Value New Business van Separate Accounts: Tijdreeksanalyse voor Vermogensbeheer

Alex Roubos

Stageverslag

Value New Business van Separate Accounts: Tijdreeksanalyse voor Vermogensbeheer

Alex Roubos

Stageverslag

Vrije Universiteit Amsterdam
Faculteit der Exacte Wetenschappen
Studierichting Bedrijfswiskunde en Informatica
De Boelelaan 1081a
1081 HV Amsterdam

Stagebedrijf:
Zwitserleven
Burgemeester Rijnderslaan 7
1185 MD Amstelveen

26 juni 2008

Voorwoord

Met dit stageverslag sluit ik de studie Bedrijfswiskunde en Informatica aan de Vrije Universiteit te Amsterdam af. Bedrijfswiskunde en Informatica is een multidisciplinaire studie, gericht op het verbeteren van bedrijfsprocessen door het toepassen van methoden uit de wiskunde, informatica en bedrijfskunde. De stage heeft plaatsgevonden bij Zwitserleven te Amstelveen. De drie disciplines van de studie zijn goed tot hun recht gekomen in de stageopdracht en het eindproduct draagt bij tot een verbetering van het bedrijfsproces.

Ik wil graag een aantal mensen bedanken voor hun hulp en ondersteuning gedurende mijn stage: Thomas Trompert en Chantal de Veen voor de begeleiding vanuit Zwitserleven; Marianne Jonker en René Bekker voor de begeleiding vanuit de universiteit; Ineke Hegeman en Ronald Gort voor het voorzien van data; Steve Djadoenath, Tonny Verbaken, Kees Smit, Frank van de Weteringe Buijs, Harmen Hessels, Gilbert Wijgergangs, Ambika Gangapersadsing, Patrick Tijmstra, Gerben Voogd, Jitske Meijering, Rianne Lurvink, Alex Punt, Raymond Roussou en niet te vergeten mijn collega's van de afdeling Products Edward Devilee, Jan Jochems, Wim van Kouteren, Floris Prakke, Rob Quant, Tom Veerman en Marianne Zonneveld voor hun hulp anderszins, de leuke gesprekken en de gezellige tijd.

Alex Roubos
Aalsmeer, 2008

Samenvatting

VERTROUWELIJK

Inhoudsopgave

Hoofdstuk 1

Inleiding

In dit inleidende hoofdstuk wordt de probleemstelling uiteengezet. Dit is een korte beschrijving van het probleem waarnaar onderzoek wordt verricht. Vervolgens wordt de probleemstelling in het kader van Zwitserleven geplaatst en wordt Zwitserleven geïntroduceerd, tezamen met overige relevante achtergrondinformatie. Ten slotte wordt in dit hoofdstuk de structuur van het verslag beschreven.

1.1 Probleemstelling

De Value New Business (VNB) van een contract is de contante waarde van de positieve of negatieve resultaten gedurende de contractperiode. Het is van belang dat een goede indicatie van deze VNB bekend is vóór het afsluiten van een contract om zodoende de juiste beslissingen te kunnen nemen. Het probleem in de markt van de Separate Accounts (SA's) is dat elk contract op maat gemaakt is en dat over elk component in zo'n contract, bijvoorbeeld de administratiekostenvergoeding, onderhandeld kan worden. Het is echter onduidelijk in welke mate ieder component de VNB beïnvloedt. In deze stage staan daarom de volgende onderzoeksvragen centraal.

- Welke componenten zijn van invloed op de VNB?
- Wat is de invloed van deze componenten op de VNB?

1.2 Doelstelling

In dit kader worden alleen de componenten behandeld die in het commerciële traject te beïnvloeden zijn. Voor de betreffende componenten ontvangt Zwitserleven een vergoeding van de klant. De kosten die Zwitserleven zelf maakt worden de normkosten genoemd. Als de vergoeding lager ligt dan de normprijs van een bepaald component wordt er op deze component verlies gemaakt. Zwitserleven is best bereid om componenten onder de normprijs te verkopen, mits de totale winstverwachting maar positief is. Het ontbreekt Zwitserleven echter aan een snel inzetbare tool om de invloed van de hoogte van de vergoeding van een component op de VNB te bepalen. Zodoende is Zwitserleven soms onvoldoende in staat om commerciële beslissingen bedrijfseconomisch te onderbouwen.

Het resultaat van de stage is tweeledig. Ten eerste wordt er een wiskundig model gemaakt waaruit de invloed van de componenten duidelijk blijkt. Ten tweede wordt er op basis van dit model een tool gemaakt die makkelijk in gebruik is. Op dit moment zijn er geen geschikte tools voorhanden die een beslissing om al dan niet een contract af te sluiten ondersteunen. De tool die gedurende deze stage ontwikkeld wordt, levert daarom een waardevolle bijdrage aan het beslissingstraject en wordt zeker gebruikt om commerciële beslissingen te ondersteunen. Verder kan de tool gebruikt worden om de minimale vergoedingen van de componenten, die momenteel gehanteerd worden, te toetsen op zinvolheid.

1.3 Zwitserleven

Zwitserleven is de Nederlandse vestiging van het Swiss Life Concern, een internationale pensioenverzekeraar. De levensverzekeraar, gevestigd in Amstelveen, bestaat in Nederland sinds 1901 en is de grootste buitenlandse levensverzekeraar. Het concern beschikt over een wereldwijde netwerkstructuur waardoor Swiss Life, en dus ook Zwitserleven, internationale ondernemingen kan ondersteunen. Sinds begin 2008 is bekend dat Zwitserleven wordt overgenomen door SNS Reaal. De verwachting is dat medio 2008 de overname rond is. In 2007 realiseerde Zwitserleven een premie-inkomen van €1,4 miljard en een nettowinst van €92 miljoen. Daarmee behoort Zwitserleven ook in Nederland tot de top tien van pensioenverzekeraars [9].

Zwitserleven heeft ongeveer 795 medewerkers in dienst, georganiseerd in vier divisies en ondersteund door een aantal stafafdelingen. Deze divisies zijn: Marketing & Sales, Customers Services, Technologie & Implementatie en Financiën. Een van de ondersteunende stafafdelingen is Products. Products is als afdeling verantwoordelijk voor de ontwikkeling van nieuwe producten en productonderdelen. De verantwoordelijkheid van Products behelst het gehele traject van uitdenken tot aan de implementatie en het monitoren van het voortbrengingsproces.

1.4 Pensioenen

Het Nederlandse stelsel van toekomstvoorzieningen is opgebouwd uit drie pijlers. Onder de eerste pijler vallen de sociale zekerheidsregelingen, zoals de Algemene Ouderdomswet (AOW), de Algemene Nabestaandenwet (ANW) en de Wet Werk en Inkomen naar Arbeidsvermogen (WIA). Deze wetten vormen de basispensioenvoorzieningen voor alle inwoners van Nederland. Het onderscheid met de tweede en derde pijler is voornamelijk dat de AOW en ANW uit deze eerste pijler een collectief en wettelijk karakter hebben. De AOW is op omslagbasis gefinancierd, dat wil zeggen dat de werkende generatie via premiebetalingen de voorziening voor de op dat moment AOW-ontvangende generatie moet opbrengen. De AOW is een opbouwverzekering voor alle in Nederland woonachtigen; de opbouw bedraagt 2% van de uitkering per verzekerd jaar.

De pensioenregelingen die vallen onder de tweede pijler zijn een aanvulling op de wetten uit de eerste pijler. Het gaat hier om pensioenregelingen van ondernemingen, bedrijfstakken of beroepsgroepen. Dit zijn de toekomstvoorzieningen die door werkgevers aan werknemers zijn toegezegd. Deze vloeien voort uit de arbeidsverhouding en komen tot stand in een individuele arbeidsovereenkomst of worden collectief geregeld via onderhandelingen en contracten tussen werkgeversorganisaties en werknemersorganisaties. Dit kan zowel op ondernemersniveau plaatsvinden als op bedrijfstakniveau. In het laatste geval vloeit de deelname doorgaans voort uit de Wet betreffende verplichte deelneming in een bedrijfstakpensioenfonds 2000 (Wet Bpf 2000). Deelname aan zo'n regeling geldt voor alle personeelsleden of voor een bepaalde groep van werknemers binnen de onderneming.

Voorzieningen uit de derde pijler zijn de aanvullende voorzieningen (privéverzekeringen) die een inkomensbron vormen voor de verzekerde wanneer deze met pensioen is, of voor de nabestaanden als de verzekerde komt te overlijden. Onder deze aanvullende voorzieningen vallen de bij een levensverzekeringsmaatschappij afgesloten contracten voor lijfrenten, levensverzekeringen en spaarregelingen voor de oude dag. Dit zijn regelingen die niet voortvloeien uit een arbeidsverhouding. Deze voorzieningen zorgen voor een aanvulling op de eerste pijler wanneer een collectieve voorziening ontbreekt. Daarnaast kunnen ze ook een aanvulling zijn op de eerste en de tweede pijler.

Een collectieve pensioenregeling vanuit de werkgever valt derhalve in de tweede pijler. Binnen zo'n pensioenregeling kunnen verschillende pensioenvormen onderscheiden worden:

- ouderdomspensioen;
- nabestaandenpensioen;
- nabestaandenoverbruggingspensioen (ANW-hiaat);
- wezenpensioen;
- prepensioen;
- (tijdelijk) overbruggingspensioen;
- arbeidsongeschiktheidspensioen (AOP);
- WGA-hiaat;
- vrijstelling van premiebetaling wegens arbeidsongeschiktheid.

De collectieve regelingen die beschouwd worden, zijn in bijna alle gevallen salaris-/diensttijdregelingen waarbij een aantal van deze pensioenvormen wordt verzekerd. Een salaris-/diensttijdregeling is een regeling waarbij de opbouw afhankelijk is van het salaris en de (totale) diensttijd.

Elk jaar wordt er een percentage van de pensioengrondslag (het salaris minus de AOW van de overheid) ingekocht. Dit percentage is vaak 1,75% waardoor er in het gunstigste scenario na 40 dienstjaren 70% ($= 40 \cdot 1{,}75\%$) van de pensioengrondslag is ingekocht. Samen met de AOW uit de eerste pijler en een eventueel pensioen in de lijfrentesfeer (de derde pijler) geeft dit het totale inkomen na pensionering.

Het soort regeling en de te verzekeren pensioenvormen worden door de werkgever vastgelegd in een pensioenreglement, een onderdeel van de arbeidsvoorwaarden. De uitvoering van dit reglement vindt plaats in een ondernemings- of bedrijfstakpensioenfonds, of het reglement wordt rechtstreeks door een levensverzekeraar uitgevoerd. Indien de uitvoering rechtstreeks plaatsvindt door een verzekeraar wordt dit in een contract vastgelegd (met een duur van vijf of tien jaar) tussen de werkgever en de uitvoerder, in dit geval Zwitserleven.

VERTROUWELIJK

1.5 Een separate account

VERTROUWELIJK

1.6 Structuur van het verslag

Het vervolg van het verslag is op de volgende manier opgebouwd. Hoofdstuk 2 behandelt de vraag hoe de VNB berekend wordt. Hierin worden de componenten besproken die invloed hebben op de VNB. In hoofdstuk 3 wordt vervolgens van de administratiekostencomponent in detail uitgelegd hoe deze de VNB beïnvloedt. Hoofdstuk 4 behandelt de theorie die nodig is voor het modelleren van het vermogensbeheercomponent. In hoofdstuk 5 wordt deze theorie toegepast om tot een wiskundig model van het vermogensbeheer te komen. Hoofdstuk 6 staat in het teken van de solvabiliteitsopslag. Dit hoofdstuk beschrijft het effect van de solvabiliteitsopslag op de VNB. In hoofdstuk 7 wordt de invloed van ieder component op de VNB besproken aan de hand van een voorbeeld en de ontwikkelde tool. Hoofdstuk 8 ten slotte geeft de conclusies en aanbevelingen weer naar aanleiding van de resultaten voortkomend uit deze stage.

Hoofdstuk 2

Value New Business

VERTROUWELIJK

Hoofdstuk 3

Administratiekosten

VERTROUWELIJK

Hoofdstuk 4

Tijdreeksanalyse

Dit hoofdstuk staat in het teken van de theorie omtrent tijdreeksanalyse. Een tijdreeks is een verzameling observaties van een bepaalde variabele in de tijd. Een van de doelen van het modelleren van een tijdreeks is om toekomstige waarden te voorspellen op basis van de geobserveerde waarden. Een karakteristieke eigenschap van een tijdreeks is dat de observaties niet onafhankelijk zijn: de spreiding varieert in de tijd, ze vertonen trends en ze bevatten vaak cyclische componenten. Statistische methoden die uitgaan van onafhankelijke en gelijk verdeelde data kunnen daarom niet worden gebruikt voor het analyseren van tijdreeksen. Hiervoor zijn geschikte methoden noodzakelijk die onder de noemer *tijdreeksanalyse* vallen. De theorie die in dit hoofdstuk wordt besproken, is voornamelijk gebaseerd op [5], [6], [7] en [8].

4.1 Het additieve model

Het klassieke additieve model gaat uit van de aanname dat voor een gegeven tijdreeks $y_1, \ldots, y_n$ de data realisaties zijn van de stochastische variabelen Y_t, die op hun beurt weer zijn opgebouwd uit vier componenten

$$Y_t = T_t + C_t + S_t + R_t, \quad t = 1, \ldots, n. \tag{4.1}$$

Hierin is T_t een (deterministische) functie van t, de *trend* genoemd. C_t is een (deterministische) lange-termijn cyclische component, waarvan de periode groter is dan de tijdspanne van de observaties. Te denken valt aan de conjunctuur met perioden van relatief snelle groei en perioden van stagnatie of daling. S_t beschrijft een (deterministische) korte-termijn cyclische component, zoals een *seizoencomponent*, met een bekende periode d. De stochastische variabele R_t bevat daarentegen alle afwijkingen van het niet-stochastische model $T_t + C_t + S_t$. De variabelen T_t en C_t kunnen bijeen genomen worden als

$$M_t = T_t + C_t, \tag{4.2}$$

dat het lange-termijn gedrag van de tijdreeks beschrijft. In het vervolg wordt met de term trend de functie M_t aangeduid. Er wordt aangenomen dat de verwachting $\mathbb{E}R_t$ van de afwijkingen bestaat en gelijk is aan nul, zodat afwijkingen boven of onder het niet-stochastische model elkaar gemiddeld genomen opheffen. De eigenschap $\mathbb{E}R_t = 0$ kan altijd worden gerealiseerd door een van de niet-stochastische componenten op een geschikte manier aan te passen.

De decompositie van een tijdreeks in een trend, een seizoencomponent en het resterende stochastische proces heeft veel aandacht in de literatuur gekregen. De wens om in eerste instantie de deterministische trend en seizoencomponent uit de tijdreeks te verwijderen, komt voort uit de noodzaak om met een *stationaire* tijdreeks te werken. Kort gezegd, stationariteit betekent dat de (stochastische) eigenschappen van een tijdreeks niet veranderen in de tijd. Stationariteit is vereist om één model op de gehele steekproef te kunnen bepalen in plaats van verschillende modellen voor elke observatie, hetgeen duidelijk onmogelijk is. Er bestaan verschillende methoden om de trend en seizoencomponent uit de tijdreeks te verwijderen.

4.1.1 Trend en seizoencomponent

Het doel is om de deterministische componenten M_t en S_t te schatten, zodat na verwijdering van Y_t een stationair proces $\{\hat{R}_t\}$ overblijft. Voor dit proces kan er dan een geschikt model worden gevonden om samen met $\{\hat{M}_t\}$ en $\{\hat{S}_t\}$ toekomstige waarden van $\{Y_t\}$ te voorspellen. De methoden die worden besproken gaan uit van het generieke additieve model

$$Y_t = M_t + S_t + R_t, \tag{4.3}$$

met als extra aannamen dat de invloed van de seizoencomponent over de perioden gelijk is

$$S_{t+d} = S_t, \tag{4.4}$$

en dat

$$\sum_{k=1}^{d} S_k = 0. \tag{4.5}$$

Small trend methode

Voor deze methode is het handig om de data te labelen naar het nummer van het seizoen en het nummer binnen het seizoen. Bijvoorbeeld, voor maandelijkse data ($d = 12$) voor b jaar worden de data aangeduid door

$$Y_{j,k}, \quad j = 1, \ldots, b, \; k = 1, \ldots, 12. \tag{4.6}$$

Deze methode is geschikt wanneer de tijdreeks een kleine trend heeft, waardoor het niet onredelijk is om aan te nemen dat de trend in elke periode constant is. Dan is het gemiddelde over een periode een zuivere schatter voor de trend. Oftewel,

$$\hat{M}_j = \frac{1}{d}\sum_{k=1}^{d} Y_{j,k}, \quad j = 1, \ldots, b. \tag{4.7}$$

De schatter voor de seizoencomponent wordt dan gegeven door

$$\hat{S}_k = \frac{1}{b}\sum_{j=1}^{b}(Y_{j,k} - \hat{M}_j), \quad k = 1, \ldots, d. \tag{4.8}$$

Door vervolgens de schatters voor de trend en seizoencomponent van de tijdreeks te verwijderen, kunnen de residuen worden verkregen

$$\hat{R}_{j,k} = Y_{j,k} - \hat{M}_j - \hat{S}_k, \quad j = 1, \ldots, b,\ k = 1, \ldots, d. \tag{4.9}$$

Moving average

In tegenstelling tot de vorige methode maakt deze methode niet de aanname dat de trend in elke periode constant is. Om deze reden is deze methode te prefereren. De data $Y_1, \ldots, Y_n$ zijn beschikbaar. De volgende stappen worden dan achtereenvolgens uitgevoerd.

- Schat de trend door het toepassen van een moving average filter ter lengte van de periode d. Op deze manier is uit de data de seizoencomponent verwijderd en zijn de afwijkingen gedempt (aangezien $\mathbb{E}R_t = 0$). De schatter voor de trend is

$$\hat{M}_t = \begin{cases} \frac{1}{d}\left(\frac{1}{2}Y_{t-q} + \sum_{i=-q+1}^{q-1} Y_{t+i} + \frac{1}{2}Y_{t+q}\right), & d = 2q,\ q < t \le n-q, \\ \frac{1}{d}\sum_{i=-q}^{q} Y_{t+i}, & d = 2q+1,\ q < t \le n-q. \end{cases} \tag{4.10}$$

- De tweede stap is om de seizoencomponent te schatten. Voor elke $k = 1, \ldots, d$ wordt het gemiddelde V_k over de waarden $\{(Y_{k+jd} - \hat{M}_{k+jd})\}, q < k+jd \le n-q, j = 1, \ldots, b$ genomen. De schatter voor de seizoencomponent S_k is dan

$$\hat{S}_k = \begin{cases} V_k - \frac{1}{d}\sum_{i=1}^{d} V_i, & k = 1, \ldots, d, \\ \hat{S}_{k-d}, & k > d. \end{cases} \tag{4.11}$$

- Verwijder de invloed van de seizoencomponent zodat de volgende data verkregen worden

$$D_t = Y_t - \hat{S}_t, \quad t = 1, \ldots, n. \tag{4.12}$$

- Herschat de trend $\hat{M}_t$ van de data $\{D_t\}$, om de reden dat dit een betere schatting oplevert. Hiervoor kunnen ook weer een aantal methoden gebruikt worden, waarvan het schatten van een functie met bijvoorbeeld de kleinste-kwadratenmethode het meest geschikt is voor het doel om toekomstige waarden te voorspellen. In deze procedure wordt er een geparametriseerde familie van functies $M_t = f(a,t)$ gefit op de data door het kiezen van de parameter(vector) a zo, dat $a \mapsto \sum_{t=1}^{n}(D_t - f(a,t))^2$ wordt geminimaliseerd naar a. Het resultaat is de schatting $\hat{M}_t = f(\hat{a},t)$.
- Bereken de residuen

$$\hat{R}_t = Y_t - \hat{M}_t - \hat{S}_t. \tag{4.13}$$

Differencing met lag d

Deze methode haalt de seizoencomponent uit de data door het kijken naar de verschillen met een *lag* (vertraging) van d, de periode van de seizoencomponent. De operator ∇_d is gedefinieerd als

$$\nabla_d Y_t = Y_t - Y_{t-d}, \quad t > d. \tag{4.14}$$

Als deze operator wordt toegepast op het generieke model (4.3), wordt het volgende verkregen

$$\begin{aligned} \nabla_d Y_t &= (M_t + S_t + R_t) - (M_{t-d} + S_{t-d} + R_{t-d}) \\ &= M_t - M_{t-d} + R_t - R_{t-d}. \end{aligned} \tag{4.15}$$

Dit geeft een decompositie in een trend $(M_t - M_{t-d})$ en een stationair proces $(R_t - R_{t-d})$. De trend kan vervolgens geschat worden door bijvoorbeeld de kleinste-kwadratenmethode toe te passen.

4.1.2 Modellen voor stationaire tijdreeksen

Als de trend en seizoencomponent uit de tijdreeks zijn verwijderd, blijft een stationair proces over. Dit is ook weer een tijdreeks van de vorm $\{X_t\}$. De formele definitie van een stationair proces is als volgt.

Definitie 4.1. *Een tijdreeks is (tweede orde) stationair als, voor alle waarden van* k, $\mathbb{E}X_t$ *en* $\mathbb{E}X_tX_{t+k}$ *bestaan en onafhankelijk zijn van* t.

Een tijdreeks die nog een trend of seizoencomponent bevat, voldoet niet aan deze definitie omdat $\mathbb{E}X_t$ niet onafhankelijk van t is. Vandaar dat de trend en seizoencomponent eerst moeten worden verwijderd. De volgende definities komen van pas bij het modelleren van stationaire tijdreeksen.

De meeste tijdreeksen vertonen enige vorm van afhankelijkheid tussen (X_s, X_t). De *autocovariantiefunctie* beschrijft dit concept van afhankelijkheid.

Definitie 4.2. *Als $\{X_t\}$ een proces is met eindige variantie, dan is de autocovariantiefunctie $\gamma_X(\cdot,\cdot)$ gedefinieerd als*

$$\gamma_X(s,t) = \mathrm{Cov}(X_s, X_t) = \mathbb{E}\left[(X_s - \mathbb{E}X_s)(X_t - \mathbb{E}X_t)\right]. \tag{4.16}$$

Voor een stationair proces voldoet de autocovariantiefunctie aan de eigenschap $\gamma_X(s,t) = \gamma_X(s-t, 0)$ voor alle s en t. Om deze reden wordt de autocovariantiefunctie voor een stationair proces geherdefinieerd als een functie van één variabele,

$$\gamma_X(h) = \gamma_X(h, 0) = \mathrm{Cov}(X_{t+h}, X_t), \quad h \in \mathbb{Z}, \tag{4.17}$$

welke onafhankelijk is van t. De *autocorrelatiefunctie* (ACF) van $\{X_t\}$ is een geschaalde versie van de autocovariantiefunctie en is als volgt gedefinieerd.

Definitie 4.3. *De autocorrelatiefunctie voor lag h van een stationair proces $\{X_t\}$ is*

$$\rho_X(h) = \mathrm{Cor}(X_{t+h}, X_t) = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad h \in \mathbb{Z}, \tag{4.18}$$

onafhankelijk van t.

Een laatste nuttige hulpmiddel is de *partiële autocorrelatiefunctie* (PACF). Daar waar de ACF een maat is voor de correlatie tussen X_t en X_{t+h}, is de PACF een maat voor de correlatie tussen X_t en X_{t+h} gecorrigeerd voor de correlatie van de tussenliggende observaties $X_{t+1}, \ldots, X_{t+h-1}$. De formele definitie van de PACF wordt niet gegeven. Zowel de ACF als de PACF worden gebruikt om tussen de modelvarianten te kiezen.

MA processen

Een veel gebruikt model voor stationaire processen is het zogenoemde *Moving Average* (MA) proces. Voor de definitie van een MA proces is eerst het concept van *witte ruis* noodzakelijk. Witte ruis is een reeks van stochastische variabelen die onafhankelijk en gelijk verdeeld zijn met verwachting nul en eindige variantie σ^2. Laat $\{Z_t\}$ een witte ruis proces zijn, laat q een positief getal zijn en laat $\beta_0, \beta_1, \ldots, \beta_q$ constanten zijn. Dan is het proces $\{X_t\}$ gedefinieerd door

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \tag{4.19}$$

een MA(q) proces. Een speciale eigenschap van een MA(q) proces is dat de ACF een duidelijke scheiding van significante waarden laat zien bij een lag van q. MA processen worden in veel gebieden gebruikt, waaronder in de econometrie. Bijvoorbeeld, economische variabelen worden door een verscheidenheid van gebeurtenissen beïnvloed, zoals stakingen en regeringsbeslissingen. Zulke gebeurtenissen hebben niet alleen een onmiddellijk effect, maar ook in mindere mate op economische variabelen in latere periodes. Een MA proces kan hiervoor geschikt zijn.

AR processen

Een ander type model is het zogenoemde *Autoregressive* (AR) proces. Laat opnieuw $\{Z_t\}$ een witte ruis proces zijn, laat p een positief getal zijn en laat $\alpha_1, \ldots, \alpha_p$ constanten zijn. Dan is het proces $\{X_t\}$ gedefinieerd door

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t \tag{4.20}$$

een AR(p) proces. Een AR proces kan worden beschouwd als een meervoudige regressiemodel. Alleen is X_t dan niet geregresseerd op onafhankelijk variabelen, maar op zijn eigen voorafgaande waarden. Een AR proces kan in theorie worden geschreven als een MA proces waarvoor de orde q oneindig is. De ACF laat om deze reden geen duidelijke scheiding, maar juist een geleidelijke afname zien. Daarentegen laat de PACF wel een scheiding zien bij een lag van p. AR processen worden gebruikt voor situaties waarbij het redelijk is om aan te nemen dat de huidige waarde van een tijdreeks afhankelijk is van de voorafgaande waarden, samen met een willekeurige afwijking.

ARMA processen

Het zogenoemde *Autoregressive Moving Average* (ARMA) proces is een combinatie van zowel een AR als een MA proces. Een ARMA(p, q) proces bevat p AR termen en q MA termen. Het wordt gegeven door

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \tag{4.21}$$

met $\{X_t\}$ een stationair proces. ARMA processen zijn belangrijk omdat een stationaire tijdreeks vaak door een ARMA model met minder parameters kan worden beschreven dan wanneer een zuivere MA of AR proces zou worden gebruikt.

ARIMA processen

Een model waarbij de data niet stationair verondersteld worden, is het zogenoemde *Autoregressive Integrated Moving Average* (ARIMA) proces. In dit model worden de niet-stationaire data stationair gemaakt door het toepassen van een soort differencing methode. De difference operator

$$\nabla X_t = X_t - X_{t-1} \tag{4.22}$$

wordt meerdere malen achtereenvolgens toegepast, zeg d keer, zodat een reeks $\{\nabla^d X_t\}$ is gevonden die gemodelleerd kan worden als een stationair proces. In het ARMA model (4.21) wordt de variabele X_t vervangen door $\nabla^d X_t$. Dit wordt dan een ARIMA(p, d, q) proces genoemd. ARIMA modellen worden vaak gebruikt om economische tijdreeksen te beschrijven.

4.1.3 Exponential smoothing

Een methode met een compleet andere filosofie om de toekomstige waarden van een tijdreeks te voorspellen, is door het nemen van een gewogen som van de voorafgaande observaties

$$\hat{X}_{t+1} = \lambda_0 X_t + \lambda_1 X_{t-1} + \cdots . \tag{4.23}$$

Het ligt voor de hand om de meest recente observaties zwaarder te wegen dan de oudere observaties. Een mogelijkheid om dit te realiseren is om de gewichten op de volgende geometrische manier te kiezen

$$\lambda_i = \alpha(1-\alpha)^i, \quad 0 < \alpha < 1. \tag{4.24}$$

De term 'exponential' komt vanwege het feit dat de gewichten exponentieel verminderen. In deze basisvorm kan exponential smoothing alleen worden gebruikt voor stationaire tijdreeksen. Een generalisatie is de *Holt-Winters* methode, die om kan gaan met tijdreeksen die nog een trend of seizoencomponent bevat. Hiervoor zijn drie parameters nodig: $\alpha \in [0,1]$ voor het niveau, $\beta \in [0,1]$ voor de trend en $\gamma \in [0,1]$ voor de seizoencomponent. De Holt-Winters methode bestaat in twee vormen, afhankelijk van hoe de seizoencomponent wordt gemodelleerd.

- *Additieve vorm.* In de additieve vorm worden de toekomstige waarden als volgt voorspeld

  $$\hat{X}_{t+h} = a_t + h \cdot b_t + c_{t-d+h}, \tag{4.25}$$

  met

  $$\begin{aligned} a_t &= \alpha(X_t - c_{t-d}) + (1-\alpha)(a_{t-1} + b_{t-1}), \\ b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}, \\ c_t &= \gamma(X_t - a_t) + (1-\gamma)c_{t-d}. \end{aligned} \tag{4.26}$$

- *Multiplicatieve vorm.* In de multiplicatieve vorm worden de toekomstige waarden als volgt voorspeld

  $$\hat{X}_{t+h} = (a_t + h \cdot b_t)c_{t-d+h}, \tag{4.27}$$

  met

  $$\begin{aligned} a_t &= \alpha\frac{X_t}{c_{t-d}} + (1-\alpha)(a_{t-1} + b_{t-1}), \\ b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}, \\ c_t &= \gamma\frac{X_t}{a_t} + (1-\gamma)c_{t-d}. \end{aligned} \tag{4.28}$$

In deze vergelijkingen heeft α de betekenis van de globale smoothing parameter. Een grote waarde van α geeft meer gewicht aan observaties in het recente verleden, terwijl een kleine waarde juist meer gewicht geeft aan observaties verder in het verleden. De parameter β duidt het effect van de trend aan. Een grote waarde van β geeft meer gewicht aan het verschil tussen de laatste paar gladgemaakte observaties, terwijl een kleine waarde meer informatie uit het verleden gebruikt. De parameter γ heeft betrekking op de seizoencomponent. Een grote waarde geeft meer gewicht aan de huidige relatie tussen de observatie en gladgemaakte observatie, terwijl een kleine waarde meer gewicht geeft aan deze relatie in het verleden. De a_t, b_t en c_t hebben de bijbehorende betekenis van gladgemaakte observatie, trend en seizoencomponent op tijd t. Vereiste gegevens hierbij zijn initiële waarden. Op $t = 1$ bijvoorbeeld zijn a_0, b_0 en c_{1-d} nodig die nog niet zijn geïnitialiseerd. Als startwaarden kunnen hiervoor zekere gemiddelden genomen worden. De parameters α, β en γ kunnen geschat worden door het toepassen van de kleinste-kwadratenmethode.

Als de parameters eenmaal zijn geschat, worden de voorspellingen via het deterministische model verkregen. Om een betrouwbaarheid aan de voorspellingen te hangen, kunnen er voorspellingsintervallen worden bepaald. Dit gebeurt door het bepalen van de variantie van de voorspellingsfouten op basis van de geschatte parameters. De fout in voorspellingsinterval die wordt gemaakt door met schattingen te werken in plaats van de werkelijke parameters is meestal verwaarloosbaar, vooral als er veel waarnemingen beschikbaar zijn [1].

De Holt-Winters methode wordt in de praktijk vaak gebruikt voor allerlei soorten tijdreeksen vanwege de eenvoudigheid en vanwege de intuïtieve manier waarop toekomstige waarden worden voorspeld. Het is dan ook niet verbazingwekkend dat er met deze methode goede resultaten zijn behaald.

4.2 Heteroscedasticiteit

In het bijzonder door het modelleren van aandelenkoersen is het idee ontstaan dat de *volatiliteit* van een tijdreeks $\{Y_t\}$ niet constant hoeft te zijn. Het is eerder een stochastische variabele die afhankelijk is van voorafgaande waarden. Het blijkt een typische eigenschap van economische data te zijn dat perioden met een hoge volatiliteit afgewisseld worden door perioden met een lage volatiliteit, en omgekeerd. Dit wordt *heteroscedasticiteit* genoemd. Om deze reden is er een aantal modellen bedacht die rekening houden met een tijdvariërende variantie.

4.2.1 GARCH processen

Het zogenoemde *Autoregressive Conditional Heteroscedasticity* (ARCH) proces heeft zijn naam te danken aan het feit dat de conditionele variantie van de tijdreeks, gegeven alle informatie over de verleden waarden van het proces, zich ontwikkelt volgens een soort van autoregressief proces. Een uitbreiding hierop is het *Generalized ARCH* (GARCH) proces dat waarschijnlijk het meest gebruikte model is voor economische tijdreeksen. De definitie is als volgt [3], [4].

Definitie 4.4. *Het GARCH(p, q) model wordt gegeven door*

$$Y_t = \mu + \epsilon_t, \quad t \in \mathbb{Z}, \tag{4.29}$$

met

$$\epsilon_t = \sigma_t Z_t, \quad t \in \mathbb{Z}, \tag{4.30}$$

waar de Z_t onafhankelijke en gelijk verdeelde stochastische variabelen zijn met

$$\mathbb{E}Z_t = 0 \text{ en } \mathbb{E}Z_t^2 = 1, \quad t \in \mathbb{Z}. \tag{4.31}$$

De schaal σ_t wordt gegeven door

$$\sigma_t^2 | \mathcal{F}_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}, \tag{4.32}$$

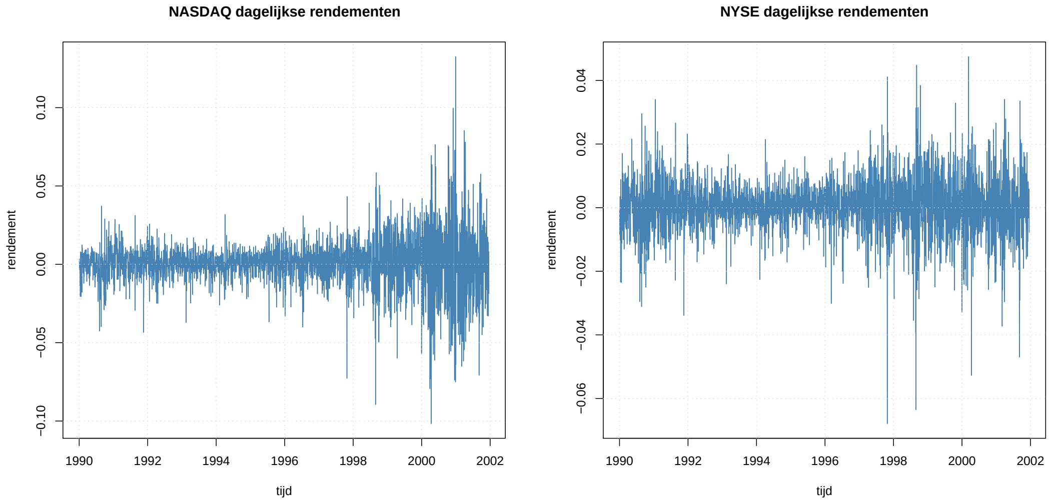
met $\mathcal{F}_t$ de verzameling informatie over ϵ_k en σ_k, $k < t$, beschikbaar op tijd t en $\omega > 0$, $\alpha_i \geq 0$ en $\beta_j \geq 0$ constanten.

Het idee is dat $\sigma_t^2 | \mathcal{F}_t$, de conditionele variantie van de tijdreeks gegeven de beschikbare informatie tot en met tijd $t-1$, een autoregressieve structuur bevat en positief gecorreleerd is met zijn eigen recente verleden en met recente waarden van de gekwadrateerde afwijkingen. Dit omvat het idee dat de volatiliteit (de conditionele variantie) tijdvariërend is en in perioden afwisselend hoog of laag is. Een GARCH(p, q) model met $p = 0$ reduceert tot een ARCH(q) model.

Een vluchtige blik op economische data (figuur 4.1) suggereert al dat in sommige tijdperioden het risico groter is dan in andere perioden. Dit wil zeggen dat de verwachtingswaarde van de grootte van de afwijkingen op sommige tijden groter is dan op andere tijden. Bovendien zijn deze risicovolle tijden niet willekeurig verspreid. In plaats daarvan is er een bepaalde mate van autocorrelatie in de risicovolheid van de rendementen. De ARCH en GARCH modellen zijn speciaal ontworpen om met deze zaken om te gaan. Zelfs in de meest simpele vorm heeft het GARCH model zich bewezen om toekomstige conditionele varianties succesvol te kunnen voorspellen.

4.3 Werkwijze

Nu een aantal modellen is besproken, is het niet onbelangrijk om te weten welk van de modellen gekozen moet worden voor het modelleren van een gegeven tijdreeks; kortom, wat de werkwijze is. In eerste instantie is het maken van een plot zeer belangrijk. Door met gezond verstand naar zo'n plot te kijken, kunnen een trend en seizoencomponent meestal wel worden ontdekt indien deze aanwezig zijn. Uit de plot kan ook duidelijk worden of de tijdreeks zich als een



Figuur 4.1. Dagelijkse rendementen van de NASDAQ en NYSE over een periode van 2 januari 1990 tot 31 december 2001.

additief of multiplicatief model gedraagt. Dit blijkt uit het al dan niet veranderen van de seizoencomponent ten opzichte van de trend. Uiteraard kunnen er hiervoor ook statistische toetsen worden toegepast. De *Augmented Dickey-Fuller* (ADF) toetst kan bijvoorbeeld worden gebruikt om te testen of een tijdreeks stationair is. De ADF toets beschouwt het volgende regressiemodel

$$\nabla Y_t = \mu + \gamma t + \alpha Y_{t-1} + \sum_{i=1}^{p-1} \beta_i \nabla Y_{t-i} + Z_t, \tag{4.33}$$

waarbij $\{Y_t\}$ het beste kan worden gerepresenteerd door een AR(p) proces en met $\{Z_t\}$ witte ruis. De coëfficiënten in dit model kunnen met behulp van de standaard kleinste-kwadratenmethode worden geschat. De nulhypothese van de ADF toest is

$$H_0 : \alpha = 0, \tag{4.34}$$

hetgeen betekent dat de tijdreeks niet stationair is. De alternatieve hypothese

$$H_1 : \alpha < 0 \tag{4.35}$$

betekent dat de tijdreeks juist wel stationair is. Nadat de coëfficiënten zijn geschat, is het mogelijk om de volgende Dickey-Fuller t-toetsingsgrootheid te berekenen

$$\tau = \frac{\hat{\alpha}}{\sqrt{\mathrm{Var}(\hat{\alpha})}}. \tag{4.36}$$

Deze toetsingsgrootheid volgt echter niet de standaard t-verdeling, maar de zogenoemde Dickey-Fuller t-verdeling. Als τ in absolute waarde groter is dan de kritieke waarde wordt de nulhypothese verworpen en wordt daarom aangenomen dat de tijdreeks stationair is [2]. Als een tijdreeks niet stationair blijkt te zijn, zal een trend of seizoencomponent verwijderd moeten worden, of zal er een transformatie op de tijdreeks moeten plaatsvinden. Het is in sommige gevallen ook mogelijk om een multiplicatief model additief te maken door het nemen van de logaritme.

Vervolgens is het een kwestie van alle modellen proberen om te onderzoeken welke het meeste geschikt is en de beste voorspellingen kan leveren. Voor alle modellen moeten er zekere parameters worden geschat. In eerste instantie komt dit neer op het bepalen van de orde van de modellen. Voor MA en AR processen is het bijvoorbeeld mogelijk om de orde (respectievelijk q en p) te bepalen aan de hand van de empirische ACF en PACF. Voor generieke ARIMA en GARCH processen zal er gekeken kunnen worden naar de *Akaike Information Criterion* (AIC), of soortgelijke maatstaven. De AIC wordt gegeven door

$$\text{AIC}(k) = -2\log(L) + 2k, \tag{4.37}$$

waarbij k het aantal parameters is en L de maximale likelihood van het geschatte model met k parameters. De AIC is gebaseerd op het principe dat door het verhogen van het aantal parameters enerzijds de likelihood wordt verhoogd, maar dat anderzijds ook het model complexer wordt en dat het gevaar van overfitting toeneemt. Overfitting betekent dat het model heel goed presteert op de bekende data maar minder goed kan generaliseren. Ook worden de parameters minder goed geschat hetgeen zorgt voor een minder goede voorspelling. De optimale waarde van k is dan die waarde waarvoor de AIC het kleinst is.

Tegelijkertijd met het bepalen van de orde van een model op basis van de AIC, moeten de parameters worden geschat. Dit is verre van eenvoudig. Voor een AR(p) proces bijvoorbeeld kunnen de schattingen $\hat{\alpha}_i$, $i = 1, \dots, p$ en $\hat{\sigma}^2$ bepaald worden met behulp van de zogenoemde *Yule-Walker* vergelijkingen

$$\begin{aligned} \hat{\rho}(i) &= \sum_{j=1}^{p} \hat{\alpha}_j \hat{\rho}(i-j), \quad i = 1, \dots, p, \\ \hat{\sigma}^2 &= \hat{\gamma}(0)\left(1 - \sum_{j=1}^{p} \hat{\alpha}_j \hat{\rho}(j)\right), \end{aligned} \tag{4.38}$$

waar de $\hat{\rho}(h)$ kunnen worden verkregen volgens

$$\begin{aligned} \hat{\gamma}(h) &= \frac{1}{n}\sum_{t=1}^{n-h}(Y_{t+h} - \bar{Y})(Y_t - \bar{Y}), \quad h \geq 0, \\ \hat{\rho}(h) &= \hat{\gamma}(h)/\hat{\gamma}(0), \end{aligned} \tag{4.39}$$

met n het aantal observaties. Voor generieke ARIMA en GARCH processen kunnen er echter geen expliciete uitdrukkingen worden gevonden. In plaats daarvan zal er een vorm van numerieke iteratie uitgevoerd moeten worden. Het softwarepakket R bevat procedures om de meest aannemelijke schattingen numeriek te benaderen.

Vervolgens zullen de modelaannamen gecontroleerd moeten worden. In bijna alle modellen wordt er aangenomen dat de afwijkingen ten opzichte van een deterministisch model worden veroorzaakt door witte ruis. Als het geschatte model en de waargenomen tijdreeks met elkaar worden vergeleken, dan moeten de residuen dezelfde verdeling volgen als deze witte ruis. Dit kan bijvoorbeeld worden getest met de *Ljung-Box* toetst. Deze toetst gebruikt de toetsingsgrootheid

$$Q = n(n+2)\sum_{h=1}^{m} \hat{\rho}_R(h)^2/(n-h), \tag{4.40}$$

met $\hat{\rho}_R(h)$ de autocorrelaties van de residuen bij een lag van h en m het aantal lags waarop getest wordt. Onder de nulhypothese dat de residuen afkomstig zijn van een witte ruis proces heeft Q een chi-kwadraatverdeling met m vrijheidsgraden. Als de nulhypothese verworpen wordt, en dus blijkt dat de modelaannamen niet kloppen, kan het betreffende model niet op die manier worden gebruikt.

Van de overgebleven modellen moet er tenslotte één worden gekozen om toekomstige waarden mee te voorspellen. Deze keuze wordt gemaakt door het voorspellend vermogen van een model te onderzoeken. Dit kan op de volgende manier. Van de bekende observaties wordt een deel achtergehouden en dit deel wordt door het model voorspeld alsof het er niet was. Deze voorspelling wordt dan vergeleken met de werkelijke waarden, hetgeen kan worden uitgedrukt in een gemiddelde relatieve fout. Aangezien de modellen stochastisch van aard zijn, zegt één getal niet zoveel. Vandaar dat deze procedure een aantal maal wordt herhaald waarna met behulp van data-analyse het beste model op een statistisch correcte wijze kan worden verkregen.

Hoofdstuk 5

Vermogensbeheer

VERTROUWELIJK

Hoofdstuk 6

Solvabiliteitsopslag

VERTROUWELIJK

Hoofdstuk 7

De VNB-tool

VERTROUWELIJK

Hoofdstuk 8

Conclusies en aanbevelingen

VERTROUWELIJK

Bijlage A

Figuren

VERTROUWELIJK

Bijlage B

Gebruikershandleiding

VERTROUWELIJK

Bibliografie

[1] C. Chatfield. *Prediction Intervals*. University of Bath, 1998.

[2] J. Cromwell, W. Labys, and M. Terraza. *Univariate Tests for Time Series Models*. Sage Publications, Thousand Oaks, California, 1994.

[3] R. Engle. GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15(4):157–168, 2001.

[4] P. Fryzlewicz. *Lecture notes: Financial time series, ARCH and GARCH models*. University of Bristol, 2007.

[5] M.C.M. de Gunst. *Statistical Models*. Vrije Universiteit Amsterdam, 2006.

[6] Chair of Statistics. *A First Course on Time Series Analysis: Examples with SAS*. University of Würzburg, 2006.

[7] H. Steehouwer. *Macroeconomic Scenarios and Reality: A Frequency Domain Approach for Analyzing Historical Time Series and Generating Scenarios for the Future*. Optima Grafische Communicatie, Rotterdam, Netherlands, 2005.

[8] W. Zucchini and O. Nenadić. *Time Series Analysis with R: Part I*.

[9] Zwitserleven. *Jaarverslag*. 2007.