# Het Markov switching model met variabele overgangskansen toegepast op de S&P 500

**V.M. van Rooijen (309165)**

**Abstract**

**Dit onderzoek bekijkt of Markov switching modellen een goede manier zijn om bull en bearmarkten te voorspellen. Hierbij neem ik Markov switching modellen met vaste en variabele overgangskansen en vergelijk deze met het onderzoek van Hogerwerf et al(2009). De variabelen waar de overgangskans het meest door beïnvloed werd bleken de industriële productie en de wisselkoersindex te zijn. Zowel qua identificatie, als qua voorspellingen bleken de Markov swichting modellen onder te doen voor de resultaten van Hogerwerf et al(2009).**

## 1 Inleiding

Sinds het bestaan van financiële markten heeft men geprobeerd koersen te voorspellen. Door te kopen wanneer de prijs laag is en te verkopen wanneer de prijs hoog is, kan men veel geld verdienen. Voor deze strategie is het echter wel vereist te weten wanneer de markt zal stijgen en wanneer deze zal dalen.

Omdat korte termijn rendementen zich slecht laten voorspellen, zal dit onderzoek zich richten op de lange termijn. In plaats van op dagelijkse of wekelijkse basis stijgingen en dalingen van de markt te bekijken, zal ik de globale staat van de markt onderzoeken op maandbasis. De economie bevindt zich ofwel in een bull (stijgende) markt, ofwel in een bear (dalende) markt. Door een lange termijn visie is de belegger minder gevoelig voor transactiekosten, die zeker bij veel kopen en verkopen een groot deel van de winst ongedaan zullen maken.

Om geen last te hebben van sector- en bedrijfsgerelateerde effecten, heb ik een brede index genomen, namelijk de S&P 500. Bijkomend voordeel is dat macro-economische en financiële data in de Verenigde Staten ruim voorhanden zijn.
Aan de hand van een scala aan financiële en macro-economische data kan een voorspelling gemaakt worden van de staat van de markt.
De hoofdvraag van dit onderzoek luidt: Laat de globale trend in de aandelenmarkt zich beter door een Logit model of door een Regime Switching model beschrijven en welk model levert het meeste op?
Een voor de handliggende manier om dit te doen is een Markov Switching model. Volgens dit model kan de markt zich in verschillende regimes bevinden, in dit geval een bull- of een bearmarkt. Het verwachte rendement in de huidige periode evenals de kans dat het model van regime switcht de volgende periode is afhankelijk van het huidige regime. Dit zijn aantrekkelijke eigenschappen, omdat in bullmarkten er een hoger rendement is dan in bearmarkten. Bulls houden langer aan dan bears , dus de kans dat er van regime geswitcht wordt is kleiner in een bullmarkt. Bijkomend geeft een Markov Switching model zelf aan in welk regime de markt zich bevindt, waardoor er vooraf geen definitie van een Bull of Bearmarkt vereist is. Hamilton (1989) gebruikte voor het eerst Markov Switching technieken om business cycles te beschrijven en analyseren. Deze manier van aanpak is sindsdien vaak herhaald en uitgebreid. In dit onderzoek zal de kans dat er van regime veranderd wordt ook afhankelijk zijn van macro-economische en financiële variabelen.
Een andere methode is het voorspellen van de markt door eerst de status van de markt bepalen, en vervolgens met een logitmodel voorspelling doen van de markt. Uit Hogerwerf et al (2009) is gebleken dat hier succesvol mee gehandeld kan worden.
Dit onderzoek zal de Regime Switching methode vergelijken met de logitregressies.
Om een antwoord op deze vraag te geven is het allereerst van belang te definiëren wanneer de markt zich globaal stijgt of daalt. De korte termijn effecten moeten uit de data gefilterd worden om goede voorspellingen te kunnen doen over de lange termijn.
Het Regime Switching model geeft zelf de kans dat de aandelen markt bull dan wel bear is. De identificatie en vorspellingen van van bull en bearmarkten van het Regime Switching model zal vergeleken worden met die uit Hogerwerf et al (2009).
Om het onderzoek vergelijkbaar te houden met dat over logitregressies, zullen dezelfde macro-economische en financiële variabelen gebruikt worden in dit

onderzoek.  De parameterschattingen van de twee methodes kunnen vervolgens vergeleken worden.
De resultaten van de verschillende modellen moeten vervolgens geëvalueerd en met elkaar vergeleken worden. Er zal voornamelijk rekening gehouden worden met de economische kwaliteit van de resultaten.

## 2 Literatuur

Om gedegen uitspraken te kunnen doen over de staat van de markt moeten Markov Switching modellen goede identificatie van bull- en bearmarkten geven en de omslagpunten van bull naar bear en vice versa goed kunnen aangeven.
Harding & Pagan (2002) onderzoekt of Markov Switching modellen een goede manier zijn om cyclische omslagpunten te bepalen. Harding & Pagan richten zich op het dateren van de US business cycle. Het onderzoek vergelijkt de niet-parametrische methode van Bry & Boschan  met parametrische methode van Hamilton (1989) op basis van Markov Switching. Harding & Pagan geven de voorkeur aan de niet-parametrische methode, omdat deze robuuster en transparanter is. Een voordeel van Markov Switching is wel dat de kans op een recessie voorspeld kan worden. Een nadeel is dat er aannames moeten worden gedaan over het data generating process.
Hamilton (2003) reageert vervolgens op de kritiek van Harding & Pagan. Hamilton voert aan dat de Markov Switching aanpak een statisch model formuleert over het onderwerp en de Harding-Pagan aanpak onduidelijk laat wat het algoritme precies aangeeft. Een dateringsmethode zou volgens Hamilton gebruik moeten maken van eigenschappen van de data.
Harding & Pagan (2003) reageren vervolgens op het commentaar van Hamilton. Zij blijven kiezen voor de transaparantie  van niet-parametrische aanpak, die bovendie dichter blijft bij wat de normale opvatting over wat een recessie is.  Wel geven ze aan dat de Markov Switching methode een betere methode van voorspellen is. Deze eigenschap is belangrijk omdat in mijn onderzoek voorspellingen zal gaan maken over het regime van de aandelenmarkt en goede identificatie niet het hoofddoel van mijn onderzoek is.
Een niet-parametrische methode om bull- en bearmarkten te indentificeren is die van Lunde & Timmermann(2004). Een markt is bullish(bearish) als deze na een lokaal minimum(maximum) de markt met minimaal een bepaald percentage stijgt(daalt).

Voordeel van deze methdoe is dat er in een bear markt ook stijgingen kunnen voorkomen en dalingen in een bullmarkt.
Naast het juist kunnen aangeven van omslagpunten, is het vereist dat de aanames die gemaakt moeten worden om een Markov Switching model te kunnen gebruiken correct zijn. Van Norden en Schaller (1993) gebruiken Markov Switching technieken om de rendementen van de aandelenmarkt te analyseren. Van Norden en Schaller kijken naar de maandelijkse rendementen van de S&P index.
Ze tonen aan dat het zeer waarschijnlijk is dat er verschillende regimes zijn in returns op de aandelenmarkt in gemiddeld rendement, volatiliteit of een combinatie van deze twee. Verder blijkt dat de aandelenmarkt te modeleren is als een eerste orde Markovketen. Nieuw aan het onderzoek is dat de kans van verandering van regime afhankelijk is van economische variabelen, in dit geval de price-dividend ratio.
De basisaanname van mijn onderzoek, namelijk dat de aandelenmarkt te beschrijven is met een eerste order Markov keten, kan dus gebruikt worden.
Maheu en McCurdy(2000) ontwikkelen een model waarbij niet alleen de overgangskansen afhankelijk zijn van de duur van het huidige regime, maar ook het conditionele gemiddelde en de conditionele variantie. De grootste stijgingen van de aandelenmarkt blijken plaats te vinden aan het begin van bullmarkten en naar mate een bearmarkt langer duurt wordt de variantie groter. Tevens neemt de kans van verandering van regime af als de markt zich langer een bepaald regime bevindt.

Diebold, Lee en Weinbach(1994) geven een uitgebreide beschrijving van een Markov Switching model met over de tijd afhankelijke overgangskansen. Het door hen beschreven EM-algoritme heb ik grotendeels voor mijn onderzoek gebruikt. Het EM-algoritme maakt eerst een voorspelling van de het regime op tijdstip *t*. Die voorspelling wordt vervolgens aangepast met de kennis van de tijdstip *t* en vervolgens met die van de gehele sample. Hierna worden de parameters geschat. De methode zoals beschreven door Diebold, Lee en Weinbach is echter traag en geeft af en geeft problemen met schatten van de parameters van de overgangskansen. Wel heb ik de likelihoodfunctie uit dit onderzoek gebruikt voor de uiteindelijke schattingen van deze parameters.
Frances&Van Dijk(1999), volgen grotendeels dezelfde aanpak als Diebold et al., door eerst de forecast te bereken, daarna de inference probabilities en daarna de smoothed inference probabilities, maar het model is eenvoudiger. De methode om de smoothed

inference te bereken is veel sneller dan die van Diebold et al. De in dit onderzoek beschrevn schattingsmethode gaat echter van vaste overgangskansen uit. Met enkele kleine aanpassingen was het echter mogelijk de overgangskansen afhankelijk te maken van externe factoren.
Kole(2009) beschrijft hoe de benodigde overganskansen geschat kunnen worden.

Hogerwerf et al.(2009) kijken naar de financiële markt met een logit model. Allereerst bepalen ze op basis van de methode van Lunde & Timmerman(2004) wanneer de markt bullish of bearish is. Daarna voeren ze hier een logitregressie op uit. De set verklarende variabelen wordt ieder tijdstip opnieuw bepaald aan de hand van AIC, SIC of in-sample hit-rate. De uitkomsten van dit onderzoek zijn positief, omdat de voorspelkwaliteit van de gebruikte methode goed is en een hiermee opgezette handelsstrategie vele malen meer oplevert dan de marktportfolio. Dit onderzoek zal als benchmark gebruikt worden voor het eigen onderzoek.

## 3 Het Model

Het Regime Sitching model gaat uit van het bestaan van twee regimes in de aandelenmarkt. Het regime op tijdstip $t$ , dat we zullen aanduiden als $s_t$ , kan niet direct worden waargenomen.
Ik neem aan dat $s_t$ een eerste orde Markov proces volgt. De overgangskansen volgen uit de logistische functies van $\Lambda( x_{t-1}'\beta_i),\ i = 1,2$ waar de $(k\times 1)$ vector $x_{t-1}$ bestaat uit een constante en macro-economische en financiële variabelen op tijdstip *t*-1. Deze variablen beïnvloeden de overgangskansen. Het effect van de variablen is afhankelijk van het regime.
Dit leidt tot de volgende overgangskansen:

$$P(s_t = 1 | s_{t-1} = 1) = \mathrm{p}_\mathrm{t}^{11} = \Lambda( x_{t-1}\beta_1)$$
$$P(s_t = 2 | s_{t-1} = 1) = \mathrm{p}_\mathrm{t}^{12} = 1 - \mathrm{p}_\mathrm{t}^{11}$$
$$P(s_t = 1 | s_{t-1} = 2) = \mathrm{p}_\mathrm{t}^{21} = 1 - \mathrm{p}_\mathrm{t}^{22}$$
$$P(s_t = 2 | s_{t-1} = 2) = \mathrm{p}_\mathrm{t}^{22} = \Lambda(x_{t-1}\beta_2)$$

waarbij $\Lambda(\mathrm{z}) = \dfrac{\exp(z)}{1+\exp(z)}$ de logistische functie is.

Laat $y_t$, het maandelijks rendement van de index op tijdstip *t,* normaal verdeeld zijn zodat:

$$y_t = \begin{cases} \mu_1 + \sigma_1 \varepsilon_t & \text{if } s_t = 1 \\ \mu_2 + \sigma_2 \varepsilon_t & \text{if } s_t = 2 \end{cases}$$

$\varepsilon_t \sim$ N(i.i.d)

Dan geldt voor $j$ = 1,2 dat de contitionele dichtheid van $y_t$ is:

$$f(y_t \mid s_t = j, \mu_j, \sigma_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{ \frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right\},$$

## 4 Het schatten van het model met het EM-algoritme

Nu er een model is geformuleerd, dient de set parameters $\beta_1, \beta_2, \mu_1, \mu_2, \sigma_1, \sigma_2$ geschat te worden. Diebold et al.(1994) doen dit met behulp van het EM algortime. Aan de set parameters wordt allereerst nog één parameter, $\rho$, toegevoegd.
$\rho = P(s_1 = 1)$ Zodat de set parameters $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \beta_1, \beta_2, \rho)'$ een $(2 \times k + 5)$ vector is.

Het schattten van $\theta$ gaat dan als volgt:

(1) Kies $\theta^{(0)}$

(2) Verkrijg de kans dat de markt zich in een bepaald regime bevindt op tijdstip *t*, gegeven alle waarnemingen $\underline{y}_T$, de regressoren $\underline{x}_T$ en de parameterschattingen $\theta^{(i)}$ voor iedere *t*

$P(s_t = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(i)}) \quad \forall t,$

$P(s_t = 2 \mid \underline{y}_T, \underline{x}_T; \theta^{(i)}) \quad \forall t,$

Verkrijg de kans dat de markt zich in bepaalde regimes bevindt op tijdstip *t* en tijdstip *t*-1

$P(s_t = 1, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(i)}) \quad \forall t,$

$P(s_t = 2, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(i)}) \quad \forall t,$

$P(s_t = 1, s_{t-1} = 2 \mid \underline{y}_T, \underline{x}_T; \theta^{(i)}) \quad \forall t,$

$P(s_t = 2, s_{t-1} = 2 \mid \underline{y}_T, \underline{x}_T; \theta^{(i)}) \quad \forall t,$

(3) Laat $\theta^{(1)} = \arg\max_{\theta} E\left[\log f(\underline{y}_T, \underline{s}_T \mid \underline{x}_T; \theta^{(0)})\right],$

(4) Herhaal tot convergentie

Stappen (2) en (3) staan in appendix 1 beschreven. Door vervolgens te itereren tot convergentie, wordt de optimale waarde van $\theta$ benaderd.

## 5 Variabelen

Voor dit onderzoek heb ik gekozen voor een brede index, namelijk de S&P 500. Hiervan heb ik de maandelijkse rendementen $y_t$ bekeken van januari 1961 tot en met februari 2009. Tabel 1 laat zien dat het gemiddelde rendement over de sample positief is . Er zijn meer maanden met een posief dan maanden met een negatief rendement en het gemiddelde positieve rendement is iets hoger dan het gemiddelde negatieve rendement. Verder is de variantie in maanden met een negatief rendement iets hoger dan in maanden met een positief rendement.

**Tabel 1:** eigenschappen maandelijkse rendementen S&P 500

| | Aantal maanden | μ | $\sigma^2$ |
|---|---|---|---|
| $y_t$ | 578 | 0.0053 | 0.0019 |
| $y_t<0$ | 252 | -0.0317 | 0.00096 |
| $y_t>0$ | 326 | 0.0338 | 0.00072 |

Het blijkt dat positieve en negatieve rendementen elkaar vaak afwisselen. De kans op hetzelfde teken van rendement de volgende periode als deze periode is ongeveer 50%. Een positief rendement de vorige periode wordt net iets vaker gevolgd door een positief rendement de volgende periode en een negatief rendement zal ook net iets vaker gevolgd worden door een positief redement.

$$P(y_t < 0 \mid y_{t-1} < 0) = 0.4127$$
$$P(y_t > 0 \mid y_{t-1} > 0) = 0.5429$$

Het model van Diebold, Lee & Weinbach(1994) geeft de mogelijkheid tot keuze welke variabelen invloed hebben op de overgangskansen. Om vergelijking met het onderzoek van Hogerwerf et al.(2009) mogelijk te maken, zullen dezelfde macro-economische en financiële variabelen gebruikt worden. Meer uitleg over deze variabelen en de wijze waarop ze getransormeerd zijn is te vinden in Hogerwerf et al. (2009). In tabel 2 zijn de opgenomen variabelen weergegeven.

| **Tabel 2:** gebruikte macro-economische en financiële variabelen | | |
|---|---|---|
| **Variabele** | **Toegepaste transformatie** | **Type** |
| Term Spreads(3M-10M) | Geen | Financieel |
| Inflatie | Percentage op jaarbasis | Macro |
| Industriële productie | Procentuele groei op jaarbasis | Macro |
| M1 | Procentuele groei op jaarbasis | Macro |
| M2 | Procentuele groei op jaarbasis | Macro |
| Werkloosheidspercentage | Geen | Macro |
| Federal funds rate | Geen | Financieel |
| Wisselkoersindex | Procentuele groei op maandbasis | Macro |
| Staatsschuld | Procentuele groei op jaarbasis | Macro |
| Dividend-payout ratio | Verandering op jaarbasis | Financieel |

Voor ieder tijdstip zal de beste combinatie van verklarende variablen gezocht worden. In tegenstelling tot Hogerwerf et al.(2009), zal niet iedere combinatie van verklarende variabelen vergeleken worden. Dit omdat er dermate veel berekeningen moeten worden gemaakt om het Markov Switching model te schatten, dat de rekentijd te groot wordt. Voor ieder tijdstip is het afgaan van een groot aantal combinaties vereist. Aanpassen van het model wordt dan vrijwel onmogelijk.

In een dergelijk geval kunnen variabelen op een aantal manieren geselecteerd worden voor opname in het model. Voor de hand liggen de bottom-up approach en de top-down approach. Aan de hand van een selectiecriterium als $R^2$, SIC of AIC kan een model gewaardeerd worden.

Bij de bottom-up approach voegt men de variabele toe aan de set geselecteerde variabelen, die volgens het gebruikte selectiecriterium de waarde van het model het meest verbeterd. Men begin met alleen een constante en voegt variabelen toe tot er geen verbetering meer is.

De keuze voor bottom-up heb ik gemaakt, omdat voor sommige combinaties variabelen de parameterschattingen niet convergeren. Voor het kleinst mogelijke

model, dat waarbij de overgangskansen alleen afhankelijk zijn van een constante, is dit wel het geval.
Hoewel deze aanpak niet garandeeert dat de beste combinatie van variabelen gevonden wordt, zal er vrij dicht in de buurt komen. Tevens is de rekentijd niet in de orde van $2^N$ afhankelijk van het aantal variabelen, maar in de orde van $\frac{1}{2}N^2$
Het selectiecriterium dat gebruikt zal worden om te bepalen of er nog variablen worden toegevoegd, zal het Akaike Information Criterion (AIC) zijn, waarbij $p$ het aantal regressors is en $s_p^2$ de maximum likelihood schatter van de variantie in het model met $p$ regressors.

$$\text{AIC}(p) = \log(s_p^2) + \frac{2p}{n},$$

Het model met de kleinste AIC-waarde wordt gekozen.

## 6 Geschatte Modellen

Allereerst zal ik gaan onderzoeken hoe goed het Markov swichting model in staat is identificatie te geven van bull- en bearmarkten. Ik zal over het model over de hele sample schatten met vast overgangskansen. Die kan ik vervolgens vergelijken met het model met variabele overgagnskansen ook geschat over de hele sample.
De identificatie op basis van Markov switching kan ik dan naast het identificatiealgoritme van Lunde&Timmermann(2004) leggen.
Na over de gehele sample te schatten, zal ik met een expanding window door de sample heen lopen. Door ieder keer het geschatte regime van de markt in de laatste maand van het sample te nemen, kan ik kijken hoe goed dit overeen komt met de identificatie van dezelfde maand door schatting over de gehele sample. Tevens wordt nu zichtbaar hoe de parametersschattingen variëren over de tijd. Ik doe dit wederom met het model met alleen een constante en het model met variabele overgangskansen.
Met deze laatste schattingen kan ik ook een handelsstrategie opzetten. Door te in de marktportfolio te investeren wanneer het model een bullmarkt voorspelt en het geld veilig weg te zetten op de bank bij een voorspelde bearmarkt, probeer ik de marktindex te verslaan. Omdat ik het regime swiching model de kans geeft dat de markt zich in een bepaald regime bevindt, kan ik ook het in de marktportfolio

geïnvesteerde deel van mijn portefeuille evenredig laten zijn met de voorspelde kans dat de markt bull is.

## 7 Resultaten

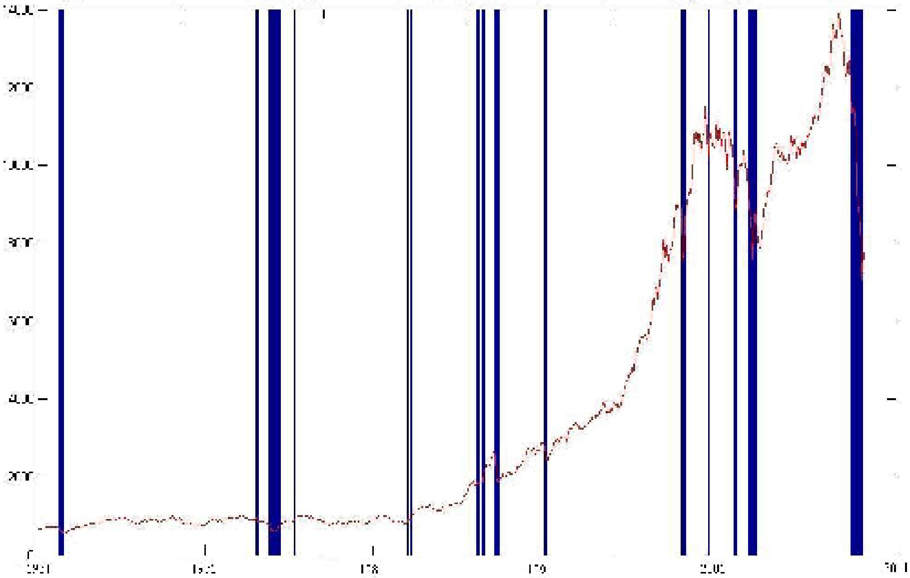
Na het EM-algoritme te hebben getest op diverse zelf geconstrueerde reeksen bleek het helaas niet altijd stabiel. Wanneer de overgangskansen alleen afhankelijk zijn van een constante, geeft het algoritme wel de goede waardes terug. Als er echter meerdere variabelen een rol spelen bij de overgangskansen, dan heeft het algoritme erg veel waarnemingen nodig om tot betrouwbare schattingen te komen, aantallen vele malen groter dan de maximale samplegrootte van 578 maanden. Daarom heb ik gekozen om een expanding window te gebruiken, in plaats van het in Hogerwerf et al.(2009) gebruikte moving window. Helaas komt het voor dat het optimalisatiealgoritme uit Matlab binnen het maximale aantal iteraties niet convergeert en dus geen schatting geeft.

## 8 Evaluatie Identificatie

Om te beginnen kijk ik naar het Regime Switching model, met constante overgangskansen over de gehele sample. Vervolgens zal ik een de overgangskansen afhankelijk laten zijn van macro-economische en financiële variabelen .
De grafieken waarbij de beurskoers is weergegeven met op de achtergrond het voorspelde regime, zullen een eerste indicatie zijn van de kwaliteit van de bull/bear-identificatie doormiddel van een Markov Switching model.

Figuur 1 toont de identificatie van bull en bearmarkten op basis van vaste overgangskansen, waarbij geschat is over de gehele sample. Te zien is dat, hoewel het algorithme nog lang niet perfect is, de belangrijkste dalen als de crisis in eind jaren 80, evenals de internetbuble en de kredietcrisis. Markov Switchingmodellen zijn dus in staat de belangrijkste bearmarkten te vinden en is dus bruikbaar als identificatiealgoritme.

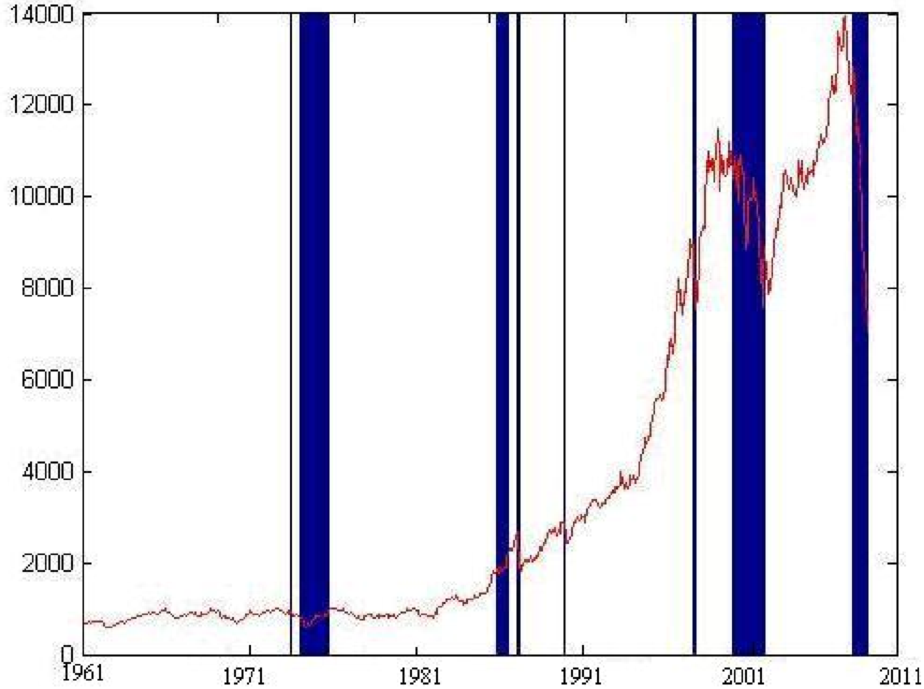
**Figuur 1: identificatie gehele sample met constante overgangskansen**



De parameterschatttingen staan weergegeven in tabel 3. In een bearmarkt is er een negatief rendement van bijna een procent en in een bullmarkt een positief rendement va bijna een procent. De variantie in een bearmarkt is een stuk groter dan in een bullmarkt. De bullmarkten duren gemiddeld langer en zullen dus ook vaker voorkomen.

Met de bottom-up aanpak heb ik vervolgens variabelen toegevoegd die een verbetering van de AIC van het model opleveren. Er worden in dit geval, naast de constante, twee variabelen opgenomen in het model. De extra opgenomen variabelen zijn de industriële productie en de wisselkoersindex. De schattingen zijn nu minder gevoelig en geven langere regimes aan dan met vaste overganskansen. Omdat het Markov switching model niet alleen op basis van rendement, maar ook op basis van variantie maanden als bull dan wel bear classificeert, kan een stijgende markt ook als bear aangemerkt worden door grote variantie. Dit is bijvoorbeeld te zien bij de voorspelde recessie rond 1974.

**Figuur 2**: identificatie gehele sample met variabele overgangskansen

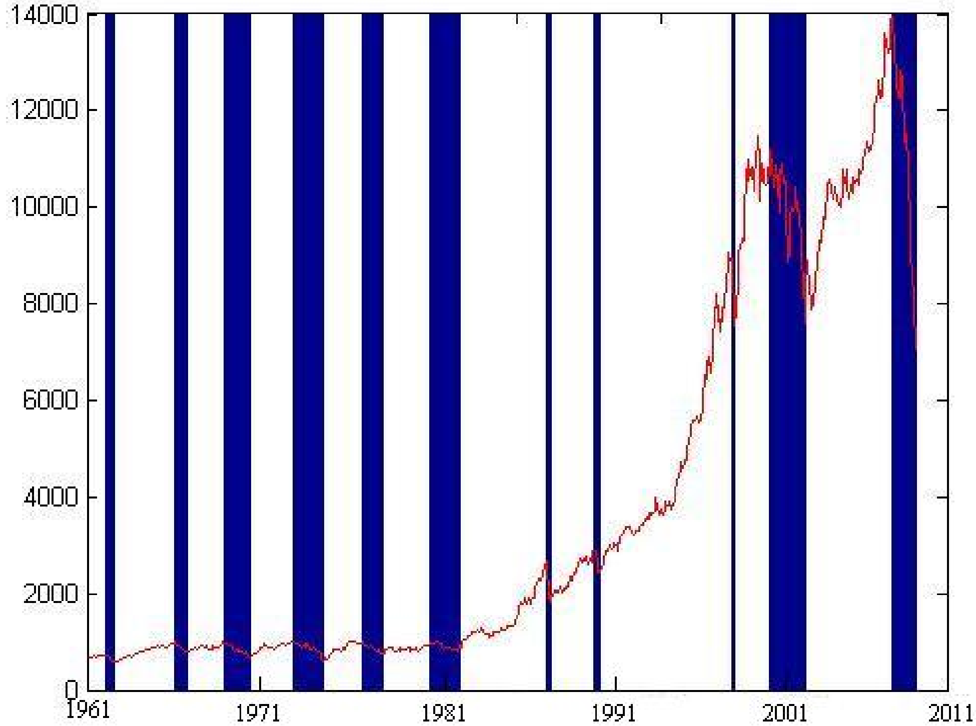



Uit de parameterschattingen, weergegeven in tabel 3, komt naar voren dat de invloed van de industriële productie en de wisselkoersindex op de overgangskansen groter is in een bearmarkt dan in een bullmarkt. Daarnaast zorgt een hogere groei van de idustriële productie er voor dat het minder waarschijnlijk is dat de markt in het huidige regime blijft. Een hogere wisselkoersindex zorgt er juist voor dat het waarschijnlijker is dat huidige regime gehandhaafd blijft. Zoals verwacht is tijdens de bearmarkten de variantie hoger dan tijdens de bullmarkten

Als we het het model met de vaste overgangskansen naast dat met de variabele leggen, blijkt dat het verschil tussen de gemiddelde rendementen groter is bij het model met de variabele overgangskansen. Dit betekent dat het uitgebreidere model beter in staat is om maanden met een slecht rendement te onderscheiden van maanden met een goed rendement. Het geringere aantal switches tussen de regime is ook gunstig. Het geringe aantal swichtes aan het begin van de sample is waarschijnlijk te wijten aan het nog niet bekend zijn van de verklarende variabelen aan het begin van de sample. Deze variabelen werden pas na 1961 bijgehouden.

Als we deze indentificatiemethode vergelijken met die van Lunde &Timmerman(2004), weergegeven in figuur 3, blijkt echter wel dat er een aantal bearmarkten minder worden aangegeven en dat de gevonden bearmarkten korter duren.

**Figuur 3:** identifiactie gehele sample volgens Lunde & Timmermann




Het identificatie volgens Lunde&Timmermann(2004) leidt tot een groter verschil van het rendement in een bullmarkt en dat in een bearmarkt. De variantie in een bearmarkt is aanzienlijk lager dan de variantie geschat met het Markov switching model. De variantie in de bullmarkt is vrijwel vergelijkbaar.

**Tabel 3**

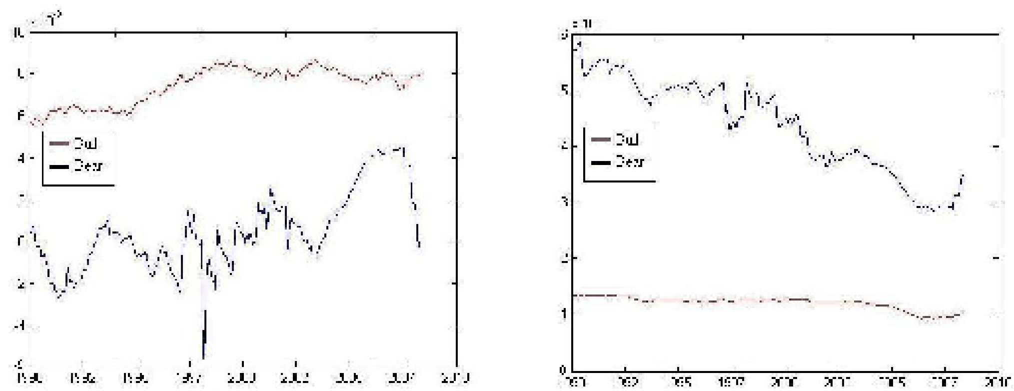
| | Bear | | | | | Bull | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu$ | $\sigma^2$ | $\beta$ | | | $\mu$ | $\sigma^2$ | $\beta$ | | | $\rho$ |
| | | | constant | Ind | wissel | | | constant | Ind | wissel | |
| Constant | - 0.0084 | 0.0048 | 1.00 | - | - | 0.0082 | 0.0012 | 2.81 | - | - | 0.177 |
| Variabel | - 0.0094 | 0.0047 | 18.55 | -4.90 | 6.38 | 0.0081 | 0.0013 | 3.91 | -0.230 | 0.189 | 0.163 |
| L&T | -0.017 | 0.0029 | - | - | - | 0.0123 | 0.0014 | - | - | - | - |

Na over de gehele sample geschat te hebben, loop ik met een expanding window van mimaal 350 maanden door de sample heen. Omdat de wisselkoersindex pas vanaf februari 1982 beschikbaar is, begin ik dus later met schatten dan Hogerwerf et al.(2009). Het wordt nu zichtbaar hoe de parameterschattingen
over de tijd variëren. Ook kan ik voorspellingen maken die ik kan vergelijken met de identificatie en waarmee ik de opbrengst van een strategie op basis van het Markov switching model kan bepalen.

Als mijn algoritme zelf voor ieder tijdstip het model met de beste AIC koos, leverde dit niet altijd de gewenste resultaten op. Daarom heb ik er voor gekozen de best werkende combinatie verklarende variablen over de gehele sample ook te gebruiken voor het schatten met het expanding window. Naast een model met een constante, de industriële productie en de wisselkoersindex, heb ik ook met alleen een constante geschat.

In figuur 4 staan de ontwikkeling van de gemiddelde rendementen en de variantie van de bull en de bearmarkt over de tijd, wanneer er vaste overgangskansen zijn. Het gemiddelde rendement tijdens bullmarkten lijken redelijk stabiel. Het gemiddelde rendement tijdens bearmarkten vertoont veel meer pieken en dalen. Het lijkt er op dat individuele gebeurtenissen veel invloed hebben. Goed voorbeeld is de kredietkrisis aan het einde van de sample die het rendement tijdens bearmarkten sterk naar beneden haalt. De variantie tijdens bearmarkten is veel hoger dan tijdens bullmarkten, en varieert ook veel meer. Periodes die bekend staan als economische krisis zorgen voor grotere variantie.

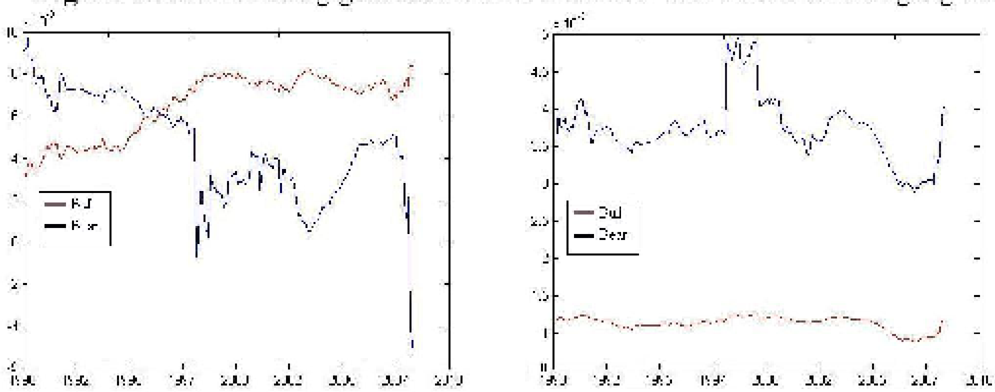
**Figuur 4:** ontwikkeling gemiddelden en varianties met constante overgangskansen




gemiddelden varianties

Na het schatten met alleen een constante is het interessant te weten wat de invloed van externe factoren op de schattingen van de gemiddelde rendementen en de variantie is. Dit leidt tot opvallende resultaten, weergegeven in figuur 5. Het is niet meer mogelijk om van bull- en bearmarkten in de klassieke zin te spreken, omdat het regime dat het hoogste gemiddelde rendement haalt, later in de sample het laagste gemiddelde rendement heeft. Het Markov switching model gebruikt bij de identificatie niet alleen de gemiddelde waarde, maar ook de variantie. De keuze voor de verklarende variabelen lijkt een grote invloed te hebben op het verlopen van de gemiddeldes en variantie. Op de gehele sample geven de industiële productie en de wisselkoersindex een goede identificatie, maar zoals uit de grafiek van de gemiddelde rendementen te zien is, is wordt het verschil tussen de rendementen in de regimes pas groot.. Het is zeer waarschijnlijk dat hiervoor andere combinaties van variabelen een veel betere indicatie geven van het regime waarin de markt zich bevindt. Zo is bij het model met de vaste overgangskansen het verschil tussen de rendementen van de verschillende regimes over de hele sample hoger, behalve over de laatste paar maanden.

**Figuur 5:** ontwikkeling gemiddelden en varianties met variabele overgangskansen

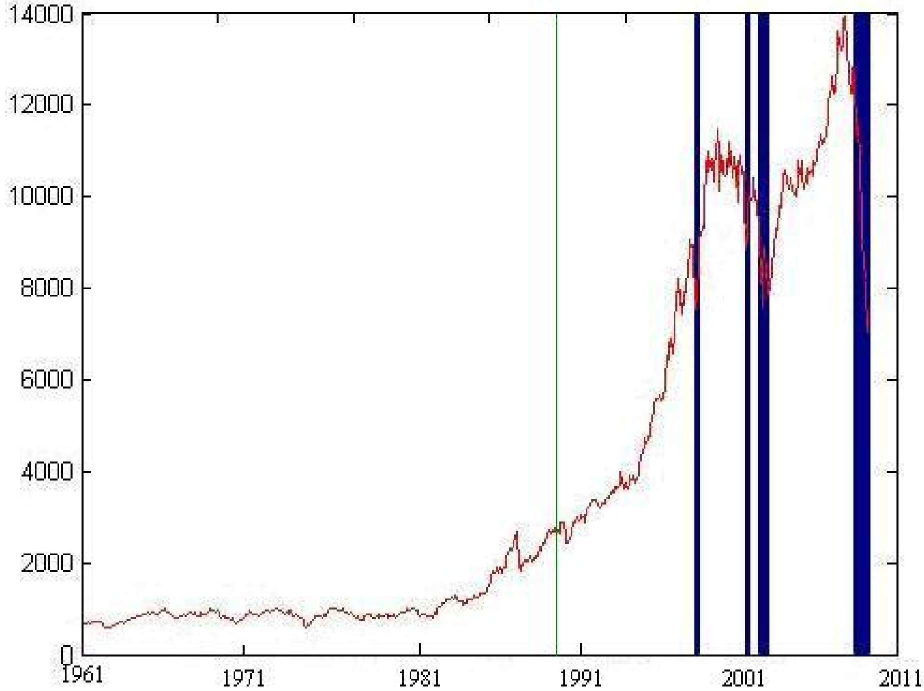



gemiddelden varianties

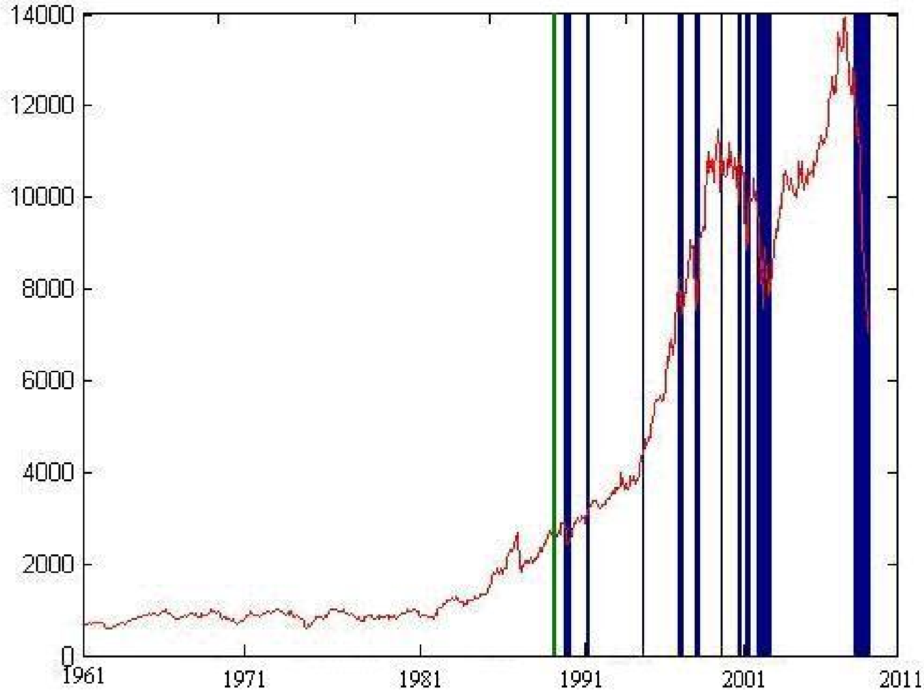
Door voor ieder tijdstip een voorspelling te maken van het regime van de volgende maand, kan het behaalde rendement bekeken worden van iedere strategie. Voorspelde bearmarkten zijn met een blauwebalk aangegeven in figuur 6 en figuur 7. De eerste, groene balk geeft aan vanaf wanneer er begonnen is met de identificatie. De voorspellingen van het model met alleen een constante lijken de grote dalingen als de krediet- en dotcomkrisis goed te pakken. Wanneer de overgangskansen afhankelijk zijn van externe variablen, worden er meer en kortere bearmarkten waargenomen.

voorspeld.

**Figuur 6:** identificatie met expanding window met constante overgangskansen




**Figuur 7:** identificatie met expanding window met variabele overgangskansen



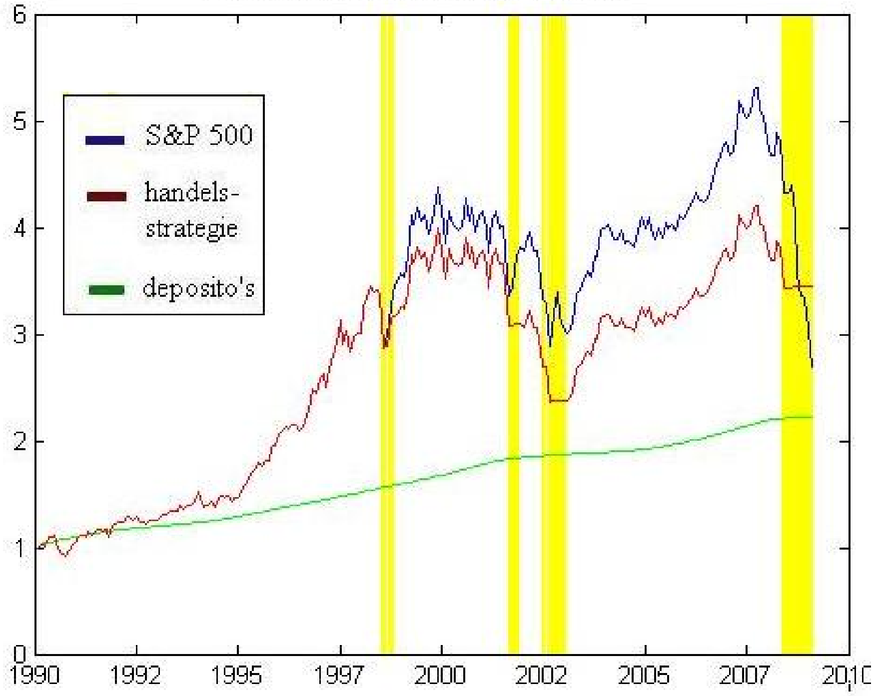
## 9 Handelsstrategieën

In figuur 8 en figuur 9 is de cummulatieve opbrengst van een handelsstrategie gebaseerd op het de voorspellingen van het model met vaste respectievelijk variabele overgangskansen. Wanneer de modellen een kans van meer dan 50% op een bullmarkt voorspellen, wordt het gehele vermogen in de marktportfolio geïnversteerd. Bij een voorspelde bearmartkt wordt het geld veilig op de bank gezet.

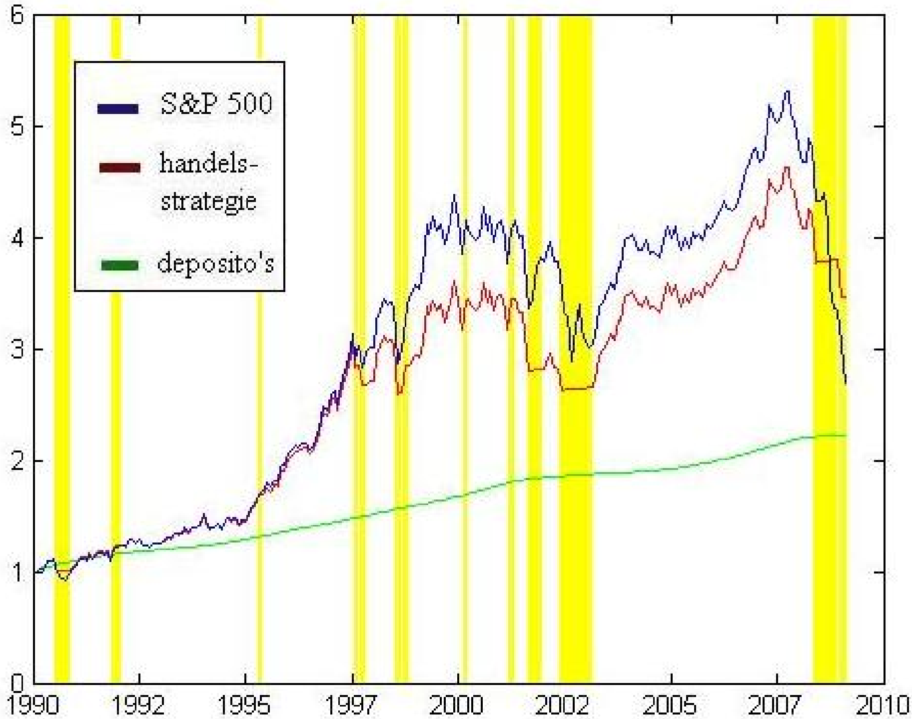
Het eindresultaat van de twee strategieën is vrij vergelijkbaar. Dat beide strategieën de markt verslaan is louter te wijten aan het tijdig uitstappen bij de kredietkrisis. Doordat het model de regimes niet alleen aan de hand van gemiddeld rendement, maar ook aan de hand van variantie beoordeelt, worden grote stijgingen snel aangezien als bearmarkt. Hierdoor wordt veel verloren en het effect van tijdig uitstappen veelal ongedaan gemaakt.

Over de gehele linie doen voorspellingen gebaseerd op variable overgangskansen het beter, maar doordat er tijdens de kredietkrisis een maand bull wordt voorspelt, drukt dit het resultaat. Ik heb bij de berekening van de opbrengst geen rekening gehouden met transactiekosten. Zouden deze er wel geweest zijn, dan zou de uitkomst van het model met constante overgangskansen waarschijnlijk beter presteren door het geringe aantal switches.

**Figuur 8:** cummulatieve opbrengst met handelsstrategie op basis van constante overgangskansen

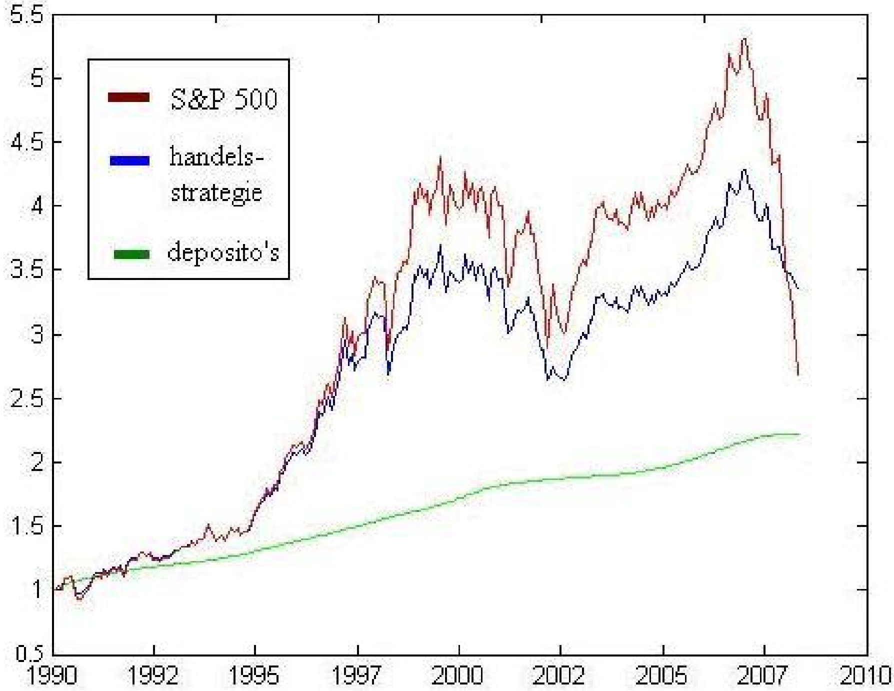



**Figuur 9:** cummulatieve opbrengst met handelstrategie op basis van variabele overgangskansen

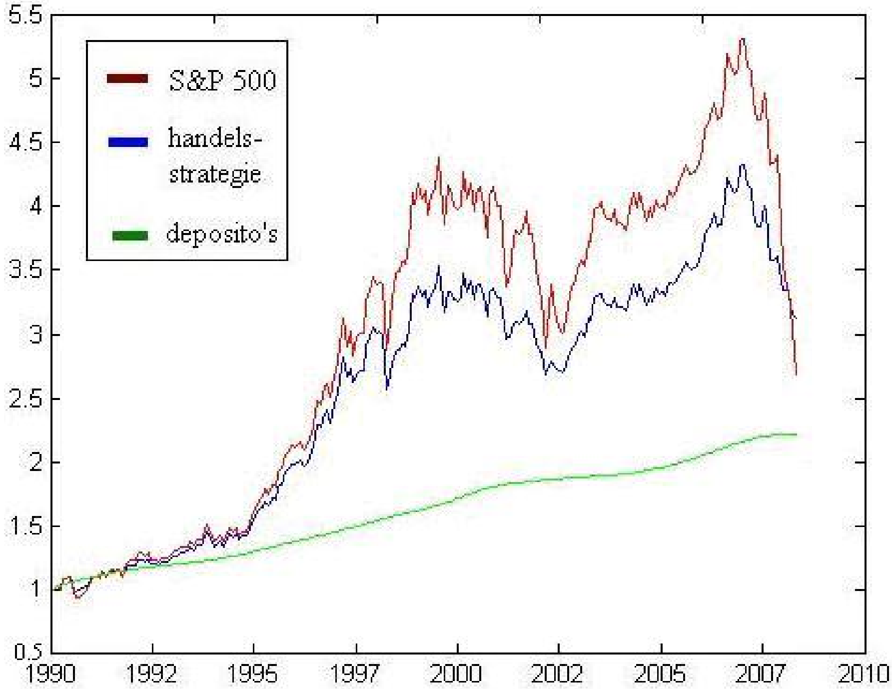


Een mooie eigenschap van het Markov switching model is dat het de kans geeft dat de markt zich in een bepaald regime bevindt. We kunnen dus ook het geïnvesteerde vermogen laten afhangen van de kans op een bullmarkt, zodat er een gemengde portfolio ontstaat. In figuur 10 en figuur 11 staan de opbrengsten van een handelstrategie die het percentage van het vermogen in de markt investeert als de voorspelde kans op een bullmarkt in de volgende periode. Wederom geldt dat beide strategieën het alleen beter doen dan de markt door de kredietkrisis. Verder doet model met de vaste overgangskansen het iets beter, maar nog steeds net iets slechter dan de eerder beschreven voorspellingen.. Het model met alleen een constante lijkt de markt iets trouwer te volgen.

**Figuur 10**: Cummulatieve opbrengst van gemengde handelsstrategie op basis van constant overgangskansen

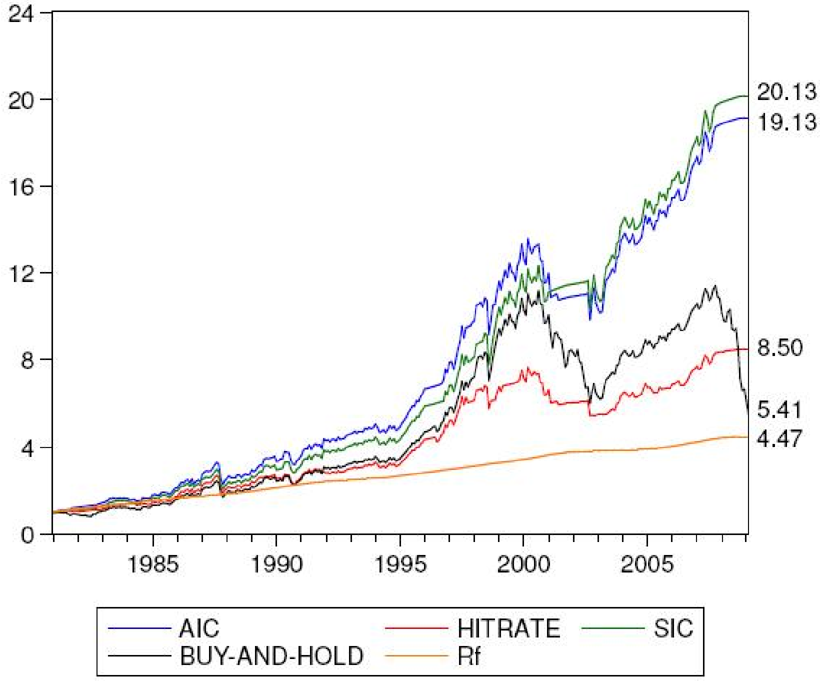


**Figuur 11**: cummulatieve opbrengst van handelsstrategie
op basis van variabele overgangskansen




Het onderzoek van Hogerwerf et al (2009). heeft over de gehele periode waarover ze hun modellen schatten een hoger opbrengst dan de markt. Al hoewel de periode waarop ze hun handelsstrategieën testen langer is dan die van mijn onderzoek, is duidelijk te zien dat door het correct voorspellen van de dotcom krisis en de kredietkrisis, de uiteindelijke opbrengst vele malen hoger is dan wat er met de marktportfolio verdiend zou zijn. Dit in tegenstelling tot de strategieën op basis van Markov switching.

**Figuur 12:** cummulatieve opbrengst handelsstrategieën
Hogerwerf et al (2009).




In tabel 4 zijn de resultaten van de verschillende strategieën weergegeven. Van de Markov sitchingmodellen doen de binaire modellen, de modellen die geheel in de markt investeren bij een voorspelde bullmarkt en geheel in deposito's bij een voorspelde bearmarkt, het beter dan de gemengde modellen. Handelen met de gemengde strategieën geeft echter wel de laagste variantie. Het logitmodel van Hogerwerf et al(2009) geeft het hoogste eindbedrag en een grote variantie, omdat de markt vaak gevolgd wordt. Als benchmark staan de marktportfolio en deposito's ook weergegeven.

| Tabel 4: Opbrengesten en varianties handelsstrategieën | | | |
|---|---|---|---|
| Methode | Eindbedrag | Maandelijks rendement | Variantie |
| constante overgangskansen binair | 3.445 | 0.54% | 0.0014 |
| Constante overgangskansen gemengd | 3.360 | 0.53% | 0.0010 |
| Variabele overgangskansen Binair | 3.458 | 0.54% | 0.0013 |
| Variabele overgangskansen Gemengd | 3.130 | 0.50% | 0.00098 |
| Hogerwerf et al. AIC | 5.386 | 0.74% | 0.0015 |
| S&P 500 | 2.688 | 0.43% | 0.0019 |
| Deposito's | 2.223 | 0.35% | 0.000 |

## 10 Conclusie

Met dit onderzoek heb ik geprobeerd voorspellingen te maken van de S&P 500 index. Dit heb ik gedaan door met Markov Switching modellen de maandelijkse rendementen in twee categoriën in te delen, elk met een eigen gemiddelde en eigen variantie. Het regimeswitching model maakt deze indeling zelf, zonder verdere definitie van bull of bear markten vooraf.

Het eerste wat ik onderzocht heb is de identificatie van regimes. Zowel het model met constante overgangskansen als het model met variabele overgangskansen gaven de belangrijkste bull en bearmarkten weer. Het model met de constante overgangskansen gaf veel korte bearmarkten, terwijl het model met de variabele overgangskansen langere aaneengeschakelde periodes als bear markeerde. Omdat Markov switching modellen niet alleen op basis van gemiddeld rendement, maar ook op basis van variantie indelen, kan een periode met veel variantie, maar een positief rendement als bear aangemerkt worden. Vergeleken met de methode van Lunde & Timmermann is

het rendement in een bearmarkt hoger en in een bullmarkt lager. Verder is de variantie iets hoger. Lunde & Timmermann geven dus een beter identificatiealgoritme.
Wanneer we door de sample heen lopen met een expanding window, levert dit een ander beeld op. De identificatie op basis van vaste overgangskansen geeft nu minder bearmarkten en alleen de heftige dalingen als de internetzeepbel en de sub-prime crisis worden gevonden. Het model met de variabele overgnagskansen vindt hier juist meer veranderingen van regime en lijkt iets beter. Wel zijn de schattingen van de gemiddelde rendementen van de regimes erg variabel.
Handelsstrategieën op basis van Markov Switchingmodellen doen het over het algemeen slechter dan de markt. Slechts door uit te stappen bij de kredietkrisis, verslaan ze de markt. Het mooie resultaat van Hogerwerf et al(2009). wordt bij lange na niet gehaald.
Markov switching modellen delen de sample niet alleen in naar gemiddeld rendement, maar vooral naar variantie. Grote plotselingen stijgingen van de markt worden hierdoor snel als bearmarkt aangemerkt. Daar komt bij dat de gemiddelde rendementen van bull- en bearmarkten erg dicht bij elkaar liggen. Hierdoor is het lastig een goed onderscheid te maken tussen de verschillende regimes. Zeker wanneer de overgangskansen variabel zijn, zijn er erg veel waarnemingen nodig om tot goede schattingen te komen, veel meer dan in een sample van 578 maanden. Schattingen worden dan te afhankelijk van individuele gebeurtenissen. Er zijn dan ook betere methoden om bull- en bearmarkten te identificeren en voorspellen. Het algoritme van Lunde & Timmerman(2004) zorgt voor heldere binaire identificatie, iets wat het Markov switching model mist.
Alles samenvattend kan ik concluderen dat de methode van Hogerwerf et al(2009). valt te preferen boven schattingen met Markov Swiching modellen, zoals door mij uitgevoerd.
Er blijft echter nog wel veel liggen voor verder onderzoek. Zo is het interessant te bekijken of het gebruiken van een andere verdeling van de rendementen, als de t-verdeling, in plaats van de normale verdeling, betere resultaten oplevert. Daarnaast zou voor ieder tijdstip de beste combinatie variabelen gebruikt kunnen worden, in plaats één combinatie over de hele sample. Ook zou frequenter schatten, bijvoorbeeld met weekdata, verbetering kunnen opleveren.

## Appendix 1

Stap (2) schatten overgangskansen

1. Bereken de contidionele dichtheden van $y_t$ en de overgangskansen gegeven $\theta^{(0)}$

2. Bereken de smoothed inference probabilities door de volgende stappen te herhalen voor $t = 2,...,T$ :

bereken de forecast $\xi_{t|t-1}$ gegeven de initiële kansen $\rho = P(s_1 = 1)$ en $1 - \rho = P(s_1 = 2)$

$$\xi_{t|t-1} = \begin{pmatrix} P(s_t = 1 \mid x_{t-1}, y_{t-1}, \theta^{(j)}) \\ P(s_t = 2 \mid x_{t-1}, y_{t-1}, \theta^{(j)}) \end{pmatrix}$$

$$\xi_{t|t-1} = P_t \xi_{t-1}$$

Waarbij $P_t$ de matrix met de overgangskansen op tijdstip *t* is met

$p_{ii} = P(s_t = i \mid s_{t-1} = i, \theta^{(j)}), i = 1,2$ :

$$P_t = \begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix}$$

Vervolgens kunnen de inference probabilities $\xi_{t|t}$ verkregen worden:

$$\xi_{t|t} = \begin{pmatrix} P(s_t = 1 \mid x_t, y_t, \theta^{(j)}) \\ P(s_t = 2 \mid x_t, y_t, \theta^{(j)}) \end{pmatrix}$$

$$\xi_{t|t} = \frac{\begin{pmatrix} f(y_t \mid s_t = 1, \theta^{(j)})\xi_{t|t-1}(1) \\ f(y_t \mid s_t = 2, \theta^{(j)})\xi_{t|t-1}(2) \end{pmatrix}}{f(y_t \mid s_t = 1, \theta^{(j)})\xi_{t|t-1}(1) + f(y_t \mid s_t = 2, \theta^{(j)})\xi_{t|t-1}(2)}$$

Door de berekende inference probabilities als input te gebruiken voor het berekenen van de nieuwe forecast, kunnen de nieuwe inference probabilities worden berekend. Nadat voor alle *t* de forecast en inference probabilities zijn berekend, kunnen de smoothed inference probabilities berekend worden.

$$\xi_{t|T} = \begin{pmatrix} P(s_t = 1 \mid \underline{x}_T, \underline{y}_T, \theta^{(j)} \\ P(s_t = 2 \mid \underline{x}_T, \underline{y}_T, \theta^{(j)} \end{pmatrix}$$

De smoothed inference probabilities voor $t = T$ zijn gelijk aan de inference probabilities voor $t = T$ . Door vanaf $t = T$-1 terug te lopen,

$P(s_t = j; s_{t-1} = i \mid \underline{x}_T, \underline{y}_T, \theta^{(j)}) = \xi_{t|t}(i) \cdot \xi_{t+1|T}(j) / \xi_{t+1|t}(j) \cdot p_{ij}$

waarna geldt:

$P(s_t = j \mid \underline{x}_T, \underline{y}_T, \theta^{(j)}) = P(s_t = j; s_{t-1} = 1 \mid \underline{x}_T, \underline{y}_T, \theta^{(j)}) + P(s_t = j; s_{t-1} = 2 \mid \underline{x}_T, \underline{y}_T, \theta^{(j)})$

kunnen alle smoothed inference probabilities berekend worden.

Stap (3) maximalisatie

Zoals beschreven in Dieboldet al.(1994) geeft dit de volgende maximum likelihood schatters:

$$\mu_i = \frac{\sum_{t=1}^{T} y_t P(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)})}{\sum_{t=1}^{T} P(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)})}$$

$$\sigma_i^2 = \frac{\sum_{t=1}^{T} (y_t - \mu_i^{(j)})^2 P(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)})}{\sum_{t=1}^{T} P(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)})}$$

$$\rho = P(s_1 = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)})$$

Voor het schatten van de parameters die invloed hebben op de overgangskansen moeten eerst de overgangskansen geschat worden:

Vervolgens volgen de parameters van de overgangskansen uit de optimalisatie van de volgende log-likelihood functie:

$$\sum \left\{ \begin{array}{l} P(s_t = 1; s_{t-1} = 1 \mid \underline{x}_T, \underline{y}_T, \theta^{(j)}) \log(p_{11}^{(t)}) \\ + P(s_t = 2; s_{t-1} = 1 \mid \underline{x}_T, \underline{y}_T, \theta^{(j)}) \log(1 - p_{11}^{(t)}) \\ + P(s_t = 1; s_{t-1} = 2 \mid \underline{x}_T, \underline{y}_T, \theta^{(j)}) \log(1 - p_{22}^{(t)}) \\ + P(s_t = 2; s_{t-1} = 2 \mid \underline{x}_T, \underline{y}_T, \theta^{(j)}) \log(p_{22}^{(t)}) \end{array} \right\}$$

## Referenties


F.X. Diebold J.-H. Lee & G.C. Weinbach (1994). Regime switching with time-varying transitionprobabilities. 283-303

P.H. Frances & D. van Dijk (1999). *Nonlinear Time Series Models in Empirical Finance.*

J.D. Hamilton (1989). A new approach to the economic analysis of non-stationary time series subject to changes in regime. *Econometrica*, 57: 357-384.

J.D. Hamilton (2003). Comment on “Comparison of Two Business Cycle Dating Methods”. *Journal of Economic Dynamics and Control*, 27(9):1691-1693

D. Harding & A.R. Pagan (2002). A Comparison of Two Business Cycle Dating Methods. *Journal of Economic Dynamics and Control*, 27(9):1681-1690.

D. Harding & A.R. Pagan (2003). Rejoinder to James Hamilton. *Journal of Economic Dynamics and Control*, 27(9):1695-1698.

J. Hogerwerf, J. van Opdurp, T. Ovaa & V.M. van Rooijen (2009). Chasing Bulls, Fleeing Bears.

H.J.W.G. Kole (2009). Updating transition probabilities in a Markov regime switching model.

A. Lunde & K.A. Timmermann (2004). Duration dependence in stock prices: An analysis of bull and bear markets. *Journal of Applied Econometrics,* 18(1):23-46.

J.M. Mahue & T.H. McCurdy (2000). Identifying Bull and Bear Markets in Stock Returns. *Journal of Business & Economic Statistics*, 18:100-112.