is gelijk aan P+1 min het aantal hyperparameters. Heeft een Chi-kwadraat(q) verdeling. Bij 5% 12,592 dus $H_0$ (geen seriële correlatie) niet verwerpen.
Rs^2	0.39002	Deze is positief, dus goed.

Opvallend is dat de residuen niet aan de normality test voldoen. De overige toetsen zijn wel positief. Het model is (uiteraard) niet goed genoeg. Niet verrassend, want het betreft hier een heel simpel model. Meer onderzoek is nodig om dit te verbeteren. Zo is gebleken dat de resultaten beter zijn als er gemodelleerd wordt met vaste seizoensinvloeden. Dit komt overeen met de literatuur [Huisman 1999].

NL\_index = Trend + Fixed seasonal + Irregular

Summary statistics

	NL_index
Std.Error	11.293
Normality	4.4196
H( 32)	4.3711
DW	2.1497
Q( 9, 7)	14.506
Rs^2	0.40333

Nagenoeg dezelfde resultaten, waarvan het merendeel met een lichte verbetering. Door eerst cyclische en vervolgens verklarende variabelen toe te voegen aan het model, kan je het ook aanzienlijk verbeteren. Dit is ook gebleken, want de resultaten werden alsmat beter (zie wederom de uitdraai in 4.9 van de bijlage).

Een veelvoorkomende CBS indicator die als verklarende variabele zou kunnen worden toegevoegd, is die van Industriële Productie (vanaf nu kortweg IP). Deze gebruikt de Rabobank ook bij zijn conjunctuurindicator (welke landelijk erkend wordt en goed presteert).

NL\_index = Trend + 1 Cycle(s) + Fixed seasonal + Expl vars + Irregular

Summary statistics

	NL_index
Std.Error	10.458
Normality	7.0505
H( 32)	3.6434
r( 1)	-0.063537
r(11)	-0.016615
DW	2.0979
Q(11, 6)	11.231
Rs^2	0.49185

Bijna alle toetsen zijn positief. Normality is groter dan 5.991, dus de nul-hypothese (normale verdeling) niet verwerpen. H(32) is groter dan 2.014, dus de nul-hypothese (geen heteroskedasticiteit) niet verwerpen. DW zit dicht bij 2, dus geen sprake van seriële correlatie. Q(11,6) is kleiner dan 12.592, dus volgens Box-Ljung is er sprake van een lichte seriële correlatie. De Rs<sup>2</sup> is gestegen, wat weer positief is.

Door verder te 'spelen' met de verklarende variabelen ontstaat het volgende model wat nog betere schattingen maakt. Nu is de NL\_index\_1 (dit houdt in de NL reeks met een lag van 1) en de IP\_1 (de Industriële Productie met lag 1) toegevoegd bij het model bestaande uit een stochastische trend met een stochastische slope, een cyclisch, een seizoenscomponent die gefixeerd is en een irreguliere component. De resultaten zijn als volgt:

NL\_index = Trend + 1 Cycle(s) + Fixed seasonal + Expl vars + Irregular

Summary statistics

	NL_index
Std.Error	10.035
Normality	6.6737
H( 32)	4.2442
r( 1)	-0.014685
r(11)	-0.00049516
DW	2.0159
Q(11, 6)	11.068
Rs^2	0.53212

Je kan zo in het oneindige wel doorgaan met het toevoegen van verklarende variabelen, maar meer onderzoek kost ook veel meer tijd. De werkwijze en toepassing van STAMP moet hierdoor wel al duidelijk genoeg zijn. Zie verder in de bijlage de uitdraai van 'Further Output'.

### 3.3.3 Voorspellingen van Monitor reeksen



Na het maken van decomposities van tijdreeksen is de vervolgstap natuurlijk het maken van voorspellingen. Dit is ook de derde vraag van de probleemstelling die de Rabobank had. Door middel van het Kalman filter en met gebruik van een programma zoals STAMP, kan een 'Final State Vector' berekend worden. Een voorbeeld van een uitdraai waarbij de parameters geschat zijn, is hieronder te zien. Voor het gemak is er aan het einde van elke regel de parameter toegevoegd zoals deze in het model voorkomt.

Eq 1 : Estimated coefficients of final state vector.

Variable	Coefficient	R.m.s.e.	t-value	
Lvl	114.31	8.6342	13.239 [0.0000]	$\mu$
Slp	0.15304	1.4159	0.10808 [0.9142]	$\beta$
Cy2_1	7.7400	7.2199		
Cy2_2	-1.6829	7.4754		
Sea_1	17.003	8.2213	2.0682 [0.0423]	$Y_1$
Sea_2	-10.053	8.0611	-1.2471 [0.2165]	$Y_2$
Sea_3	0.92937	8.0334	0.11569 [0.9082]	$Y_3$
Sea_4	-0.26583	8.0147	-0.033168 [0.9736]	$Y_4$
Sea_5	-9.8248	7.9973	-1.2285 [0.2233]	$Y_5$
Sea_6	7.2794	7.9858	0.91155 [0.3651]	$Y_6$
Sea_7	12.136	7.9864	1.5195 [0.1331]	$Y_7$
Sea_8	9.6260	8.0010	1.2031 [0.2329]	$Y_8$
Sea_9	2.2820	8.0251	0.28436 [0.7770]	$Y_9$
Sea_10	-4.3576	8.0519	-0.54119 [0.5901]	$Y_{10}$
Sea_11	-14.929	8.0742	-1.849 [0.0686]	$Y_{11}$

Deze vector geeft de geschatte parameters van het model. Hiermee is het mogelijk om voorspellingen te doen. Om te checken of de voorspellingen wel goed zijn, is het verstandig om eerst binnen de tijdreeks te voorspellen. Voorbeeld uitdraai voor een voorspelling binnen de reeks:

Eq 1 : 12 post-sample predictions for Industrie.

Period	Error	R.m.s.e.	Residual	Cusum	Cusum2
2005.1	-3.581	17.70	-0.2023	-0.2023	0.04091
2005.2	0.6294	17.68	0.03559	-0.1667	0.04218
2005.3	2.927	17.68	0.1656	-0.001118	0.06959
2005.4	2.611	17.68	0.1477	0.1466	0.09141
2005.5	-5.180	17.67	-0.2932	-0.1466	0.1774
2005.6	6.595	17.66	0.3735	0.2270	0.3169
2005.7	-0.4111	17.65	-0.02330	0.2037	0.3174
2005.8	-8.290	17.64	-0.4700	-0.2663	0.5383
2005.9	-19.73	17.64	-1.118	-1.385	1.789
2005.10	-11.08	17.64	-0.6282	-2.013	2.184
2005.11	14.24	17.64	0.8076	-1.205	2.836
2005.12	14.41	17.64	0.8173	-0.3881	3.504

Dan heb je namelijk de mogelijkheid om de geschatte waarden te vergelijken met de werkelijke waarden. Indien deze goed blijken te zijn, mag je er wel van uit gaan dat dit voor toekomstige voorspellingen ook het geval is. Voorbeeld udraai van een voorspelling buiten de reeks voor de komende 12 maanden:

Eq 1 : Forecasts for F-Industrie.

Period	Forecast	R.m.s.e.	- Rmse	+ Rmse
2006.1	116.30	32.895	83.408	149.20
2006.2	106.90	34.665	72.231	141.56
2006.3	112.29	36.726	75.564	149.02
2006.4	113.82	38.755	75.064	152.57
2006.5	117.01	40.653	76.360	157.67
2006.6	117.02	42.386	74.630	159.40
2006.7	111.64	43.968	67.675	155.61
2006.8	95.987	45.450	50.537	141.44
2006.9	108.60	46.902	61.696	155.50
2006.10	113.85	48.392	65.457	162.24
2006.11	107.23	49.958	57.267	157.18
2006.12	138.28	51.451	86.830	189.73

## H.4 Bijlagen

Dit laatste hoofdstuk zal alle achterliggende informatie bevatten. De eerste paragraaf geeft uitleg over structurele tijdreeksmodellen (4.1). De paragraaf erna zal ingaan op de State Space Form (4.2). Dan volgt in de derde paragraaf de theorie over het Kalman filter (4.3). Vanwege de hoge moeilijkheidsgraad van de berekeningen kan men dit niet zonder computers en de juiste software doen, dus is daar ook een paragraaf aan gewijd (4.4). Paragraaf 5 zal achtergrondinformatie van de structuur van Rabobank Nederland bevatten als ook van de afdeling waar de stage heeft plaatsgevonden (4.5). De paragraaf die dan volgt gaat in zijn geheel over het instrument de 'Lokale Activiteiten Monitor', zowel de oude als de nieuwe versie (4.6). Uitleg van de gebruikte reeksen is ook aanwezig (4.7). Vanwege verschillende indelingen van bedrijfstakken tussen ondernemingen, is ook dit ter sprake (4.8). Een uitdraai van een aanpak van een reeks in STAMP is vervolgens te lezen (4.9). Voor het leesgemak zijn een aantal grote tabellen en figuren apart gezet in paragraaf 10 (4.10). Dit hoofdstuk zal afsluiten met een woordenlijst (4.11) en als laatste de literatuurlijst (4.12).

### 4.1 Structurele tijdreeksmodellen

Een structureel tijdreeksmodel is gebaseerd op een additieve decompositie van de componenten trend, seizoen, cyclus, verklarende variabelen en irregulier (in de literatuur vaak ook bekend als ruis of storing). Je noemt het structureel, omdat elke component direct interpreteerbaar is.

#### Trend

Is er een tendentie aanwezig van een reeks om op de lange termijn te stijgen of te dalen?

Er zijn in totaal vier trendmodellen te onderscheiden [Harvey 1989]:

- (i)  $\sigma_{\eta}^2 \neq 0$  en  $\sigma_{\zeta}^2 \neq 0$ : zowel de trend als de helling zijn stochastische series; dit geval is bekend als een lokaal lineair trend model.
- (ii)  $\sigma_{\eta}^2 \neq 0$  en  $\sigma_{\zeta}^2 = 0$ : De trend is stochastisch en de helling is constant; dit geval is de trend een random walk met (constante) drift.
- (iii)  $\sigma_{\eta}^2 = 0$  en  $\sigma_{\zeta}^2 \neq 0$ : De helling is stochastisch, terwijl het trend niveau vast staat; De trend heeft de eigenschappen van een geïntegreerde serie van 2<sup>e</sup>-orde. Een model als deze karakteriseert een trend met een smooth verandering.
- (iv)  $\sigma_{\eta}^2 = 0$  en  $\sigma_{\zeta}^2 = 0$ : Zowel de helling als de trend staan vast; In dit geval wordt de trend deterministisch, dus  $\mu_t = \mu_0 + \beta_t$ .

## Seizoen

Zijn er invloeden aanwezig die betrekking hebben op regelmatige fluctuaties binnen het tijdsverloop van een jaar optreden (zoals maandelijks of per kwartaal)?

De seizoenscomponent,  $\gamma_t$ , kan op meerdere manieren gespecificeerd worden:

- Als dummyvariabele  $\gamma_t = \sum_{j=1}^s \gamma_j^* z_{jt}$ , waarbij  $\gamma_j^*$  de seizoenscoëfficiënten zijn die niet afhangen van  $t$  en dus vast zijn over verschillende jaren.

$$z_{jt} \text{ zijn dummyvariabelen met waarden } \begin{cases} 1 & \text{als } t = j, j+s, j+2s, \dots \\ 0 & \text{als } t \neq j, j+s, j+2s, \dots \\ -1 & \text{als } t = s, 2s, 3s, \dots \end{cases}$$

$s$  is de periodiciteit van de serie in een jaar. In geval van maandelijkse cijfers (zoals hier het geval is)

is  $s$  gelijk aan 12. Door de seizoenscoëfficiënten te laten sommeren tot 0,  $\sum_{j=1}^s \gamma_j^* = 0$ , worden de

seizoenscomponenten vastgelegd binnen een jaar. Als een storingsterm  $\omega_t \sim \text{NID}(0, \sigma_\omega^2)$  wordt toegevoegd aan deze som, kunnen de seizoenscomponenten variëren over tijd en is hun som niet meer gelijk aan 0.

- Als een som van golfbewegingen met behulp van cosinus functies (trigonometrische vorm).

$$\gamma_t = \sum_{j=1}^{\lfloor s/2 \rfloor} (\gamma_j \cos \lambda_j t + \gamma_j^* \sin \lambda_j t)$$

Als het aantal jaren in de data klein is (zoals hier het geval is), is het verstandig om de seizoenscomponenten te fixeren omdat er niet genoeg data is om het veranderende patroon te schatten [Huisman 1999].

## Cyclus

Gedrag hiervan is gelijk aan die van seizoenen, alleen dan betrekking op schommelingen met een periode langer dan een jaar. Bekendst voorbeeld is natuurlijk de economische conjunctuur.

Het cyclische component is geacht van een stationair 2<sup>e</sup>-orde autoregressief proces te zijn.

Een 2<sup>e</sup>-orde autoregressief proces, in de literatuur vaak geschreven als AR(2), heeft als algemene vorm

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \xi_t \quad t = 1, \dots, T.$$



## Verklarende variabelen

Het kan zijn dat er een variabele is die invloed heeft op de tijdreeks. Denk hierbij bijvoorbeeld aan rente, werkloosheid, inkomsten, enzovoort.

Onder verklarende variabelen vallen ook de zogenaamde interventievariabelen voor het corrigeren van onregelmatigheden (uitbijters en structurele breuken) in de data. Deze onregelmatigheden kunnen worden gemodelleerd met behulp van dummyvariabelen  $w_{jt}$  die de waarde 1 hebben voor de betreffende gegevens en 0 elders.

## Irregulier

Is er verder een bepaalde structuur te herkennen in de uitgaven die niet te classificeren valt in de voorgaande 4 categorieën?

Vaak wordt de irreguliere component gespecificeerd als een witte ruis. Witte ruis bestaat uit een reeks van serieel ongecorrleerde willekeurige variabelen met een constant gemiddelde, in dit geval 0, en een constante variantie.

Dit geschiedt zodanig dat deze afzonderlijke componenten direct interpreteerbaar zijn:

$$y_t = \text{trend} + \text{seizoen} + \text{cyclus} + \text{verklarende variabelen} + \text{irregulier}$$

Deze afzonderlijke componenten zijn niet direct observeerbaar. Zij worden geschat uit een te definiëren model en de metingen  $y_t$ .

Zoals hierboven al eerder genoemd, gaat het structurele tijdreeksmodel uit van de additieve decompositie [Visser 2002]. Als blijkt dat hiermee het model niet correct geschat wordt, dan kan geprobeerd worden om hier met behulp van een geschikte transformatie naar toe te werken. Zo kan je bijvoorbeeld een multiplicatief model met behulp van een logaritmische transformatie eenvoudig omzetten naar een additief model:

$$y_t = \text{trend} * \text{seizoen} * \text{cyclus} * \text{verklarende variabelen} * \text{irregulier}$$



$$\log(y_t) = \log(\text{trend}) + \log(\text{seizoen}) + \log(\text{cyclus}) + \log(\text{verklarende variabelen}) + \log(\text{irregulier}).$$

Bij het modelleren van tijdreeksen kan het een beperking zijn dat de parameters voor de componenten over het hele tijdsinterval constant moeten worden gekozen. Harvey (1989) heeft daarom een structureel tijdreeksmodel voorgesteld waarbij deze parameters geleidelijk mogen veranderen. Om analyse hiervan eenvoudig mogelijk te maken, moet het structurele model in een andere vorm gebracht worden. Deze vorm heet de toestandsruimte vorm (of beter bekend als de engelse term State Space Form). In de volgende paragraaf zal deze verder uitgeschreven worden.

## 4.2 State Space Form ▲

De directe vertaling van state space naar het Nederlands is 'toestandsruimte', maar over het algemeen wordt deze term niet veel gebruikt. Bijna iedereen praat over state space, omdat dit het meest gebruikelijke is en bekend in de literatuur. Sommige statistici kunnen zich simpelweg gewoon niet vinden in deze vertaling (zoals ook de heer Koopman, 2001).

De algemene vorm van een state space model (ook bekend als SSF, of State Space Form) heeft betrekking tot multivariate tijdreeksen,  $y_t$ , die  $N$  elementen bevat.

Geobserveerde variabelen die gerelateerd zijn aan een  $m \times 1$  vector,  $\alpha_t$ , bekend als de 'state' vector, via de meetvergelijking:

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad t = 1, \dots, T. \quad (1)$$

Waar  $Z_t$  een  $N \times m$  matrix is,  $d_t$  een  $N \times 1$  vector en  $\varepsilon_t$  een  $N \times 1$  vector van seriële ongecorrleerde storingen met een verwachting nul en een covariantie matrix  $H_t$ . Dus  $E(\varepsilon_t) = 0$  en  $\text{Var}(\varepsilon_t) = H_t$ .

Het univariate model ( $N = 1$ ) ziet er als volgt uit

$$y_t = z_t' \alpha_t + d_t + \varepsilon_t, \quad \text{Var}(\varepsilon_t) = h_t, \quad t = 1, \dots, T. \quad (2)$$

Over het algemeen zijn de elementen van  $\alpha_t$  niet observeerbaar, maar worden ze gegenereerd door een 1<sup>e</sup> orde Markov proces.

De transitie vergelijking:

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t, \quad t = 1, \dots, T. \quad (3)$$

Waar  $T_t$  een  $m \times m$  matrix is,  $c_t$  een  $m \times 1$  vector,  $R_t$  een  $m \times g$  matrix en  $\eta_t$  een  $g \times 1$  vector van seriële ongecorrleerde storingen met  $E(\eta_t) = 0$  en  $\text{Var}(\eta_t) = Q_t$ .

Door toevoeging van de volgende twee assumpties is de specificatie van een state space systeem compleet:

(a) De startwaarde vector (initial state vector),  $\alpha_0$ , heeft  $E(\alpha_0) = a_0$  en  $\text{Var}(\alpha_0) = P_0$ .

(b) De storingen  $\varepsilon_t$  en  $\eta_t$  zijn ongecorrleerd met elkaar in alle tijdperiodes, en ongecorrleerd met de

startwaarde vector, dat houdt in:

$$\begin{aligned} E(\varepsilon_t \eta_s') &= 0 \quad \text{voor alle } s, t = 1, \dots, T. \\ E(\varepsilon_t \alpha_0') &= 0, \quad E(\eta_t \alpha_0') = 0 \quad \text{voor } t = 1, \dots, T. \end{aligned} \quad (4)$$

De matrices  $Z_t$ ,  $d_t$  en  $H_t$  van de meetvergelijking en de matrices  $T_t$ ,  $c_t$ ,  $R_t$  en  $Q_t$  van de transitie vergelijking zijn de systeem matrices. Tenzij anders vermeld, wordt er aangenomen dat ze niet-stochastisch zijn. Ondanks dat ze wel mogen veranderen in de loop van de tijd, doen ze dit op een manier die van tevoren vastgesteld stond. Dit heeft als resultaat dat het systeem lineair is en voor alle

waarden van  $t$ ,  $y_t$  uitgedrukt kan worden als een lineaire combinatie van huidige en historische  $\varepsilon_t$ 's en  $\eta_t$ 's met de startwaarde vector,  $\alpha_0$ . Als de systeem matrices  $Z_t$ ,  $d_t$ ,  $H_t$ ,  $T_t$ ,  $c_t$ ,  $R_t$  en  $Q_t$  niet veranderen over de tijd, dan is het model tijdsonafhankelijk (tijdinvariant) [Harvey 1989].

Het doel van de state space form is om een  $\alpha_t$  op te zetten op een manier dat deze alle relevante informatie van het systeem bevat op tijdstip  $t$  en dat dit gedaan is door zo min mogelijk elementen te gebruiken. Een state space form die de lengte van de state vector minimaliseert, heet een minimale realisatie.

De voordelen van het gebruik van een state space model zijn [Koopman 2000]:

- Dat deze gebaseerd is op een structurele analyse van tijdreeks problemen. Een tijdreeks bestaat uit verschillende componenten (trend, seizoen, cyclische bewegingen, kalender effecten en exogene verklarende variabelen en interventies) die allemaal apart kunnen worden gemodelleerd voordat zij aaneen worden geschakeld in een state space model om daarna simultaan te kunnen worden geanalyseerd (hier komt het Kalman filter aan te pas).
- Dat het state space model zeer flexibel is. Het is vrij eenvoudig om veranderingen in de structuur van het model te introduceren (zoals het combineren van waarnemingen met verschillende tijdsvakken).
- Dat het zonder problemen rekening kan houden met niet-waargenomen observaties in de reeks. Soms mis je door storingen in het systeem of onbeholpenheid van de gebruikers wel eens belangrijke data. Deze techniek weet daar mee om te gaan.

Een algemeen toegepast model is het zogenaamde basisstructureel model (Basic Structural Model - BSM) [Harvey 1989]. Om het concept van een structureel model in State space form wat duidelijker te maken, zal de BSM hier uitgeschreven worden. Dit voorbeeld heeft een seizoenseffect met een lengte van 4 tijdstappen.

$$\text{Meetvergelijking} \quad : y_t = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_{1,t} \\ \gamma_{2,t} \\ \gamma_{3,t} \end{pmatrix} + \xi_t \Leftrightarrow y_t = z_t' \alpha_t + \xi_t$$

$$\xi_t \approx NID(0, h_t \sigma^2)$$

$$\begin{aligned}
 \text{Transitievergelijking} & : \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_{1,t} \\ \gamma_{2,t} \\ \gamma_{3,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{1,t-1} \\ \gamma_{2,t-1} \\ \gamma_{3,t-1} \end{pmatrix} + \begin{pmatrix} \lambda_t \\ \zeta_t \\ \omega_t \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow \alpha_t = T\alpha_{t-1} + \eta_t \\
 \eta_t \approx NID(0, \sigma^2 Q) & \quad Q = \begin{pmatrix} \sigma_\lambda^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_\zeta^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\omega^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}
 \end{aligned}$$

### 4.3 Kalman filter

Als een model in een state space form is geschreven, zoals in de vorige paragraaf, dan kunnen een aantal belangrijke algoritmes toegepast worden. Het Kalman filter is daar de voornaamste van.

Het is een recursieve procedure voor het berekenen van de optimale voorspeller van de state vector op tijdstip  $t$ , gebaseerd op de informatie die beschikbaar is op tijdstip  $t$ . Deze informatie bestaat uit de waarnemingen tot en met  $y_t$  (dus  $y_1, y_2, \dots, y_t$ ). De systeem matrices ( $Z_t, H_t, T_t, R_t$  en  $Q_t$ ) samen met de  $a_0$  en  $P_0$  worden verondersteld om ten allen tijde bekend te zijn en hoeven daarom niet expliciet vermeld te worden in de informatieset.

Het Kalman filter kan in 3 situaties onderscheiden worden, dat zijn filteren, smoothen en voorspellen [Visser 2002].

#### Filteren

Het filter stelt zich bij door de voorspelde waarden te vergelijken met de werkelijke, dit gebeurt on-line (houdt in dat het direct gebruikt maakt van de beschikbare data om vervolgens hiermee nieuwe schattingen te maken). Zo ontstaan voor alle waarnemingen  $y_t$ , één-staps-voorspelfouten (aangeduid als innovaties). Het Kalman filter genereert schattingen zodanig dat de som van gekwadrateerde innovaties minimaal is. Dit recursieve proces heet filteren.

#### Smoothen

Betere schattingen kunnen voor een waarneming  $y_t$  verkregen worden door het smoothen (te vertalen in het Nederland als 'gladstrijken', maar smoothen is algemeen bekend) van alle waarnemingen. Dit gebeurt in off-line (alle metingen voor de analyse zijn al aanwezig) situaties. Dit betekent dat een voorspelling voor  $y_t$  niet alleen gebaseerd is op waarnemingen  $y_i, i = 1, \dots, t$ , maar ook op de daaropvolgende waarnemingen  $y_i, i = t+1, \dots, n$ .

In een lineair model zijn er drie smoothing algoritmes: Fixed-point smoothing, Fixed-lag smoothing en Fixed-interval smoothing.

#### Voorspellen

Schattingen zullen gegenereerd worden voor  $y_t$ , met  $t = n+1, n+2, \dots, n+J$ . Het Kalman filter kan hierbij ook betrouwbaarheidsintervallen berekenen. Hoe verder de voorspelling, hoe wijder de betrouwbaarheidsintervallen natuurlijk zullen zijn.

De belangrijkste redenen waarom het Kalman filter het voornaamste algoritme is die gebruikt kan worden, zullen hier nog even genoemd worden:

- In bepaalde systeem applicaties is het Kalman filter van groot belang, vanwege de on-line schattingen. De huidige waarde van een state vector is extreem belangrijk (het zou bijvoorbeeld de coördinaten van een ruimteraket kunnen voorstellen) en het Kalman filter maakt het mogelijk om de state vector continu aan te passen als er nieuwe waarnemingen beschikbaar komen.
- Wanneer de storingen en de startwaarde normaal verdeeld zijn, het de mogelijkheid geeft om de aannemelijkheid functie te berekenen via gebruik van de voorspelde fout decompositie (prediction error decomposition). Dit maakt het mogelijk om elke onbekende parameter in het model te schatten. Tevens verschaft dit de basis voor de statistische toetsing en specificatie van het model.
- Het kan eenvoudig omgaan met ontbrekende observaties.

Nu volgt een omschrijving van de algemene vorm van een Kalman filter [Harvey 1989].

Neem het eerder beschreven state space model.

Laat  $\hat{\alpha}_{t-1}$  de optimale schatter voorstellen van  $\alpha_{t-1}$  gebaseerd op de waarnemingen tot en met  $y_{t-1}$ .

Laat  $P_{t-1}$  de  $m \times m$  covariantie matrix van de schattingsfout voorstellen, oftewel

$$P_{t-1} = E \left[ (\alpha_{t-1} - \hat{\alpha}_{t-1})(\alpha_{t-1} - \hat{\alpha}_{t-1})' \right]. \quad (5)$$

$$\text{Gegeven } \hat{\alpha}_{t-1} \text{ en } P_{t-1}, \text{ dan is de optimale schatter van } \alpha_t \text{ bepaald door } \hat{\alpha}_{t|t-1} = T_t \hat{\alpha}_{t-1} + c_t \quad (6)$$

terwijl de covariantie matrix van de schattingsfout gelijk is aan

$$P_{t|t-1} = T_t P_{t-1} T_t' + R_t Q_t R_t' \text{ voor } t = 1, \dots, T. \quad (7)$$

De vergelijkingen (6) en (7) worden de voorspellingsvergelijkingen genoemd.

Wanneer de nieuwe waarneming,  $y_t$ , bekend is, kan de schatter van  $\alpha_t$ ,  $\hat{\alpha}_{t|t}$ , geupdate worden.

De update vergelijkingen zijn:

$$\begin{aligned} \hat{\alpha}_t &= \hat{\alpha}_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} (y_t - Z_t \hat{\alpha}_{t|t-1} - d_t) \text{ en} \\ P_t &= P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1} \text{ waar } F_t = Z_t P_{t|t-1} Z_t' + H_t, \quad t = 1, \dots, T. \end{aligned} \quad (8)$$

De vergelijkingen (6), (7) en (8) vormen samen het Kalman filter.

Indien gewenst kan je dit ook schrijven als een enkele set van recursies die direct van  $a_{t-1}$  naar  $a_t$  ( $a_{t|t-1}$ ) gaat.

$$\mathbf{a}_{t+1|t} = (\mathbf{T}_{t+1} - \mathbf{K}_t \mathbf{Z}_t) \mathbf{a}_{t|t-1} + \mathbf{K}_t \mathbf{y}_t + (\mathbf{c}_{t+1} - \mathbf{K}_t \mathbf{d}_t) \quad (9)$$

waar de gain matrix,  $K_t$ , gegeven is door  $\mathbf{K}_t = \mathbf{T}_{t+1} \mathbf{P}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1}$ ,  $t = 1, \dots, T$  (10)

De recursie voor de fout covariantie matrix is dan

$$\mathbf{P}_{t+1|t} = \mathbf{T}_{t+1} (\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_{t|t-1}) \mathbf{T}_{t+1}' + \mathbf{R}_{t+1} \mathbf{Q}_{t+1} \mathbf{R}_{t+1}', \quad t = 1, \dots, T \quad (11)$$

Dit is bekend als de Riccati vergelijking.

Een handige notatie van het Kalman filter is in de vorm van (met de dimensies gegeven tussen de haken) [Koopman, Shephard en Doornik 1998, Koopman en Ooms 2001]:

$$\begin{aligned} \mathbf{v}_t &= \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t && (N \times 1) \\ \mathbf{F}_t &= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{G}_t \mathbf{G}_t' && (N \times N) \\ \mathbf{K}_t &= (\mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t \mathbf{G}_t) \mathbf{F}_t^{-1} && (m \times N) \\ \mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t \mathbf{v}_t && (m \times 1) \text{ 1-stap-vooruit voorspelling van de toestandsvector} \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{H}_t \mathbf{H}_t' - \mathbf{K}_t \mathbf{F}_t \mathbf{K}_t' && (m \times m) \text{ geproduceerd door het Kalman filter.} \end{aligned} \quad (12)$$

voor  $t = 1, \dots, n$  en met startwaarde  $\mathbf{a}_1 = \bar{\mathbf{a}}$  en  $\mathbf{P}_1 = \mathbf{P}$  en waar  $\mathbf{v}_t$  de innovatie (voorspellingsfout) is en  $\mathbf{F}_t$  (=var( $\mathbf{v}_t$ )) de variantie ervan. De afgeleide van de voorspellingsfunctie voor de toestand met respect voor de huidige innovatie is het Kalman gain  $\mathbf{K}_t$ . De begin toestands variantie matrix  $\mathbf{P}$  is gegeven door  $\mathbf{P} = \mathbf{P}^* + \kappa \mathbf{P}_\infty$ , waar  $\kappa$  heel groot is (bijvoorbeeld  $10^6$ ). De matrix  $\mathbf{P}^*$  bevat de varianties en covarianties tussen de stationaire elementen van de toestandsvector (nul voor de rest) en  $\mathbf{P}_\infty$  is een diagonale matrix met één voor niet-stationaire en deterministische elementen van de toestand en nul op de overige plekken.

Even terugkomen op het smoothen. Zoals al eerder vermeld, zijn er drie basis smoothing algoritmes [Harvey 1989]:

- Fixed-point smoothing      Deze berekent gesmoothde schattingen van de state vector op een vast tijdstip. Het geeft de  $\mathbf{a}_{\tau|t}$  voor een specifieke waarde van  $\tau$  op alle tijdstippen  $t > \tau$ . Dit algoritme loopt parallel met het Kalman filter.
- Fixed-lag smoothing      Deze berekent gesmoothde schattingen van de state vector met een vaste vertraagde periode, dat is  $\mathbf{a}_{t-j|t}$  voor  $j = 1, \dots, M$  waar  $M$  een maximale vertraagde periode is.

- Fixed-interval smoothing      Deze berekent de volledige set van gesmoothde schattingen, gebruik makend van alle informatie uit  $y_n$ . Het is een off-line techniek welke  $a_{t|T}$  oplevert waar  $t = 1, \dots, T$ . Deze loopt niet parallel aan het Kalman filter, maar is een terugwaartse recursie.

De laatste smoothing techniek wordt het meest gebruikt. Het geeft betere optimale schattingen, omdat het gebruikt maakt van alle informatie die aanwezig is.

In formule vorm is de toestand smoothing algoritme gegeven door [Koopman en Ooms 2001]:

$$\hat{a}_t = a_t + P_t r_{t-1}, \quad V_t = P_t - P_t N_{t-1} P_t, \quad t = n, \dots, 1, \quad (13)$$

waar  $r_{t-1}$  en  $N_{t-1}$  geëvalueerd zijn door de terugwaartse recursie

$$\begin{aligned} e_t &= F_t^{-1} v_t - K_t' r_t && (N \times 1) \\ D_t &= F_t^{-1} + K_t' N_t K_t && (N \times N) \\ r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t && (m \times 1) \\ N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t && (m \times m) \quad \text{met } L_t = T_t - K_t Z_t \text{ en voor } t = n, \dots, 1. \end{aligned} \quad (14)$$

Wanneer observaties  $y_t$  voor  $t = t_0, \dots, t_1 - 1$  ontbreken, zal  $v_t = 0$  en  $K_t = 0$  zijn en de Kalman updates worden dan:

$$a_{t+1} = T_t a_t, \quad P_{t+1} = T_t P_t T_t' + H_t H_t', \quad t = t_0, \dots, t_1 - 1; \quad (15)$$

en terugwaartse smoothing recursie wordt dan:

$$r_{t-1} = T_t' r_t, \quad N_{t-1} = T_t' N_t T_t, \quad t = t_1 - 1, \dots, t_0. \quad (16)$$

Als derde onderscheid van het Kalman filter is de voorspeller genoemd. Ook deze zal hier in formule vorm kort behandeld worden.

Als je de huidige dataset  $y_1, \dots, y_n$  verlengd met ontbrekende waarden, dan kan je met gebruik van het Kalman filter voorspellingen maken samen met hun MSE (mean squared error). Wanneer  $y_{n+j}$  ontbreekt, zal het Kalman filter reduceren tot

$$a_{n+j+1} = T_{n+j} a_{n+j}, \quad P_{n+j+1} = T_{n+j} P_{n+j} T_{n+j}' + H_{n+j} H_{n+j}' \quad (17)$$

welke de state space voorspellings vergelijkingen zijn voor  $j = 1, \dots, J$  waar  $J$  de voorspellingshorizon is. De  $j$ -stappen-vooruit voorspelling van  $y_{n+j}$  is simpelweg gegeven door

$$\hat{y}_{n+j} = Z_{n+j} a_{n+j}, \quad \text{Var}(\hat{y}_{n+j}) = Z_{n+j} P_{n+j} Z_{n+j}' \quad j = 1, \dots, J. \quad (18)$$

Een rij van ontbrekende waarden aan het einde van een steekproef zal daarom een set van  $j$ -stappen-vooruit voorspellingen produceren.



Om na te gaan of het model goede schattingen maakt, zijn er een aantal diagnostische toetsen om naar te kijken. Deze toetsen hebben betrekking op de residuen [Harvey 1989].

### Seriële correlatie

De berekening van de residuen steekproef autocorrelaties is gegeven door

$$r_v(\tau) = \frac{\sum_{t=d+1+\tau}^T (v_t - \bar{v})(v_{t-\tau} - \bar{v})}{\sum_{t=d+1}^T (v_t - \bar{v})^2}, \quad \tau = 1, 2, \dots$$

Als je dit in een correlogram plaatst, geeft deze een indicatie of er seriële correlatie aanwezig is.

Een gecombineerd significantie toets voor de eerste P residuen autocorrelaties is gegeven door een portmanteau toets die gebaseerd is op de gekwadrateerde sommatie van deze autocorrelaties.

De Box-Ljung notatie van de toetsingsgrootte is

$$Q^* = T^*(T^* + 2) \sum_{\tau=1}^P (T^* - \tau)^{-1} r_v^2(\tau), \quad \text{waar } T^* = T - d.$$

In een structureel model heeft deze een asymptotische  $\chi^2$  verdeling met het aantal vrijheidsgraden gegeven door P-(n-1).

### Niet-lineairiteit

De steekproef autocorrelaties van de gekwadrateerde residuen,  $rsq_v(\tau)$ , kan een indicatie geven of er bepaalde vormen van niet-lineairiteit aanwezig zijn.

$$rsq_v(\tau) = \frac{\sum_{t=d+1+\tau}^T (v_t - \sigma_*^2)(v_{t-\tau} - \sigma_*^2)}{\sum_{t=d+1}^T (v_t - \sigma_*^2)^2}, \quad \tau = 1, 2, \dots$$

Dit kan gebruikt worden om een Box-Ljung toets te doen als je  $r_v(\tau)$  zou vervangen door  $rsq_v(\tau)$ .

Mcleod en Li (1983) hebben laten zien dat de asymptotische verdeling van deze

toetsingsgrootte gelijk is aan een  $\chi_p^2$  verdeling.

### Heteroskedasticiteit

Kies  $h$  als de integer die het dichtste bij  $T^*/3$  ligt en pas de berekening

$$H(h) = \frac{\sum_{t=T-h+1}^T v_t^2}{\sum_{t=d+1}^{d+1+h} v_t^2} \text{ toe. Als de enige onbekende parameter in het model } \sigma_*^2 \text{ is, dan zou deze}$$

toets een  $F(h,h)$  verdeling hebben onder de nul hypothese. Dat is de reden waarom in de meeste gevallen deze toetsingsgrootheid  $H(h)$  tegen een  $F(h,h)$  verdeling wordt getest.

### Normaliteit

De gestandaardiseerde derde en vierde momenten van de gemiddelde residuen zijn

$$\sqrt{b_1} = \sigma_*^{-3} \sum \left( v_t - \bar{v} \right)^3 / T^* \text{ en } b_2 = \sigma_*^4 \sum \left( v_t - \bar{v} \right)^4 / T^*, \text{ waar}$$

$$\sigma_*^2 = (T-d-1)^{-1} \sum_{t=d+1}^T \left( v_t - \bar{v} \right)^2.$$

toetsingsgrootheden zijn de basis metingen van de scheefheid en kurtosis. Voor een normale verdeling zouden deze gecentreerd moeten zijn rond 0 en 3 respectievelijk.

Ze zijn asymptotisch normaal verdeeld met  $\sqrt{b_1} \sim AN(0, 6/T^*)$  en  $b_2 \sim AN(3, 24/T^*)$  wanneer het model correct gespecificeerd is (Bowman en Shenton 1975).

Een toets voor niet-normaliteit mag daarom gebaseerd worden op de toetsingsgrootheid

$N = (T^*/6)b_1 + (T^*/24)(b_2 - 3)^2$ . Onder de nul hypothese heeft deze een  $\chi_2^2$  verdeling (mits de steekproef groot genoeg is, anders kan er beter gebruik gemaakt worden van de Shapiro-Wilk toets).

## 4.4 Bruikbare programmatuur



OxMetrics is een familie van software pakketten die een geïntegreerde oplossing leveren voor econometrische analyses van tijdreeksen, voorspellingen, financieel econometrisch modelleren en statistische analyses van cross-sectie en panel data. De kern pakketten zijn GiveWin™, welke de user interface, data handling, en grafieken levert, en Ox Professional™, welke de geïmplementeerde taal levert.

Ox-Professional (J.A. Doornik) is een objectgeoriënteerde programmeeromgeving. Het bestaat uit een krachtige matrixtaal, welke voorzien is van een uitgebreide statistische en econometrische bibliotheek. De specifieke kenmerken van Ox zijn snelheid (Ox wordt daarvoor alom geroemd), een goed gestructureerde syntax en uitgebreide grafische mogelijkheden. Ox kan de belangrijkste bestandsformaten lezen en schrijven, waaronder spreadsheetformaten en GiveWin bestanden. Ox is beschikbaar onder Windows, Linux en verschillende andere Unix systemen. In C geschreven procedures kunnen eenvoudig aan Ox worden toegevoegd.

Software voor maximum likelihood, simulatie, Monte Carlo-experimenten, en econometrische modellen (zoals vector-autoregressie en cointegratie-analyses) zijn onderdeel van Ox. Additionele modules voor fractionele ARIMA modellen, "state space analyses", dynamische panel data analyse, stochastische volatiliteit analyses en "Markov-switching" modellen zijn ook zonder extra kosten beschikbaar. Er zijn verder ook veel bruikbare gratis te downloaden pakketten die je dan in Ox kan aanroepen en gebruiken. Een daarvan is het pakket SsfPack, geschreven door dhr. Koopman, Shephard en Doornik.

Een ander simpel te gebruiken pakket heet loess. Het is ook te downloaden van de oxmetrics.net site. Het is onduidelijk door wie en hoe het geschreven is. De opbouw is vrij simpel, hoewel er wel een flink aantal mogelijkheden zijn om toe te passen. Wat loess doet is decomposities maken van tijdreeksen in trend, seizoen en irregulier componenten. Het werkt snel en gemakkelijk. Het enige wat het mist, is een functie voor voorspellingen. Om deze reden is het beter om SsfPack te gebruiken, want die bevat wel alles.

SsfPack is eigenlijk niks anders dan een pakketje dat vol zit met allemaal handige C routines. Deze C routines zijn berekeningen op het gebied van statistische analyses van univariate en multivariate modellen in een state space form. SsfPack is bruikbaar voor een brede verzameling state space forms: van een simpel tijdsafhankelijk model tot een gecompliceerd tijdsafhankelijk model. Functies kunnen gebruikt worden om standaard modellen zoals ARIMA en cube spline modellen in state space form te plaatsen. Basis functies zijn geïmplementeerd voor filteren, moment smoothing en simulation smoothing. Er zijn meerdere voorgeprogrammeerde functies in SsfPack te gebruiken voor standaard

taken. Voorbeelden hiervan zijn likelihood evaluation, forecasting en signal extraction. Het pakket kan tevens handig gebruikt worden voor het implementeren, fitten en analyseren van Gaussian modellen die relevant zijn voor veel gebieden van de econometrie en statistiek.

De andere elementen van de familie zijn PcGets, PcGive, PcNaive, TSP/GiveWin, G@rch en STAMP. Het zijn allen interactief en eenvoudig te gebruiken tools die kunnen helpen om de gewenste specifieke modellen en voorspellingsbehoeftes op te lossen. STAMP is een pakket dat ontworpen is om structurele tijdreeksen te modelleren en te voorspellen. Dit komt overeen met wat Rabobank Nederland wil. Dit is uiteindelijk ook de software die gebruikt is om deze opdracht te maken en daarom zal ik deze even toelichten.

## **STAMP**

STAMP (geschreven door S.J. Koopman, A.C. Harvey, J.A. Doornik en N. Shephard) staat voor "Structural Time Series Analyser, Modeller and Predictor". STAMP schat de trend- en seizoencomponenten van tijdreeksen met moderne statistische tijdreeksmodellen en genereert punt- en intervalvoorspellingen. De interface van STAMP is eenvoudig van opzet, waardoor de gebruiker zich kan concentreren op modelselectie en interpretatie.

STAMP kan zowel univariate als multivariate tijdreeksen modelleren. Interventies en andere verklarende variabelen kunnen eenvoudig worden toegevoegd. De multivariate modellen bieden opties voor gezamenlijke (trend-)componenten en co-integratie. STAMP maakt het mogelijk om een model interactief te kiezen en geeft uitgebreide diagnostische informatie. STAMP biedt ook mogelijkheden voor automatische verwerking in de achtergrond (in batch).

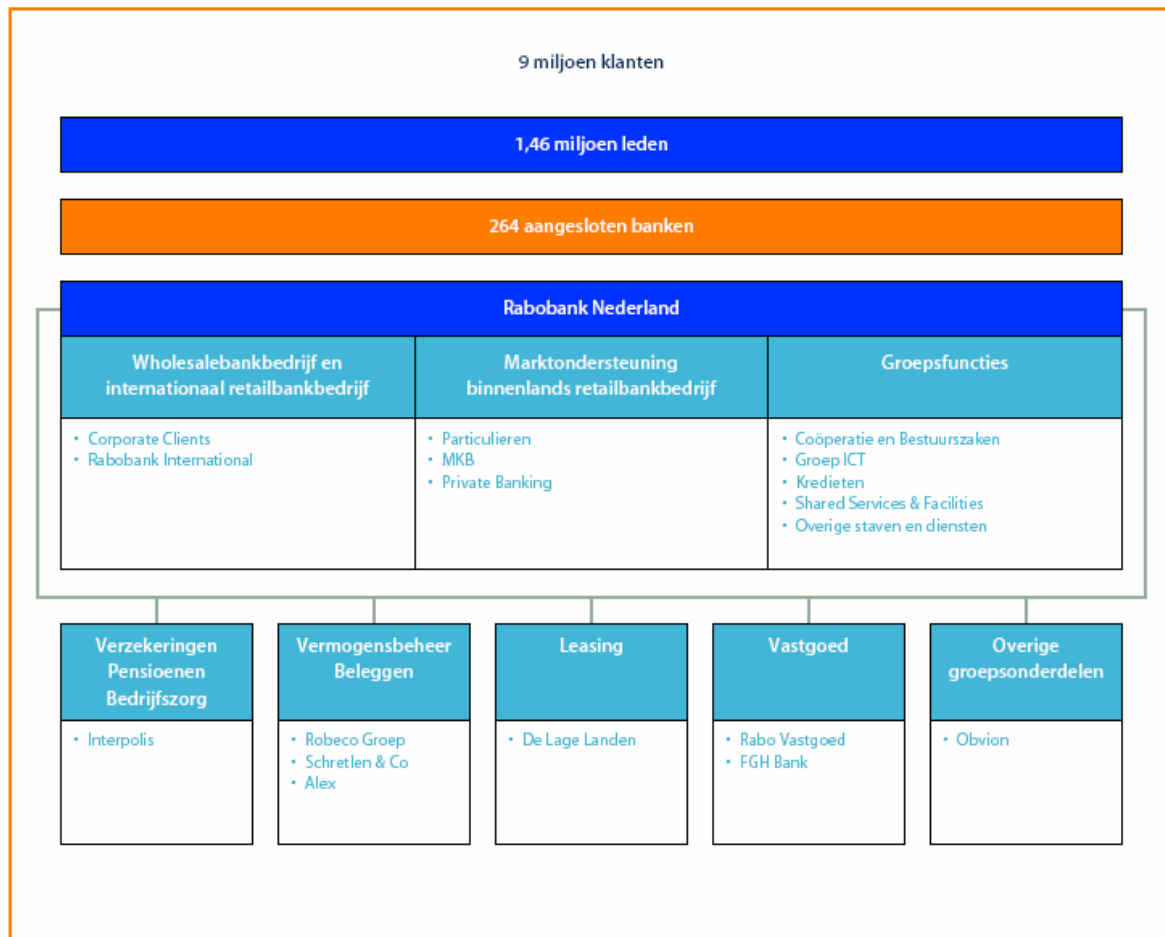
Het STAMP handboek introduceert structurele tijdreeksmodellen en illustreert deze in een brede reeks uitgewerkte toepassingen.

## 4.5 Achtergrondinformatie Rabobank Nederland



### Structuur Rabobank Groep

Allereerst zal de organisatie in kaart gebracht worden aan de hand van uitleg bij het organigram van de Rabobank Groep (zie figuur 11).



Bron: Rabobank

Figuur 11.

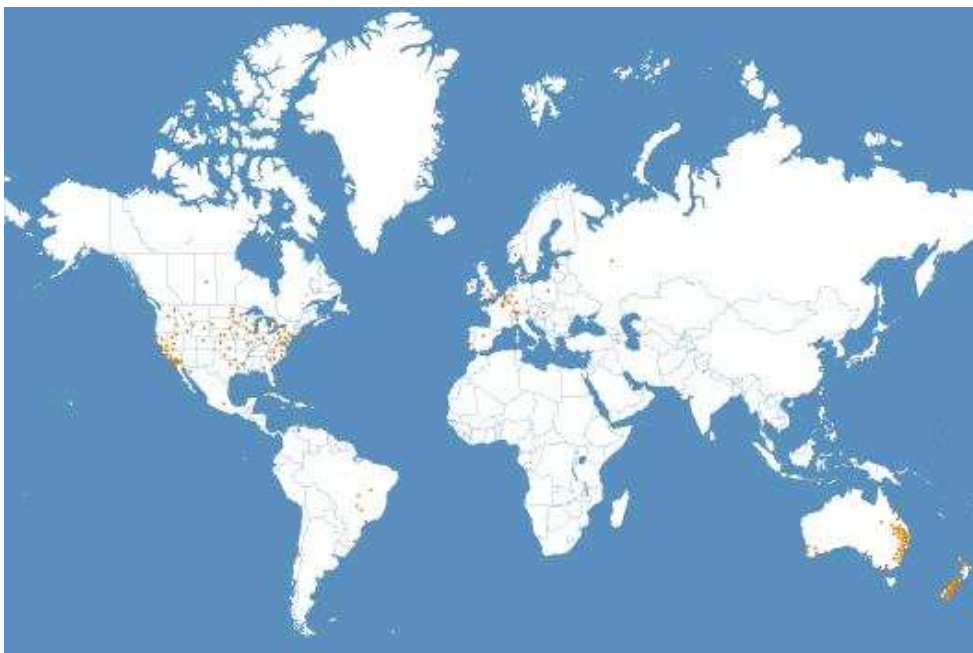
Zoals je ziet heeft de Rabobank Groep ongeveer 9 miljoen klanten en 1,46 miljoen leden. Klanten zijn alle personen die een rekening hebben bij een van de 264 aangesloten lokale banken. Deze lokale banken zijn verspreid over heel Nederland (zie figuur 12).



Bron: Rabobank

Figuur 12.

Verder is de Rabobank Groep met 109 internationale banken gevestigd in 37 verschillende landen (zie figuur 13).



Bron: Rabobank

Figuur 13.

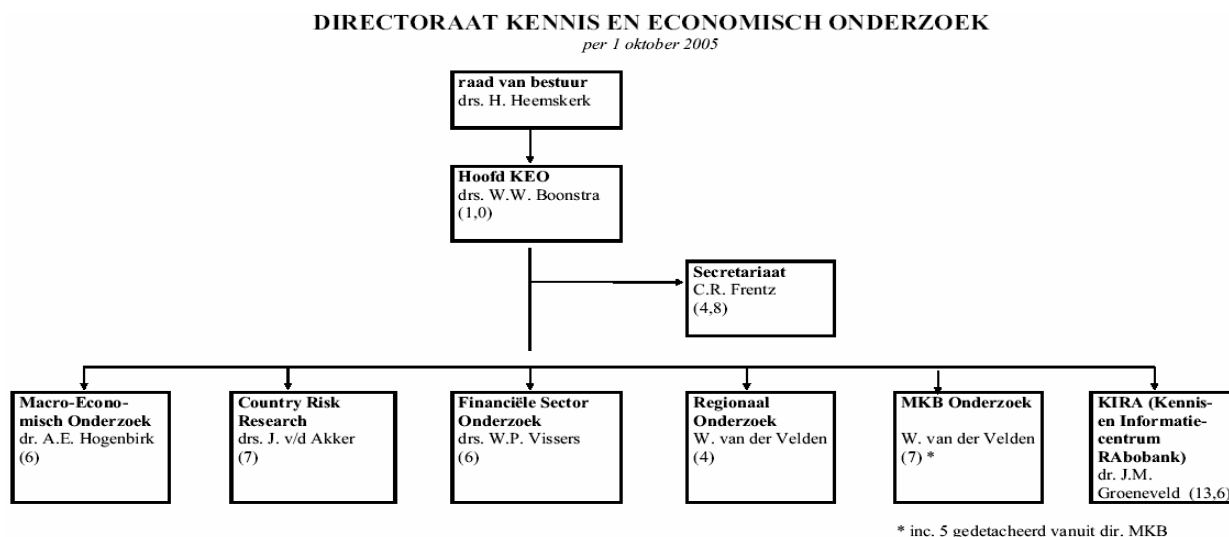
Leden zijn personen die, zoals het woord al aangeeft, lid zijn van de Rabobank. Dit geeft een stemrecht dat vergelijkbaar is als bij aandeelhouders van een bedrijf. Leden kunnen profiteren van kortingen en hebben de mogelijkheid tot het aanschaffen van ledencertificaten (soort obligaties).

Rabobank Nederland is het overkoepelende orgaan van Wholesalebankbedrijf en internationaal retailbankbedrijf, Marktondersteuning binnenlands retailbankbedrijf en Groepsfuncties. Onder groepsfuncties bij overige staven en diensten bevindt zich dan het directoraat Kennis en Economisch

Onderzoek. Nu duidelijk is geworden waar de stageopdracht heeft plaatsgevonden, zal nu kort even omschreven worden hoe de structuur van afdeling KEO er uit ziet.

### Structuur KEO

Deze is opgedeeld in 6 subcategorieën: Macro-Economisch Onderzoek, Country Risk Research, Financiële Sector Onderzoek, Regionaal Onderzoek, MKB Onderzoek en KIRA (zie ook figuur 14). Deze afstudeerstage was onderdeel van de sectie Regionaal Onderzoek.



Bron: Rabobank

Figuur 14.

De bezigheden van regionaal onderzoek zal nu in het kort samengevoegd en onderbouwd worden aan de hand van praktische voorbeelden.

Het regionaal-economisch onderzoeksteam is gespecialiseerd in onderzoek en advisering op regionaal sociaal-economisch gebied en op het gebied van thema's die een duidelijke ruimtelijke dimensie hebben. Het onderzoek wordt uitgevoerd vanuit een gedegen kennis van de bancaire omgeving en met een hoge mate van gespecialiseerde kennis en ervaring op regionaal en ruimtelijk sociaal-economisch gebied. Waar nodig werken ze samen met deskundigen op velerlei gebied, zowel 'intern' als 'extern'.

Het Directoraat adviseert bestuur en directie van de Rabobank in- en externen met diverse publicaties, zoals het economisch kwartaalbericht, de themaberichten en de jaarlijkse macro-economische verkenning 'Visie op...'

Enkele voorbeelden zijn:

#### Plan de Venen

Het gebied 'De Venen', onderdeel van het Groene Hart, gaat ingrijpend op de schop. Dit heeft gevolgen voor de aanwezige bedrijvigheid, en daarmee ook voor de betrokken Rabobanken. Op verzoek van de Rabobanken hebben wij de gevolgen van het plan in kaart gebracht en een analyse gemaakt van de bedreigingen en kansen die het plan heeft voor het gebied en de spelers in het gebied. Dit heeft zeer verhelderend gewerkt voor de banken zelf, de leden en klanten van de banken en de overige bij het gebied betrokken organisaties.

#### Kusttoerisme

De economie in het werkgebied van Rabobank Bergen-Egmond-Schoorl is sterk afhankelijk van recreatie en toerisme. Voldoende reden voor de lokale bank om hieraan extra aandacht te geven. In dit kader hebben wij onderzoek gedaan naar de ontwikkelingen en perspectieven van recreatie en toerisme voor het werkgebied, waarbij ook een raming is gemaakt van de indirecte werkgelegenheid die hiervan afhankelijk is. De gebleken grote afhankelijkheid van de detailhandel was een belangrijk argument om deze sector er meer bij te betrekken. Door het onderzoek en de aanverwante activiteiten heeft Rabobank Bergen-Egmond-Schoorl zich neergezet als natuurlijke partner voor ondernemers en overheid in Bergen, Egmond en Schoorl.

#### Nota Wonen

In de 'Nota Wonen' geeft de rijksoverheid nieuwe kaders voor het toekomstige woningbouwbeleid, een onderwerp dat duidelijk regionale gevolgen heeft. Wij hebben deze nota samengevat, van commentaar voorzien en tevens aangegeven wat de gevolgen van het vernieuwd beleid is voor de Rabobank.



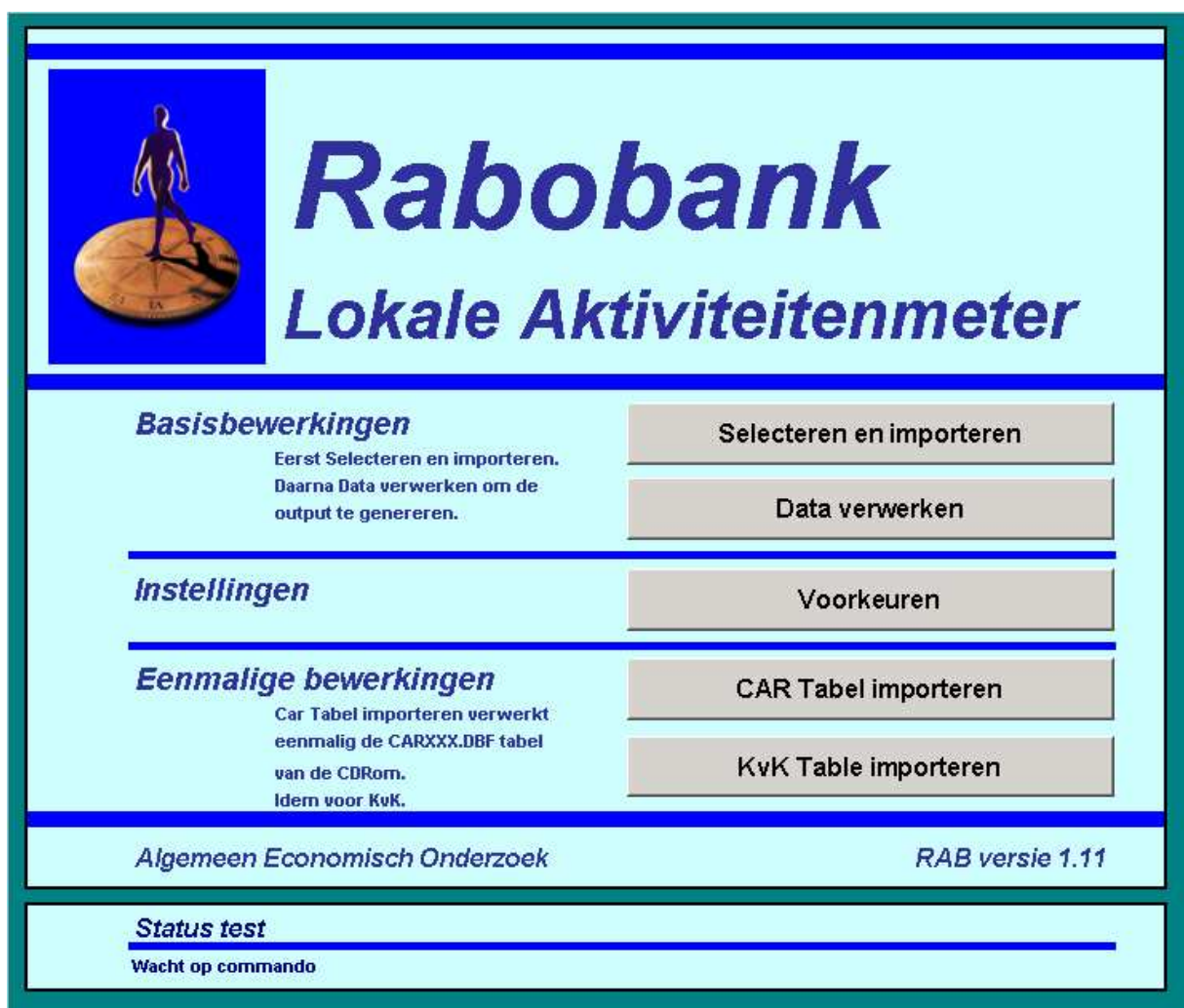
## 4.6 Beschrijving Lokale Activiteiten Monitor



### Oude LAM (Visual Basic)

Het programma is in opdracht van Rabobank Nederland geschreven door Ronald Buining in 1998. Er was gekozen voor Windows, dus de gebruikte programmatuur is in Excel. Excel heeft een programmeertaal op de 'achtergrond' genaamd Visual Basic. Over het algemeen gebruikt men dit voor het maken van Macro's.

Als je de LAM opstart, is het eerste wat je ziet het invoerscherm van de LAM waar je een aantal keuzes kan maken (zie figuur 15).



Bron: Rabobank

Figuur 15.

De selectieknoppen zijn 'Selecteren en importeren', 'Data verwerken', 'Voorkeuren', 'CAR Tabel importeren' en 'KvK Table importeren'. Een voor een zal ik de selectieknoppen behandelen.

Om te beginnen zal uitleg gegeven worden over de eenmalige bewerkingen (Tabellen importeren), omdat deze maar nauwelijks gebruikt worden. Alleen als er aan het eind van de maand weer een nieuwe cd met de CAR Tabellen geleverd wordt, dient men deze te importeren met behulp van de knop 'CAR Tabel importeren'. Eenmaal hier op gedrukt, verschijnt er een nieuwe window (zie figuur 16).



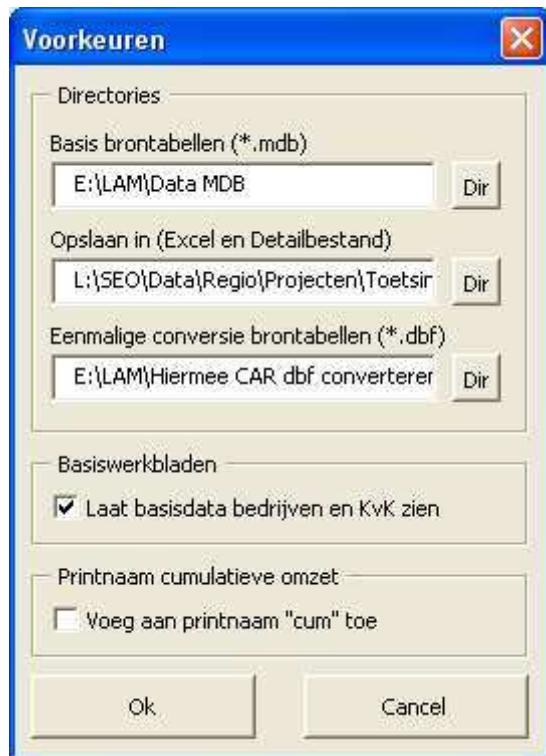
Bron:Rabobank

Figuur 16.

Hier voer je dan het nummer in van de CAR Tabel die je wilt importeren. Het moeten altijd 3 cijfers zijn, dus 73 voer je in als 073, 74 als 074, enz. Het laatst uitgebrachte nummer is 139. Zodra 140 uitkomt, open je de LAM, klikt op CAR Tabel importeren en voert in 140. Na het invoeren van een getal druk je op 'Ok' en het programma zal dan de \*.dbf bestanden omzetten naar \*.mdb bestanden. Dit is nodig zodat de LAM ermee kan werken.

De tweede selectieknop (KvK Table importeren) van de eenmalige bewerkingen werkt nagenoeg hetzelfde. Ditmaal is het enige verschil dat je niet een nieuwe window krijgt waarbij je een getal moet invoeren. Als je op deze knop geklikt hebt, zal het direct gaan beginnen met importeren. Je hebt namelijk opgegeven waar de file KvK.dbf staat die geconverteerd moet worden. Dit brengt ons tot het volgende, de selectieknop 'Voorkeuren'.

Nadat je hier op geklikt hebt, verschijnt er een nieuwe window (zie figuur 17).



Bron: Rabobank

Figuur 17.

Zoals je ziet is dit verdeeld in een aantal invoervakken en onderaan zijn er nog 2 selecties mogelijk. Het bovenste invoervak is bedoeld om aan te geven waar het programma de brontabellen (de CAR Tabellen in \*.mdb formaat) vandaan moet halen. Het tweede invoervak is om aan te geven waar het programma de uitvoer mag wegzetten. Het laatste invoervak is om aan te geven waar de bestanden zich bevinden die het programma moet converteren van \*.dbf naar \*.mdb. Uiteraard moet men dit opgeven alvorens over te gaan op het importeren van de CAR Tabellen. Gebeurt dit niet, dan verschijnt er een foutmelding dat de benodigde bestanden niet gevonden zijn.

Dan zijn er nog 2 selecties mogelijk. Je kunt er voor kiezen om de basisdata (dit is de data uit de CAR Tabellen) van de bedrijven en de KvK (Kamer van Koophandel) te laten zien als aparte tabbladen in de uitvoer. Als laatste heb je de keuze om standaard "cum" aan de printnaam toe te voegen als je te maken hebt met cumulatieve omzet. Als je hier verder klaar bent, druk je op 'Ok'.

Nu je alle benodigde tabellen en voorkeuren heb ingevuld, kun je verder met het selecteren en importeren van de data. Eenmaal klikken, geeft wederom een nieuwe window (zie figuur 18).

Bron: Rabobank

Figuur 18.

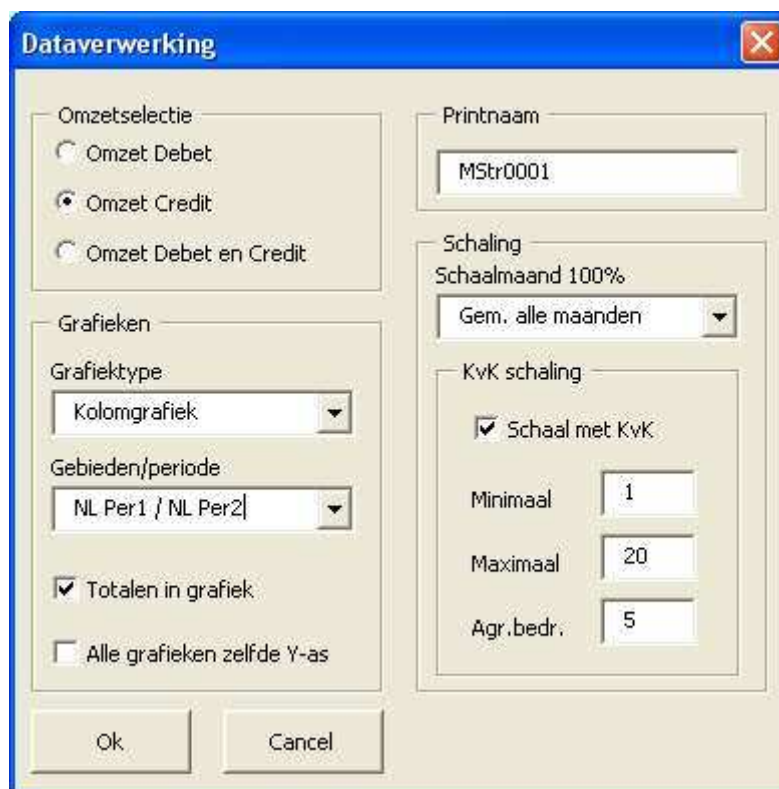
Hier zijn een aantal mogelijkheden. Linksboven zie je Periode 1 start en Periode 2 start staan. Daaronder kun je het aantal maanden invoeren en dan bestaat er de mogelijkheid om te selecteren voor een periode 2 of niet. Doe je dit niet, dan zal je ook geen periode 2 start op kunnen geven. Het invoervakje zal dan geblokkeerd worden. Eerst zal ik even uitleggen waarom je wel of niet met een periode 2 zal werken.

Wat je hier doet, is aangeven over welke periode je wilt dat het programma gaat kijken. Over het algemeen kiest men voor een periode van een jaar (dus bij invoer aantal maanden staat bijna altijd standaard 12 ingevuld). Als je kiest voor 1 periode, dan zal het programma het 'bedrijvenpanel' bepalen aan de hand van dat jaar. Mocht een bedrijf in de loop van het jaar geen uitgaven gedaan hebben (failliet of iets dergelijks), dan zal deze niet meegenomen worden in de data. Valt dan buiten beschouwing. Kies je wel voor 2 periodes, dan zal het programma het 'bedrijvenpanel' aan de hand van 2 achtereenvolgende jaren bepalen. Of er een groot verschil tussen deze methodes zal zitten, is nog niet bekend. Wel kan er gezegd worden, dat het kiezen voor 2 periodes de voorkeur heeft, omdat je dan een panel van bedrijven kiest die bestaat uit bedrijven met meer continuïteit. Ze bestaan immers al 2 jaar achter elkaar. Over het algemeen zullen dit de grotere bedrijven zijn die uiteindelijk ook meer invloed op de resultaten zullen hebben dan de kleinere. Naast de invoer van de periode en maanden, bevinden zich vier keuzes. Men kan ervoor kiezen om de individuele bedrijven tabel voor zowel Selecties lokaal als Nederland aan te vinken. Eigenlijk worden ze allebei niet echt meer gebruikt, dus worden deze verder buiten beschouwing gelaten. Hieronder heb je dan de keuze voor de

maandomzet of cumulatief vanaf de eerste maand. Standaard werk ik meestal met gewoon de maandomzet, omdat je hiermee meer kan vergelijken.

De volgende stap is dan om het gebied in te vullen waarnaar je onderzoek wil doen. Dit kan je selecteren door de postcodes van het betreffende gebied op te geven. Gescheiden door een komma en meerdere opeenvolgende postcodes mogen met behulp van een verbindingstreepje worden ingevoerd. Dan is er nog de keuze om bepaalde bedrijfscode uitzonderingen mee te geven. Als je bijvoorbeeld van een bedrijfstak geen data wil hebben. Dit bevordert namelijk de snelheid van het programma.

Als laatste invoer kan je opgeven onder welke naam je de uitvoer wil opslaan. Heb je dit allemaal gedaan, dan druk je op 'Ok' en beland je direct in de window die je ook krijgt als je op 'Data verwerken' klikt (zie figuur 19).



Bron: Rabobank

Figuur 19.

Hier zijn wederom weer een aantal keuzes en selecties te doen. Linksboven kan je de omzetselectie kiezen. Over het algemeen vind je meestal reeksen van uitgaven, dus is het verstandig om Omzet Credit als selectie te nemen. Op die manier heb je meer om mee te vergelijken. Dan heb je de keuze voor het soort grafiek dat je door het programma automatisch geplott wil hebben. Je kunt kiezen uit een aantal variaties. Vervolgens heb je de keuze om de gebieden en periode waarnaar je wilt kijken te selecteren. De keuzes zijn Lok.Per1 / NL Per1, Lok.Per 2 / NL Per2, Lok.Per1 / Lok.Per2 en NL Per1 / NL Per2. Geheel aan de gebruiker zelf om te kiezen wat juist is. Persoonlijk denk ik dat je het meeste hebt

aan de verhouding van het lokaal gekozen gebied in periode 1 en periode 2 (dus Lok.Per1 / Lok.Per2). Onder deze selectie kun je nog 2 keuzes maken. De keuze om de totalen in de grafiek weer te geven en de andere keuze is om alle grafieken dezelfde y-as te geven. Dit heeft zijn voor- en zijn nadelen, aan de gebruiker de keuze om hiervan gebruik te maken of niet.

Dan kan je rechtsboven nog een printnaam meegeven. Deze komt dan voor de omzet te staan. Vervolgens heb je de keuze om te kiezen voor een schaling. Mogelijkheden zijn om te kiezen voor een gemiddelde van alle maanden of voor een bepaalde maand afzonderlijk. Het is aan de gebruiker om te kiezen wat het beste lijkt voor de doeleinden die hij/zij voor ogen heeft met de uitvoerdata. Als laatste van dit scherm is er de mogelijkheid om te schalen met de KvK gegevens. Na uitvoerig getest te hebben, is er gekozen voor een minimum van 1 en een maximum van 20. De ophogingfactor bij agrarische bedrijven (Heeft KvK niet in het bestand) is vastgesteld op 5. Na dit alles drukt men op 'Ok' en het programma begint heftig te werken. Linksonder is te zien dat het nu al de nodige CAR bestanden invoegt.

Een CAR bestand heeft een aantal opgeslagen waarden die gebruikt worden voor berekeningen. De lay-out van een dergelijk bestand is te zien in de volgende tabel (zie tabel 1).

KNT_NR	PER_CD	CUST_NR	POST_CD_NR	BDRT_NR	ARC_DB	ARC_CR
1030	139	1030000042	5826	22	378044	378185
.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....

Enz. tot bijna 600000 records lang.

Tabel 20.

Een voor een zal ik de afkorting en de betekenis van de kolommen uitleggen.

- KNT\_NR : Staat voor Kantoor Nummer. Sommige bedrijven kunnen immers in hetzelfde 4-cijferige postcode gebied meerdere kantoren hebben.
- PER\_CD : Dit staat voor de periode code. Elke periode heeft zijn eigen code. Zo betekent een periode code van 139 de maand juli van 2005.
- CUST\_NR : Customer nummer oftewel klantnummer. Elk bedrijf heeft bij de Rabobank een eigen vaste klantnummer.
- POST\_CD\_NR : Dit is natuurlijk van groot belang voor dit onderzoek. Het staat namelijk voor postcode nummer. Hiermee kunnen we op regionaal gebied de nodige waarden opvragen.
- BDRT\_NR : Staat voor bedrijfstak nummer. Elke bedrijfstak heeft een nummer

---

toegewezen gekregen. De betreffende nummers met de bedrijfstaknamen zijn te vinden in de bijlage.

- ARC\_DB : Deze afkorting is van activiteit rekening courant debet. Dit zijn dus de inkomsten van het bedrijf in de desbetreffende periode.
- ARC\_CR : Dit is de afkorting voor activiteit rekening courant credit. Hier worden de uitgaven van de bedrijven in opgeslagen.

Aan de hand van de invoer zoals te zien is in figuur 18 en 19 zal nu de werking uitgelegd worden. Eerst kiest men de periode. Start periode 1 zal gezet worden we op 73 en start periode 2 op 85. Men wil bijvoorbeeld de omzet weten voor de komende 12 maanden, dus voor het jaar 2000. Het gekozen postcode gebied is 6211-6229. Al deze postcodes beslaat het gebied Maastricht. Nu gaat het programma alle CAR bestanden inlezen die hij nodig heeft. Dit is dus van CAR073 tot en met CAR096. Vervolgens onthoud het de ARC\_CR van elke regel waarbij PER\_CD overeenkomt met de opgegeven periode en POST\_CD\_NR met het opgegeven gebied. Als dan blijkt dat een bedrijf in een maand geen uitgaven heeft gedaan (inactief voor minimaal een maand of failliet), dan wordt deze buiten beschouwing gelaten. Op deze manier wordt er een bedrijvenpanel gevormd. Door deze methode wordt vertroebeling van de cijfers door acquisitie en vertrek van Rabobank klanten voorkomen. Na een uurtje ratelen is de computer klaar en is te zien dat er een extra tabblad toegevoegd zal zijn. Deze heet 'samenvatting' en zal naast het tabblad invoer te vinden zijn (zie tabel 21 voor een voorbeeld).

Gebied: MStr0001													
Periode 1: Jan 0 - Dec 0													
Periode 2: Jan 1 - Dec 1													
Omzet: Credit													
Schaling: met KvK gegevens													
Bedrijfstak (%)	Lok. 1	NL 1	Lok. 2	NL 2	Lok 2/1	NL 2/1	Lok/NL	Bedr. Lok	Bedr. NL	KvK. Lok	KvK. NL	Fact. Lok	Fact. NL
Akkerbouw	0	0	0	0	47	6	41	11	4063			5	5
Melkveehouderij	0	1	0	1		2	-2	0	17023				5
Intensieve veehouderij	0	1	0	1		10	-10	0	6423				5
Gecombineerd bedrijf	0	0	0	0	13	7	7	7	2960			5	5
Tuinbouw	0	1	0	1	45	3	41	3	8065			5	5
Visserij	0	0	0	0		-4	4	0	228				5
Loonbedrijf	0	0	0	0	7	2	5	1	1442			5	5
Overig Agr.Bedr.	0	0	0	0		-2	2	0	2137				5
Industrie	9	5	10	5	13	3	10	29	11955	356	70883	12	6
Bouwnijverheid	4	5	2	5	-39	2	-41	62	21414	540	111395	9	5
Groothandel	7	11	7	12	3	5	-2	26	13113	516	105984	20	8
Detailhandel Food	1	1	2	1	16	-1	16	30	6798	246	28141	8	4
Detailhandel Non-Food	5	3	5	3	2	2	0	85	18572	850	104187	10	6
Garages/autoreparatie	5	2	4	2	-10	2	-12	10	8220	178	34024	18	4
Horeca en recreatie	5	2	4	1	-10	-2	-8	53	14150	774	81268	15	6
Transport	1	2	1	3	-6	5	-12	12	6229	158	40426	13	6
Commerciële dienstverlening	54	52	46	55	-15	6	-20	155	43045	3697	672732	20	16
Medisch vrije beroepen	0	0	0	0	4	1	3	16	4103	84	10727	5	3
Non-profit dienstverlening	9	10	10	10	5	4	1	27	9867	824	105419	20	11
Overheid	0	1	0	1		5	-5	0	316		1729		5
Recreatieve verenigingen	1	2	0	2	-72	4	-76	10	5406	895	103073	20	19
Totaal	100	100	91	105	-9	5	-14	537	205529	9118	1469988		
Stand.dev.	11,64	11,29	10,06	11,94	28,42	3,25	24,65	37,48	9705,83	970,01	172702,14	6,20	3,98
Scheefheid	4,11	4,09	3,89	4,10	-0,82	-0,12	-1,04	2,43	2,17	2,92	3,27	0,22	2,35
Kurtosis	17,87	17,65	16,38	17,74	2,26	0,46	3,72	6,71	6,24	9,31	11,34	-1,65	5,08
Gebied: MStr0001													
Periode 1: Jan 0 - Dec 0													
Periode 2: Jan 1 - Dec 1													
Omzet: Credit													
Schaling: met KvK gegevens													
Bedrijfstak (%)	Lok. 1	NL 1	Lok. 2	NL 2	Lok 2/1	NL 2/1	Lok/NL	Bedr. Lok	Bedr. NL	KvK. Lok	KvK. NL		
Agrarisch	0	4	0	4	39	4	35	22	42341	0	0		
Industrie	9	5	10	5	13	3	10	29	11955	356	70883		
Bouwnijverheid	4	5	2	5	-39	2	-41	62	21414	540	111395		
Handel	17	18	17	18	0	4	-3	151	46703	1790	272336		
Horeca en recreatie	5	2	4	1	-10	-2	-8	53	14150	774	81268		
Transport	1	2	1	3	-6	5	-12	12	6229	158	40426		
Commerciële dienstverlening	54	52	46	55	-15	6	-20	155	43045	3697	672732		
Overige dienstverlening	11	13	10	13	-3	4	-7	53	19692	1803	220948		
Totaal	100	100	91	105	-9	5	-14	537	205529	9118	1469988		
Stand.dev.	17,60	16,99	15,11	17,98	22,49	2,24	22,25	55,69	15923,08	1244,20	220805,46		
Scheefheid	2,32	2,30	2,10	2,31	0,45	-1,58	0,45	1,05	0,35	1,34	1,92		
Kurtosis	5,73	5,54	4,77	5,63	1,63	3,36	1,45	-0,46	-1,91	1,60	3,91		

Bron:Rabobank

Tabel 21.



---

Linksboven op het tabblad zie je een korte samenvatting met de informatie over welk gebied het gaat, welke periode, of er gekeken wordt naar Debet en/of Credit en of er schaling met de KvK is gebruikt. Vervolgens geeft de eerste kolom de bedrijfstakken weer. Eerst uitgebreid naar 21 varianten en daaronder samengevoegd tot 8 verschillende (hoofd-)bedrijfstakken (Agrarisch, Industrie, Bouwnijverheid, Handel, Horeca en Transport, Commerciële Dienstverlening en als laatste de Overige Dienstverlening). CBS en andere instellingen hanteren ook ongeveer dezelfde indeling. Er kan altijd een verschillende interpretatie bestaan, maar dat is aan de gebruiker om dat goed in de gaten te houden (zie 4.8 in de bijlagen voor de interpretatie hoe de indeling van de bedrijfstakken is van het CBS t.o.v. de Rabobank).

De 2<sup>e</sup> kolom (Lok.1) geeft de verdeling van de bedrijfstakken weer in het lokaal gekozen gebied (Maastricht) voor de gekozen periode 1. De volgende 3 kolommen geven hetzelfde weer, maar dan voor NL 1, Lok.2 en NL 2. Dit staat natuurlijk voor Nederland in periode 1, Lokaal periode 2 en Nederland periode 2.

Kolom 6 (Lok2/1) geeft de verdeling van de bedrijfstakken weer van het lokale gebied in periode 2 tot verhouding van het lokale gebied in periode 1. De kolom erna (NL2/1) geeft hetzelfde weer alleen dan voor Nederland. De 8<sup>e</sup> kolom is me niet helemaal duidelijk, maar het is gelijk aan kolom 6 – kolom 7.

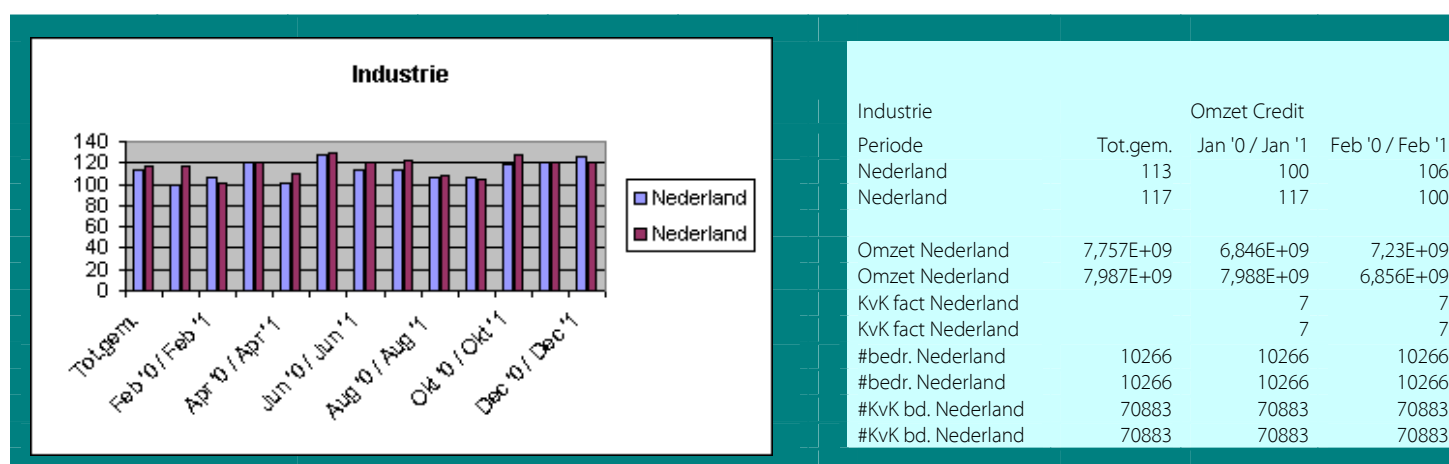
Dan volgen er 4 kolommen die het aantal bedrijven meegeeft. Kolom 9 (Bedr.Lok) geeft het aantal bedrijven per bedrijfstak weer dat in het lokaal gekozen gebied in de data van de Rabobank gevonden was. Kolom 10 (Bedr.NL) doet hetzelfde, maar dan voor Nederland. De 11<sup>e</sup> kolom (KvK.Lok) geeft het aantal bedrijven per bedrijfstak neer per lokaal gekozen gebied die bekend zijn bij de Kamer van Koophandel. De 12<sup>e</sup> kolom (KvK.NL) doet hetzelfde, alleen voor geheel Nederland.

De laatste 2 kolommen (Fact.Lok en Fact.NL) geven de ophogingfactor mee die berekend is aan de hand van de kolom 11 en 9 respectievelijk 12 en 10. Zoals al eerder vermeld, beschikt de KvK niet over gegevens van de agrarische sector. Er is gekozen voor een standaard ophogingfactor van 5. Voor de rest geldt een ophogingfactor met een minimum van 1 en een maximum van 20. Deze waarden geven een meest betrouwbare benadering van de werkelijkheid, zonder de schaling kan je heel vertekende beelden krijgen.

Onder Totaal staan nog 3 statistieken (Stand.dev., de Kurtosis en de Scheefheid). Eerlijk gezegd zeggen deze waarden niet zo veel. Je kunt nu alleen vergelijken of bepaalde gegevens uit eenzelfde verdeling bestaan. Zo is het zien dat Lok.1, NL1, Lok.2 en NL2 vrijwel eenzelfde verdeling vertonen. De waarden liggen dicht bij elkaar.

De kolom eronder zijn een aantal bedrijfstakken samengevoegd. Zo is Akkerbouw, Melkveehouderij, Intensieve veehouderij, Gecombineerd bedrijf, Tuinbouw, Visserij, Loonbedrijf en Overig Agr.Bedr. samengevoegd tot agrarisch. Groothandel, Detailhandel Food, Detailhandel Non-Food en Garages/autoreparatie zijn samengevoegd tot Handel. Medisch vrije beroepen, Non-profit dienstverlening, Overheid en Recreatieve verenigingen zijn samengevoegd tot Overige Dienstverlening. Het voordeel hiervan is dat het iets overzichtelijker is en de meeste reeksen die te vinden zijn, om mee te vergelijken, een soortgelijke classificatie hebben.

Dit tabblad bevat weliswaar wat nuttige informatie, maar het tabblad met de meeste info heet 'Graph LOK1\_NL1'. De naam hangt af van de keuze die je invoert zoals beschreven is bij figuur 19. Kies je daar in Gebieden/periode Lok.Per1 / NL Per1 dan krijg je als naam van het tabblad 'Graph LOK1\_NL1', kies je Lok.Per 2 / NL Per2 dan krijg je 'Graph LOK2\_NL2', enzovoort. Hieronder is een deel te zien van hoe de lay-out is (zie tabel 22).



Bron:Rabobank

Tabel 22.

Op het linkerdeel van het tabblad staan grafieken. De keuze van de vorm van de grafiek maak je al eerder in het programma (zoals te zien in figuur 19). Daar is gekozen voor een kolomgrafiek, met dit als resultaat. Andere mogelijkheden zijn een lijngrafiek, kolomgrafiek 100 %, staafgrafiek, staafgrafiek 100 % en een cirkeldiagram. Meest gebruikte vorm is de kolomgrafiek en de lijngrafiek. Beide geven snel een duidelijk beeld weer.

Rechts van deze figuur komt een tabel met gegevens te staan. Uiteraard loopt deze tabel door tot het aantal gekozen maanden (standaard 12) dat is opgegeven. Voor het gemak is deze even ingekort. Er zal linksboven in de hoek staan om welke bedrijfstak het gaat, hier is die industrie. Verder is er af te lezen om welke omzet het gaat (hier omzet Credit), en of het een geschaalde maandomzet betreft. Dan volgt daaronder de periodes en het gekozen gebied voor zowel periode 1 als periode 2. Hier is dat misschien een beetje onduidelijk, omdat hier voor beide periodes het gebied Nederland is gekozen. Zoals gebruikelijk volgt periode 1 voor periode 2, dus de bovenste Nederland staat voor Nederland in

periode 1. Dan staat er eerst het totaal gemiddelde indexcijfer gevolgd door de indexcijfers per maand. Hier is gekozen voor een schaling naar de maand januari 2000, dus deze staat op 100.

De 2 rijen eronder geven de maandomzetten voor het gekozen gebied en periode weer. Dan volgen er 2 rijen waarbij de KvK ophogingfactor staat. Deze waarde is al eerder berekend (zie tabel 21). De 4 laatste rijen zijn eveneens al eerder berekend (zie ook tabel 21) en geven het aantal bedrijven weer zoals opgenomen in de bestanden van de Rabobank respectievelijk de Kamer van Koophandel.

Dan volgen er een aantal tabbladen van iets mindere grote waarde, maar ze kunnen wel degelijk nuttig zijn. Het gaat om KVK\_NL, KVK\_LOK, CAR\_NL2, CAR\_LOK2, CAR\_NL1 en CAR\_LOK1. De eerste 2 genoemde van dit rijtje staan voor het aantal bedrijven per bedrijfstak bekend bij de KvK voor Nederland en het lokaal gekozen gebied. In de laatste 4 staan de berekeningen die gedaan zijn om te komen tot de ophogingfactor. De opbouw van deze bestanden zijn hetzelfde, dus zal maar een ervan toegelicht worden (zie tabel 23a en tabel 23b).

CAR_NL1		276				
PER_CD	BDRT_CD	AANTAL	DB_SUM	CR_SUM	DB2_SUM	CR2_SUM
73	0	3344	259723613	256564896	1,438E+15	1,589E+15
74	0	3344	256979339	241001185	1,208E+15	1,011E+15
75	0	3344	258702537	274022817	1,143E+15	1,315E+15
76	0	3344	224284664	233580584	6,214E+14	6,841E+14
77	0	3344	298610466	322008694	2,103E+15	2,183E+15
78	0	3344	261830262	250429849	8,669E+14	7,763E+14
79	0	3344	272446996	281613346	9,415E+14	9,855E+14
80	0	3344	252430838	251549639	1,125E+15	1,031E+15
81	0	3344	234929354	235606563	8,062E+14	7,888E+14
82	0	3344	272628853	283637430	8,626E+14	9,13E+14
83	0	3344	286265352	280633882	1,025E+15	9,689E+14
84	0	3344	345944916	350726318	2,321E+15	2,51E+15
73	11	4463	94500194	82985924	3,216E+13	1,466E+13
74	11	4463	79748188	82515484	1,618E+13	2,235E+13
75	11	4463	90993936	100449185	2,019E+13	2,194E+13
76	11	4463	97397312	90852049	1,353E+13	1,259E+13
77	11	4463	121947879	97908537	1,937E+13	1,904E+13
78	11	4463	114523305	100663776	2,164E+13	2,696E+13
79	11	4463	95069240	87894846	1,732E+13	1,7E+13
80	11	4463	78399898	79741513	1,061E+13	1,234E+13
81	11	4463	89081909	94951953	4,087E+13	4,606E+13
82	11	4463	98089043	102243925	1,477E+13	3,309E+13
83	11	4463	110812254	116343258	6,067E+13	3,505E+13
84	11	4463	164468920	168939191	8,096E+13	8,716E+13
.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....

Bron:Rabobank

Tabel 23a.

Wat je hier ziet is een klein deel van het blad CAR\_NL1 (dus de gegevens verwerkt uit CAR voor Nederland over periode 1). De waarde 276 staat voor het aantal records dat het bevat. Dan volgt op de volgende rij PER\_CD (Periode code). Hier loopt periode 1 van 73 tot en met 84 (dit is het jaar 2000). Dan staat in de volgende kolom BDRT\_CD (Bedrijfstak code, zie bijlage voor codering).

Vervolgens een kolom AANTAL die het aantal bedrijven weergeeft. DB\_SUM zijn de debet totalen opgeteld van de bedrijven in die periode en CR\_SUM uiteraard de credit totalen. De 2 kolommen daarnaast (DB2\_SUM en CR2\_SUM) staan voor de kwadraten van respectievelijk DB\_SUM en CR\_SUM

(deze kwadraten kunnen later gebruikt worden bij de berekening van de Stand.dev., de kurtosis en de scheefheid).

KvK_AANTAL	FACTOR	CORR_FACT	DB_SUM	CR_SUM	DB2_SUM	CR2_SUM
24972	7,4677033	7,4677033	1,94E+09	1,916E+09	8,02E+16	8,862E+16
24972	7,4677033	7,4677033	1,919E+09	1,8E+09	6,738E+16	5,639E+16
24972	7,4677033	7,4677033	1,932E+09	2,046E+09	6,374E+16	7,335E+16
24972	7,4677033	7,4677033	1,675E+09	1,744E+09	3,465E+16	3,815E+16
24972	7,4677033	7,4677033	2,23E+09	2,405E+09	1,173E+17	1,217E+17
24972	7,4677033	7,4677033	1,955E+09	1,87E+09	4,834E+16	4,329E+16
24972	7,4677033	7,4677033	2,035E+09	2,103E+09	5,251E+16	5,496E+16
24972	7,4677033	7,4677033	1,885E+09	1,878E+09	6,276E+16	5,751E+16
24972	7,4677033	7,4677033	1,754E+09	1,759E+09	4,496E+16	4,399E+16
24972	7,4677033	7,4677033	2,036E+09	2,118E+09	4,81E+16	5,092E+16
24972	7,4677033	7,4677033	2,138E+09	2,096E+09	5,715E+16	5,403E+16
24972	7,4677033	7,4677033	2,583E+09	2,619E+09	1,294E+17	1,4E+17
-1		5	472500971	414929619	8,04E+14	3,664E+14
-1		5	398740941	412577422	4,045E+14	5,587E+14
-1		5	454969681	502245926	5,046E+14	5,484E+14
-1		5	486986559	454260247	3,383E+14	3,148E+14
-1		5	609739396	489542687	4,842E+14	4,76E+14
-1		5	572616526	503318878	5,411E+14	6,739E+14
-1		5	475346198	439474228	4,33E+14	4,25E+14
-1		5	391999489	398707564	2,653E+14	3,085E+14
-1		5	445409546	474759766	1,022E+15	1,151E+15
-1		5	490445217	511219627	3,692E+14	8,274E+14
-1		5	554061269	581716290	1,517E+15	8,763E+14
-1		5	822344601	844695954	2,024E+15	2,179E+15
.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....

Bron:Rabobank

Tabel 23b.

KvK\_AANTAL staat voor het aantal bedrijven dat bekend is in de bestanden van de Kamer van Koophandel. Als je deze waarde deelt door het aantal bedrijven bekend in de data van de Rabobank, dan krijg je de ophogingfactor die in de kolom FACTOR komt te staan. Omdat de KvK geen gegevens bijhoudt op het agrarische gebied, zal er een -1 komen te staan op die plekken. Deze waarde heeft verder geen betekenis, maar is alleen bedoeld om het geheel wat overzichtelijker te maken. Mocht dit het geval zijn, dan is het niet mogelijk om een ophogingfactor te berekenen en om deze reden zal er niets in de kolom FACTOR komen te staan.

Zoals al eerder vermeld (zie ook figuur 19) zijn er beperkingen opgelegd aan de ophogingfactor. Zo kan deze niet kleiner zijn dan 1 en niet groter dan 20. Daarom is er een kolom toegevoegd (CORR\_FACT) die de gecorrigeerde factor aangeeft. Is het een waarde kleiner dan 1, dan geeft deze een 1 aan. Is het een waarde groter dan 20, dan geeft deze kolom een 20 aan. Zit de waarde ertussen, dan zal dat de factor zijn voor de verdere berekeningen. Een 5 is het in het geval het gaat om een bedrijfstak uit de agrarische sector. Dan volgen opnieuw de kolommen met de sommaties van de debet, credit etc. alleen nu na de ophoging. Deze opgehoogde waarden zullen uiteindelijk gebruikt worden in het tabblad 'Graph .....'.

Het tabblad 'tussenstap' heeft als functie, wat de naam al suggereert, een tussentap voor verdere berekeningen. De gegevens van het tabblad 'Graph....' kun je kopiëren en hierin plakken. Dan zal in het tabblad 'maanden en kwartalen' alles doorberekend worden (zie tabel 24). Deze kun je dan weer in grafieken uitzetten tegen bijvoorbeeld het CBS.

			Tot.gem.	Jan '0/Jan '1	Feb '0/Feb '1	Mrt '0/Mrt '1	Apr '0/Apr '1	Mei '0/Mei '1
Industrie	periode0	Omzet MStr0001	26982175,05	26076434,53	23912076,79	34189451,68	22198427,09	31401208,84
	periode1	Omzet MStr0001	27844083,13	29915421,75	25927466,27	25204355,15	22415073,8	31476793,8
	periode1/periode0		3,2%	14,7%	8,4%	-26,3%	1,0%	0,2%
	maand op maand				-13,3%	-2,8%	-11,1%	40,4%
	aantal bedrijven			35	35	35	35	35
Bron:Rabobank								Tabel 24.

Hier is een gedeelte te zien van de lay-out van het blad. Dit zijn de omzetten van beide periodes voor het gebied Maastricht. De 2<sup>e</sup> en 3<sup>e</sup> rij laten de omzetten per maand zien.

De rij eronder geeft de verhouding van de omzet van de periode een jaar geleden weer. Deze is uitgedrukt in percentages. Dit kan soms erg van pas komen, omdat sommige reeksen bij het CBS alleen in mutatie veranderingen gepubliceerd worden. Nu heb je de mogelijkheid om deze te vergelijken.

Vervolgens is er een rij opgenomen waar de mutaties van maand op maand worden berekend. Dit is eveneens uitgedrukt in percentages en kan gebruikt worden om dezelfde reden als hierboven beschreven.

De laatste rij geeft wederom het aantal bedrijven weer.

Zoals de naam van het tabblad al aangaf, staan hier ook optellingen van kwartalen. Deze kwartaaloptellingen worden berekend aan de hand van deze maanden en staan in een tabel rechts van tabel 24 (zie tabel 25).

			1e 3 maanden	2e 3 maanden	3e 3 maanden	4e 3 maanden
Industrie	periode0	Omzet MStr0001	84177963,01	82887235,49	77953051,12	78767850,99
	periode1	Omzet MStr0001	81047243,16	81142213,6	84358200,83	87581340
	periode1/periode0		-3,7%	-2,1%	8,2%	11,2%
	periode op periode			0,1%	4,0%	3,8%
aantal bedrijven			35	35	35	35
Bron:Rabobank						Tabel 25.

De indeling en betekenis zijn hetzelfde als bij tabel 23, alleen nu hebben we te maken met kwartalen in plaats van maanden.

Het laatste tabblad (variabelen, zie tabel 26) geldt enkel en alleen als controle. Hier ken men controleren of de juiste variabelen zijn gebruikt.

Variabelen blad niet veranderen, gebeurt automatisch via Invoer.					
naam	waarde	naam	waarde	previous	temp
Voor Data inlezen		Periode 1 start	73		61
Opslaan als naam	Maastricht0001	Periode 2 start	85		73
DataPath CAR	C:\LAM\Data\LAM\Basisdata.mdb	Aantal maanden	12		12
DataPath KvK	D:\Rabo\Data\KvKBAS2.MDB	Schaalmnd last	1		
DataPath DBF	D:\	Gedetail. Tbl Lok.	ONWAAR	ONWAAR	
DataPath Excel doel	C:\LAM\Output en werk folder	Postcode sel.	6211-6229	6211-6229	6211-6229
Laat basisdata sh.zien	WAAR	Bedrijfstakcd. Sel.			
Voeg "cum" toe	ONWAAR	Error in sel dialg	0		
		valid import			
		Gedetail. Tbl NL	ONWAAR	ONWAAR	
		Maandomzet/cum	0		0
		Use Period 2	WAAR	WAAR	
		Opslaan als naam	Maastricht0001	Maastricht9900	
naam	waarde	naam	waarde		
Voor Data bewerken		GrafiekBreedte	8		
Ophoog bedrt ongeldig	0	Grafiekhoopte	14		
bedrijfstakcodering		Grafiektype	2		
Aantal coderingen	22	Grafiektypenaam	Kolomgrafiek		
0 Niet bekend		GrafiekTotals	WAAR		
11 Akkerbouw		Gebied/periode nr	3		
12 Melkveehouderij		Printnaam	MStr0001		
13 Intensieve veehouderij		Schaalmaand	1		
14 Gecombineerd bedrijf		Schaal KvK	WAAR		
15 Tuinbouw					
16 Visserij		Omzetsel. Nr.	1		
17 Loonbedrijf		KvK Min Factor	1		
18 Overig Agr.Bedr.		KvK Max Factor	20		
21 Industrie		KvK Agr. Factor	5,00		
22 Bouwnijverheid					
23 Groothandel		Alle graf. Zelfde Yas	ONWAAR		
24 Detailhandel Food					
25 Detailhandel Non-Food					
26 Garages/autoreparatie					
27 Horeca en recreatie					
28 Transport					
29 Commerciële dienstverlening					
31 Medisch vrije beroepen					
32 Non-profit dienstverlening					
33 Overheid					
34 Recreatieve verenigingen					
98 Totaal					
Aantal coderingen 2e tbl	8	volgnummerbereik 2e tabel			
	Agrarisch	1	8		
	Industrie	9	9		
	Bouwnijverheid	10	10		
	Handel	11	14		
	Horeca en recreatie	15	15		
	Transport	16	16		
	Commerciële dienstverlening	17	17		
	Overige dienstverlening	18	21		

Bron: Rabobank

Tabel 26.





Bron: Rabobank

Figuur 21.

Standaard staat deze op de 'Strikt: Geen verh., bdrtwissel en  $Db > 0$  en  $Cr > 0$ ', want deze is gelijk aan de methode die de oude LAM als voorkeur gebruikte. Er zijn voor de bedrijven geen verhuizingen toegestaan, dit houdt in dat de bedrijven niet de ene maand bij de ene gemeente in de berekening worden meegenomen en vervolgens in de andere maand bij de andere. Het is ook niet toegestaan dat de bedrijven van bedrijfstak wisselen. Zou dit wel gebeuren, dan vertroebel je de data. Verder is de voorwaarde dat de bedrijven wel actief zijn, dus moeten er op zowel de Debet als de Credit activiteit gemeten worden. Dit is meegenomen, omdat je op deze manier de bedrijven eruit filtert die failliet zijn gegaan. Nadeel is natuurlijk dat je ook de bedrijven die een maand inactief zijn geweest (vanwege bijvoorbeeld persoonlijke zaken, vakanties, etc.) niet meeneemt in je data.

Om deze reden is er bedacht om een iets minder strikte regel toe te passen. Dit is de middelste keuze van de selecties bij 'Voorkeuren', 'Strikt: Geen verhuizing en bedrijfstakwisseling'. Je houdt nu met meer bedrijven rekening dus zouden de data nauwkeuriger zijn. Dit laatste moet natuurlijk nog onderzocht worden door correlaties te berekenen tussen deze data en andere bronnen. In hoofdstuk 3 is dit al ter sprake gekomen.

De derde en laatste mogelijkheid is: 'Ruim: Alle data'. Nu maak je gebruik van alle data die aanwezig is in elk CAR bestand. Er zullen geen bedrijven gefilterd worden.

Als je hier je keuze heb gemaakt, klik je daarna op 'Selecteren'. Dan opent er een nieuwe window zoals te zien is in figuur 22.



Bron: Rabobank

Figuur 22.

Dit komt overeen met de 'Selectie en importeren' uit de eerste LAM (zie figuur 18.). Linksboven heb je de mogelijkheid om de start en einde van de gewenste periode op te geven. Handmatig het CAR nummer invoeren of via de pijltjes ernaast. Bij het betreffende CAR nummer zal rechts van de pijltjes ook direct de datum in maand-jaar aangegeven staan. Dit is over het algemeen makkelijker werken, omdat je natuurlijk uit het hoofd niet altijd weet welk nummer voor welke periode staat. De mogelijkheid om direct een tweede periode op te geven om deze te vergelijken is niet meer aanwezig. Dit is onnodig veel extra rekentijd voor de computer terwijl deze vaak niet gebruikt werd. Je kunt altijd nog zelf een run draaien met het volgende jaar als periode en vervolgens deze twee reeksen op elkaar delen om toch tot de mutatie veranderingen van jaar op jaar te komen. Je zal zien dat dit veel sneller gaat.

Er is nog steeds de mogelijkheid om te kijken naar de Debet omzet, de credit omzet of de gecombineerde omzet. Rechts daarvan is de selectie voor wel of niet gebruik maken van de schaling ten opzichte van de KvK gegevens.

Wat nieuw is bij de nieuwe LAM, is de keuze voor verkorte bedrijfstakken of alle bedrijfstakken. Aangezien de reeksen van externe bronnen, waarmee je over het algemeen met grote regelmaat wil vergelijken, vaak alleen maar met de acht grote bedrijfstakken werkt, is het wel handig om deze selectie te kunnen maken. Dit scheelt wederom weer in de rekentijd van het programma.

---

Bij de selectie van bedrijven zijn ook wat vernieuwingen. Standaard staat deze op 'Gehele periode bestaan'. Als je dan een periode van 2 jaar zou opgeven, krijg je dezelfde resultaten als bij de oude LAM. Die werkte namelijk met een 2-jarig bedrijven panel. Nu is er ook de mogelijkheid om alleen de bedrijven mee te tellen die in december nog bestaan. Als derde en laatste keuze is het mogelijk alle bedrijven mee te laten tellen.

Als je in de invoer bij 'Postcode selectie' niks invoert, produceert de reeks op Nationaal niveau. Mocht je een postcode gebied berekenen, dan kan je gebruik maken van het Excel bestand waar alle postcodes en COROP's in staan. Copy paste dit vervolgens naar deze invoer en het resultaat zal betrekking hebben op de betreffende selecties die je heb opgegeven.

Als je nu op de knop 'Bereken' drukt, zal het programma de CAR bestanden inlezen en de uitvoer is na enkele tientallen seconden klaar. Die kan je vervolgens wegschrijven naar een Excel bestand waarvan de naam in het invoergebied 'Naam' kan opgeven. Het programma zet hem dan weg in de map Results. Als je het programma wil eindigen of je niet helemaal tevreden bent met de instelling, druk je op 'Cancel'. De overige knoppen 'Open' en 'Save' spreken voor zich.

In de main window (zie figuur 20.) zal tijdens het programma de berekeningen in beeld komen. Dit is meer om na te gaan of er niet rare waarden in voor komen, want eigenlijk werk je verder met de opgeslagen Excel bestand.

Ga naar de map Results waar de Excel bestanden zijn opgeslagen. Als je dan een van de bestanden opent, krijg je een Excel bestand met meerdere tabbladen te zien. De tabbladen hebben de namen 'Var100%', 'Var Total', 'Details1', en 'Details2' (de overige 3 zijn leeg en mag je vergeten).

In 'Var100%' staan de indices voor de gekozen bedrijfstakken en periode, waarbij de beginmaand gelijk staat aan 100 (zie tabel 27.).

Het volgende tabblad 'Var Total' geeft nagenoeg hetzelfde weer, alleen is nu de hele maand gelijkgesteld aan 100 (zie tabel 28.). Het totaal van de bedrijfstakken is in de eerste maand van de periode gelijk aan 100.

CarNr	Period	11	21	22	23	27	28	29	31
		Agrarisch	Industrie	Bouwnijverhe	Handel	Horeca en rec	Transport	Commerciële	Overige diens
49	jan-98	100	100	100	100	100	100	100	100
50	feb-98	85,8524523	92,1858953	89,434111	93,3159135	87,9454933	91,1512705	92,3343093	92,902544
51	mrt-98	93,7072552	110,12797	105,085987	108,922974	94,0727119	104,161283	106,797215	87,9513052
52	apr-98	111,000049	109,25548	105,068007	105,993743	97,4278165	105,832077	92,2916083	91,2420848
53	mei-98	107,698745	106,222103	107,577287	110,342013	105,45008	114,203414	112,495619	91,5604594
54	jun-98	105,031489	108,728704	117,24928	113,255636	114,387174	119,373667	96,886067	88,6335995
55	jul-98	101,380726	115,566152	136,194711	99,6788377	117,556229	129,810035	128,403871	101,751502
56	aug-98	82,3740643	89,859382	80,8629834	93,6935967	111,886199	105,832407	82,4621692	83,3772968
57	sep-98	97,0995814	100,082908	101,783437	101,520798	99,1844136	107,248692	108,36271	78,7522361
58	okt-98	100,321944	106,587503	114,779316	105,424067	92,2236907	109,342735	118,419919	94,1424479
59	nov-98	98,7107945	109,725501	119,840277	104,625599	102,225933	108,047812	96,6059848	83,0795206
60	dec-98	122,815604	127,575502	138,638757	115,473366	108,554014	119,618388	142,887257	98,0715148
61	jan-99	95,22168	97,8247342	92,8043966	100,950511	87,9800689	96,8267876	109,422205	116,0677
62	feb-99	100,780508	108,630179	98,9969258	105,384293	110,476813	104,297621	123,172748	116,252234
63	mrt-99	111,884823	123,068799	121,747795	125,390813	109,833684	121,027082	140,922066	110,300578
64	apr-99	106,740192	113,200959	111,269966	112,145838	101,830205	105,911661	111,874969	100,939271
65	mei-99	132,036505	124,559772	123,865518	127,498924	126,124852	122,861026	147,469893	118,972187
66	jun-99	124,601259	133,083963	138,803228	112,215344	128,821143	134,165183	127,37875	113,509431
67	jul-99	114,739571	126,630298	143,485433	122,616096	126,813064	127,506618	127,476957	72,7989473
68	aug-99	95,1053209	107,026254	96,0204402	108,655155	127,604391	111,201013	115,308155	61,0550742
69	sep-99	113,327183	116,175677	115,984203	112,889128	111,639157	113,979903	136,757907	97,0095787
70	okt-99	105,608986	114,998516	125,873329	114,596038	109,445648	115,469659	121,752053	117,675792
71	nov-99	121,950278	125,953842	143,43242	121,238906	110,174862	126,723021	118,403203	110,666091
72	dec-99	151,683144	138,93164	167,705266	116,969905	113,865663	129,585398	124,898427	127,377465

Bron: Rabobank

Tabel 27.

CarNr	Period	11	21	22	23	27	28	29	31
		Agrarisch	Industrie	Bouwnijverhe	Handel	Horeca en rec	Transport	Commerciële	Overige diens
49	jan-98	11,9402324	8,46727204	9,28017001	28,4994905	2,51050684	3,23450464	25,1705229	10,8973007
50	feb-98	10,2509823	7,80563054	8,29963754	26,5945599	2,20787763	2,94829207	23,2410285	10,1238695
51	mrt-98	11,188864	9,32483485	9,75215822	31,0424926	2,36170187	3,36910153	26,8814176	9,58431816
52	apr-98	13,2536638	9,25095869	9,75048969	30,2076767	2,445932	3,42314343	23,2302804	9,9429243
53	mei-98	12,8594804	8,99411445	9,98335516	31,4469116	2,64733148	3,69391473	28,3157356	9,97761854
54	jun-98	12,5410039	9,20635512	10,8809325	32,2772793	2,87169783	3,86114678	24,3867297	9,65866982
55	jul-98	12,1050943	9,7853005	12,6391008	28,4079609	2,95125718	4,19871159	32,3199259	11,0881671
56	aug-98	9,83565468	7,60863833	7,50422233	26,7021977	2,80891067	3,42315411	20,7561592	9,08587471
57	sep-98	11,5939156	8,47429206	9,44567603	28,9329103	2,49003149	3,4689639	27,2754607	8,58186794
58	okt-98	11,9786732	9,02505384	10,6517156	30,045322	2,31528207	3,53669582	29,8069128	10,2589856
59	nov-98	11,7862982	9,29075665	11,1213814	29,8177628	2,56638903	3,49481148	24,3162316	9,05342514
60	dec-98	14,6644684	10,8021648	12,8659124	32,9093211	2,72525594	3,86906231	35,9654698	10,6871478
61	jan-99	11,3696898	8,28308637	8,61240578	28,7703812	2,20874565	3,13186693	27,5421411	12,6482462
62	feb-99	12,0334269	9,19801276	9,18708302	30,0339866	2,77352796	3,37351139	31,0032246	12,6683554
63	mrt-99	13,3593079	10,42057	11,2984024	35,7357428	2,75738214	3,91462656	35,4708209	12,0197856
64	apr-99	12,745027	9,58503314	10,326042	31,9609924	2,55645427	3,42571758	28,1595146	10,9996559
65	mei-99	15,7654654	10,5468147	11,4949306	36,3365437	3,16637305	3,97394559	37,1189432	12,9647569
66	jun-99	14,8776798	11,2685812	12,8811756	31,9808012	3,2340636	4,33957907	32,0618975	12,369464
67	jul-99	13,7001714	10,7221318	13,3156921	34,9449626	3,18365065	4,12420748	32,0866167	7,93312016
68	aug-99	11,3557963	9,06220408	8,91086009	30,9661657	3,20351698	3,59680191	29,0236655	6,653355
69	sep-99	13,5315289	9,83691064	10,7635312	32,1728263	2,80270867	3,68668523	34,4226804	10,5714255
70	okt-99	12,6099583	9,7372372	11,6812589	32,659287	2,74764049	3,73487148	30,6456284	12,8234849
71	nov-99	14,5611465	10,6648545	13,3107725	34,5524705	2,76594745	4,09886199	29,8027053	12,0596167
72	dec-99	18,1113199	11,7637199	15,5633338	33,335827	2,85860527	4,19144569	31,4375871	13,8807053

Bron: Rabobank

Tabel 28.

De tabbladen die dan volgen, Details1 en Details2, bevatten de ruwe data. 'Details1' is geordend naar bedrijfstak en 'Details2' is geordend naar periode (Zie tabel 29 en 30).

Period	Bdrt	NrBedrSel	NrBedrDel	Db	Cr	Db2	Cr2	DbCr2	
49	jan-98	11 Agrarisch	52556	32223	1240157333	1280343791	3,05933E+14	3,60169E+14	1,20251E+15
50	feb-98	11 Agrarisch	52556	35238	2323434949	2379550335	1,00939E+15	9,89544E+14	3,84303E+15
51	mrt-98	11 Agrarisch	52556	22541	3579326307	3579325358	2,19093E+15	2,17077E+15	8,48255E+15
52	apr-98	11 Agrarisch	52556	36580	4862502224	5000507599	3,76224E+15	3,76542E+15	1,47563E+16
53	mei-98	11 Agrarisch	52556	37366	6184674105	6379421798	5,79157E+15	5,87762E+15	2,29598E+16
54	jun-98	11 Agrarisch	52556	37748	7634990296	7724185949	8,30269E+15	8,36091E+15	3,28698E+16
55	jul-98	11 Agrarisch	52556	37971	8975667359	9022207783	1,13172E+16	1,12216E+16	4,44685E+16
56	aug-98	11 Agrarisch	52556	38168	10070973334	10076879002	1,42215E+16	1,39702E+16	5,56926E+16
57	sep-98	11 Agrarisch	52556	38204	11267851343	11320087463	1,76997E+16	1,75155E+16	6,96221E+16
58	okt-98	11 Agrarisch	52556	38189	12537410918	12604553244	2,19628E+16	2,19148E+16	8,68109E+16
59	nov-98	11 Agrarisch	52556	38261	13789117445	13868390773	2,66037E+16	2,61009E+16	1,04366E+17
60	dec-98	11 Agrarisch	52556	38288	15495585622	15440852730	3,28445E+16	3,19233E+16	1,2832E+17
61	jan-99	11 Agrarisch	52556	34395	1176938437	1219164868	4,3354E+14	4,49748E+14	1,6898E+15
62	feb-99	11 Agrarisch	52556	35331	2528704335	2509501850	1,47641E+15	1,56311E+15	5,88579E+15
63	mrt-99	11 Agrarisch	52556	36149	3968713053	3942012238	3,23932E+15	3,23743E+15	1,27141E+16
64	apr-99	11 Agrarisch	52556	36650	5267147293	5308653663	5,39385E+15	5,24184E+15	2,09527E+16
65	mei-99	11 Agrarisch	52556	36650	6862670030	6999174852	8,36634E+15	8,18339E+15	3,27324E+16
66	jun-99	11 Agrarisch	52556	36349	8509019662	8594499333	1,21135E+16	1,18436E+16	4,75438E+16
67	jul-99	11 Agrarisch	52556	39207	10048806248	10063560310	1,60621E+16	1,58521E+16	6,34415E+16
68	aug-99	11 Agrarisch	52556	39024	11318420485	11281235382	1,99089E+16	1,95277E+16	7,8469E+16
69	sep-99	11 Agrarisch	52556	38734	12648944205	12732212929	2,47087E+16	2,42497E+16	9,75166E+16
70	okt-99	11 Agrarisch	52556	38427	14003752738	14084371023	2,95267E+16	2,91984E+16	1,17061E+17
71	nov-99	11 Agrarisch	52556	38254	15528911411	15645753834	3,59372E+16	3,57979E+16	1,431E+17
72	dec-99	11 Agrarisch	52556	37876	17557809411	17587819554	4,47191E+16	4,44309E+16	1,77881E+17
49	jan-98	21 Industrie	10361	7930	946654730,6	907940387,1	1,90782E+15	2,64645E+15	8,62825E+15
50	feb-98	21 Industrie	10361	9381	1815218692	1744933362	6,79309E+15	8,16129E+15	2,88768E+16
51	mrt-98	21 Industrie	10361	7002	2837113777	2744829684	1,71113E+16	2,06354E+16	7,27743E+16
52	apr-98	21 Industrie	10361	10611	383795398	3736804309	3,40394E+16	4,4451E+16	1,49023E+17
53	mei-98	21 Industrie	10361	11014	4819002264	4701237685	5,07152E+16	5,98845E+16	2,1332E+17
54	jun-98	21 Industrie	10361	11381	5848940033	5688429497	7,08871E+16	7,90698E+16	2,92271E+17
55	jul-98	21 Industrie	10361	11679	6922805384	6737701268	9,88545E+16	1,02433E+17	3,94279E+17
56	aug-98	21 Industrie	10361	11902	7747788576	753570889	1,21998E+17	1,23989E+17	4,82911E+17
57	sep-98	21 Industrie	10361	12149	8685335996	8462264028	1,50039E+17	1,51898E+17	5,9342E+17
58	okt-98	21 Industrie	10361	12421	9708840890	9430015016	1,87837E+17	1,84523E+17	7,32587E+17
59	nov-98	21 Industrie	10361	12611	10729274712	10426257153	2,27741E+17	2,22333E+17	8,86721E+17
60	dec-98	21 Industrie	10361	12844	11879514618	11584566661	2,81792E+17	2,72205E+17	1,09366E+18
61	jan-99	21 Industrie	10361	11085	925947219,9	888190270,7	2,08561E+15	2,04854E+15	8,04495E+15
62	feb-99	21 Industrie	10361	11819	1932265772	1874487537	8,61491E+15	9,06419E+15	3,50377E+16
63	mrt-99	21 Industrie	10361	12415	3037172320	2991878870	2,18489E+16	2,23727E+16	8,80183E+16
64	apr-99	21 Industrie	10361	13095	4062319691	4019676095	4,48079E+16	4,56689E+16	1,79536E+17
65	mei-99	21 Industrie	10361	13318	5187102169	5150604567	7,25017E+16	7,34142E+16	2,90392E+17
66	jun-99	21 Industrie	10361	13540	6326622285	6358927614	1,04E+17	1,08514E+17	4,18721E+17
67	jul-99	21 Industrie	10361	13813	7532807923	7508655230	1,462E+17	1,45172E+17	5,81763E+17
68	aug-99	21 Industrie	10361	13948	8512912759	8480389815	1,92816E+17	1,88819E+17	7,62373E+17
69	sep-99	21 Industrie	10361	14176	9555829692	9535195709	2,42401E+17	2,42674E+17	9,69189E+17
70	okt-99	21 Industrie	10361	14371	10595976498	10579313681	2,94083E+17	2,96019E+17	1,17928E+18
71	nov-99	21 Industrie	10361	14571	11703963730	11722899484	3,50689E+17	3,52839E+17	1,40621E+18
72	dec-99	21 Industrie	10361	14805	12960221527	12984315953	4,32371E+17	4,32331E+17	1,72853E+18
49	jan-98	22 Bouwnijverhei	18048	12871	1072722602	995106938,1	1,01432E+15	1,29738E+15	4,36662E+15

Bron: Rabobank

Tabel 29.

Dit is een gedeelte van het tabblad 'Details1'. De eerste kolom geeft de opgegeven periode weer (dit zijn de CAR nummers). De tweede kolom laat de datum zien in maanden voor de gekozen periode. De derde kolom geeft de bedrijfstak nummers weer met in de vierde kolom de bijbehorende bedrijfstak. Vervolgens worden het aantal geselecteerde bedrijven weergegeven en het aantal die er uit gehaald zijn. Dan komen we bij 2 belangrijke kolommen aan. De kolom met de Debet waarden en die met de Credit waarden. Let wel op dat dit cumulatieve waarden per jaar zijn, want zo worden ze in de CAR bestanden aangeleverd. De laatste zijn van minder belang (Db2, Cr2 en DbCr2). Dit zijn gekwadraterde waarden die je zou kunnen gebruiken bij de berekening van de standaard deviaties, kurtosis en

scheefheid. Officieel volgen er nog 9 kolommen die nodig waren als je gebruik maakte van een ophoging, maar dat is na nader onderzoek niet nodig gebleken.

Period	Bdrt	NrBedrSel	NrBedrDel	Db	Cr	Db2	Cr2	DbCr2	
49	jan-98	0 Onbekend	4879	19665	338112348,4	349446631	9,64055E+14	1,36416E+15	4,18109E+15
49	jan-98	11 Agrarisch	52556	32223	1240157333	1280343791	3,05933E+14	3,60169E+14	1,20251E+15
49	jan-98	21 Industrie	10361	7930	946654730,6	907940387,1	1,90782E+15	2,64645E+15	8,62825E+15
49	jan-98	22 Bouwnijverhei	18048	12871	1072722602	995106938,1	1,01432E+15	1,29738E+15	4,36662E+15
49	jan-98	23 Handel	43580	32856	3130648058	3055982891	9,25119E+15	1,43103E+16	4,47389E+16
49	jan-98	27 Horeca en rec	12718	11522	258804876,9	269200108,9	3,36764E+14	1,38743E+15	3,06387E+15
49	jan-98	28 Transport	5452	4159	371311584	346833948,1	1,77125E+14	1,74148E+14	6,83877E+14
49	jan-98	29 Commerciele	36187	28106	2729567646	2699019734	4,87322E+16	4,71843E+16	1,90847E+17
49	jan-98	31 Overige diens	18749	17635	1075016978	1168510865	5,85569E+15	1,07386E+16	3,03325E+16
50	feb-98	0 Onbekend	4879	23647	633911166,4	662138188,2	3,07326E+15	3,81897E+15	1,30915E+16
50	feb-98	11 Agrarisch	52556	35238	2323434949	2379550335	1,00939E+15	9,89544E+14	3,84303E+15
50	feb-98	21 Industrie	10361	9381	1815218692	1744933362	6,79309E+15	8,16129E+15	2,88768E+16
50	feb-98	22 Bouwnijverhei	18048	15102	1957888799	1885071981	2,83581E+15	2,83756E+15	1,11147E+16
50	feb-98	23 Handel	43580	37545	6031020645	5907701240	2,7682E+16	4,32526E+16	1,34266E+17
50	feb-98	27 Horeca en rec	12718	13557	481644974,2	505949472,7	9,93031E+14	3,28272E+15	7,82765E+15
50	feb-98	28 Transport	5452	4791	694168686,2	662977498,2	6,34724E+14	6,27668E+14	2,49333E+15
50	feb-98	29 Commerciele	36187	37699	5164914365	5191140963	1,92496E+17	1,89974E+17	7,60974E+17
50	feb-98	31 Overige diens	18749	24528	2121871823	2254087185	2,57237E+16	3,23899E+16	1,11228E+17
51	mrt-98	0 Onbekend	4879	20768	994896714,9	1010524421	7,75145E+15	7,96716E+15	3,06736E+16
51	mrt-98	11 Agrarisch	52556	22541	3579326307	3579325358	2,19093E+15	2,17077E+15	8,48255E+15
51	mrt-98	21 Industrie	10361	7002	2837113777	2744829684	1,71113E+16	2,06354E+16	7,27743E+16
51	mrt-98	22 Bouwnijverhei	18048	10603	3013149587	2930789925	6,22804E+15	6,10004E+15	2,43238E+16
51	mrt-98	23 Handel	43580	26898	9356992810	9236368677	6,71324E+16	9,94877E+16	3,17069E+17
51	mrt-98	27 Horeca en rec	12718	10142	822861246,5	759193315,6	1,70144E+16	5,93491E+15	4,29373E+16
51	mrt-98	28 Transport	5452	3262	1057238616	1024244189	1,43981E+15	1,46266E+15	5,75349E+15
51	mrt-98	29 Commerciele	36187	30866	8044678496	8073618880	4,61574E+17	4,59505E+17	1,83761E+18
51	mrt-98	31 Overige diens	18749	21522	3155367411	3281807743	5,1252E+16	5,63017E+16	2,09544E+17
52	apr-98	0 Onbekend	4879	27561	1328784475	1346786054	1,40539E+16	1,39818E+16	5,50995E+16
52	apr-98	11 Agrarisch	52556	36580	4862502224	5000507599	3,76224E+15	3,76542E+15	1,47563E+16
52	apr-98	21 Industrie	10361	10611	3837795398	3736804309	3,40394E+16	4,4451E+16	1,49023E+17
52	apr-98	22 Bouwnijverhei	18048	17030	4055593464	3976328954	1,12635E+16	1,10063E+16	4,40721E+16
52	apr-98	23 Handel	43580	41427	12603310263	12475519324	1,20226E+17	1,79003E+17	5,68725E+17
52	apr-98	27 Horeca en rec	12718	15356	1083973348	1021469104	2,13977E+16	9,52413E+15	5,93476E+16
52	apr-98	28 Transport	5452	5318	1426378443	1391305759	2,79735E+15	3,01155E+15	1,14752E+16
52	apr-98	29 Commerciele	36187	46680	10545884381	10564587603	3,67557E+17	3,47391E+17	1,42365E+18
52	apr-98	31 Overige diens	18749	30782	4191990802	4347981417	8,39824E+16	9,95682E+16	3,58585E+17
53	mei-98	0 Onbekend	4879	29236	1684915013	1694907199	2,279E+16	2,28515E+16	9,03073E+16
53	mei-98	11 Agrarisch	52556	37366	6184674105	6379421798	5,79157E+15	5,87762E+15	2,29598E+16
53	mei-98	21 Industrie	10361	11014	4819002264	4701237685	5,07152E+16	5,98845E+16	2,1332E+17
53	mei-98	22 Bouwnijverhei	18048	17689	5112300533	5046838004	1,79888E+16	1,7474E+16	7,03755E+16
53	mei-98	23 Handel	43580	42702	15883271380	15847552369	1,83135E+17	2,842E+17	8,83404E+17
53	mei-98	27 Horeca en rec	12718	16014	1352585219	1305340834	2,642E+16	1,37915E+16	7,82513E+16
53	mei-98	28 Transport	5452	5454	1815875884	1787401970	5,27822E+15	5,64109E+15	2,15807E+16
53	mei-98	29 Commerciele	36187	49846	13581579313	13600866564	7,38787E+17	7,37133E+17	2,94436E+18
53	mei-98	31 Overige diens	18749	32829	5264709628	5417875333	1,27689E+17	1,38315E+17	5,21955E+17
54	jun-98	0 Onbekend	4879	30007	2051284805	2066623900	3,44481E+16	3,47858E+16	1,37468E+17
54	jun-98	11 Agrarisch	52556	37748	7634990296	7724185949	8,30269E+15	8,36091E+15	3,28698E+16
54	jun-98	21 Industrie	10361	11381	5848940033	5688429497	7,08871E+16	7,90698E+16	2,92271E+17
54	jun-98	22 Bouwnijverhei	18048	18276	6262180483	6213593727	2,65167E+16	2,58928E+16	1,04177E+17

Bron:Rabobank

Tabel 30.

Op de volgorde in de rijen na, is deze identiek aan tabel 29. De uitleg bij tabel 29 geldt daarom ook voor tabel 30.

## 4.7 Uitleg bij de gebruikte reeksen



Alle reeksen van de Rabobank die hier zijn gebruikt en uitgeschreven, hebben betrekking op de maandelijkse credit transacties van de zakelijke klanten van de Rabobank over heel Nederland in de periode januari 1998 tot en met augustus 2005. De reeks is geconstrueerd aan de hand van de LAM met nog de oude bestanden voor de crash (gebeurde tijdens het switchen van het Windows bestuursysteem). Deze is aangevuld met uitdraaien van de LAM van latere jaren. De Rabobank heeft natuurlijk niet de beschikking over de gegevens van alle bedrijven, vandaar dat er gebruik is gemaakt van een ophogingfactor. Deze ophogingfactor wordt berekend aan de hand van de informatie waar de Kamer van Koophandel (KvK) over beschikt. Van de bedrijfstak agrarisch hebben ze geen gegevens, dus is besloten om standaard een ophogingfactor van 5 toe te passen. Uit uitvoerig testen bleek dit een redelijke waarde te zijn. De overige bedrijfstakken werken met een ophogingfactor tussen de 1 en 20, die berekend wordt door het aantal bedrijven wat aanwezig is in de bestanden van de KvK te delen door het aantal bedrijven dat de Rabobank in zijn bestand heeft.

Dit document is om wat meer duidelijkheid te geven in de gevonden resultaten. Er zal ingegaan worden op de verschillen in reeksen tussen die van de Rabobank en het CBS. Tevens zal er geprobeerd worden om meer duidelijkheid te verschaffen betreffende uitschieters en andere vreemde verschijnselen.

Elke gebruikte reeks wordt nu volgens alfabetische volgorde behandeld:

- Agrarisch

Rabo: De gegevens die de Rabobank hanteert als agrarisch, zijn alle transacties van bedrijven die onderdeel zijn van Akkerbouw, Melkveehouderij, Intensieve veehouderij, Gecombineerd bedrijf, Tuinbouw, Visserij, Loonbedrijf of Overig Agr.Bedr.

CBS: Het CBS heeft weinig gegevens op agrarisch gebied. De indexcijfers die ik nu gebruik, zijn prijsindexcijfers uit het bedrijfsleven landbouw en visserij en die worden verzameld door de LEI. Dit steekproefkader is gedefinieerd als land- en tuinbouwbedrijven met een omvang tussen 16 en 1200 Europese grootte-eenheden (ege). Bij visserij bestaat de actieve kottervloot bestaat uit ongeveer 400 vaartuigen, waarvan er ruim 100 deel uitmaken van het LEI-panel.

Conclusies: De correlatie is redelijk te noemen, maar lager dan bij andere bedrijfstakken. Het is een aanzienlijke verbetering als je correlaties neemt van het 3-maanden-voortschrijdend gemiddelde. Dit kan haast niet anders zijn dan dat het anders wordt gemeten. Kleine veranderingen in de definities kunnen leiden tot flinke afwijkingen in de reeksen.

Als je de grafiek van de index omzet bekijkt, zie je wel dat beide reeksen een bepaalde conjuncturele cyclus vertonen. De bewegingen van de LAM reeks zijn grilliger. CBS reeks is behoorlijk constant, dit komt waarschijnlijk omdat de indices al een soort gemiddelde zijn. Bij de grafiek van de 3-mnd-vg zie je goed dat er bij de LAM reeks sprake is van een seizoensschommeling. De laatste grafiek toont de procentuele verandering per reeks van maand op maand. Opvallend is dat de veranderingen bij de Rabobank heftiger zijn dan die van de CBS gegevens. De oorzaak hiervan is waarschijnlijk dat de reeks van het CBS al geneutraliseerd is tijdens de gegevensverzameling.

- Bouw

Rabo: Alle bedrijven die bij de Rabobank bekend staan als een bouwbedrijf, of iets gerelateerd daaraan, komen in de berekening bij de bedrijfstak Bouw. Het zijn voornamelijk midden- en kleinbedrijven.

CBS: Deze reeks is opgebouwd uit 2 reeksen. Vanaf 2000 hebben ze namelijk iets anders gemeten. Dan maken ze geen onderscheid meer tussen bedrijven met 10 of meer werknemers en bedrijven met 100 of meer werknemers. Om een lange reeks te vormen, zijn beide geschakeld en is er gelet op deze indeling. Verder waren ze identiek. Bij beide gaat het om de omzet maandindices van de bedrijfstak '45 bouwnijverheid'.

Conclusies: De correlatie is niet verkeerd. Aan de grafiek is ook te zien dat ze een zelfde trend volgen. Pieken in jun/jul en dec en dalen in jan en aug. Dit is te verklaren. De bouwvakvakantie valt altijd eind juli, begin augustus. Januari heeft meestal te kampen met vorstverlet, dus kan men dan ook niet bouwen. Dit heeft tot gevolg dat er geen omzet wordt gemaakt. De pieken zijn te wijten aan de hoge omzet vanwege het tijdig willen afronden voor de vakantie c.q. vorstverlet.

Opvallend is dat de index van de Rabobank tijdens de jaren 98 en 99 hoger liggen dan die van het CBS. Vervolgens volgen er 2 jaren waarbij ze nagenoeg gelijk zijn en de laatste 3 jaar zijn de Rabobank indices wederom hoger. Het is ook opvallend dat de pieken van de Rabobank heviger zijn dan die van het CBS en juist de dalen van het CBS heviger zijn dan die van de Rabobank.

- Bouw - Conjunctuurgegevens - Bouwnijverheid Index Omzet Geschakeld

Deze toont alleen de reeksen en hoe deze geschakeld is.

- Bouw (ZONDER ophoging)

De interpretatie hiervan is gelijk aan de eerdere reeks Bouw. Het enige verschil is dat deze maal er niet gebruik gemaakt is van een ophoging van de Rabobank reeks aan de KvK gegevens. De

correlatie op maandniveau is lichtjes gedaald, maar op jaarniveau is deze gestegen. De waarden liggen dicht bij elkaar, je kunt je dus afvragen of het wel nodig is om op te hogen.

- Bouw Kwartaal

Rabo: Idem als bij Bouw

CBS: Idem als bij Bouw

Conclusies: Bij gebruik van kwartaalindices zijn in principe de seizoensinvloeden deels niet meer zichtbaar. Dat blijkt ook wel uit de hogere correlatie en de grafiek. Je ziet nu duidelijker dat ze een gelijke cyclus vertonen en trend volgen. Verder treedt nog hetzelfde op als bij de grafiek in het werkblad Bouw, waarbij de ene reeks hoger ligt dan de andere enzovoort.

- CBS jaarreeksen 98-05

Rabo: De jaarreeksen zijn de gemiddeldes van de maandindices per bedrijfstak. Mocht de maandindices niet aanwezig zijn, dan is de jaarreeks berekend aan de hand van de kwartaalindices.

CBS: Sommige bedrijfstakken hebben niet de beschikking over maandindices, maar zijn alleen op jaarbasis te verkrijgen. In deze werkmap staat even een handig en kort overzicht met de correlaties tussen Rabo en CBS gegevens. Op het gebied van dienstverlening is er jammer genoeg geen juiste informatie te vinden op Statline.

Conclusies: Wat als eerste opvalt, is dat de correlatie van industrie als enige vrij laag te noemen is. Oorzaak hiervan is mogelijk dat ze in de industrie met meerdere banken werken en dat wij erg klein zijn in dat specifieke wereldje. De grote bedrijven doen over het algemeen meer zaken met andere banken dan met ons.

De grafieken van de index zijn niet al te duidelijk. Wel is goed te zien dat industrie als enige een behoorlijk afwijkend patroon vertoont dan de andere. De rest volgen nog een gezamenlijke trend.

Bij de j-o-j groei grafieken zijn wel nog interessante dingen waar te nemen. Hier zie je bij industrie en bouw erg afwijkend gedrag. De overige grafieken zijn niet zo extreem afwijkend als deze.

- Credit Omzet NL tm augustus 2005

In dit Excel bestand zijn per bedrijfstak de credit omzetten, maand- kwartaal- en jaarindices uitgerekend. Deze gegevens komen voort uit de LAM waarbij gewerkt is met een 2-jarig panel.



- 
- Credit Omzet NL tm augustus 2005 ZONDER ophoging  
Dit Excel bestand is identiek aan de reeks hierboven, alleen nu zonder ophoging van de data tegen de KvK gegevens.
  
  - Debet Omzet NL tm augustus 2005  
In dit Excel bestand zijn per bedrijfstak de debet omzetten, maand- kwartaal- en jaarindices uitgerekend. Deze gegevens komen voort uit de LAM waarbij gewerkt is met een 2-jarig panel.
  
  - Debet Omzet NL tm augustus 2005 ZONDER ophoging  
Dit Excel bestand is identiek aan de reeks hierboven, alleen nu zonder ophoging van de data tegen de KvK gegevens.
  
  - Een lange reeks construeren  
Dit gebruik je om van de jaarlijkse reeksen een lange reeks te maken die van 1998 tot en met 2005 gaat.
  
  - EIM jaarreeksen 98-05  
Rabo: Idem als bij CBS jaarreeksen 98-05  
EIM: Zelfde als bij het CBS alleen nu van het EIM.

Conclusies: Net als bij de CBS reeksen, geldt hier ook hetzelfde als bij de CBS jaarreeksen. Industrie is veel lager dan de andere. Dit komt ook duidelijk in de index grafieken naar voren. Omdat je bij de database van EIM in tegenstelling tot CBS wel onderscheid kan maken in grootte van bedrijven, is er nog een tabblad aangemaakt waarbij de correlatie per omvang bedrijf berekend is. Hier is te zien dat de gegevens van de Rabobank beter correleren met de kleinere industrie bedrijven, maar toch is de correlatie vergeleken met andere bedrijfstakken nog laag te noemen. Dit kan meerdere oorzaken hebben. Een ervan is mogelijk dat industrie bedrijven vaak met meerdere banken werken. Transactie worden dan niet bij een en dezelfde bank volbracht waardoor in wezen gegevens verloren gaan.

Als laatste heb is er nog een aantal j-o-j groei grafieken geconstrueerd. Hier vallen ook een aantal dingen op. Natuurlijk weer de afwijkende lijnen bij bouw en industrie (net als bij CBS), maar dit zijn dit maal niet de enige. Handel, transport en de dienstverlening vertonen ook afwijkend gedrag (hoewel dit niet naar voren kom in de correlatie, want die zijn allen behoorlijk hoog).

- Handel

Rabo: De maandelijkse credit omzet van handel bestaat uit omzet in de groothandel, detailhandel food, detailhandel non-food en garages/autoreparatie.

CBS: De definitie voor handel volgens het CBS is '52 Detailhandel (incl. reparatie)'. Dit betreft de totale detailhandel, exclusief SBI 5231 Apotheken. Onder detailhandel wordt verstaan het verhandelen van goederen, niet in de eigen onderneming vervaardigd, aan gezinnen resp. aan particulieren. De omzet van detailhandel in en reparatie van consumentenartikelen omvat; de door winkels (food en non-food) behaalde omzet, de omzet van de postorderbedrijven en de ambulante handel en de omzet van de reparatiebedrijven.

Conclusies: Redelijk goede correlatie. Correlatie van 3 maanden voortschrijdend gemiddelden nog even beter, want dit neutraliseert de seizoensinvloeden. Wat direct opvalt in de grafiek van indices, is dat de lijn van de Rabobank een andere stijgende trend volgt dan het CBS. Dit is nog duidelijker zichtbaar in de grafiek van de 3 maanden voortschrijdend gemiddelden.

De meeste pieken zijn in mei en dec. Dit is te verklaren, omdat in mei iedereen extra te besteden heeft vanwege vakantiegeld wat bovenop het loon komt. In december hebben de meeste mensen ook meer te besteden vanwege een 13<sup>e</sup> maand en geeft men meer uit voor cadeaus (sinterklaas en kerst). Het merendeel van de dalen zijn in feb en aug. De verklaring voor februari kan zijn dat men vanwege de uitgaven in december even de hand op de knip moet houden. Het kan ook te maken hebben met de wintersport of misschien het op de markt brengen van de nieuwe collectie zomerkleding (wat meestal vanaf april te koop is). Voor augustus geldt ongeveer hetzelfde. Veel mensen zijn dan op zomervakantie of sparen hun geld op om de nieuwe wintercollectie kleding aan te schaffen.

Het grote verschil in index omzet vanaf 2004 is mogelijk te wijten aan de enorme exportstijging van Nederland (zie bestand wederuitvoer.xls). In de definitie van de Rabobank is groothandel opgenomen in de handel. Bij het CBS is dit niet het geval. Dit kan de reden zijn waarom de omzet index van Rabobank aanzienlijk hoger ligt dan het CBS vanaf januari 2004. Het kan ook zijn dat het marktaandeel van de Rabobank in zijn geheel is gestegen.

- Handel - Conjunctuurgegevens - Detailhandel Maandstatistiek Geschakeld

Deze toont de reeksen van Statline CBS. Een heeft het jaar 1995 als index 100 en de ander heeft jaar 2000 als index 100. Hierin staan beide uitgezet en zijn ze geschakeld tot een lange reeks.

- Handel (ZONDER ophoging)

De interpretatie hiervan is gelijk aan de eerdere reeks Handel. Het enige verschil is dat deze maal er niet gebruik gemaakt is van een ophoging van de Rabobank reeks aan de KvK gegevens. Op maandniveau is de correlatie licht gedaald, maar op jaarniveau licht gestegen. Wederom kan je concluderen dat het misschien onnodig is om op te hogen.

- Horeca

Rabo: Maandelijkse credit mutaties van horeca en recreatie.

CBS: De indices bevatten alle opbrengsten uit verkopen van goederen en/of diensten (SBI 55). Ontvangsten uit hoofde van nevenactiviteiten zijn bij deze omzet inbegrepen. Omzet inclusief BTW en logiesbelasting (voor zover van toepassing). 5510b Hotels, eetgelegenheden, cafés (Het totaal van hotels, pensions, conferentieoordens(551), restaurants, cafetaria's, snackbars(553) en cafés(554)), 55301 Restaurants (Verstrekken van maaltijden voor directe consumptie ter plekke met al dan niet daarnaast afzonderlijk verstrekken van dranken of kleine eetwaren voor directe consumptie ter plekke. Restaurants in combinatie met afhaalrestaurant) en 55302 Cafetaria's, lunchrooms, snackbars (Verstrekken van al dan niet zelf bereide kleine eetwaren voor directe consumptie ter plekke met daarnaast afzonderlijk verstrekken van dranken voor directe consumptie ter plekke. Afhaalrestaurants, fastfood-restaurants, lunchrooms, snackbars, eetkramen, exploitatie van automatieken voor snacks en andere kleine eetwaren.

Conclusies: Grafiek van index omzet oogt niet echt samenhangend, maar de correlatie is toch behoorlijk. De reeks van het CBS toont een duidelijke cyclus (wat ook erg goed is te zien als je de grafiek van het 3 maanden voortschrijdend gemiddelden bekijkt) terwijl de reeks van de Rabobank dat minder heeft. De extremen zijn bij de gegevens van de Rabobank groter dan bij het CBS. Hoe is dit te verklaren?

Bij het CBS zijn in de zomermaanden duidelijk pieken te traceren en in de wintermaanden dalen. Bij de gegevens van de Rabobank komt dit minder duidelijk naar voren. Waarschijnlijk zit dit hem in een andere definitie benadering of is het marktaandeel van de Rabobank te klein in deze bedrijfstak.

- Horeca - Dienstverlening - Kortlopende ontwikkelingen  
Dit is de reeks zoals deze te downloaden is vanaf Statline van het CBS. Hier is eveneens te zien hoe de nodige reeks berekend is voor verder gebruik.
  
- Horeca (ZONDER ophoging)  
De interpretatie hiervan is gelijk aan de eerdere reeks Horeca. Het enige verschil is dat deze maal er niet gebruik gemaakt is van een ophoging van de Rabobank reeks aan de KvK gegevens. Deze maal is de correlatie zowel op maandniveau als op jaarniveau gedaald. Bij beide gaat het om een minimale daling, dus zelfde conclusies als bij eerdere reeksen zijn mogelijk.
  
- Industrie  
Rabo: Alle industrie gerelateerde bedrijven vormen de data voor de bedrijfstak industrie van de Rabobank. Dit zijn de maandelijkse credit transacties van de zakelijke klanten van Rabobank.  
CBS: De omzet betreft de factuurwaarde (exclusief BTW) van de aan derden verkochte zelfvervaardigde of geassembleerde goederen en opbrengsten van diensten. De gegevens worden verzameld bij industriële ondernemingen met minstens 20 werknemers. De meest recente indices van de omzet hebben een voorlopig karakter. Als gevolg van nagekomen respons van bedrijven kunnen deze nog wijzigen. In zijn algemeenheid wijzigen de indices van de omzet, hoewel ze nog als voorlopig aangeduid worden, na zes maanden nauwelijks.

Conclusies: De correlatie tussen beide reeksen is laag. Zeker in vergelijking tot de correlaties van de andere bedrijfstakken is industrie niet goed te noemen. Dit is grotendeels te wijten aan de manier van gegevens verzamelen. Het CBS kijkt alleen maar naar de omzet van industriële ondernemingen waarbij er 20 of meer werknemers werkzaam zijn. Dit zijn de midden- en grootbedrijven, terwijl de Rabobank voornamelijk midden- en kleinbedrijven als klant heeft. Andere reden kan zijn dat ze in de industrie vaak met meerdere accounts werken bij verschillende banken.

In de grafiek van de omzetindices is weinig samenhang te bespeuren (zoals de lage correlatie al aangeeft). In de maanden feb en aug vallen vrijveel dalen, maar niet altijd. Soms is het weer feb en sep en dan weer jan en jul. De grafiek met het 3 maanden voortschrijdende gemiddelden geeft eveneens weinig informatie, behalve als je de grafiek samenstel van wederom het 3 maanden voortschrijdende gemiddelden waarbij de gegevens van de Rabobank een maand verschoven zijn, lijkt er enig verband te zijn. De correlatie is dan ook behoorlijk gestegen en de pieken/dalen vallen met grote regelmaat samen. Maar wat houdt dit in? Wat geeft dit voor informatie?

- Industrie - Productie en verwerking Omzetindices  
Hier staat de reeks van omzetindices zoals deze van Statline te downloaden is onder Bedrijfsleven -> Industrie -> Productie en verwerking -> indexcijfers; productie, omzet en orders.
- Industrie (ZONDER ophoging)  
De interpretatie hiervan is gelijk aan de eerdere reeks Industrie. Het enige verschil is dat deze maal er niet gebruik gemaakt is van een ophoging van de Rabobank reeks aan de KvK gegevens. Industrie is de enige bedrijfstak die zowel op maandniveau als op jaarniveau een stijging in de correlatie heeft. Op jaarniveau is deze stijging best behoorlijk te noemen.
- NatRekRegio Jaar - Productiestructuur Regio  
Rabo: Is nog niet gebruikt  
CBS: Geeft van de jaren 1998 tot met 2002 de productie weer van elke COROP. De gegevens zijn bij Statline te halen onder het volgende: Macro-economie, financiële instellingen -> Nationale rekeningen -> Nationale rekeningen regionaal -> Productiestructuur per regio of via Nederland regionaal -> Macro-economie, financiële instellingen -> Nationale rekeningen -> Productiestructuur per regio of via Nederland regionaal.

Conclusies: Nog geen vergelijking kunnen maken, dus ook nog geen conclusies.

- Reeks per jaar  
Hier is er gekeken naar de omzetten die ontstaan als je werkt met een 1-jarig panel en een 2-jarig panel. De conclusie is dat het in werkelijkheid niet heel veel uitmaakt. Je ziet dat de correlatie (uiteraard) erg hoog zijn. De grafiek van verschillen ligt ook vlak op elkaar. Omdat het toch meer gaat over de informatie die je eruit kan halen (wanneer daling, stijging, afwijkend gedrag, enz.) maakt het in wezen niet uit welke je kiest. De storing is te verwaarlozen. Verder is dit bestand handig om lange reeksen te maken van meerdere jaren. Copy/paste naar het betreffende jaar en je krijgt vanzelf in het tabblad Hor de reeksen in zijn geheel. Hiermee kan je dan weer berekeningen doen.
- Transport CBS Jaarreeks  
Rabo: Transport is bij de Rabobank al een aparte bedrijfstak. Het CBS heeft helaas niet meer dan jaarcijfers van transport, dus zijn hier alleen maar jaarcijfers gebruikt die weer berekend zijn uit de maandcijfers geproduceerd door de LAM.  
CBS: De beste reeks die qua definitie gelijk is aan die van de Rabobank, bestaat helaas alleen maar uit jaarindices van de jaren 2001 tot met 2004. Voor het gemak is 2001 hier op 100 gezet en de rest is daarmee berekend. Op Statline staat deze onder het

volgende: Macro-economie, financiële instellingen -> Nationale rekeningen -> Nationale rekeningen jaarcijfers -> Goederen en diensten; productgroepen

Conclusies: Ook al is de correlatie maar berekend over 4 waarden, toch is deze aanzienlijk hoog. Blijkbaar is de transport niet zo wispelturig als andere bedrijfstakken zijn. Verder is er nog niet zoveel te concluderen, maar dat is eerlijk gezegd ook niet zo erg. Het is niet een van de 'belangrijkste' bedrijfstakken.

- **Wederuitvoer**

In deze file komt duidelijk naar voren dat er vanaf 2004 een flinke stijging in de wederuitvoer aanwezig was. Deze heeft in grote mate de verandering van Rabobank t.o.v. het CBS tot stand gebracht. Dit komt ook duidelijk naar voren bij de berekening van de correlaties.

## 4.8 Indeling van bedrijfstakken



### Acht samengestelde bedrijfstakken

Agrarisch

=

### Uitgebreide indeling bedrijfstakken

Akkerbouw

Melkveehouderij

Intensieve veehouderij

Gecombineerd bedrijf

Tuinbouw

Visserij

Loonbedrijf

Overig Agr.Bedr.

Industrie

=

Industrie

Bouwnijverheid

=

Bouwnijverheid

Handel

=

Groothandel

Detailhandel Food

Detailhandel Non-Food

Garages/autoreparatie

Horeca en recreatie

=

Horeca en recreatie

Transport

=

Transport

Commerciële dienstverlening

=

Commerciële dienstverlening

Overige dienstverlening

=

Medisch vrije beroepen

Non-profit dienstverlening

Overheid

Recreatieve verenigingen

### Indeling bij Rabobank LAM

Agrarisch

=

A+B Landbouw, bosbouw en visserij

(n.v.t.)

=

C Delfstoffenwinning

Industrie

=

D+E Industrie, openbare nutsbedrijven

Bouwnijverheid

=

F Bouwnijverheid

Handel

=

G Handel, reparatie consumentenart

Horeca en recreatie

=

H Horeca

Transport

=

I Vervoer, opslag en communicatie

Commerciële dienstverlening

=

J Financiële instellingen

K Verhuur, zakelijke dienstverlening

---

Overige dienstverlening	=	L	Overheid, defensie, sociale verz.
		M	Gesubsidieerd Onderwijs
		N	Gezondheids- en welzijnszorg
		O+P	Milieu, cultuur en ov. Diensten

Indeling bij Rabobank LAM

Agrarisch

Industrie

Bouwnijverheid

Handel

Horeca en recreatie

Transport

Commerciële dienstverlening

Overige dienstverlening

Indeling bij EIM

Agrarische producten

Industrie (exclusief 36631)

Bouw

Groothandel

Detailhandel

Toeristisch-recreatieve diensten

Transportdiensten

Financiële diensten

Zakelijke diensten

Onderhouds- en reparatiediensten

Persoonlijke diensten

Communicatiediensten



## 4.9 Uitdraai STAMP



---- GiveWin 2.20 session started at 15:19:52 on Wednesday 30 August 2006 ----

---- STAMP 6.30 session started at 15:19:52 on Wednesday 30 August 2006 ----

Please cite STAMP as:

Koopman S.J., Harvey, A.C., Doornik, J.A. and Shephard, N. (2000).

Stamp: Structural Time Series Analyser, Modeller and Predictor,

London: Timberlake Consultants Press.

Method of estimation is Maximum likelihood

The present sample is: 1998 (1) to 2006 (4)

MaxLik initialising...

it 1 f= -2.58010 e0= 0.64712 step= 1.00000

it 2 f= -2.47521 e0= 0.19332 step= 1.00000

it 3 f= -2.45344 e0= 0.15177 step= 1.00000

it 4 f= -2.44853 e0= 0.12618 step= 1.00000

it 5 f= -2.43961 e0= 0.03139 step= 1.00000

MaxLik iterating...

it 4 f= -2.43872 df= 0.00001 e1= 0.00002 e2= 0.00027 step= 1.00000

it 9 f= -2.43872 df= 0.00000 e1= 0.00000 e2= 0.00001 step= 1.00000

Equation 55.

NL\_index = Trend + Trigo seasonal + Irregular

Estimation report

Model with 4 parameters ( 3 restrictions).

Parameter estimation sample is 1998. 1 - 2006. 4. (T = 100).

Log-likelihood kernel is -2.438718.

Very strong convergence in 9 iterations.

( likelihood cvg 2.872437e-012

gradient cvg 8.661516e-007

parameter cvg 7.311283e-006 )

Eq 55 : Diagnostic summary report.

Estimation sample is 1998. 1 - 2006. 4. (T = 100, n = 87).

Log-Likelihood is -243.872 (-2 LogL = 487.744).

Prediction error variance is 130.384

Summary statistics

	NL_index
Std.Error	11.419
Normality	4.5336
H( 29)	7.1624
r( 1)	-0.091424
r( 9)	-0.12795
DW	2.1502
Q( 9, 6)	13.117
Rs^2	0.39002

Eq 55 : Estimated variances of disturbances.

Component	NL_index (q-ratio)
Irr	107.87 ( 1.0000)
Lvl	10.093 (0.0936)
Slp	0.00000 ( 0.0000)
Sea	0.00000 ( 0.0000)

Method of estimation is Maximum likelihood

The present sample is: 1998 (1) to 2006 (4)

MaxLik initialising...

it 1 f= -2.44874 e0= 0.34144 step= 1.00000  
 it 2 f= -2.37955 e0= 0.15892 step= 1.00000  
 it 3 f= -2.36191 e0= 0.13764 step= 1.00000  
 it 4 f= -2.35343 e0= 0.07802 step= 1.00000  
 it 5 f= -2.34985 e0= 0.02771 step= 1.00000

MaxLik iterating...

it 4 f= -2.34913 df= 0.00000 e1= 0.00002 e2= 0.00029 step= 1.00000  
 it 6 f= -2.34913 df= 0.00000 e1= 0.00000 e2= 0.00000 step= 0.00000

Equation 56.

NL\_index = Trend + Fixed seasonal + Irregular

Estimation report

Model with 3 parameters ( 2 restrictions).

Parameter estimation sample is 1998. 1 - 2006. 4. (T = 100).

Log-likelihood kernel is -2.34913.

Very strong convergence in 6 iterations.

( likelihood cvg 1.701397e-015

gradient cvg 4.884981e-009

parameter cvg 5.24858e-014 )

Eq 56 : Diagnostic summary report.

Estimation sample is 1998. 1 - 2006. 4. (T = 100, n = 98).

Log-Likelihood is -234.913 (-2 LogL = 469.826).

Prediction error variance is 127.537

Summary statistics

	NL_index
Std.Error	11.293
Normality	4.4196
H( 32)	4.3711
r( 1)	-0.089610
r( 9)	-0.12574
DW	2.1497
Q( 9, 7)	14.506
Rs^2	0.40333

Eq 56 : Estimated variances of disturbances.

Component	NL_index (q-ratio)
Irr	107.87 ( 1.0000)
Lvl	10.093 ( 0.0936)
Slp	0.00000 ( 0.0000)

Method of estimation is Maximum likelihood

The present sample is: 1998 (2) to 2006 (4)

MaxLik initialising...

it 1 f= -2.44267 e0= 0.29285 step= 1.00000  
 it 2 f= -2.39299 e0= 0.16876 step= 1.00000  
 it 3 f= -2.37467 e0= 0.09089 step= 1.00000  
 it 4 f= -2.37337 e0= 0.08750 step= 0.50000  
 it 5 f= -2.36911 e0= 0.03142 step= 1.00000

MaxLik iterating...

it 4 f= -2.36430 df= 0.00002 e1= 0.01012 e2= 0.40306 step= 1.00000  
 it 9 f= -2.35761 df= 0.00002 e1= 0.00019 e2= 0.02400 step= 1.00000  
 it 14 f= -2.35721 df= 0.00004 e1= 0.00336 e2= 0.08169 step= 1.00000  
 it 19 f= -2.35697 df= 0.00000 e1= 0.00000 e2= 0.00187 step= 1.00000  
 it 24 f= -2.35697 df= 0.00000 e1= 0.00000 e2= 0.00001 step= 1.00000

Equation 69.

NL\_index = Trend + 1 Cycle(s) + Fixed seasonal + Expl vars + Irregular

Estimation report

Model with 6 parameters ( 2 restrictions).

Parameter estimation sample is 1998. 2 - 2006. 4. (T = 99).

Log-likelihood kernel is -2.35697.

Very strong convergence in 24 iterations.

( likelihood cvg 1.025168e-012

gradient cvg 1.828648e-007

parameter cvg 8.607709e-006 )

Eq 69 : Diagnostic summary report.

Estimation sample is 1998. 2 - 2006. 4. (T = 99, n = 97).

Log-Likelihood is -233.34 (-2 LogL = 466.68).

Prediction error variance is 118.962

Summary statistics

	NL_index
Std.Error	10.907
Normality	4.4790
H( 32)	3.8730
r( 1)	-0.0065410
r(11)	-0.016717
DW	1.9826

Q(11, 6)	11.514
Rs^2	0.44728

Eq 69 : Estimated variances of disturbances.

Component	NL_index (q-ratio)
Irr	109.75 ( 1.0000)
Lvl	3.2011 ( 0.0292)
Slp	0.00000 ( 0.0000)
Cy3	0.53854 ( 0.0049)

Method of estimation is Maximum likelihood  
The present sample is: 1998 (2) to 2006 (4)

MaxLik initialising...

it 1 f= -2.44750 e0= 0.29533 step= 1.00000  
it 2 f= -2.40384 e0= 0.18622 step= 1.00000  
it 3 f= -2.38110 e0= 0.10556 step= 1.00000  
it 4 f= -2.37803 e0= 0.08115 step= 0.50000  
it 5 f= -2.37412 e0= 0.03482 step= 1.00000

MaxLik iterating...

it 4 f= -2.36633 df= 0.00125 e1= 0.00786 e2= 0.25212 step= 0.45197  
it 9 f= -2.35338 df= 0.00004 e1= 0.00386 e2= 0.07205 step= 1.00000  
it 14 f= -2.35303 df= 0.00002 e1= 0.00029 e2= 0.03191 step= 1.00000  
it 19 f= -2.35300 df= 0.00000 e1= 0.00000 e2= 0.00018 step= 1.00000  
it 22 f= -2.35300 df= 0.00000 e1= 0.00000 e2= 0.00001 step= 1.00000

Equation 70.

NL\_index = Trend + 1 Cycle(s) + Fixed seasonal + Expl vars + Irregular

Estimation report

Model with 6 parameters ( 3 restrictions).  
Parameter estimation sample is 1998. 2 - 2006. 4. (T = 99).  
Log-likelihood kernel is -2.352998.  
Very strong convergence in 22 iterations.  
( likelihood cvg 3.281885e-012  
gradient cvg 1.201409e-007  
parameter cvg 7.901114e-006 )

Eq 70 : Diagnostic summary report.

Estimation sample is 1998. 2 - 2006. 4. (T = 99, n = 97).  
Log-Likelihood is -232.947 (-2 LogL = 465.894).  
Prediction error variance is 109.37

Summary statistics

	NL_index
Std.Error	10.458
Normality	7.0505

H( 32)	3.6434
r( 1)	-0.063537
r(11)	-0.016615
DW	2.0979
Q(11, 6)	11.231
Rs^2	0.49185

Eq 70 : Estimated variances of disturbances.

Component	NL_index (q-ratio)
Irr	113.29 ( 1.0000)
Lvl	0.00000 ( 0.0000)
Slp	0.00000 ( 0.0000)
Cy3	1.1607 (0.0102)

Method of estimation is Maximum likelihood  
The present sample is: 1998 (2) to 2006 (4)

MaxLik initialising...

it 1 f= -2.42315 e0= 0.27148 step= 1.00000  
it 2 f= -2.38249 e0= 0.12609 step= 1.00000  
it 3 f= -2.37543 e0= 0.16806 step= 0.50000  
it 4 f= -2.35903 e0= 0.06067 step= 1.00000  
it 5 f= -2.35691 e0= 0.02935 step= 1.00000

MaxLik iterating...

it 4 f= -2.35164 df= 0.00000 e1= 0.00433 e2= 0.00000 step= 0.00000  
it 9 f= -2.32828 df= 0.00157 e1= 0.00560 e2= 0.28716 step= 1.00000  
it 14 f= -2.32403 df= 0.00001 e1= 0.00029 e2= 0.01855 step= 1.00000  
it 19 f= -2.32391 df= 0.00000 e1= 0.00001 e2= 0.00037 step= 1.00000  
it 24 f= -2.32391 df= 0.00000 e1= 0.00000 e2= 0.00000 step= 0.00001

Equation 74.

NL\_index = Trend + 1 Cycle(s) + Fixed seasonal + Expl vars + Irregular

Estimation report

Model with 6 parameters ( 3 restrictions).

Parameter estimation sample is 1998. 2 - 2006. 4. (T = 99).

Log-likelihood kernel is -2.323908.

Very strong convergence in 24 iterations.

( likelihood cvg 5.732875e-016

gradient cvg 2.280102e-007

parameter cvg 5.261938e-011 )

Eq 74 : Diagnostic summary report.

Estimation sample is 1998. 2 - 2006. 4. (T = 99, n = 97).

Log-Likelihood is -230.067 (-2 LogL = 460.134).

Prediction error variance is 100.702

## Summary statistics

	NL_index
Std.Error	10.035
Normality	6.6737
H( 32)	4.2442
r( 1)	-0.014685
r(11)	-0.00049516
DW	2.0159
Q(11, 6)	11.068
Rs^2	0.53212

Eq 74 : Estimated variances of disturbances.

Component	NL_index (q-ratio)
Irr	97.471 ( 1.0000)
Lvl	0.00000 ( 0.0000)
Slp	0.00000 ( 0.0000)
Cy3	2.8424 ( 0.0292)

Eq 74 : Estimated standard deviations of disturbances.

Component	NL_index (q-ratio)
Irr	9.8727 ( 1.0000)
Lvl	0.00000 ( 0.0000)
Slp	0.00000 ( 0.0000)
Cy3	1.6859 ( 0.1708)

Eq 74 : Estimated parameters of Cy3.

The cycle variance is 80.2024.

The rho coefficient is 0.98212.

The cycle period is 33.633 ( 2.80275 'years').

The frequency is 0.186816.

Eq 74 : Estimated coefficients of final state vector.

Variable	Coefficient	R.m.s.e.	t-value
Lvl	-111.83	97.371	-1.1485 [ 0.2536]
Slp	1.0855	0.13185	8.2326 [ 0.0000]
Cy3_1	5.5919	4.4880	
Cy3_2	-13.506	5.5513	
Sea_1	-4.8309	3.3837	-1.4277 [ 0.1566]
Sea_2	4.3071	3.3064	1.3026 [ 0.1958]
Sea_3	1.2479	3.3131	0.37666 [ 0.7073]
Sea_4	5.9169	3.4398	1.7201 [ 0.0886]
Sea_5	5.8008	3.5051	1.655 [ 0.1012]
Sea_6	6.9267	3.4829	1.9888 [ 0.0495]
Sea_7	-11.773	3.4830	-3.3801 [ 0.0010]
Sea_8	-12.217	3.6345	-3.3614 [ 0.0011]
Sea_9	-7.1934	3.5977	-1.9994 [ 0.0484]
Sea_10	-10.804	3.4752	-3.1089 [ 0.0025]
Sea_11	9.2527	3.6175	2.5577 [ 0.0121]

Eq 74 : Estimated coefficients of explanatory variables.

Variable	Coefficient	R.m.s.e.	t-value
NL_index_1	-0.17547	0.10426	-1.683 [0.0956]
IP	0.78479	0.81192	0.96659 [0.3362]
IP_1	2.3011	0.82097	2.8029 [0.0061]

Eq 74 : Cycle analysis for Cy3.

The amplitude of the cycle is 14.6178.

Normality test for Residual NL\_index

Sample Size 97

Mean	-0.038319
Std.Devn.	0.924230
Skewness	0.143508
Excess Kurtosis	1.024512
Minimum	-2.731630
Maximum	2.478328
Skewness Chi <sup>2</sup> (1)	0.33295 [0.5639]
Kurtosis Chi <sup>2</sup> (1)	4.2422 [0.0394]
Normal-BS Chi <sup>2</sup> (2)	4.5752 [0.1015]
Normal-DH Chi <sup>2</sup> (2)	6.6737 [0.0355]

Goodness-of-fit results for Residual NL\_index

Prediction error variance (p.e.v)	100.702227
Prediction error mean deviation (m.d)	68.737724
Ratio p.e.v. / m.d in squares	1.366369
Coefficient of determination R <sup>2</sup>	0.920910
... based on differences RD <sup>2</sup>	0.675726
... based on diff around seas mean RS <sup>2</sup>	0.532122
Information criterion of Akaike AIC	5.036410
... of Schwartz (Bayes) BIC	5.586890

Serial correlation statistics for Residual NL\_index.

Durbin-Watson test is 2.01591.

Asymptotic deviation for correlation is 0.101535.

Lag	dF	SerCorr	BoxLjung	ProbChi2(dF)
-----	----	---------	----------	--------------

1	0	-0.0147		
2	0	-0.0515		
3	0	0.1450		
4	0	-0.1783		
5	0	0.0056		
6	0	0.0552		
7	1	-0.1633	8.8911	[0.0029]
8	2	0.0706	9.4287	[0.0090]
9	3	-0.1122	10.8033	[0.0128]
10	4	0.0489	11.0675	[0.0258]
11	5	-0.0005	11.0675	[0.0501]
12	6	-0.1823	14.8210	[0.0217]
13	7	0.0581	15.2074	[0.0334]
14	8	-0.1644	18.3331	[0.0189]
15	9	-0.0736	18.9679	[0.0255]

---

16	10	-0.0020	18.9684	[0.0407]
17	11	0.0294	19.0721	[0.0598]
18	12	0.0746	19.7482	[0.0720]
19	13	-0.1054	21.1153	[0.0707]

The loss in the degrees of freedom takes account of the number of relative parameters. It does not take account of any lagged dependent variables and their presence necessitates further adjustment.

#### Normality test for IrrRes NL\_index

Sample Size	99
Mean	-0.001058
Std.Devn.	1.001858
Skewness	-0.105214
Excess Kurtosis	1.909492
Minimum	-3.763628
Maximum	2.880877
Skewness Chi <sup>2</sup> (1)	0.18265 [0.6691]
Kurtosis Chi <sup>2</sup> (1)	15.04 [0.0001]
Normal-BS Chi <sup>2</sup> (2)	15.223 [0.0005]
Normal-DH Chi <sup>2</sup> (2)	16.281 [0.0003]

Eq 75 : Large values in IrrRes NL\_index.

Period	Value
2001.12	-3.7636 [0.0001]

#### Normality test for LvlRes NL\_index

Sample Size	99
Mean	-0.003805
Std.Devn.	0.726732
Skewness	0.272131
Excess Kurtosis	1.634789
Minimum	-2.341383
Maximum	2.496619
Skewness Chi <sup>2</sup> (1)	1.2219 [0.2690]
Kurtosis Chi <sup>2</sup> (1)	11.024 [0.0009]
Normal-BS Chi <sup>2</sup> (2)	12.246 [0.0022]
Normal-DH Chi <sup>2</sup> (2)	12.219 [0.0022]

#### Normality test for SlpRes NL\_index

Sample Size	99
Mean	-0.166717
Std.Devn.	0.368008
Skewness	-2.023775
Excess Kurtosis	4.552415
Minimum	-1.757741
Maximum	0.203122
Skewness Chi <sup>2</sup> (1)	67.578 [0.0000]
Kurtosis Chi <sup>2</sup> (1)	85.488 [0.0000]
Normal-BS Chi <sup>2</sup> (2)	153.07 [0.0000]
Normal-DH Chi <sup>2</sup> (2)	108.55 [0.0000]



## 4.10 Tabellen en figuren



1 jarig panel										
	Correlatie	Agrarisch	Industrie	Bouwnij-verheid	Handel	Horeca en recreatie	Transport	Commerciële dienst-verlening	Overige dienst-verlening	
CBS	Strikt per	0,630	0,657	0,649	0,551	0,617				3,103
	Strikst per	0,626	0,657	0,648	0,550	0,616				3,096
	Ruim per	0,569	0,662	0,672	0,552	0,599				3,054
	Strikt dec	0,420	0,626	0,683	0,593	0,618				2,940
	Strikst dec	0,420	0,626	0,683	0,593	0,618				2,940
	Ruim dec	0,396	0,616	0,686	0,584	0,614				2,897
	Schaling Strikt per	0,626	0,518	0,630	0,478	0,631				2,882
	Schaling Strikt per	0,630	0,458	0,571	0,385	0,604				2,647
	Schaling Ruim per	0,569	0,417	0,563	0,377	0,588				2,515
	Schaling Strikt dec	0,420	0,379	0,531	0,365	0,419				2,114
	Schaling Ruim dec	0,396	0,387	0,523	0,357	0,406				2,069
	Schaling Strikt dec	0,420	0,374	0,517	0,361	0,395				2,066
EIM	Strikt per	0,817	0,821	0,979	0,973	0,964	0,975	0,939	0,886	7,356
	Strikt per	0,815	0,821	0,979	0,972	0,964	0,975	0,940	0,887	7,353
	Ruim per	0,803	0,792	0,970	0,967	0,958	0,962	0,948	0,898	7,298
	Strikt dec	0,902	0,765	0,974	0,946	0,952	0,944	0,927	0,881	7,292
	Strikt dec	0,902	0,765	0,974	0,946	0,952	0,944	0,927	0,881	7,292
	Ruim dec	0,897	0,763	0,976	0,942	0,954	0,943	0,929	0,872	7,276
	Schaling Strikt dec	0,902	0,584	0,957	0,935	0,960	0,917	0,808	0,640	6,703
	Schaling Strikt dec	0,902	0,478	0,936	0,919	0,947	0,888	0,739	0,605	6,415
	Schaling Ruim dec	0,897	0,492	0,927	0,909	0,941	0,889	0,710	0,491	6,258
	Schaling Strikt per	0,817	0,391	0,862	0,860	0,963	0,852	0,413	0,512	5,671
	Schaling Strikt per	0,815	0,123	0,599	0,758	0,949	0,734	0,048	0,403	4,428
	Schaling Ruim per	0,803	-0,029	0,540	0,749	0,946	0,694	-0,068	0,088	3,723

Bron: CBS, EIM; Bewerking: Rabobank

Tabel 1.

2 jarig panel										
	Correlatie	Agrarisch	Industrie	Bouwnij-verheid	Handel	Horeca en recreatie	Transport	Commerciële dienst-verlening	Overige dienst-verlening	
CBS	Strikt per2	0,648	0,674	0,678	0,589	0,652				3,240
	Strikt per2	0,644	0,674	0,678	0,588	0,651				3,236
	Ruim per2	0,567	0,662	0,700	0,595	0,628				3,153
	Schaling Strikt per2	0,644	0,580	0,646	0,510	0,673				3,054
	Strikt dec2	0,422	0,597	0,720	0,661	0,653				3,054
	Strikt dec2	0,422	0,597	0,720	0,661	0,653				3,053
	Ruim dec2	0,393	0,575	0,717	0,643	0,640				2,968
	Schaling Strikt per2	0,648	0,495	0,576	0,427	0,644				2,790
	Schaling Ruim per2	0,567	0,407	0,556	0,416	0,630				2,576
	Schaling Strikt dec2	0,422	0,425	0,577	0,481	0,480				2,385
	Schaling Strikt dec2	0,422	0,398	0,577	0,478	0,496				2,372
	Schaling Ruim dec2	0,393	0,387	0,566	0,466	0,480				2,292
EIM	Strikt per2	0,859	0,864	0,980	0,982	0,988	0,987	0,948	0,892	7,500
	Strikt per2	0,862	0,863	0,980	0,982	0,988	0,987	0,945	0,890	7,497
	Strikt dec2	0,915	0,790	0,988	0,971	0,982	0,972	0,945	0,890	7,453
	Strikt dec2	0,914	0,790	0,988	0,971	0,981	0,972	0,945	0,890	7,450
	Ruim dec2	0,906	0,777	0,987	0,965	0,980	0,969	0,942	0,875	7,402
	Ruim per2	0,847	0,797	0,967	0,977	0,968	0,969	0,956	0,905	7,388
	Schaling Strikt dec2	0,914	0,600	0,945	0,944	0,986	0,934	0,849	0,696	6,869
	Schaling Strikt dec2	0,915	0,497	0,920	0,920	0,979	0,909	0,781	0,668	6,590
	Schaling Strikt per2	0,862	0,560	0,879	0,871	0,988	0,934	0,698	0,642	6,434
	Schaling Ruim dec2	0,906	0,483	0,909	0,909	0,978	0,908	0,734	0,561	6,388
	Schaling Strikt per2	0,859	0,318	0,683	0,765	0,941	0,855	0,433	0,649	5,504
	Schaling Ruim per2	0,847	0,025	0,598	0,724	0,933	0,762	0,260	0,384	4,533

Bron: CBS, EIM; Bewerking: Rabobank

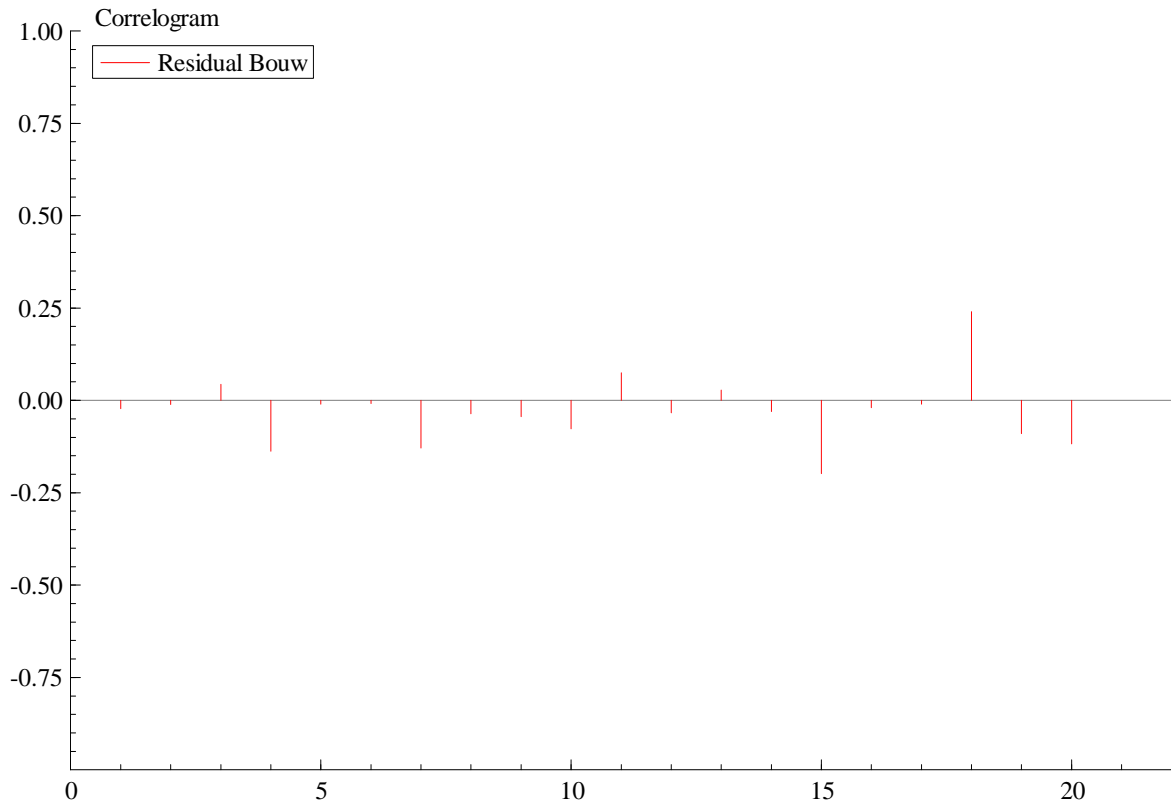
Tabel 2.

C O R O P	Provincie							Commer-	Overige
		Agrarisch	Industrie	Bouwnij- verheid	Handel	Horeca	Transport	dienst- verlening	dienst- verlening
1	Oost-Groningen	0,807	0,795	0,696	0,756	0,975	0,665	0,993	0,885
2	Delfzijl e.o.	0,464	0,611	0,832	0,386	-0,479	-0,303	0,708	0,969
3	Overig Groningen	0,756	0,610	0,954	0,866	0,932	0,822	0,974	0,682
4	Noord-Friesland	0,735	0,729	0,863	0,957	0,976	0,605	0,829	0,934
5	Zuidwest-Friesland	-0,107	-0,110	0,197	0,913	0,627	-0,570	0,881	0,833
6	Zuidoost-Friesland	0,752	0,968	0,965	0,975	0,912	0,757	0,920	0,963
7	Noord-Drenthe	-0,594	0,759	0,780	0,944	0,967	0,618	0,995	0,858
8	Zuidoost-Drenthe	0,875	0,817	0,867	0,973	0,782	0,149	0,764	-0,784
9	Zuidwest-Drenthe	0,890	-0,078	0,732	0,857	0,344	0,825	0,485	-0,142
10	Noord-Overijssel	-0,474	0,037	0,905	0,807	0,901	0,853	0,960	0,596
11	Zuidwest-Overijssel	-0,549	0,846	0,611	-0,265	0,571	0,692	0,907	-0,589
12	Twente	0,869	0,810	0,915	0,962	0,981	0,954	0,904	0,814
13	Veluwe	-0,631	0,656	0,965	0,969	0,981	0,536	0,960	0,942
14	Achterhoek	-0,498	0,451	0,641	0,939	0,952	0,701	0,766	0,895
15	Arnhem/Nijmegen	-0,500	0,607	0,961	0,533	0,834	0,397	0,994	0,948
16	Zuidwest-Gelderland	0,776	0,803	0,858	0,965	0,828	0,817	0,932	0,929
17a	Utrecht West	0,427	0,345	0,841	0,985	0,489	0,908	0,724	0,904
17b	Utrecht	0,812	0,757	0,868	0,955	0,984	0,882	0,936	0,827
17c	Amersfoort	-0,459	0,842	0,607	0,931	0,978	0,953	0,978	0,695
17d	Overig Utrecht	0,816	0,978	0,967	0,915	0,936	0,909	0,877	0,919
18	Kop van Noord-Holland	0,819	0,973	0,907	0,986	0,940	0,961	0,943	0,871
19	Alkmaar e.o.	0,823	0,941	0,964	0,961	0,844	0,469	0,843	0,977
20	IJmond	0,787	0,291	0,881	0,942	0,887	-0,655	0,914	0,808
21	Agglomeratie Haarlem	-0,038	0,837	0,960	0,972	0,930	-0,331	0,929	0,976
22	Zaanstreek	-0,322	-0,003	0,953	0,967	0,824	0,872	0,884	0,872
23	Groot-Amsterdam	-0,815	0,877	0,955	0,979	0,836	0,686	0,901	0,891
24	Het Gooi en Vechtstreek	-0,038	-0,178	0,974	0,951	0,988	0,586	0,912	0,857
25	Agglomeratie Leiden en Bollenstreek	0,301	0,941	0,982	0,982	0,770	0,518	0,871	0,777
26	Agglomeratie 's-Gravenhage	-0,247	0,844	0,929	0,970	0,973	0,998	0,828	0,860
27	Delft en Westland	-0,259	0,727	0,985	0,785	0,861	0,188	0,875	0,920
28	Oost-Zuid-Holland	0,859	0,865	0,848	0,940	0,776	0,866	0,947	0,854
29	Groot-Rijnmond	0,347	0,853	0,927	0,878	0,949	0,964	0,944	0,621
30	Zuidoost-Zuid-holland	-0,765	0,651	0,873	0,882	0,820	0,932	0,701	0,884
31	Zeeuwsch-Vlaanderen	0,770	0,553	0,735	0,858	0,741	0,902	0,724	0,823
32	Overig Zeeland	0,868	0,775	0,849	0,945	0,865	0,684	0,749	0,910
33	West-Noord-Brabant	0,961	0,933	0,972	0,992	0,937	0,899	0,948	0,954
34	Midden-Noord-Brabant	0,072	0,804	0,908	0,993	0,965	0,872	0,973	0,942
35	Noordoost-Noord-Brabant	0,849	0,778	0,874	0,988	0,989	0,974	0,975	0,942
36	Zuidoost-Noord-brabant	0,529	0,846	0,953	0,987	0,968	0,956	0,943	0,951
37	Noord-Limburg	0,816	-0,388	0,758	0,973	0,918	0,984	0,977	0,963
38	Midden-Limburg	0,686	0,764	0,883	0,989	0,968	0,996	0,941	0,960
39	Zuid-Limburg	0,869	0,750	0,904	0,977	0,991	0,984	0,956	0,954
40	Flevoland	0,932	0,968	0,963	0,998	0,894	0,962	0,943	0,949

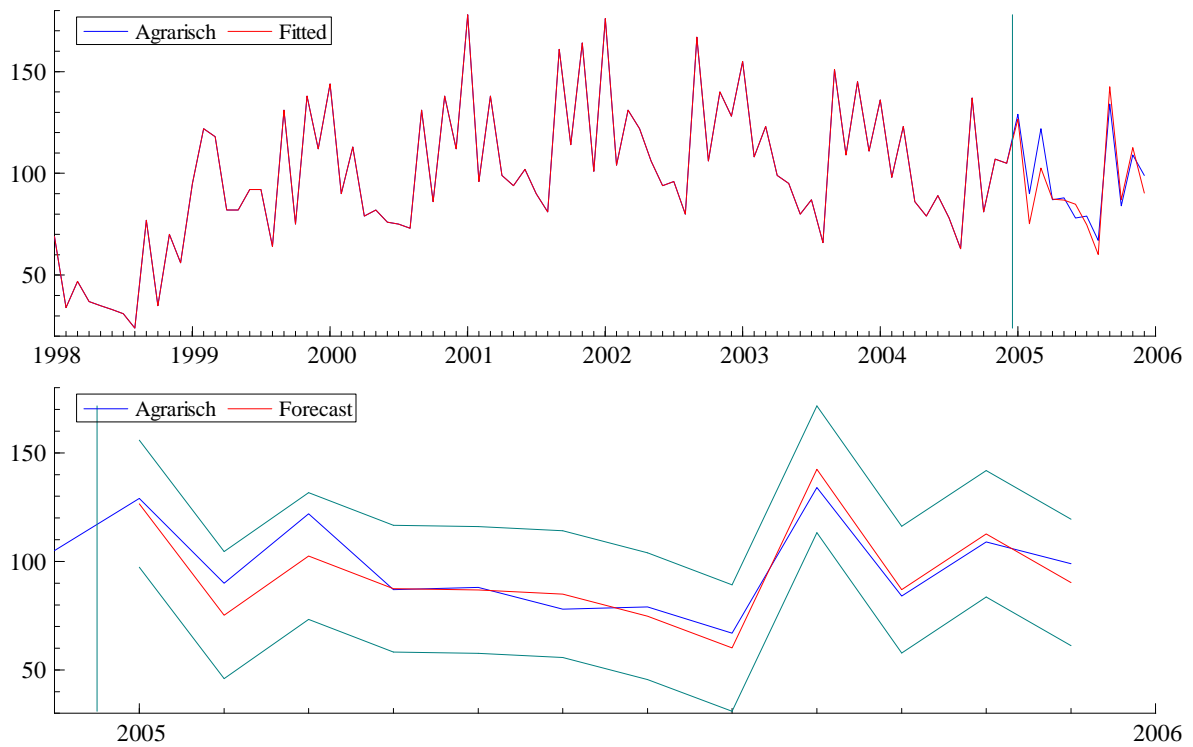
Sterk	Waarde tussen 0,8 en 1.
Redelijk	Waarde tussen 0,5 en 0,8.
Zwak	Waarde kleiner dan 0,5.

Bron: CBS;Bewerking: Rabobank

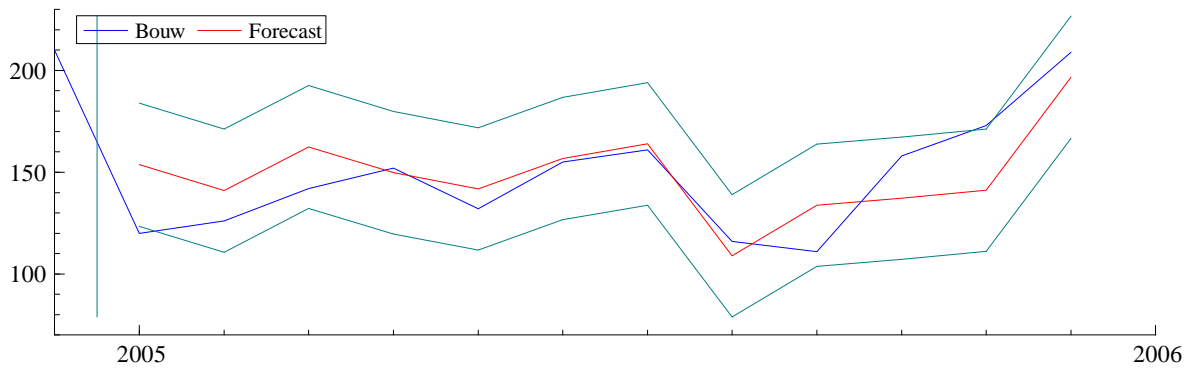
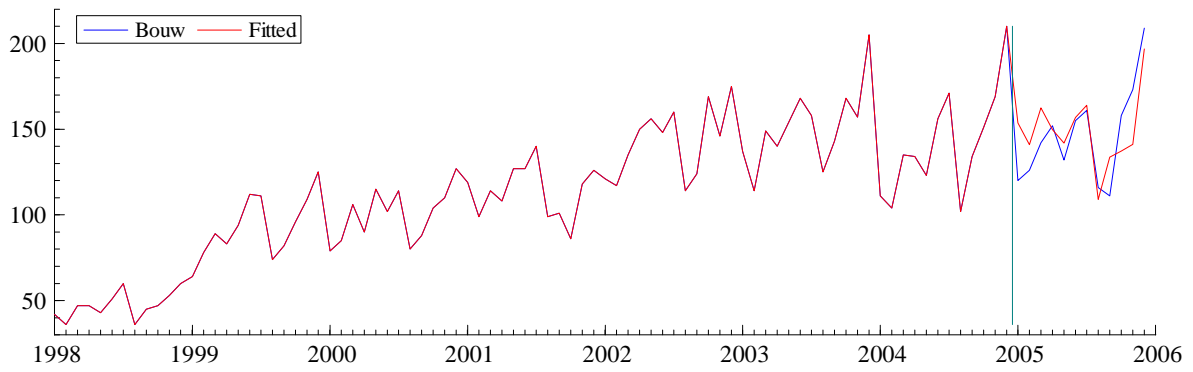
Tabel 12.



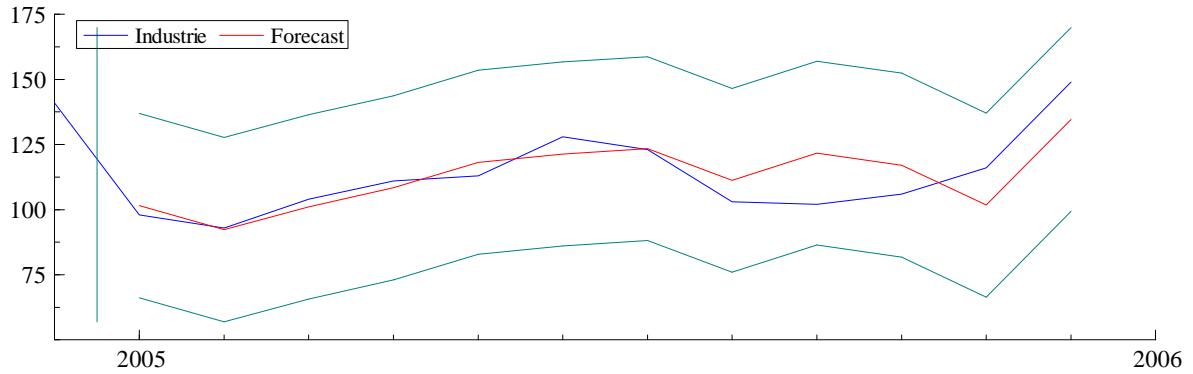
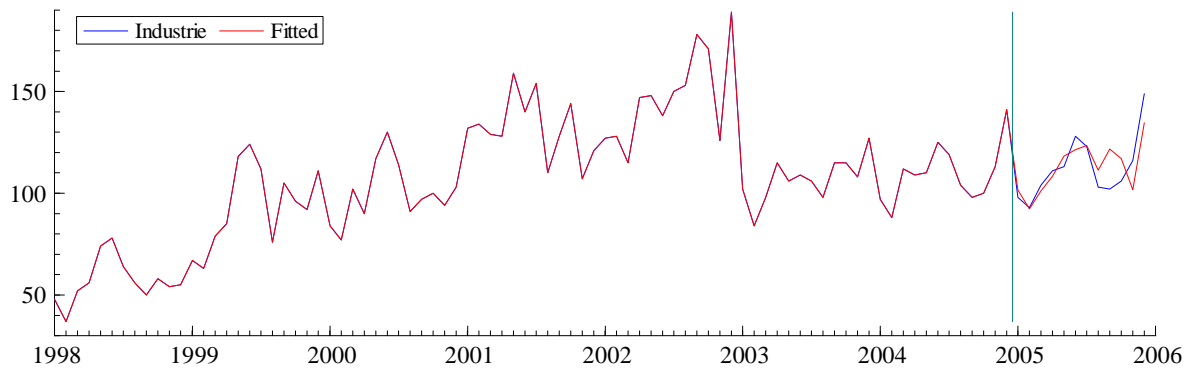
Figuur 23



Figuur 24



Figuur 25



Figuur 26

## 4.11 Woordenlijst



ARMA	=	Auto-regressive Moving average.
CAR	=	Courant Activiteit Rekening.
CBS	=	Centraal Bureau voor de Statistiek.
Conjunctuur	=	Tijdsomstandigheden die van invloed zijn op vraag en aanbod.
COROP	=	Coördinatie Commissie Regionaal Onderzoekprogramma.
Correlatie	=	Verhouding van onderlinge afhankelijkheid of beïnvloeding.
Covariantie	=	Correlatiemaat voor de relatieve standaardafwijking van twee verdelingen.
Delphi	=	Programmaartaal ontwikkeld door Borland.
Deterministisch	=	Waarden zijn bepaald (deze staan vast).
DNB	=	De Nederlandsche Bank.
EIM	=	Economisch Instituut voor Midden- en kleinbedrijf.
Kalman filter	=	Recursief algoritme bedacht door Kalman in de jaren 60.
KvK	=	Kamer van Koophandel.
LAM	=	Lokale Activiteiten Monitor.
Macro	=	Reeks instructies onder een naam of toets(combinatie) om geregeld terugkerende handelingen te verrichten, bv. in een tekstverwerker.
Markov eigenschap	=	De verblijftijd in een gegeven toestand moet een "geheugenloze" verdeling hebben.
Markov proces	=	Een stochastisch proces dat aan de Markov-eigenschap voldoet.
Matrix	=	Geordend systeem van waarden.
Monitor bedrijfsontwikkeling	=	Het instrument van de Rabobank, ook bekend als LAM.
Multivariaat	=	Verklarende variabelen voorspellen een doelvariabele.
Recursief	=	Als iets onbeperkt herhalend toepasbaar is.
Smoothing	=	Gladstrijken van variabelen.
State space	=	Toestandsruimte.
Stationair	=	Wil zeggen dat de statistische eigenschappen, zoals de verwachte waarde en variantie, gelijk zijn voor elk tijdstip t.
Statline	=	Online databank van CBS.
Stochastisch	=	Veranderen met een bepaalde kans.
Tijdreeks	=	Laat de ontwikkeling in de tijd zien van een bepaald verschijnsel.

Trend	=	Is de structurele component in een tijdreeks die de geleidelijke beweging in één richting aangeeft.
Univariaat	=	Alleen data van te voorspellen variabele wordt gebruikt.
Vector	=	Een waarde gespecificeerd door grootte en richting.
Visual basic	=	Programmeertaal in Office Excel.

---

## 4.12 Literatuurlijst



- Bain, L. J. en Engelhardt, M. (1992). "Introduction to probability and mathematical statistics".
- Harvey, A. C. (1989). "Forecasting, structural time series models and the Kalman Filter".  
Cambridge, UK: Cambridge University Press.
- Huisman, M. (1999). "Fluctuaties in de gerigistreeerde werkloosheid en de respons in de Enquete Beroepsbevolking". Centraal Bureau voor de Statistiek.
- Kalman R.E. (1960). "A new approach to linear filtering and prediction problems".
- Koopman, S. J. (2000). "Met het Kalman filter vooruit".
- Koopman, S. J., Harvey, A. C., Doornik, J. A. en Shephard, N. (2000). Stamp: Structural Time Series Analyser, Modeller and Predictor, London: Timberlake Consultants Press.
- Koopman, S. J. en Ooms, M. (2001). "Time series modelling of daily tax revenues". *Statistica Neerlandica* 57, 439–469.
- Koopman, S. J. en Shephard, N. and Doornik, J. A. (1998). "Statistical algorithms for models in state space using SsfPack 2.2". *Econometrics Journal*, volume 1, 1-55.
- Visser, H. (2002). "Detectie van milieuveranderingen". Rijksinstituut voor Volksgezondheid en Milieu.