

# Huurders Default

## Risico modellering



Wing Hong Kan  
Afstudeerverslag



# Huurders Default

## Risico modellering

Wing Hong Kan  
Afstudeerverslag

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



## Voorwoord

Dit afstudeerverslag vormt het afsluitende onderdeel van mijn masteropleiding Business Mathematics and Informatics aan de Vrije Universiteit Amsterdam.

Hierbij wil ik mijn eerste begeleider Bert Kersten bedanken voor de uitstekende begeleiding die hij mij heeft geboden. De grote mate van vrijheid in het bepalen van een koers, de aanwijzingen die hij gaf, en de aanmoediging in zwaardere tijden zijn zeer gewaardeerd. Ondanks zijn volle schema en volle privé-leven stond hij altijd voor mij klaar.

Mijn tweede lezer Sandjai Bhulai heeft tijdens mijn student-assistentenschap uitstekende begeleiding geboden. Deze samenwerking resulteerde in een paper die nu acceptatie geniet en in de nabije toekomst waarschijnlijk in een hoog aangeschreven tijdschrift komt. Ook bij dit afstudeerverslag kwam vanuit zijn hoek verfrissende kritiek die het verhaal ten goede zijn gekomen, en hiervoor wil ik hem ook bedanken.

Aan het begin van dit project heb ik kennis gemaakt met de praktijk van vastgoed. Rabo Bouwfonds stelde mij in staat een aantal maanden stage te lopen. Graag bedank ik hen daarvoor.

Verder dank aan familie en vrienden voor de support gedurende het afstuderen.

Wing Hong Kan  
Mei 2007

# Inhoudsopgave

<b>VOORWOORD .....</b>	<b>I</b>
<b>INHOUDSOPGAVE .....</b>	<b>II</b>
<b>SAMENVATTING .....</b>	<b>V</b>
<b>1. INLEIDING.....</b>	<b>1</b>
1.1 VASTGOED.....	1
1.2 VASTGOED ALS BELEGGINGSPRODUCT .....	1
1.2.1 <i>Opzet van een vastgoedleningenportefeuille</i> .....	1
1.2.2 <i>Risico's van een vastgoedleningenportefeuille</i> .....	3
1.2.3 <i>Alternatieve beleggingen</i> .....	5
1.3 AMERIKAANSE HYPOTHEEKCRISIS (SUBPRIME MORTGAGE CRISIS) .....	5
1.3.1 <i>Banken in een notendop</i> .....	5
1.3.2 <i>Aanleidingen voor de subprime mortgage crisis</i> .....	6
1.3.3 <i>Krediet crisis</i> .....	7
1.4 VASTGOED EN MODELLERING.....	7
1.5 ONDERZOEK .....	8
1.6 INDELING VERSLAG .....	8
<b>2. MODELLERING: BASIS.....</b>	<b>11</b>
2.1 INLEIDING .....	11
2.2 COLLATERAL .....	11
2.3 BASIS .....	12
2.4 GENEREREN VAN EINDTIJDSTIPPEN VAN HUURPERIODES.....	12
2.4.1 <i>Tijd tot eindtijdstip huurperiode</i> .....	12
2.4.2 <i>Tijd tot huurders default</i> .....	12
2.4.3 <i>Tijd tot (door huurder gewenste) huurperiode beëindiging</i> .....	14
2.5 GENEREREN VAN EINDTIJDSTIPPEN VAN LEEGSTANDSPERIODES.....	14
2.6 NIEUWE HUURDER.....	16
2.7 LEEGSTANDSCONTRACTEN .....	16
2.8 UITBREIDINGEN .....	17
<b>3. MODELLERING: UITBREIDING .....</b>	<b>19</b>
3.1 INLEIDING .....	19
3.2 ASSUMPTIES .....	19
3.3 UITBREIDING 1: RATING TRANSITIES (GEDURENDE NON-DEFAULT PERIODES).....	21
3.3.1 <i>Modellering van rating transitie</i> .....	21
3.3.2 <i>Analyse van rating transitie modellering</i> .....	24
3.4 UITBREIDING 2 (THEORIE): AFHANKELIJKHEID TUSSEN DEFAULTS.....	47
3.4.1 <i>Waarom geen lineaire regressie?</i> .....	48
3.4.2 <i>Introductie van copulas</i> .....	50
3.4.3 <i>Modellering en simulatie met de Clayton Copula</i> .....	53
3.5 UITBREIDING 2 (TOEGEPAST): VERWERKING VAN COPULA'S IN MODEL .....	56
3.5.1 <i>Groepering per Huurdersklasse</i> .....	56
3.5.2 <i>Verdere groepering</i> .....	58

<b>4. SIMULATIE</b> .....	<b>61</b>
4.1 INLEIDING .....	61
4.2 DATA .....	61
4.2.1 <i>Fictieve data</i> .....	61
4.2.2 <i>De dataset</i> .....	61
4.3 ASSUMPTIES .....	65
4.4 IMPLEMENTATIE .....	66
4.4.1 <i>Omgeving / programmeertaal</i> .....	66
4.4.2 <i>Beschouwing</i> .....	66
<b>5. ANALYSE</b> .....	<b>69</b>
5.1 INLEIDING .....	69
5.2 BASISPARAMETERS .....	69
5.3 GECORRELEERDE VS. ONGECORRELEERDE DEFAULTTIJDEN.....	72
5.4 GEVOELIGHEIDSANALYSE .....	74
5.4.1 <i>Output</i> .....	74
5.4.2 <i>Contract periode</i> .....	76
5.4.3 <i>Gemiddelde leegstandsperiode</i> .....	78
5.4.4 <i>Contract voortzettingkans</i> .....	80
5.5 CONCLUSIE .....	82
<b>6. CONCLUSIE</b> .....	<b>85</b>
6.1 INLEIDING .....	85
6.2 ACHTERGRONDEN .....	85
6.2.1 <i>De correlatiecoëfficiënt</i> .....	85
6.2.2 <i>Model versus werkelijkheid</i> .....	87
6.3 SUGGESTIES VOOR VERVOLGONDERZOEK .....	87
6.3.1 <i>Correlatie / regressie zijn niet de enige technieken</i> .....	88
6.3.2 <i>Contractstructuren</i> .....	88
6.3.3 <i>Opname van sociaal-economische informatie</i> .....	88
6.4 SLOTCONCLUSIE / EINDWOORD.....	89
<b>APPENDIX A – INVERSE TRANSFORMATIE METHODE</b> .....	<b>91</b>
<b>APPENDIX B – CLAYTON COPULA (KANSDICHTHEID)</b> .....	<b>92</b>
<b>APPENDIX C – CLAYTON COPULA (SIMULATIE)</b> .....	<b>93</b>
<b>BRONNENLIJST</b> .....	<b>94</b>





## Samenvatting

In dit afstudeeronderzoek is een model opgesteld voor het verhuurproces in een vastgoedportefeuille. Hierbij is onder andere afhankelijkheid van defaults opgenomen met behulp van modellering met behulp van copulas, in het bijzonder de Clayton copula. Eenvoudig gezegd vormen copulas een handig wiskundig hulpmiddel om afhankelijkheden van verschillende aard te modelleren (bijvoorbeeld meer afhankelijkheid tussen lage of tussen hoge waarden).

Naast de wiskundige aspecten van de modellering wordt er ook stilgestaan bij de positie van de modellering in een bedrijfskundig kader: wie heeft baat bij de modellering?

Aan de hand van een geïmplementeerde simulatie zijn verschillende tests uitgevoerd. Hieruit blijkt dat de opname van gecorreleerde defaults tot wezenlijk verschil leidt in vergelijking met de situatie met ongecorreleerde defaults. Verder bleek de simulatie een handige tool om gevoeligheidsanalyse uit te voeren naar verschillende input parameters (voor uitgebreide bespreking verwijs ik naar hoofdstuk 6).

Hoewel de modellering in dit afstudeerwerk een goede basis vormt, is er nog voldoende mogelijkheid tot vervolgonderzoek om tot een nog betere modellering te komen. Daarnaast zal een vertaalslag naar concrete vastgoed portefeuilles in de praktijk, bijvoorbeeld die van Bouwfonds, tot interessante inzichten kunnen leiden. Vaak is de praktijk immers lastiger dan de theorie.



# 1. Inleiding

## 1.1 Vastgoed

*"Onroerend zijn de grond, de nog niet gewonnen delfstoffen, de met de grond verenigde beplantingen, alsmede de gebouwen en werken die duurzaam met de grond zijn verenigd, hetzij rechtstreeks, hetzij door vereniging met andere gebouwen of werken" (Nederlandse B.W., boek 3, art. 3, lid 1).*

Uit het bovenstaande fragment uit het Nederlands Burgerlijk Wetboek komt naar voren dat met de term vastgoed of onroerend goed wordt verwezen naar grond en/of gebouwen. Een kenmerk van vastgoed, wat ook in de naam naar voren komt, is dat het niet verplaatsbaar is.

Vastgoed heeft waarde. Deze wordt mede bepaald door factoren zoals: vraag naar ruimte, voorzieningen in de omgeving van het vastgoed, lokale en landelijke economie, politieke besluiten. Het is mogelijk om vastgoed te verhuren om hieruit een vaste stroom van inkomsten te verkrijgen. Dit gebeurt bijvoorbeeld met huurwoningen en ook bij bedrijfsruimtes. Dit heeft ook zijn impact op de waarde van vastgoed. Een goede verhuurbaarheid, en lage leegstandsprognose leidt tot een hogere waarde van het vastgoed.

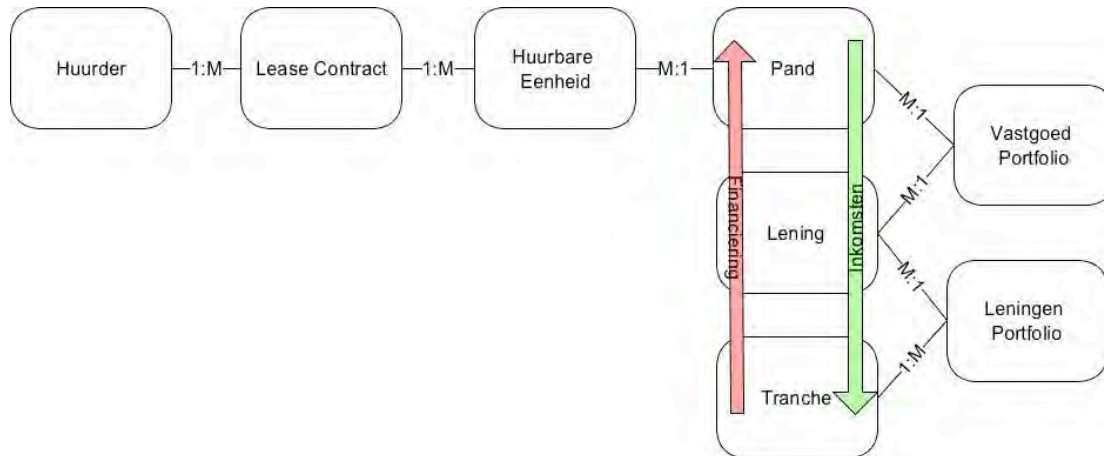
Alleen al in Nederland gaan er vele miljarden in vastgoed om. Vastgoedbeleggingen vormen een klein percentage van portefeuilles van bijvoorbeeld pensioenfondsen. Maar zelfs die paar procenten monden uit in miljarden. Investeerders zijn zeer geïnteresseerd in nauwkeurige risico-inschattingen van vastgoed.

## 1.2 Vastgoed als beleggingsproduct

### 1.2.1 Opzet van een vastgoedleningenportefeuille

Om vastgoed te kunnen financieren is het veelal nodig om leningen af te sluiten. Deze leningen verplichten de geldlener om periodiek een rente en/of een deel van de nominaal uitstaande lening te voldoen. De geldlener kan aan deze verplichting voldoen uit eigen inkomsten of, wanneer deze niet zelf een stuk vastgoed betreft maar verhuurt aan anderen, voldoen vanuit de inkomstenstroom die de verhuur hem oplevert. De lener is bij verhuur afhankelijk van de verhuurbaarheid van zijn pand en de kredietwaardigheid van zijn huurders. "Slechte" huurders hebben een hogere kans op wanbetaling (delinquency), en uiteindelijk op uitzetting (huurders default). Dit

laatste betekent verlies van inkomsten gedurende een leegstandsperiode totdat een nieuwe huurder is gevonden. De leegstandsperiode wordt onder andere bepaald door de verhuurbaarheid van de verhuurbare eenheid. Deze verhuurbaarheid is afhankelijk van factoren als prijs, locatie, vraag, etc.



**Fig. 1: Vastgoedbelegging**

In figuur 1 is schematisch de opzet van een vastgoedbelegging weergegeven. Aan de basis van het geheel staat de huurder. Dit kunnen personen, bedrijven, maar ook de overheid zijn (bij experimenten verderop in dit verslag werken we met bedrijven omdat daarvan bepaalde vrij toegankelijke data beschikbaar is op internet). De huurder heeft een of meerdere leasecontracten afgesloten en elk leasecontract heeft betrekking op een of meerdere huurbare eenheden. In een leasecontract is onder meer de huurtermijn en de periodieke betaling vastgelegd. Een of meerdere huurbare eenheden vormen samen een pand (voor simpliciteit gaan we bij experimenten later uit van simulatie op pandniveau in plaats van op huurbare eenheid niveau). Banken verstrekken leningen zodat anderen panden kunnen kopen om hiermee inkomsten te genereren.

Het geheel van bij elkaar horende panden en leningen is een vastgoed portefeuille, wat gezien kan worden als een balans met aan de ene kant panden en daartegenover de leningen (Fig. 2). De panden leveren uiteindelijk de inkomstenstromen waarmee de leningenterugbetaling wordt voldaan. De lening kan dus gezien worden als de belegging. Men financiert een lening tegen een renteopslag van een bepaald aantal basispunten (100 basispunten = 1%) en wil via deze investeringsconstructie geld verdienen. De rente is de vergoeding die tegenover het te lopen risico staat, en is hoger naarmate het risico hoger is.

<i>Panden</i>	<i>Leningen</i>
“Pand 1”	“Lening 1”
“Pand 2”	“Lening 2”
“Pand 3”	“Lening 3”
“Pand 4”	

**Fig. 2: Vastgoedportefeuille**

Om niet afhankelijk te zijn van de leningselectie in de vastgoed portefeuille is het mogelijk om een fonds op te zetten waarbij uit verschillende vastgoed portfolio's de meest veelbelovende leningen worden geselecteerd. Het geheel van leningen (opbrengsten) wordt meestal samengevoegd en opgeknipt in tranches geassocieerd met een bepaald risicoprofiel en uitkeringsmarge. Het geheel van leningen en tranches vormt de leningenportefeuille en kan ook gezien worden als een balans, met aan de ene kant de leningen en daartegenover de tranches (Fig. 3).

<i>Leningen</i>	<i>Tranches</i>
“Lening 1” van “Vastgoed Portfolio 1”	“Tranche A”
“Lening 2” van “Vastgoed Portfolio 1”	“Tranche B”
“Lening 3” van “Vastgoed Portfolio 1”	“Tranche C”
“Lening 1” van “Vastgoed Portfolio 2”	
“Lening 2” van “Vastgoed Portfolio 2”	
“Lening 2” van “Vastgoed Portfolio 4”	
“Lening 1” van “Vastgoed Portfolio 5”	

**Fig. 3: Leningenportefeuille**

Een investeerder koopt bepaalde tranches in en men kan stellen dat deze uiteindelijk bijdraagt aan de financiering. De inkomstenstromen beginnen helemaal vanuit de basis, de huurder, en vinden hun weg via de panden en leningen, terug naar de tranches van de investeerder.

### 1.2.2 Risico's van een vastgoedleningenportefeuille

Wanneer de bezitter van een pand altijd inkomsten ontvangt van zijn huurders is er niks aan de hand. Problemen ontstaan wanneer huurders hun lasten niet meer kunnen betalen. Dit heeft gevolgen voor de inkomstenstroom en kan ertoe leiden dat de bezitter van het pand op zijn beurt ook zijn lasten niet meer kan betalen; Hij kan niet voldoen aan de terugbetalingsverplichting van zijn leningen. En dit zal uiteindelijk ook invloed hebben op andere partijen binnen het systeem van figuur 1, tot aan de investeerders die in de tranches beleggen. Om dit nauwkeuriger toe te lichten neem ik eerst een stap terug.

Wanneer een huurder niet aan zijn betalingsverplichting kan voldoen zijn er twee mogelijkheden:

- De huurder wordt uitgezet.
- De situatie wordt aangekeken, in de hoop dat de huurder zich herstelt en alsnog aan achterstallige betalingsverplichtingen kan voldoen en ook aan toekomstige betalingen kan voldoen.

Uitzetten van de huurder leidt gegarandeerd tot een onbekende periode leegstand, maar het is de beste zet wanneer de prognose voor de huurder er slecht uit ziet. De huurder niet uitzetten en hopen op verbetering zou mogelijk een leegstandsperiode kunnen helpen vermijden, maar kan ook leiden tot langere periode van wanbetaling waar de verhuurder uiteindelijk ook niks aan heeft.

Wanneer een verhuurder door teveel huurders defaults en te lange leegstandsperiodes niet meer aan zijn terugbetalingsverplichting kan voldoen, genereert een lening niet de vooraf vastgelegde inkomsten voor de financier. Op dat moment is de lening in default gegaan. Er zijn hier ook weer twee mogelijkheden:

- De panden worden in beslag genomen en uitgewonnen (verkocht).
- De situatie wordt aangekeken in hoop op herstel.

Ook hier is niet eenduidig te zeggen wat de beste beslissing is. Wel moet worden opgemerkt dat bij uitwinnen verlies kan worden gemaakt maar ook alsnog winst. Default van een lening hoeft dus niet noodzakelijk een slechte ontwikkeling te zijn. Hiertoe zijn er in de risicoanalyse, niet alleen bij vastgoed, metrieken als LGD (Loss Given Default) en LGL (Loss Given Loss). LGD geeft de het verlies weer, gegeven dat er default is opgetreden. LGL geeft het verlies weer, gegeven die gevallen waar default is opgetreden én er verlies wordt gemaakt. LGD is groter of gelijk aan LGL en aan het verschil kan men verder een inzicht krijgen in de mate waarin default daadwerkelijk tot een verlies leidt.

Wanneer de inkomstenstroom niet voldoet om alle tranches uit te keren zal er via een waterval mechanisme geld worden uitgekeerd. Dat wil zeggen dat investeerders die minder risicovolle posities hebben bedongen als eerste hun aandeel krijgen uitgekeerd.

Naast risico's die voortvloeien uit de huurder zijn ook op pandniveau risico's mogelijk. Onderhoudskosten kunnen bijvoorbeeld sterker stijgen dan verwacht, of de huurprijzen in de markt kunnen tegenvallen. Dit heeft echter een minder grote impact dan huurders default wanneer men risicoanalyse uitvoert, waarbij men toch geïnteresseerd is in prijsbepaling van de meer risicovolle tranches. Bij risicovolle scenario's zijn het vaker de weggevallen inkomsten die meer pijn doen dan de hoger uitvallende kosten.

### **1.2.3 Alternatieve beleggingen**

Naast beleggen in vastgoed, zoals in paragraaf 1.2.1 is omschreven, is het ook mogelijk om in vele andere zaken en op vele andere manieren te investeren. In principe is het mogelijk om van allerlei soorten leningen een beleggingsproduct op te stellen, waarbij het onderliggende waardegenererende onderpand verschilt maar de basisopzet hetzelfde is als bij de vastgoedbeleggingen die in dit verslag aan de orde komen. De huurder in het huidige verhaal kan een particuliere huurder zijn, maar ook een bedrijf of ondernemer. En het soort pand kan variëren van huis, kantoor, winkel, tot ziekenhuis en garage. Verder kan het uiteindelijke fonds een gesloten fonds zijn of een genoteerd fonds met vrij verhandelbare papieren.

## **1.3 Amerikaanse hypotheekcrisis (subprime mortgage crisis)**

In de Verenigde Staten en in de rest van de wereld woedt op dit moment een krediet crisis. Dit is voor een groot deel te wijten aan grote verliezen die banken hebben geleden in de zogenaamde subprime mortgage crisis. Eerst volgt een korte toelichting over banken.

### **1.3.1 Banken in een notendop**

Banken zijn geldscheppende instanties. Weliswaar mag niet iedere bank fysiek geld afdrukken, in Nederland mag alleen De Nederlandsche Bank dit, ze dragen toch bij aan de totale hoeveelheid beschikbaar geld. Mensen en bedrijven zetten hun geld op de bank en deze kan hier dan vervolgens mee gaan investeren (beleggen, uitlenen, etc.). De hoeveelheid geld die de bank heeft zal over het algemeen nooit in een keer worden opgenomen door iedereen tegelijk. Hierdoor heeft de bank de mogelijkheid om geld uit te geven dat in principe gecreëerd is. De financiële autoriteit legt vast dat banken een bepaalde dekkingsgraad moeten hebben om te voorkomen dat er een oneindige hoeveelheid geld kan worden gecreëerd op deze manier. Banken zijn verplicht om hun totale hoeveelheid geld voor een bepaald percentage gedekt te hebben door “echt” geld.

Banken lenen geld uit aan bedrijven en personen, maar doen dit ook onderling. Hiervoor worden lagere rentepercentages dan voor het gewone publiek gerekend. Wat ook deels gerechtvaardigd is door de sterkere financiële positie (kredietwaardigheid) van een bank om haar verplichting na te komen, ten opzichte van een willekeurig persoon. Bank A leent tegen een laag rentepercentage geld van bank B, en bank A kan dit bijvoorbeeld uitlenen aan een persoon die een auto wil kopen tegenover een hoger rentepercentage. Zo heeft de persoon aan het eind geld om te besteden, bank A heeft geld om uit te lenen en uiteindelijk om winst mee te maken, en bank B maakt ook winst aan het door haar aan bank A uitgeleende geld. Dit gaat lange tijd goed, totdat doemscenario's zich voordoen.

### 1.3.2 Aanleidingen voor de subprime mortgage crisis

In de Verenigde Staten is een grote hoeveelheid geld uitgeleend aan mensen waarvan de kredietwaardigheid zich in de lagere segmenten bevindt, de zogenaamde subprime klasse. Deze klasse heeft een hoog risico maar dit is niet direct een probleem omdat het gecompenseerd wordt door de hogere rentepercentages die voor deze groep wordt berekend. Zo ook zijn de premiums voor investeerders die beleggen in subprime gerelateerde producten hoger. Er heeft zich echter een aantal ontwikkelingen voorgedaan waardoor beleggingen in subprime massale klappen hebben geleden en er in de wereld een liquiditeitscrisis is ontstaan. Eerst worden diverse redenen gegeven hoe het zover heeft kunnen komen, alvorens meer over de liquiditeitscrisis uit te wijden.

Het is lastig, soms misschien zelfs onzinnig, om één directe verklarende oorzaak te geven voor een ontwikkeling. Zo ook voor de subprime mortgage crisis. De volgende redenen worden echter vaak gegeven als grondoorzaken, en zijn in combinatie met elkaar aannemelijk de reden van de huidige crisis:

- In de Verenigde Staten is er een kredietcultuur. Men geeft geld uit voordat dit daadwerkelijk is verdiend en sommigen bouwen hiermee substantiële schulden op. Een dergelijke ontwikkeling is ook in Nederland te zien waarbij geld wordt geleend aan mensen die het eigenlijk niet kunnen veroorloven, en dit tegen woekerrentes. Door jarenlang stijgende huizenprijzen zijn huizenbezitters extra hypotheek af gaan sluiten om te kunnen voorzien in verdere (luxe) consumptie. Wanneer de huizenprijsontwikkeling echter tegenvalt of een huis overgewaardeerd blijkt, dan blijft de huizenbezitter met schulden zitten. Banken zien hun reguliere hypotheekaflossingen stagneren en kunnen bij voldoende harde klappen zelfs failliet gaan. En investeerders zien hun subprime investeringen dalen in waarde tot verliesgevendheid.
- Banken en rating agencies zijn (uit winstbejag?) te positief geweest in het beoordelen van de kredietwaardigheid van hun cliënten. De hogere rentepercentages die zijn berekend voor deze groep hadden nog hoger gelegen moeten hebben, of deze groep had helemaal geen lening mogen krijgen.
- Vastgoedbeleggingen “veranonimiseert” leningen. Wanneer een lening wordt afgesloten bij een bank en het met de lener even wat slechter gaat, kan de bank uit afweging van een langdurig opgebouwde relatie en wellicht andere lopende leningen, besluiten om coulant om te gaan met wanbetaling. Bij vastgoedbeleggingen is de afstand tussen lener en investeerder groot. De investeerder heeft geen band met de lener, maar bepaalt wel wat er bij wanbetaling moet gebeuren. Dit kan leiden tot snellere besluiten om bij wanbetaling direct uit te winnen om potentiële verdere verliezen uit te sluiten.



### **1.3.3 Krediet crisis**

De grote verliezen die banken hebben geleden als gevolg van de subprime mortgage crisis heeft sommige banken aan de rand van faillissement gebracht. De slechtere economische positie van de banken heeft ertoe geleid dat banken elkaar hogere rentepercentages gaan berekenen (meer risico leidt immers tot hogere premium eis voor het geleende geld). Hierdoor is er krapte op de geldmarkt ontstaan. En wanneer het met banken slecht gaat, gaat het met de hele economie slecht. Dus financiële autoriteiten zagen zich genoodzaakt om geld te injecteren, door goedkoop geld te lenen aan banken. Hiermee moeten de banken genoeg kapitaal hebben om hun positie weer te versterken. In de toekomst moet blijken of de keuze van centrale autoriteiten om goedkoop geld te injecteren de juiste was. Ook hier had degelijke modellering vroegtijdig al tot betere risico-inschatting van de situatie kunnen leiden en crisis kunnen voorkomen.

## **1.4 Vastgoed en modellering**

Een model is om de werking van een bestaande situatie te beschrijven op een versimpelde wijze. Modellen worden gebruikt omdat er antwoorden worden gezocht op vragen uit de echte wereld en deze laatste te complex is om direct te onderzoeken. Zo ook de modellering in dit afstudeerverslag.

Bij vastgoedinvesteringen kan het zeer lonend zijn om wiskundige modellering toe te passen. Investeerders kunnen een betere inschatting maken van het risico-rendement profiel. En voor fondsozettelaren is het een goede mogelijkheid om zich te onderscheiden van concurrenten; de prijs kan (verantwoord/afgewogen) laag gehouden worden en gedegen research geeft potentiële klanten het gevoel met de juiste partij in zee te gaan.

Tevens is er in internationale kringen door banken een akkoord gesloten dat er voldoende reserves moeten worden aangehouden als buffer bij tegenslagen. Vroeger was dit simpelweg een bepaald percentage over de inleg (BASEL I). Maar men heeft ingezien dat dit te simplistisch is, er moet namelijk onderscheid worden gemaakt tussen risicovolle en minder risicovolle investeringen. Het is logischer om voor minder risicovolle investeringen minder reserve te hoeven aanhouden dan voor risicovollere investeringen. Hiertoe is na jarenlang overleg een nieuw akkoord (BASEL II) overeengekomen. Wat in het verleden wel eens verkeerd ging is dat de reserves veel te laag waren, ofwel, men ging uit van te rooskleurige prognoses. Tegenwoordig is het nodig om degelijk kwantitatief onderbouwd een fonds op te zetten. Er wordt niet afgedwongen om bepaalde methodieken te hanteren, maar het is wel de bedoeling om met modellen en simulaties een fonds te onderbouwen. Dit wordt niet internationaal getoetst, maar in Nederland wordt dit bijvoorbeeld door de financiële toezichthouder De Nederlandsche Bank getoetst. NB. BASEL II bevat niet alleen het hier vermelde maar is een zeer uitgebreid document met vele richtlijnen met betrekking tot bancaire zaken.

De partij die een fonds opzet hoeft niet perse te participeren in het fonds. Het is echter niet ongewoon als deze dat toch doet, want het toont vertrouwen in het zelfgestructureerde product en biedt kansen om ook via deze weg winst te behalen uit het beleggingsproduct. Door zelf een aandeel te nemen in het product, en dus ook eventuele verliezen, krijgen derden meer vertrouwen in het product omdat de structureerder van het fonds er zelf ook belang bij heeft dat het fonds wel vaart.

Bij de grote hoeveelheden data en complexiteit die bij vastgoedbeleggingen spelen is het praktisch onmogelijk om gevoelsmatig goede beslissingen te nemen en is een kwantitatief model de uitkomst.

Een degelijke modellering heeft voor veel verschillende partijen voordelen. Zelfs op de gewone burger heeft het impact in bijvoorbeeld de pensioenpremies. Pensioenmaatschappijen investeren grote sommen geld in vastgoed. Een degelijke investering hierin, aan de hand van goede modellering, is nodig om goede investeringsbeslissingen te maken welke de continuïteit in de pensioenuitbetaling garanderen.

### **1.5 Onderzoek**

De studie in deze paper heet betrekking op vastgoedbeleggingen. De focus ligt op de **modellering van de onderliggende geldstroom genererende entiteit, de huurder.**

In het bijzonder zal er worden gekeken naar correlaties tussen het in default gaan van verschillende huurders. Met andere woorden, de mate waarin een default van huurder  $x$  bijdraagt aan de default waarschijnlijkheid van een huurder  $y$ . Een onderwerp waarin nog niet veel onderzoek is gedaan, maar waar vanuit investeerders veel vraag naar is. Gecorreleerde default kan onder andere voorkomen bij recessies, rentecrises en natuurrampen. We zijn geïnteresseerd in de modellering hiervan en de impact die dit heeft op het rendement.

### **1.6 Indeling verslag**

Dit verslag is als volgt ingedeeld. Hoofdstuk 2 beschrijft de basis van de opgestelde modellering en is vooral wiskundig van aard. Hoofdstuk 3 bevat een uitbreiding van het in hoofdstuk 2 opgestelde model. Hoofdstuk 4 beschrijft de opzet van de gemaakte simulatie. Hoofdstuk 5 geeft een analyse van simulaties die zijn uitgevoerd. Het laatste hoofdstuk bevat de conclusies en bevat tevens onderwerpen die ik interessant acht, maar die niet in de andere hoofdstukken zijn gezet omwille de flow van het verhaal niet te verstoren.





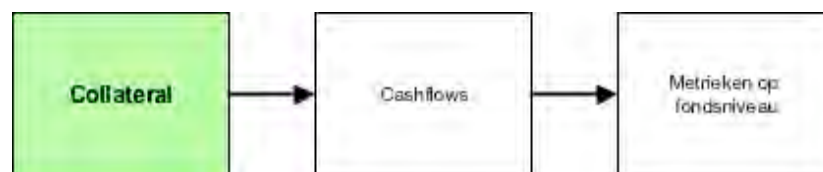
## 2. Modelling: Basis

### 2.1 Inleiding

Bij het zoeken naar bestaande literatuur viel al snel op dat over correlaties tussen huurders defaults weinig kwantitatief onderzoek was te vinden. Wel is er een grote hoeveelheid werk te vinden over correlatie tussen defaults (in de zin van faillissement) van bedrijven. De modellering in dit hoofdstuk is gebaseerd op onderzoek naar dit laatste. De aanname is dat faillissementen een goede proxy zijn voor het niet na kunnen komen aan de betalingsplicht van huur- en/of hypotheeklasten (huurders default).

### 2.2 Collateral

Het “onderliggende waardegenererende” wordt in jargon ook wel collateral genoemd. Dat is het gedeelte waar in dit verslag aandacht aan wordt besteed. Met alleen een model voor dit gedeelte kan men nog niet direct fondsen simuleren, maar dit is wel het gedeelte waar de meeste modellering (wiskunde, creativiteit) komt kijken. Overige modelgedeelten bestaan uit het vertalen van informatie van het “collateral model” naar cashflows, en van cashflows naar statistieken over uiteindelijke prestatie-maten (Fig. 4). Deze laatste delen zijn voor het grootste gedeelte, zo niet geheel, van deterministische aard, met andere woorden statische doorreken-tools. Hierbij moet opgemerkt worden dat alhoewel het cashflow model vooral doorrekent, het een significant model onderdeel is omdat er veel cashflows spelen en het programmeren hiervan een delicaat werk is.



**Fig. 4: Positie van de huidige modellering**

Het collateral gedeelte moet informatie geven over het onderliggende inkomsten genererende van waaruit het cashflow gedeelte alle inkomsten en verlies geldstromen kan doorrekenen. Dit komt neer op informatie over huurstatus per verhuurde unit, dat is, per verhuurbare eenheid een collectie van tijdstippen waarop contracten of leegstandsperioden beginnen.

## 2.3 Basis

Het model moet in ieder geval de volgende onderdelen bevatten:

- Genereren van eindtijdstippen van contracten, wat neerkomt op genereren van een potentieel default tijdstip en een potentieel eindtijdstip van een gewenste huurperiode.
- Genereren van starttijdstippen van nieuwe contracten, wat neerkomt op genereren van leegstandsperiodes.
- Random een nieuwe huurder bepalen voor een nieuw contract.
- Rekening houden met huurder – contract – verhuurbare eenheid structuur.

## 2.4 Genereren van eindtijdstippen van huurperiodes

### 2.4.1 Tijd tot eindtijdstip huurperiode

Voor iedere huurder zijn er twee mogelijkheden. Of hij gaat in huurders default, of hij zit een bepaald aantal gewenste huurtermijnen uit, alvorens zonder default zijn gecontracteerde huurbare eenheden te verlaten. Voor beide potentiële gebeurtenissen genereren we een random tijdstip waarbij, het laagste tijdstip bepalend is voor welke gebeurtenis er plaatsvindt.

$$T_{L,i} = T_{H,i} + \min\{\text{tijd tot default, tijd tot beëindiging huurperiode}\} \quad (1)$$

Voor iedere verhuurbare eenheid geven we met  $T_{H,i}$  het starttijdstip van de  $i$ -de huurperiode weer, en geven we met  $T_{L,i}$  het starttijdstip van de  $i$ -de leegstandsperiode weer.  $T_{H,i}$  kan kleiner zijn dan 0, waarbij tijdstip 0 het starttijdstip is van het gemodelleerde fonds.

### 2.4.2 Tijd tot huurders default

De tijd tot default modelleren we als een exponentiële verdeling. De reden hiervoor ligt in de geheugenloosheidseigenschap, waarbij reeds verstreken tijd geen additionele informatie geeft over verdere tijd tot default.

$$\begin{aligned}
 T &\sim \text{Exp}(\lambda) \\
 f(t) &= \lambda e^{-\lambda t}, \quad t \geq 0 \\
 F(t) &= P(T \leq t) = 1 - e^{-\lambda t}, \quad t \geq 0 \\
 E(T) &= \frac{1}{\lambda}
 \end{aligned}
 \tag{2}$$

Om de modellering te kunnen gebruiken moet de defaultrate  $\lambda$  geschat worden vanuit data. Aannemende dat huurders uit eenzelfde ratingklasse dezelfde defaultrate hebben, kunnen we deze schatten vanuit de defaultkansen (per jaar) die door ratingorganisaties als Moody's en Fitch worden opgesteld. Laat PD de kans op default zijn voor een gegeven jaar.

$$\begin{aligned}
 PD &= P(T \leq 1) = F(1) = 1 - e^{-\lambda} \\
 \Rightarrow \lambda &= -\log(1 - PD)
 \end{aligned}
 \tag{3}$$

Voor een groep huurders kunnen we de defaultrate ook schatten vanuit gerealiseerde defaultduraties. Laat  $T_i$  de tijd tot default zijn van huurder  $i$ , met  $i = 1, \dots, n$ .

$$\begin{aligned}
 \bar{T} &= \frac{\sum_{i=1}^n T_i}{n} \\
 \lambda &= \frac{1}{E(T)} = \frac{1}{\bar{T}} = \frac{n}{\sum_{i=1}^n T_i}
 \end{aligned}
 \tag{4}$$

De trekking van een willekeurig exponentieel verdeeld getal kan met behulp van de inverse transformatie methode.

$$\begin{aligned}
 U &\sim \text{Uniform}(0,1) \\
 T = F^{-1}(U) &= \frac{-\ln(1-U)}{\lambda} = \frac{-\ln(U)}{\lambda}
 \end{aligned}
 \tag{5}$$

### 2.4.3 Tijd tot (door huurder gewenste) huurperiode beëindiging

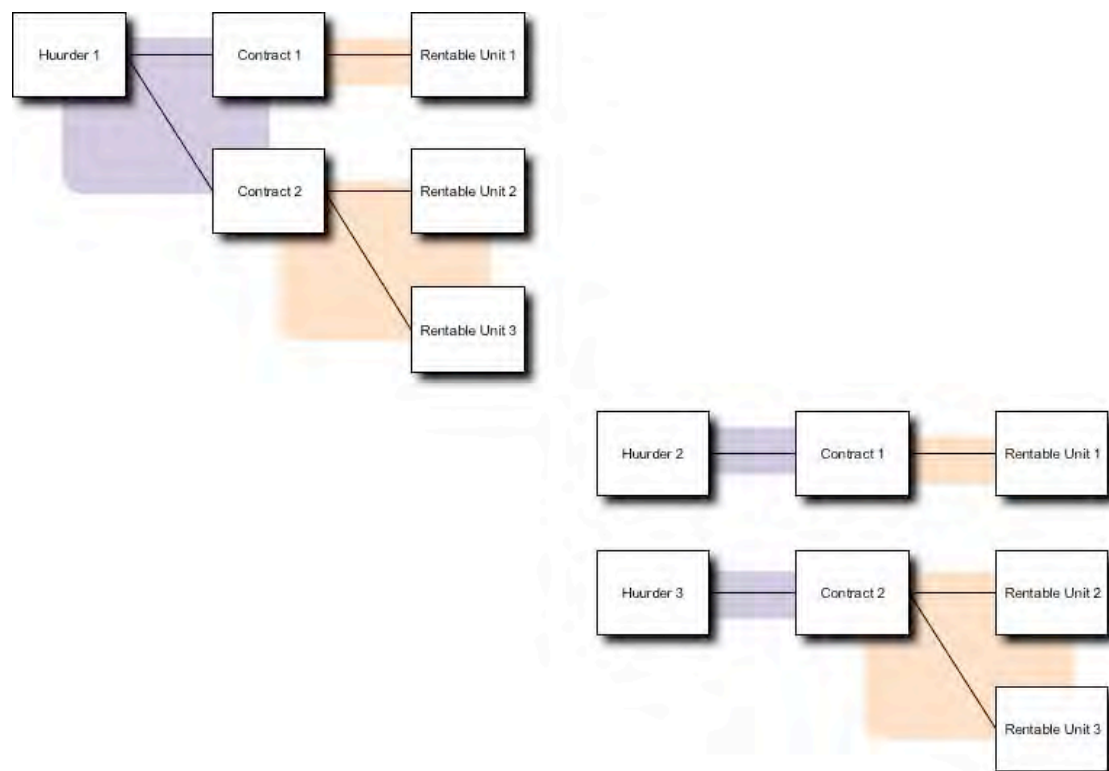
Contracten hebben meestal een vaste looptijd van bijvoorbeeld 5 jaar. Na het einde van die periode heeft een huurder de mogelijkheid om zijn contract te verlengen met weer een contractperiode. Wanneer we de kans op voortzetting van een contract ingeven (bijvoorbeeld 50%) gebaseerd op schattingen uitgevoerd op historische data, dit kan een globaal percentage zijn maar ook contractspecifiek, dan kunnen we het aantal gewenste contractperiodes van de huurder modelleren als een geometrische verdeling.

$$\# \text{ Contractperioden} \sim \text{Geo}(p)$$

$$T = \# \text{ Contractperioden} \cdot \text{Contractperiode}$$

(6)

### 2.5 Genereren van eindtijdstippen van leegstandsperiodes



**Fig. 5: Van linksboven naar rechtsonder:  
Uiteenvallen van huurder-contract structuur (paars)  
en behoud van contract-verhuurbare eenheid structuren (oranje)**



Wanneer na het begin van de simulatie de initiële huurders vertrekken, is het onwaarschijnlijk dat nieuwe huurders precies dezelfde huurder-contract structuren (in figuur 5 weergegeven met paarse achtergrond) overnemen. Nieuwe huurders zullen waarschijnlijk volgens hun eigen wensen contracten afsluiten. Het is echter onduidelijk welke verhuurbare eenheden deze nieuwe huurders in hun contracten zullen opnemen. Om dit toch op een of andere manier te kunnen modelleren gaan we er vanuit dat de huurder-contract structuren na de eerste simulatieslag (van de vorige paragraaf) vervalt, maar dat contract-verhuurbare eenheid structuren (in figuur weergegeven met oranje achtergrond) blijven bestaan. Ofwel, de verhuurbare eenheden blijven in een zelfde groepering verhuurd worden onder nieuwe contracten, en nieuwe huurders nemen niet dezelfde contracten aan maar slechts één contract. We laten de verhuurbare eenheid groepering als in het originele contract alhoewel in de werkelijkheid versplintering of hergroeperingen van eenheden kan optreden. Dit doen we omdat totale versplintering leidt tot een onderschatting van het risico, immers is het dan zo dat een huurder die in default gaat minder verhuurbare eenheden met zich meebrengt. En er is geen eenduidig algoritme op te stellen waarmee versplintering en hergroepering kan worden nagebootst.

Dus waar de simulatie eerst op huurderniveau plaatsvond, zal de simulatie vanaf dit punt overstappen op contractniveau. Voor iedere verhuurbare eenheid kan door een vastgoedexpert opgegeven worden hoelang leegstand gemiddeld duurt. De gemiddelde leegstand van alle verhuurbare eenheden binnen een contract nemen we als uitgangspunt voor het genereren van een leegstandperiode. Net zoals bij huurders default kiezen we voor de exponentiële verdeling. NB. Helaas was er bij dit onderzoek geen data beschikbaar om de aanname van een exponentiële verdeling te verifiëren.

$$\begin{aligned} \frac{1}{E(T)} &= \lambda \\ T &\sim Exp(\lambda) \end{aligned} \tag{7}$$

Om een minimale leegstandperiode te handhaven is er slechts een kleine aanpassing nodig. De minimale leegstand moet bij het random gedeelte worden opgeteld als constante, en het random gedeelte moet een aangepaste parameter krijgen, dit laatste opdat de gemiddelde leegstand gehandhaafd blijft.

$$\begin{aligned} T &\sim \text{minimale leegstand} + Exp\left(\frac{1}{E(T) - \text{minimale leegstand}}\right) \\ &\text{waarbij } E(T) > \text{minimale leegstand} \end{aligned} \tag{8}$$

Het tijdstip ten opzichte van simulatiestarttijd  $T=0$  wordt bepaald door de vergelijking:

$$T_{H,i+1} = T_{L,i} + \text{leegstandsperiode} \quad (9)$$

## 2.6 Nieuwe huurder

Een nieuwe huurder kan worden gesimuleerd door random een rating te trekken en deze, zoals in het vorige paragraaf al vermeld, aan een groep verhuurbare eenheden te verbinden dat eerder onder eenzelfde contract viel. Laat het aantal ratings, exclusief default als ratingklasse, gelijkstaan aan  $n$ . Zonder additionele informatie over de relatie tussen de rating van de vorige huurder en de rating van de volgende huurder, modelleren we de rating van de volgende huurder als een discrete uniform verdeelde verdeling.

$r$  = rating van volgende huurder

$$P(r = AAA) = \frac{1}{n}$$

$$P(r = AA) = \frac{1}{n}$$

$$P(r = A) = \frac{1}{n}$$

...

$$P(r = C) = \frac{1}{n}$$

(10)

Hierbij herhalen we om en om hetzelfde procedé waarmee eerder eindtijden van huurperiodes en leegstandsperiodes werden gegenereerd, totdat voor iedere verhuurbare eenheid genoeg tijdstippen zijn vergaard tot het einde van de simulatietijd (bijvoorbeeld 20 jaar,  $T=20$ ).

## 2.7 Leegstandscontracten

In vastgoed bestaan zogenaamde leegstandscontracten. Dit zijn niet bestaande contracten maar worden onder andere gebruikt om verhuurbare eenheden met elkaar te groeperen. Voor deze leegstandscontracten geldt dat er aan het begin van de simulatie een leegstandsperiode wordt gegenereerd, en dat er per leegstandscontract een willekeurige eerste huurder wordt getrokken.

Omdat bij leegstandscontracten wordt begonnen met leegstand moeten we voor deze groep een kleine aanpassing doorvoeren in eerdere notatie.

$$\begin{aligned}T_{L,1} &= 0 \\T_{H,i} &= T_{L,i} + \text{leegstandsperiode} \\T_{L,i+1} &= T_{H,i} + \min\{\text{tijd tot default, tijd tot beëindiging huurperiode}\}\end{aligned}\tag{11}$$

## **2.8 Uitbreidingen**

Hoewel in dit hoofdstuk een flinke hoeveelheid van de basis van het model is beschreven, zijn er altijd verbeteringen mogelijk. Twee van deze uitbreidingen worden in het volgende hoofdstuk behandeld en hebben betrekking op de rating transities en afhankelijkheid tussen defaults.



## 3. Modellering: Uitbreiding

### 3.1 Inleiding

De modellering in het vorige hoofdstuk stelt men in staat om een werkende simulatie op te bouwen. Maar zoals bij elk onderzoek en bij elke modellering is er altijd ruimte voor verbetering. In dit hoofdstuk gaan we de basis van het vorige hoofdstuk verfijnen en uitbreiden. Te beginnen met het introduceren van enkele assumpties.

### 3.2 Assumpties

Bij het opstellen van het model zijn de volgende assumpties gemaakt:

*De default van huurders vertoont een grotere mate van afhankelijkheid naarmate wordt gekeken naar huurders in lagere rating klassen.<sup>1</sup>*

Dit is gebaseerd op het feit dat huurders die er economisch slecht aan toe zijn, minder buffer hebben om tegenvallers te verwerken dan huurders die er economisch sterk voor staan. Wanneer een huurder met economisch zwakke status een hoge defaultrate heeft, kan dit wijzen op negatieve economische factoren. Dit werkt sterker in op de economisch zwakkere huurders, waardoor de afhankelijkheid tussen huurders in de zwakke rating klassen groter is dan bij solidere rating klassen. Dit kan zich bijvoorbeeld, maar niet noodzakelijk, uiten in een hogere correlatiecoëfficiënt.

*Failliet gaan van een huurder is proxy voor huurders default.*

Wanneer een huurder niet op tijd of niet volledig aan zijn (maandelijkse) betalingsverplichting kan voldoen wordt dit delinquency genoemd. Een huurder wordt bij delinquency niet altijd direct uit zijn huureenheden gezet, omdat een verhuurder nog vertrouwen kan hebben dat een huurder zich hiervan herstelt en de achterstallige en toekomstige betalingsverplichtingen kan voldoen. Dit kan voor de verhuurder aantrekkelijker zijn dan de huurder bij de eerste wanbetaling uitzetten met een zekere leegstand te worden geconfronteerd. Natuurlijk kan het ook zo zijn dat de huurder zich niet herstelt, vervolgens een x aantal maanden delinquency vertoont, en daarna in huurders default gaat. Dit proces is niet eenduidig te modelleren omdat verhuurders hun eigen criteria hanteren. In dit model gaan we er vanuit dat faillissement gelijk staat aan huurders default. Over faillissement van bijvoorbeeld bedrijven is vaker data beschikbaar dan over huurders default van bedrijven. Gegevens over delinquency is van gedetailleerder niveau dan gegevens over huurders default en is slecht

beschikbaar. Merk op dat default, als in het failliet gaan van een bedrijf, in principe minder snel voorkomt dan huurders default. Immers, huurders default kan al optreden voor een definitief faillissement, bijvoorbeeld door een tegenvallende inkomstenperiode met huurwanbetalingen.

*Geen één-op-één relaties tussen bedrijven.*

In dit model gaan we er vanuit dat er geen directe relaties zijn tussen bedrijven. In de praktijk kan dit verband er bijvoorbeeld zijn in de computerbranche. Bedrijf X levert laptops en alleen bedrijf Y levert een cruciaal onderdeel hiervoor. Wanneer bedrijf Y in default gaat heeft dit direct gevolgen voor de bedrijfsvoering en dus defaultkans van bedrijf X. Er is gekozen om deze directe relaties te negeren omdat dit alleen zinnig onderzocht kan worden op behoorlijk gedetailleerd niveau met specifieke kennis en informatie over bedrijven. Het verband tussen de default van bedrijven wordt hiermee veroorzaakt door bijvoorbeeld de gedeelde economische achtergronden.

### 3.3 Uitbreiding 1: Rating Transities (Gedurende non-default periodes)

#### 3.3.1 Modelling van rating transities

→	AAA	AA	A	BBB	BB	B	CCC	Default	Σ
AAA	66.26	22.22	7.37	2.45	0.86	0.67	0.14	0.02	100
AA	21.66	43.04	25.83	6.56	1.99	0.68	0.2	0.04	100
A	2.76	20.34	44.19	22.94	7.42	1.97	0.28	0.1	100
BBB	0.3	2.8	22.63	42.54	23.52	6.95	1	0.26	100
BB	0.08	0.24	3.69	22.93	44.41	24.53	3.41	0.71	100
B	0.01	0.05	0.39	3.48	20.47	53	20.58	2.01	100
CCC	0	0.01	0.09	0.26	1.79	17.77	69.94	10.13	100
Default	0	0	0	0	0	0	0	100	100

Tabel 1: Eén-jaar transitiematrix gebaseerd op periode 1990-1995 (in %)<sup>2</sup>

Tabel 1 bevat de overgangskansen voor de rating van een bedrijf na één jaar. Zo zien we dat de rating van een bedrijf dat oorspronkelijk rating A had, na een jaar met een kans van 20,34% een rating AA heeft. De kansen in de transitiematrix zijn bepaald door over de periode 1990-1995 ieder jaar de transitiekansen te bepalen en deze te middelen over de jaren. Dus, met  $A^{(k)}$  als transitiematrix voor jaar  $k$ ,  $n$  als het aantal geobserveerde jaren,  $i$  als de beginrating in een jaar en  $j$  als de rating aan het begin van het volgende jaar:

$$\bar{A}_{i,j} = \frac{\sum_{k=1}^n A_{i,j}^{(k)}}{n}$$

$A_{i,j}^{(k)}$  is de transitiekans van rating  $i$  naar rating  $j$  voor jaar  $k$

( 12 )

De modellering van defaultperiodes in het vorige hoofdstuk lijkt op het eerste gezicht correct, maar bevat een potentiële fout. De defaultkans per jaar wordt namelijk veelal bepaald door het aantal huurders met een rating  $x$  te tellen en hiervan het gedeelte te nemen dat na een jaar in default gaat. Formeler: laat  $PD_r$  de kans zijn dat een huurder van rating  $r$  in een gegeven jaar in default gaat, en laat  $x_{r,i}$  met 0 of 1 aangeven of de  $i$ -de huurder met ratingklasse  $r$  in een gegeven jaar in default gaat.

$$PD_r = \frac{\sum_{i=1}^n x_{r,i}}{n}$$

( 13 )

Als een huurder gedurende zijn hele bestaan in dezelfde ratingklasse zit, volstaat de modellering tot dit punt. Maar huurders kunnen over een periode van meerdere jaren meermaals van ratingklasse muteren als gevolg van positieve of negatieve economische ontwikkelingen. De defaultkans zoals hierboven beschreven is kan dan niet voor de gehele periode worden aangehouden (en analoog de defaultrate die uit de defaultkans wordt berekend ook niet). De rating transities van huurders moeten in de modellering verwerkt worden. Organisaties als Standard & Poors hebben transitiematrices opgesteld met daarin de kansen dat een huurder na een jaar van een rating  $x$  overgaat in een rating  $y$ . Hiermee kan het model aangepast worden met het volgende algoritme.

1. Laat de tijd tot default  $T$  gelijk staan aan 0.
2. Zoek de defaultkans  $PD_r$  voor de huurder met ratingklasse  $r$  op.
3. Bereken de defaultrate  $\lambda$  aan de hand van  $PD_r$ .
4. Doe een trekking  $\Delta T$  uit de exponentiële verdeling met defaultrate  $\lambda$ . Dit is de additionele tijd tot default.
5. Check of default binnen een jaar gebeurt of na een jaar gebeurt.  
 Indien  $\Delta T < 1$ :  $T := T + \Delta T$  [EINDE algoritme]  
 Indien  $\Delta T \geq 1$ :  $T := T + 1$
6. Nieuwe ratingklasse  $r$  voor de huurder bepalen (zie verderop voor volledige specificatie). Terug naar stap 2.

Om berekeningstijd uit te sparen kan ervoor worden gekozen om in plaats van defaultkansen direct alleen met defaultrates te werken. Deze zullen dan wel van tevoren gecalculeerd moeten worden. In plaats van defaultkans opzoeken en berekenen is er dan enkel één stap, namelijk defaultrate opzoeken. Dit voorkomt veelvuldig herberekenen van defaultrates, iets wat bij veelvuldig simuleren een redundante operatie is ten opzichte van eenmalig berekenen van alle defaultrates en simpelweg opzoeken.

**Stap 6** van bovenstaand algoritme is bewust niet uitgewerkt om de kern van dat algoritme duidelijk naar voren te laten komen, en niet hoofdlijnen en “details” door elkaar te laten lopen. Hieronder volgt een specificatie van de stappen die impliciet in stap 6 opgenomen dienen te worden. Laat  $M$  de matrix zijn van  $n$  bij  $n$  ( $n$  mogelijke ratings), waarbij de  $n$ -de rij en de  $n$ -de kolom horen bij de default rating. Laat verder de huidige status van de huurder in de  $i$ -de rij staan.



1. Stel een vector  $v$  ter lengte  $n-1$  op waarbij:

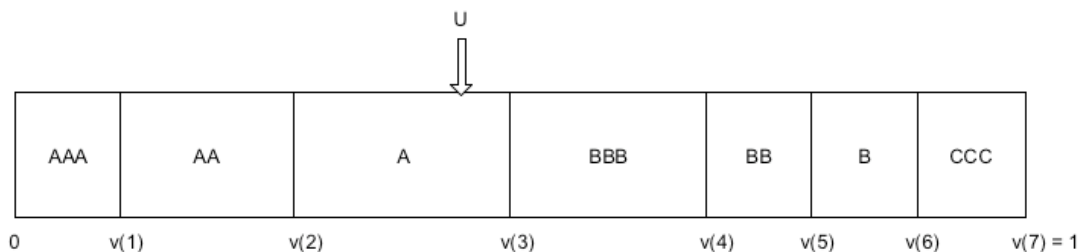
$$v(1) = M(i,1) + \frac{M(i,1)}{1 - M(i,n)} M(i,n)$$

$$v(k) = v(k-1) + M(i,k) + \frac{M(i,k)}{1 - M(i,n)} M(i,n) \quad \text{voor } k = 2, \dots, n-1$$

( 14 )

2. Doe een willekeurige trekking  $U$  uit de standaard uniforme verdeling.
3. Laat  $i = 1$ . Wanneer  $U \leq v(i)$ , dan is de volgende rating  $r$  van de huurder, de rating behorende bij index  $i$ . Indien  $U > v(i)$ , laat  $i := i + 1$  en herhaal het zoekproces totdat rating  $r$  gevonden is.
4. Ga terug naar stap 2.

Een nieuwe rating is volgens het algoritme alleen nodig wanneer er geen default optreedt in een jaar. Dus wat er in feite gebeurt voor de vector  $v$  is het berekenen van cumulatieve overgangskansen met een correctie voor de kans op transitie naar default rating. De kans op transitie naar default rating wordt herverdeeld onder de overige transitiekansen, waarbij de onderlinge verhouding tussen deze overige transitiekansen gehandhaafd blijft. Intuïtief kan  $v$  gezien worden als een balk met lengte 1, die wordt opgehakt in stukken corresponderend met de transitiekansen (Fig. 6).



**Fig. 6: Intuïtieve weergave vector  $v$**

Merk op dat het mogelijk is dat de defaultkans bepaald is uit jarenlange observatie. Huurders met een aanvangstrating  $r$  kunnen jarenlang zijn gevolgd totdat iedereen die oorspronkelijk van rating  $r$  was in default is gegaan. Hierbij zijn de ratingtransities impliciet meegenomen. Aan de hand van gerealiseerde defaultduraties kan een defaultrate worden geschat. Aan de hand van de defaultrate kan vervolgens een random exponentieel getal worden getrokken. Indien de defaultkans is bepaald uit jarenlange observatie van een groep met dezelfde rating, dan is rating transitie zoals in deze paragraaf is getoond niet vereist.

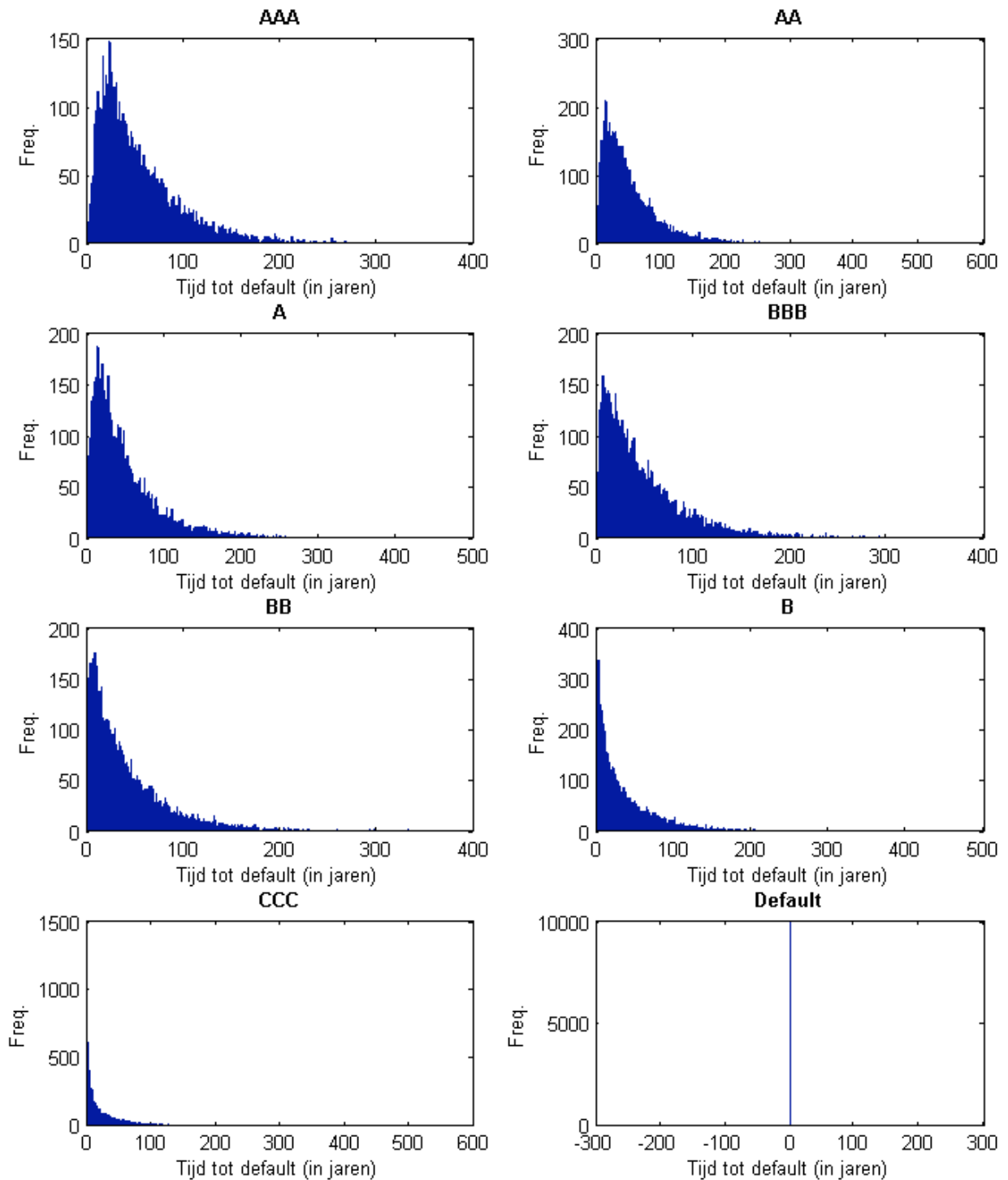
De in dit hoofdstuk beschreven transitie matrix is slechts een van meerdere manieren om een één-jaars transitie matrix te kiezen. Andere keuzen zijn onder andere het kiezen van een random één-jaars transitie matrix uit de beschikbare transitie matrices.

In de praktijk is het zo dat een langere periode van goede ratings leidt tot een verhoogde kans op goede ratings. Ofwel, wanneer een bedrijf jarenlang zijn A rating behoudt, dan is de kans dat het een volgende jaar ook een A rating heeft groter dan tabel 1 aangeeft. Er is echter besloten om deze lastige verfijning niet in het huidige model te onderzoeken.

### **3.3.2 Analyse van rating transitie modellering**

#### **Histogrammen**

De modellering uit het voorgaande paragraaf is in Matlab geïmplementeerd voor nadere analyse. Voor iedere initiële ratingklasse is 10.000 maal een simulatie uitgevoerd om inzicht te verkrijgen hoe de tijd tot default is verdeeld. Allereerst maken we histogrammen (Fig. 7) om een globaal beeld te krijgen van de simulatiedata. De hoogste ratingklassen vertonen een vorm die lijkt op de kansdichtheidsfunctie van een exponentiële verdeling. Vanaf klasse B lijkt de vorm wat minder op een exponentiële verdeling, maar valt deze toch nog terug te herkennen. En het triviale geval van klasse “Default” leidt tot default tijd van 0 (geen huurder meer, dus geen default).

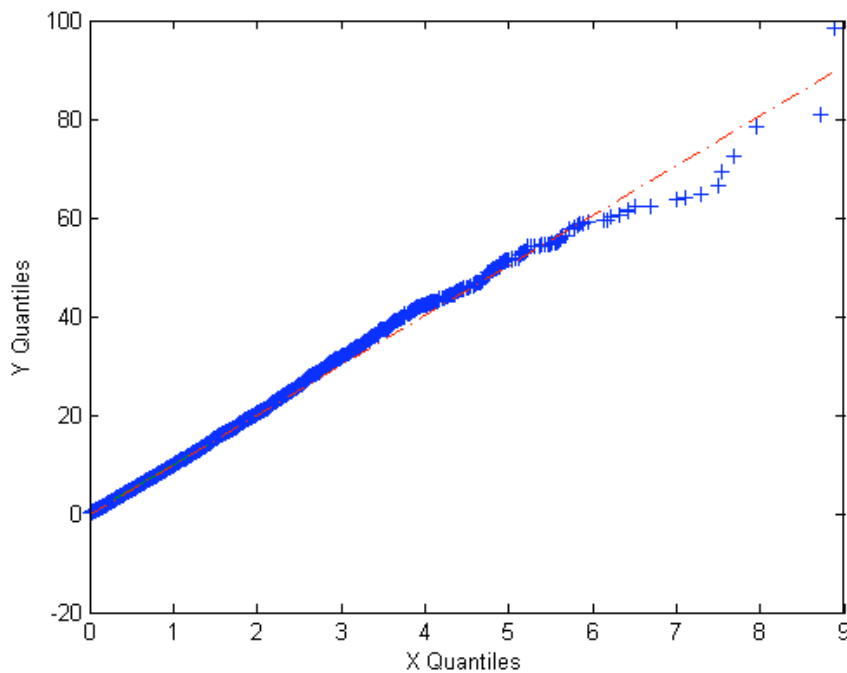


**Fig. 7: Histogrammen van default tijden**

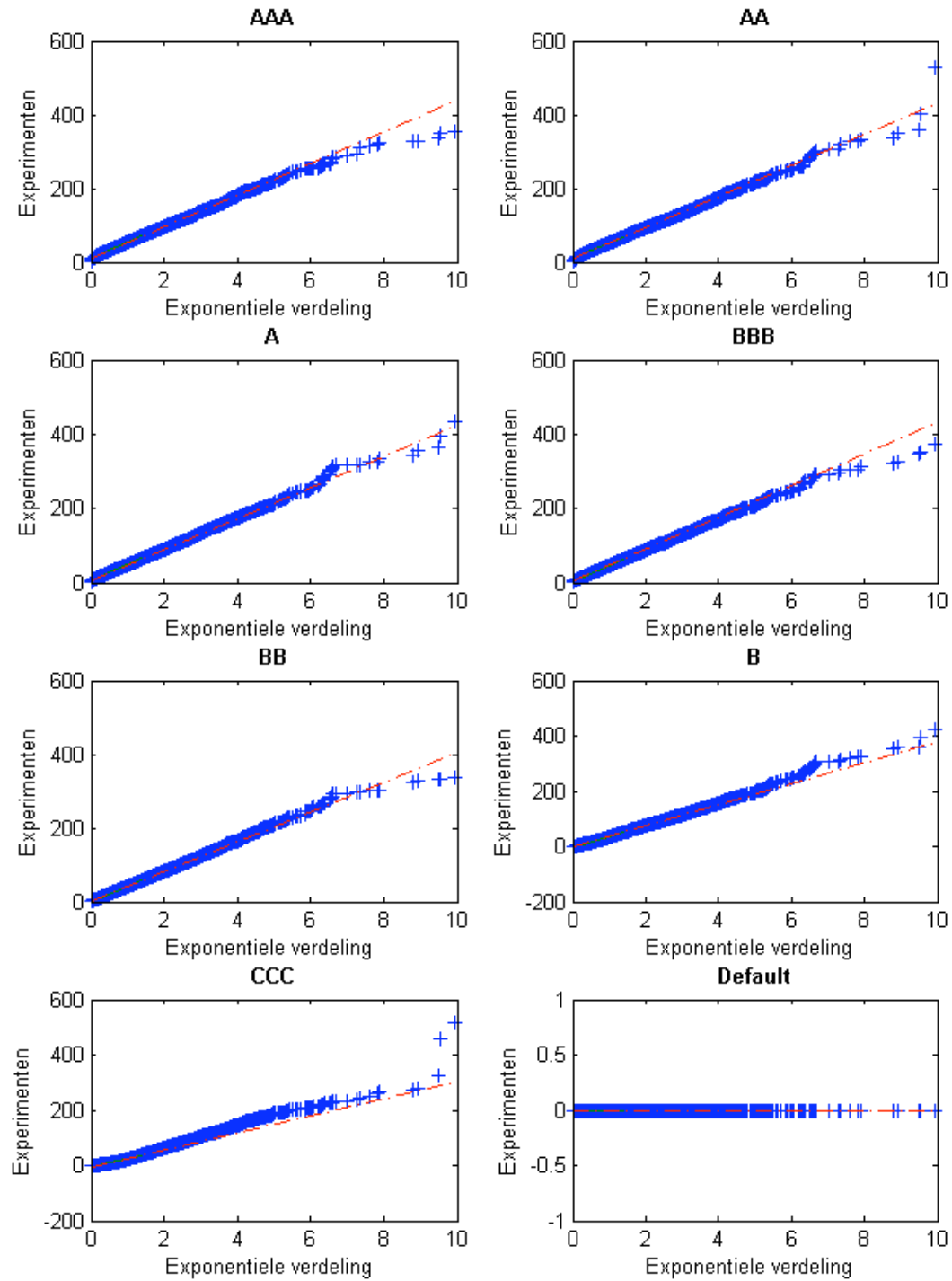
**QQ-Plots (Quantile/Quantile-Plots)**

Om het vermoeden dat er een exponentiële verdeling in het spel is te checken maken we QQ-plots van de simulatiedata tegenover random getrokken getallen uit een lid van de exponentiële familie (Fig. 9). Wat QQ-plots doen is twee vectoren ordenen, in dit geval een vector met simulatie getallen en een vector met random getrokken getallen uit een lid van de exponentiële familie. Vervolgens worden paren gevormd vanuit de geordende vectoren, dus kleinste van de ene vector in combinatie met de kleinste van de andere vector, ..., grootste van de ene vector in combinatie met de grootste van de andere vector. Indien beide data uit eenzelfde kansverdelingfamilie komen dan verwacht men van een plot van deze paren een rechte lijn te zien. Bij onze QQ-plots blijkt vrijwel overal een vrij rechte lijn uit te komen. Van de scatter die we zien constateren we dat deze een marginaal aandeel betreffen van grofweg 10 uit 10000 punten of minder, en ook valt ons op dat de scatter voorkomt bij hoge waarden. Dus deze QQ-plots zouden het modelleren van de tijd tot default als een exponentiële verdeling niet direct verwerpen.

Hieronder een QQ-plot van twee vectoren gevuld, ieder gevuld met 10.000 random exponentieel getrokken getallen (Fig. 8). Hierin is te zien dat er zelfs wanneer je weet dat het twee leden van de exponentiële familie betreft, er in het midden zeer lichte afwijking kan bevinden en aan het eind een aantal sterkere afwijkingen kunnen bevinden. Dit laatste licht waarschijnlijk aan het feit dat er aan het eind weinig waarnemingen zullen zijn gezien de kans op extreem hoge waarnemingen zeer laag zijn (wat mede is te zien door de lagere dichtheid van de punten). En bij minder waarnemingen hebben afwijkingen sneller invloed.



**Fig. 8: QQ-Plot, Exponentieel(1) tegen Exponentieel (0,1)**



**Fig. 9: QQ-plots (Experimentele data versus exponentiële random getallen)**

**Statistische Tests**

We voeren statistische tests uit om te onderzoeken of onze experimentele data significant verschilt van de exponentiële verdeling. De hypothese die gevormd is aan de hand van de histogrammen en de QQ-plots is dat de modellering van de rating transitie ook kan worden uitgevoerd aan de hand van een exponentiële verdeling. Indien dit vermoeden juist blijkt zou dit zeer welkom zijn omdat een random exponentieel getal trekken minder bewerkelijk is dan het algoritme dat is opgesteld. Het zou significant schelen in simulatietijd.

De toets die we gaan gebruiken is de (two-sample) Kolmogorov-Smirnov toets. Deze toets accepteert twee vectoren, ieder afkomstig van een eigen verdeling, en gebruikt de data van deze vectoren om te testen of beide vectoren dezelfde distributiefunctie hebben. Wanneer we in ons geval onze experimenteel verkregen data vergelijken met random exponentieel getrokken data (met gemiddelde gelijk aan gemiddelde van experimenteel verkregen data), en de test verwerpt de nul-hypothese (beide vectoren aannemelijk uit eenzelfde verdeling) niet, dan kunnen we zeggen dat onze experimenteel verkregen data aannemelijk is gedistribueerd als een exponentiële verdeling. De Kolmogorov-Smirnov toets creëert stapsgewijze distributiefuncties op basis van de data, en bepaalt op basis van de grootste afstand tussen deze twee of ze wel of niet gelijk zijn. Formeler, laat  $\underline{x}$  en  $\underline{y}$  twee vectoren zijn met respectievelijke stapsgewijze empirische distributiefuncties  $F_x(c)$  en  $F_y(c)$ :

$$H_0 := \underline{x} \text{ en } \underline{y} \text{ hebben aannemelijk dezelfde verdeling}$$

$$H_1 := \underline{x} \text{ en } \underline{y} \text{ hebben NIET dezelfde verdeling}$$

Toetsingsgrootheid:

$$\max_c |F_x(c) - F_y(c)|$$

( 15 )

In de volgende tabel staan de resultaten van de Kolmogorov-Smirnov toets.

<b>Rating</b>	<b>Hypothese</b>	<b>P-value</b>	<b>Toetsingsgrootheid</b>
<b>AAA</b>	1	2.6664E-60	0,1302
<b>AA</b>	1	9.2950E-44	0,1147
<b>A</b>	1	9.3585E-27	0,0962
<b>BBB</b>	1	6.6238E-02	0,0596
<b>BB</b>	1	3.5916E-01	0,0330
<b>B</b>	1	2.4367E-03	0,0623
<b>CCC</b>	1	2.1887E-99	0,1610
<b>Default</b>	0	1.0000E+00	0.0000E+00

**Tabel 2: Resultaten Kolmogorov-Smirnov Toets**

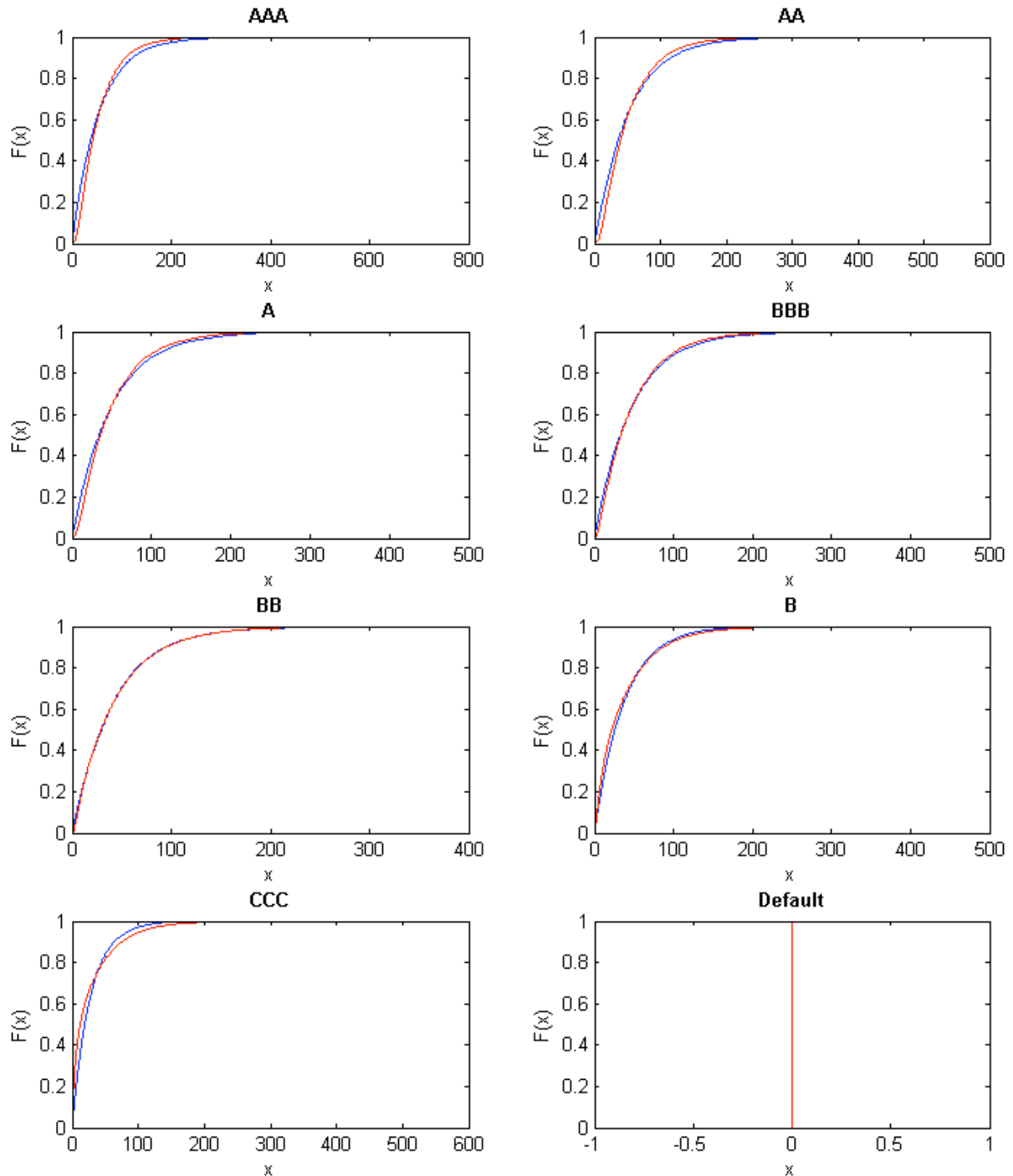
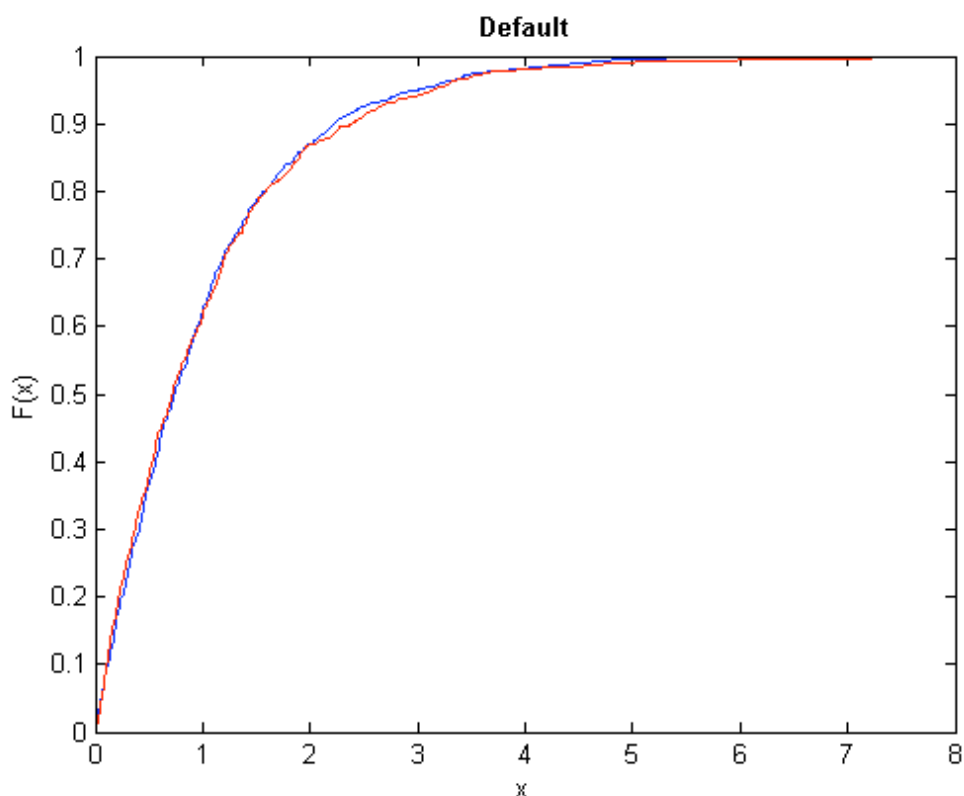


Fig. 10: Stapsgewijze distributiefuncties (rood=experimenteel, blauw=referentie)

De toets keurt in dit geval onze nul-hypothese af (afgezien van rating “Default”, maar dat is een triviale case). Om een beter inzicht te krijgen zijn ook de bijhorende plots gemaakt, immers, de Kolmogorov-Smirnov vat haar hele oordeel samen in een scalar en verradt weinig over de “vorm” van de afwijking.

Voor rating BB lijkt de fit erg goed te zijn (Fig. 10), dit blijkt ook uit de tabel met toetsingswaardes. Rating B en BB vertonen ook vrij goede fit, maar voor de overige ratings is de fit niet optimaal. Ter referentie hieronder (Fig. 11) een fit van twee rondom exponentieel getrokken vectoren (met dezelfde parameterwaarde). De distributie functies zijn in dat geval visueel nauwelijks onderscheidend



**Fig. 11: Stapsgewijze distributiefuncties (“perfecte” fit)**

### **Inzoomen op 0 t/m 50 jaar**

Op dit moment stel ik dat afwijking bij de waardes na grofweg 50 jaar van minder belang, of voor de huidige modellering zelfs irrelevant zijn. Immers, op vastgoed gebaseerde beleggingsproducten hebben meestal slechts een looptijd van een tot enkele decennia. Dus het gaat voornamelijk om de fit van 0 tot 50 jaar.

We plotten nogmaals, maar ditmaal ingezoomd op de periode van 0 tot 50 jaar (Fig. 12).



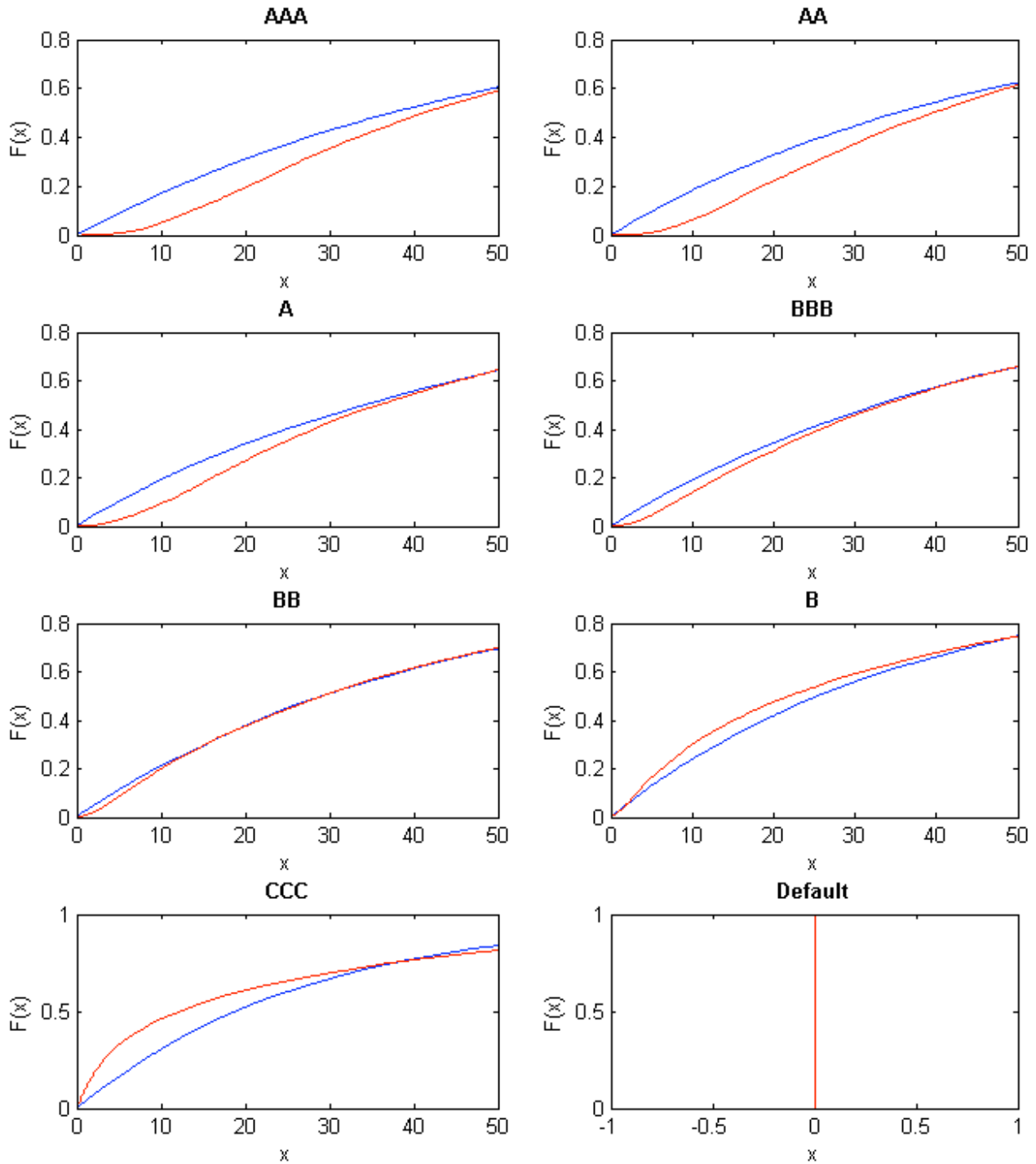
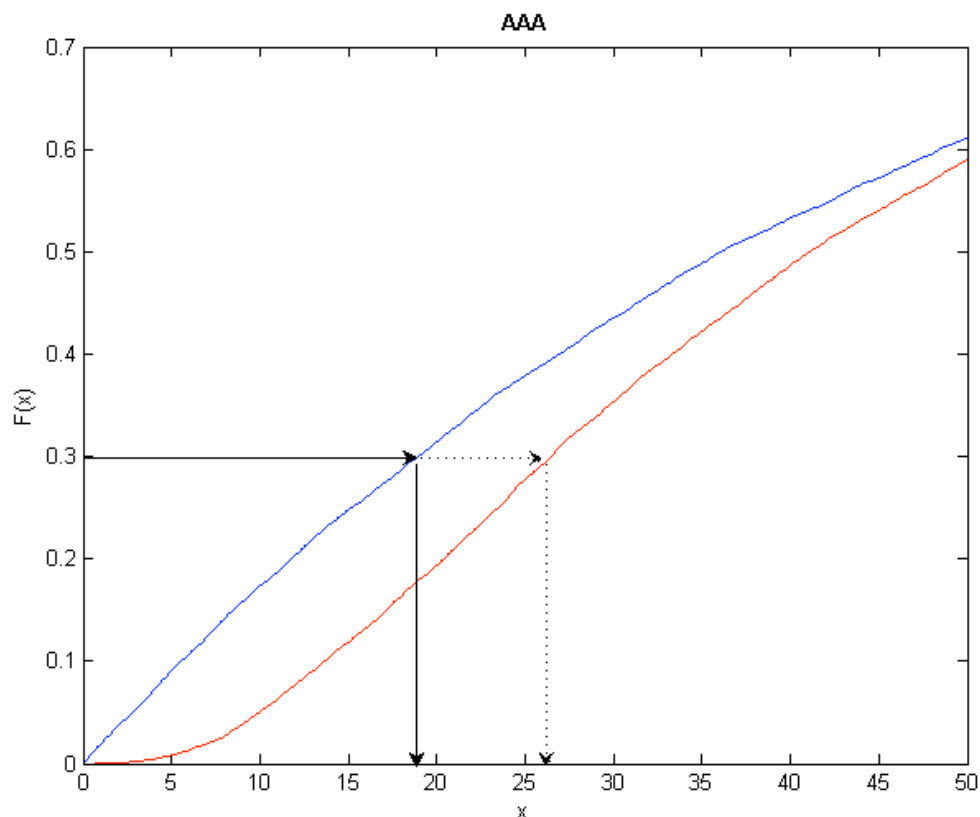


Fig. 12: Stapsgewijze distributiefuncties (ingezoomd op domein van 0 t/m 50)

Wat opvalt is dat de plot van rating BB ook ingezoomd op het domein van 0 t/m 50 jaar nog steeds erg goed fit (aan een exponentiële verdeling). Bij ratings AAA t/m BBB genereert een exponentiële verdeling een te lage getal, en bij ratings B en CCC geeft een exponentiële verdeling te hoge getallen. Vooral als de rating zich verder af bevindt van BB lijkt het probleem te verergeren. Wanneer we kijken naar rating AAA (Fig. 13), dan zien we bijvoorbeeld dat random number generatie aan de hand van een exponentiële verdeling flink kan schelen van onze gewenste distributie (die behorende bij de experimenteel verkregen getallen). Via de inverse transformatie methode zouden we bij een invoer van  $u=0.3$  uitkomen op ongeveer 19 jaar, daar waar volgens onze experimenten een waarde van 26 jaar gewenst is; Een zeer groot verschil. Voor BB is het verschil marginaal, dus deze kan goed gemodelleerd worden met een exponentiële verdeling (op het domein 0 t/m 50 jaar in ieder geval). Maar voor de overige ratings kunnen op dit moment nog niet direct een snelle simulatiemethode geven.

Noot: in paragraaf 4.3.1 is weliswaar een algoritme gegeven waarmee gesimuleerd kan worden, maar deze kan niet direct vertaald worden in termen van simulatie met behulp van inverse transformatie methode. Het belang hiervan zal later blijken wanneer we afhankelijkheid tussen default tijden modelleren.



**Fig. 13: Significante marge bij random nummer generatie**

### Een geheugenintensieve methode

Om default tijden te genereren bij de verschillende ratings (gebaseerd op een random gegenereerd standaard uniform getal) zou de volgende methode kunnen worden gebruikt:

- 1 We trekken een groot aantal, zeg 10.000, random getallen volgens het recept uit paragraaf 4.3.1 getallen.
- 2 Maak een stapsgewijze distributiefunctie aan de hand van de data uit stap 1.
- 3 Met deze empirische distributiefunctie kunnen we de werking van de inverse transformatie methode nabootsen.

Hieronder volgt een uitgebreidere uitleg over stap 3. Laat hierbij  $\underline{x}$  de vector zijn gevuld met de 10.000 random getallen van stap 1:

- 1 Allereerst sorteren we de vector  $\underline{x}$ .

Er zijn in dit geval 10.000 random getallen, dus iedere getal in de vector, van klein naar groot, representeert een toename van de cumulatieve kans van  $1/10.000$ . Bij het  $i$ -de getal in de vector is de waarde bij de distributiefunctie dus  $i/10.000$ .

- 2 Trek nu een getal  $u$  uit de standaard uniforme verdeling.
- 3 Bepaal welke index/indices horen bij de waarde van  $u$  en bepaal op basis hiervan een waarde voor de random trekking.

Wanneer  $u=0,0005$  dan is onze random trekking het vijfde getal uit de vector, dus  $\underline{x}(5)$ . Wanneer de waarde van  $u$  gedetailleerder is dan het aantal indices toelaat dan is het nodig om interpolatie toe te passen. Dus stel de waarde van  $u=0.00023$ , dit valt tussen de tweede en derde index, dan wordt de waarde van de random trekking ( $y$ ):

$$y = \underline{x}(2) + 0.00003 \cdot \frac{\underline{x}(3) - \underline{x}(2)}{0.0001}$$

Of in algemenere termen, waarbij  $n$  het aantal random getallen in de vector  $\underline{x}$  is.

$$\begin{aligned} p &= 1/n \\ i &= \left\lfloor \frac{u}{p} \right\rfloor = \lfloor u \cdot n \rfloor \\ y &= \underline{x}(i) + (u - i \cdot p) \frac{\underline{x}(i+1) - \underline{x}(i)}{p} \\ &= \underline{x}(i) + (u - i \cdot 1/n)(\underline{x}(i+1) - \underline{x}(i))n \\ &= \underline{x}(i) + (u \cdot n - i)(\underline{x}(i+1) - \underline{x}(i)) \end{aligned}$$

(16)

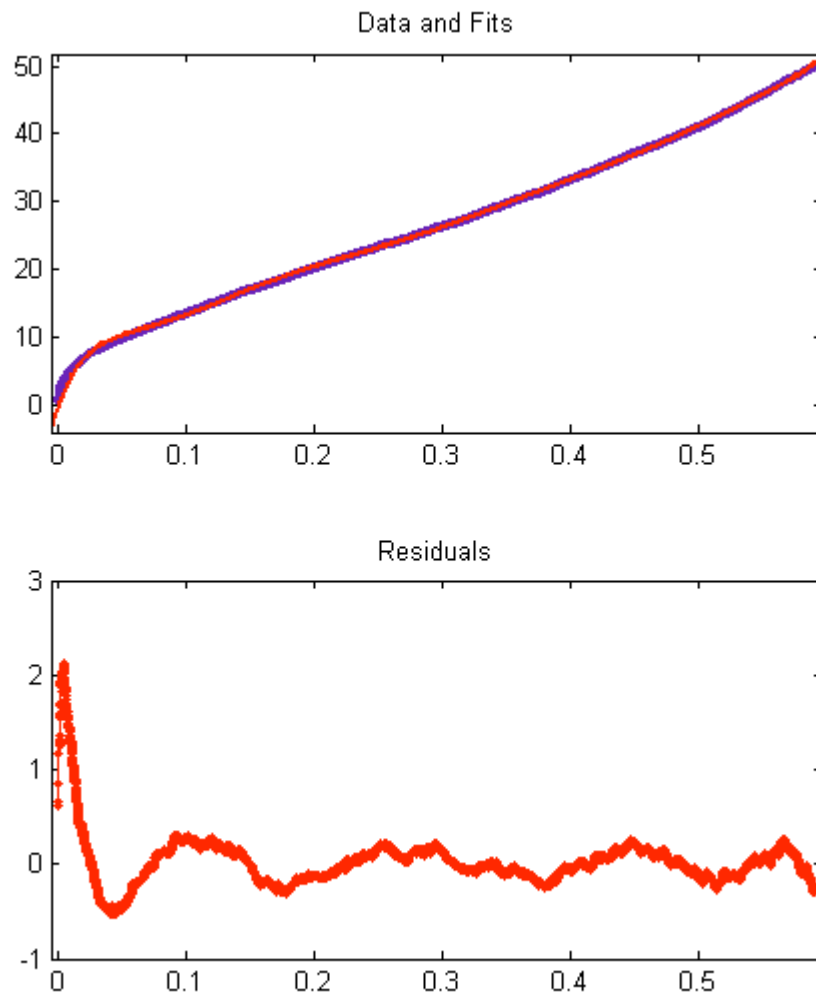
Deze methode vereist eenmalig tijd om een grote hoeveelheid random getallen te trekken (volgens het algoritme in paragraaf 4.3.1) om een stapsgewijze distributiefunctie te kunnen construeren voor iedere ratingklasse. Hoe fijner we de distributiefunctie willen hebben, hoe meer random getallen we moeten trekken. Het opslaan van de getallen kost geheugen, maar wellicht is er een efficiëntere methode. Dit wordt onderzocht in het volgende stuk.

### **Functie fitten aan empirische distributiefunctie**

Misschien is het mogelijk om een functie te vinden die het werk van de inverse transformatie beter kan uitvoeren dan de verdelingsfunctie van een exponentiële verdeling; Een functie die dus beter fit aan onze experimentele data. We gaan ons bij het fitten van deze functie concentreren op het gedeelte van 0 t/m 50 jaar. De fit van de jaren erna is in de context van onze toepassing van minder belang. Ook is het niet nodig dat de functie de eigenschappen bezit van een distributiefunctie op het gehele domein, als dit maar het geval is in de periode die we beschouwen, dus continue en strikt stijgend met een cumulatieve kans 0 voor een default tijdspanne van 0 tot en met 0. Met behulp van deze functie kunnen we aan de hand van haar inverse random trekkingen verrichten. Belangrijk: Het is uiteraard niet meer mogelijk om berekeningen uit te voeren met de functie als zijnde een distributiefunctie, bijvoorbeeld gemiddelde berekenen of iets dergelijks. Maar daarvoor gebruiken we deze functie in de huidige context ook niet voor.

Bij de Kolmogorov-Smirnov werd gekeken naar het verschil in kans voor een gegeven waarde van de gemodelleerde stochast (default tijd). Omdat we nu geïnteresseerd zijn hoeveel onze random trekking, vanuit een nog te formuleren functie, afwijkt van de experimentele data, verwisselen we de stochast en de cumulatieve kans van as. Dat representeert beter de fout minimalisatie waarmee we bezig zijn alsook de werking van random trekkingen (invoer  $F(x)$ ~std uniform op horizontale as en uitvoer random trekking  $x$  op verticale as).

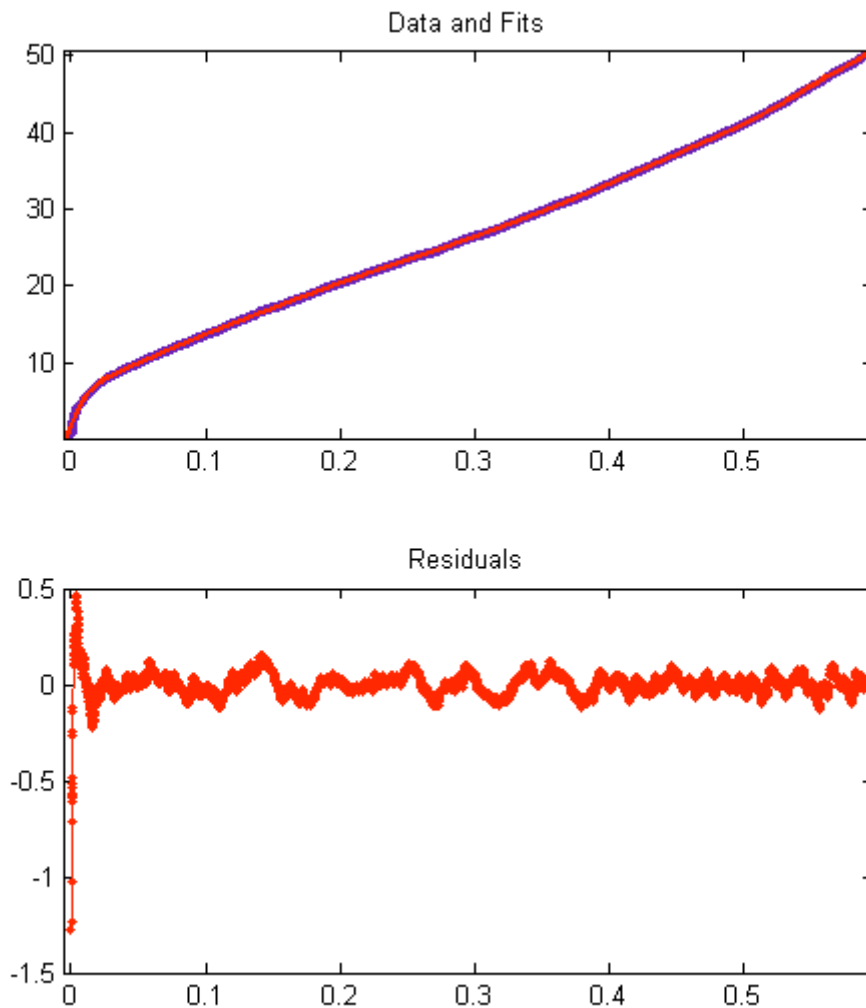
Matlab bevat een curve fitting toolbox die onderzoek naar functie fitting toelaat. Experimenterend met verschillende functies voor de hoogste ratingklasse vinden we dat veel in staat zijn een aardig fittende functie te vinden, maar moeite hebben met het sterk concave gedeelte bij de lage waarden. Ook valt op dat niet in te stellen is om een functie te forceren door de oorsprong. Behalve voor een fit door een polynomiale functie, waarbij er de mogelijkheid is om de waarde van de constante term gelijk aan nul te stellen.



**Fig. 14: Benadering door 9<sup>e</sup> graads polynoom, door oorsprong (AAA rating)**

Een benadering door een 9<sup>e</sup> graads polynoom is zeer aardig te noemen (Fig. 14). De fout blijft zeer marginaal voor hogere waarden. De residual plot dient ook zo gelezen te worden dat foutmarge aan het eind nog meer meevalt bij een vrij constant verloop van de residuals, zoals we in het figuur zien, omdat de uitkomstwaarde alsmaar toeneemt; Gelijkblijvende afwijking uitgedrukt als percentage van een steeds grotere waarde neemt af. De fout aan het begin is echter zeer hoog, oplopend tot een verschil van twee jaar. Dit is ontoelaatbaar omdat het bij een trekkingswaarde van ongeveer 5 jaar, een fout van tientallen procenten bedraagt.

NB. Merk op dat het fitten van de functie gebeurt voordat de simulatie wordt gedraaid, bijvoorbeeld met behulp van Matlab en dus niet een proces is dat tijdens simulatie gebeurt. Dit zou anders ook veel te intensief zijn voor de simulator, resulterend in hoge runtijden.

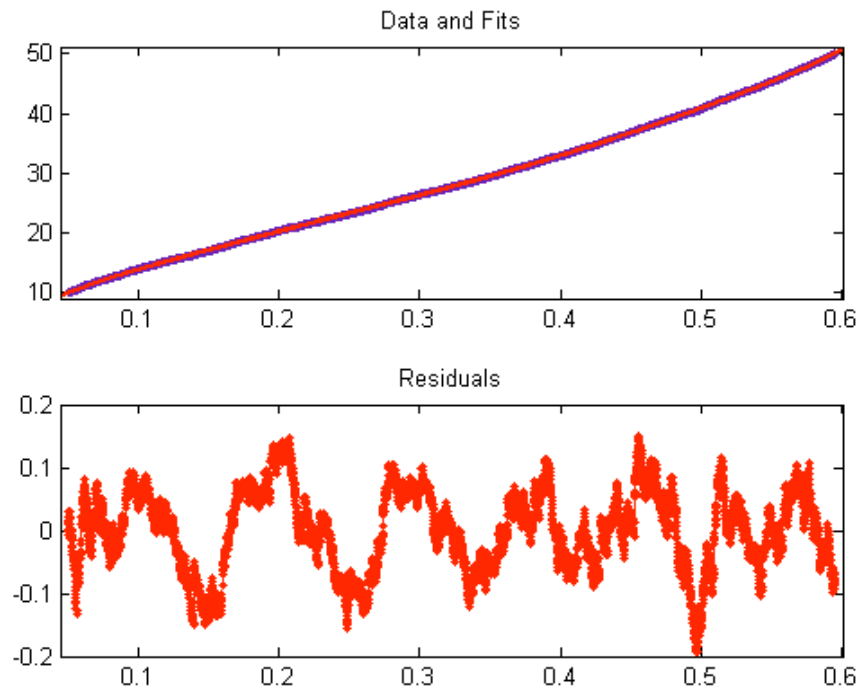


**Fig. 15: Benadering door Fourier reeks (AAA rating)**

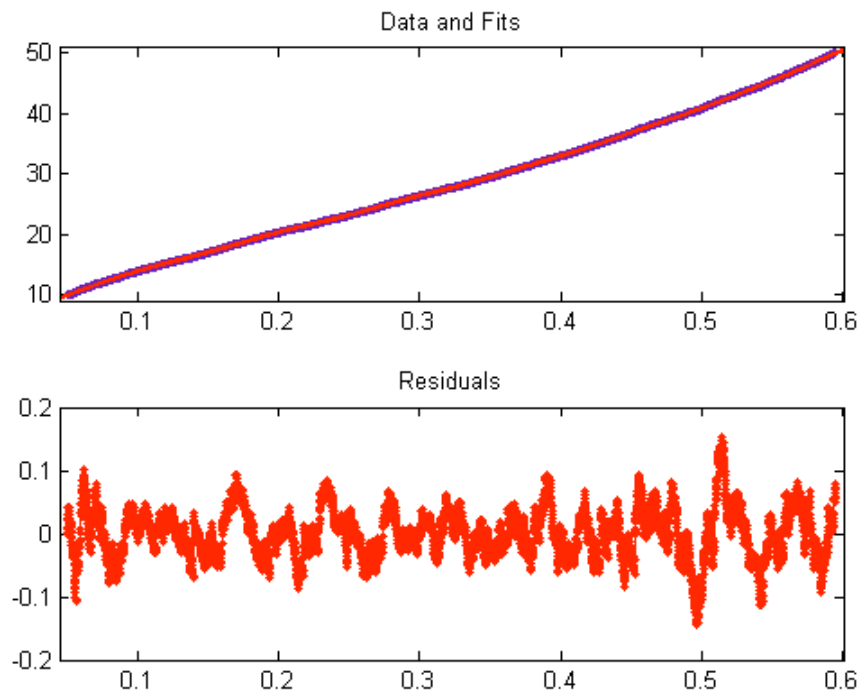
Benadering door een Fourier reeks (grofweg: een compositie van sinus en cosinus functies) geeft een nog betere fit aan de functie (Fig. 15). Toch zien we dat in een klein, maar zeker wel voorkomend, aantal gevallen de afwijking een half jaar tot iets meer dan een jaar kan bedragen.

Overige functies fitten niet zo goed als een polynoom of een Fourier reeks en zijn hier buiten beschouwing gelaten. Het probleem van negatieve waarden kan nog opgevangen door te stellen dat de waarde minimaal 0 moet bedragen, maar de afwijking voor uiterst lage waarden voor  $u$  blijft een zorg. Wellicht is het beste om een hybride functie op stellen: Dus voor waarden van  $u$  tussen zeg 0 en 0.05 de “geheugenintensieve methode” (die op dit minieme interval niet meer geheugenintensief is) en voor waarden boven 0.05 bijvoorbeeld benadering door een Fourier reeks. We herhalen de functiefitting, maar ditmaal met exclusie van waarden vóór 0.05, bijkomstig effect is dat de fit op het kleinere interval potentie heeft om nog

beter aan te sluiten. De volgende figuren krijgen we te zien bij het aangepaste experiment (Fig. 16 en Fig. 17).



**Fig. 16: Benadering door 7<sup>e</sup> graads polynoom, vanaf 0.05 (AAA rating)**



**Fig. 17: Benadering door Fourier reeks, vanaf 0.05 (AAA rating)**

De Fourier reeks, alsook de 7<sup>e</sup> graads polynoom geven goede benaderingen op het interval, waarbij de benadering door de Fourier reeks iets beter presteert dan de benadering door de polynoom. Deze methodiek doortrekkende naar de andere ratings levert een vergelijkbaar beeld. Hieronder volgen de gevonden benaderingen voor elke rating, gevolgd door een serie grafieken die de afwijking representeren.

$$\begin{array}{c}
 \textit{Rating AAA} \\
 x(u) = \begin{cases} 0, & u = 0 \\ \text{Lineaire Interpolatie,} & 0 < u < 0.05 \\ -132.6u^4 + 293.4u^3 - 160.6u^2 + 93.9u + 5.669, & 0.05 \leq u < 0.5949 \\ 50, & u \geq 0.5949 \end{cases}
 \end{array}$$

$$\begin{array}{c}
 \textit{Rating AA} \\
 x(u) = \begin{cases} 0, & u = 0 \\ \text{Lineaire Interpolatie,} & 0 < u < 0.05 \\ 143.3u^4 - 90.1u^3 + 18.02u^2 + 61.83u + 5.824, & 0.05 \leq u < 0.6140 \\ 50, & u \geq 0.6140 \end{cases}
 \end{array}$$

$$\begin{array}{c}
 \textit{Rating A} \\
 x(u) = \begin{cases} 0, & u = 0 \\ \text{Lineaire Interpolatie,} & 0 < u < 0.05 \\ 68.01u^3 - 28.9u^2 + 61.45u + 4.528, & 0.05 \leq u < 0.6395 \\ 50, & u \geq 0.6395 \end{cases}
 \end{array}$$

$$\begin{array}{c}
 \textit{Rating BBB} \\
 x(u) = \begin{cases} 0, & u = 0 \\ \text{Lineaire Interpolatie,} & 0 < u < 0.05 \\ 63u^3 - 17.39u^2 + 53.81u + 2.923, & 0.05 \leq u < 0.6668 \\ 50, & u \geq 0.6668 \end{cases}
 \end{array}$$

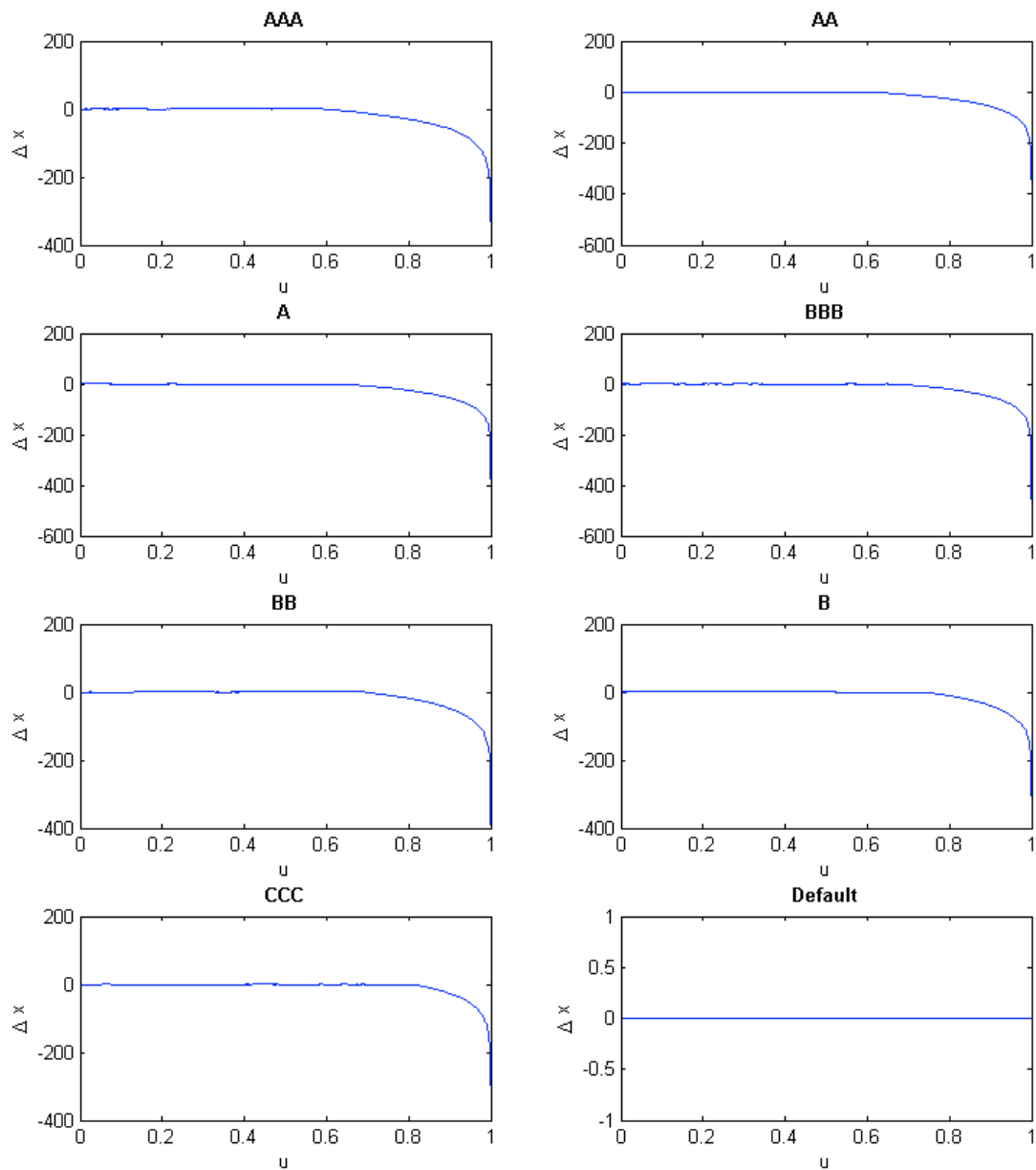


$$\begin{array}{c}
 \textit{Rating BB} \\
 x(u) = \begin{cases} 0, & u = 0 \\ \text{Lineaire Interpolatie,} & 0 < u < 0.05 \\ 53.81u^3 + 1.761u^2 + 42.01u + 1.575, & 0.05 \leq u < 0.6938 \\ 50, & u \geq 0.6938 \end{cases}
 \end{array}$$

$$\begin{array}{c}
 \textit{Rating B} \\
 x(u) = \begin{cases} 0, & u = 0 \\ \text{Lineaire Interpolatie,} & 0 < u < 0.05 \\ -193.8u^5 + 317.9u^4 - 103.1u^3 + 37.65u^2 + 24.08u + 0.5178, & 0.05 \leq u < 0.7415 \\ 50, & u \geq 0.7415 \end{cases}
 \end{array}$$

$$\begin{array}{c}
 \textit{Rating CCC} \\
 x(u) = \begin{cases} 0, & u = 0 \\ \text{Lineaire Interpolatie,} & 0 < u < 0.05 \\ 118.2u^4 - 65.29u^3 + 46.63u^2 + 2.761u^1 + 0.3449, & 0.05 \leq u < 0.8102 \\ 50, & u \geq 0.8102 \end{cases}
 \end{array}$$

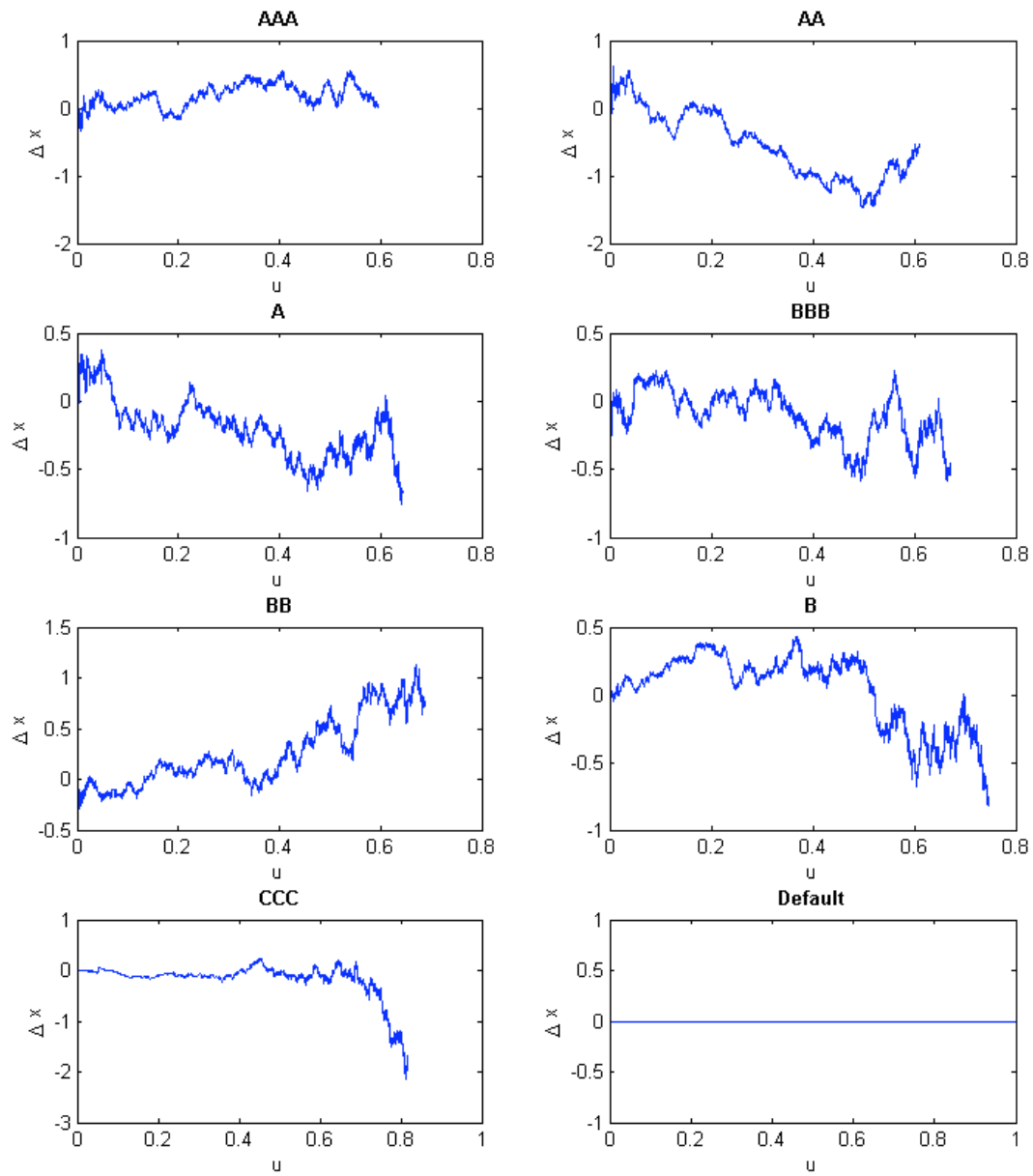
We testen de bruikbaarheid van deze gevonden recepten als volgt: we trekken voor iedere rating 10.000 random nummers volgens de benaderingen hierboven, en vergelijken dit met 10.000 random nummers getrokken volgens de ‘echte manier’. Beide vectoren sorteren we, en de grootte van de fout is het verschil in waarde van de vectoren (dus een vector ter lengte 10.000 die voor ieder benadering/echt-paar het verschil geeft.). De volgende resultaten zijn waargenomen (Fig. 18):



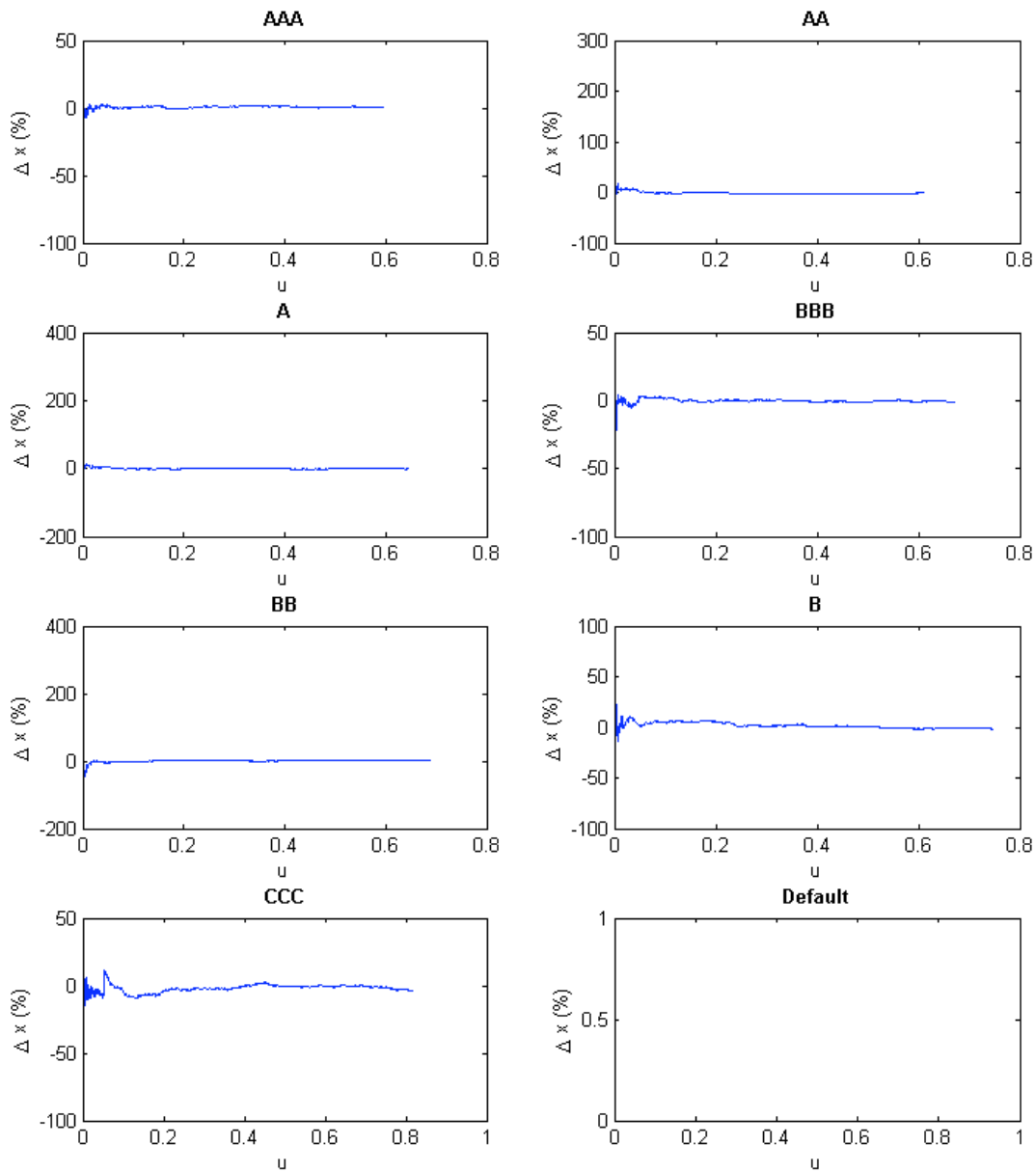
**Fig. 18: Hybride benadering op domein  $[0, 1]$  (Absolute fout)**

Het algehele plaatje ziet er goed uit. De fout loopt vanaf een bepaald punt zeer snel op, wat uiteraard wordt veroorzaakt doordat we de approximatie afkappen op 50 jaar. Zoals eerder vermeld is modellering boven de 50 jaar in de praktische situaties niet nodig. We zoomen nu in op dat gedeelte waarvoor de approximatie in de range  $[0, 50]$  jaar valt (Fig. 19 en Fig. 20).

## Huurders Default – Risico modellering

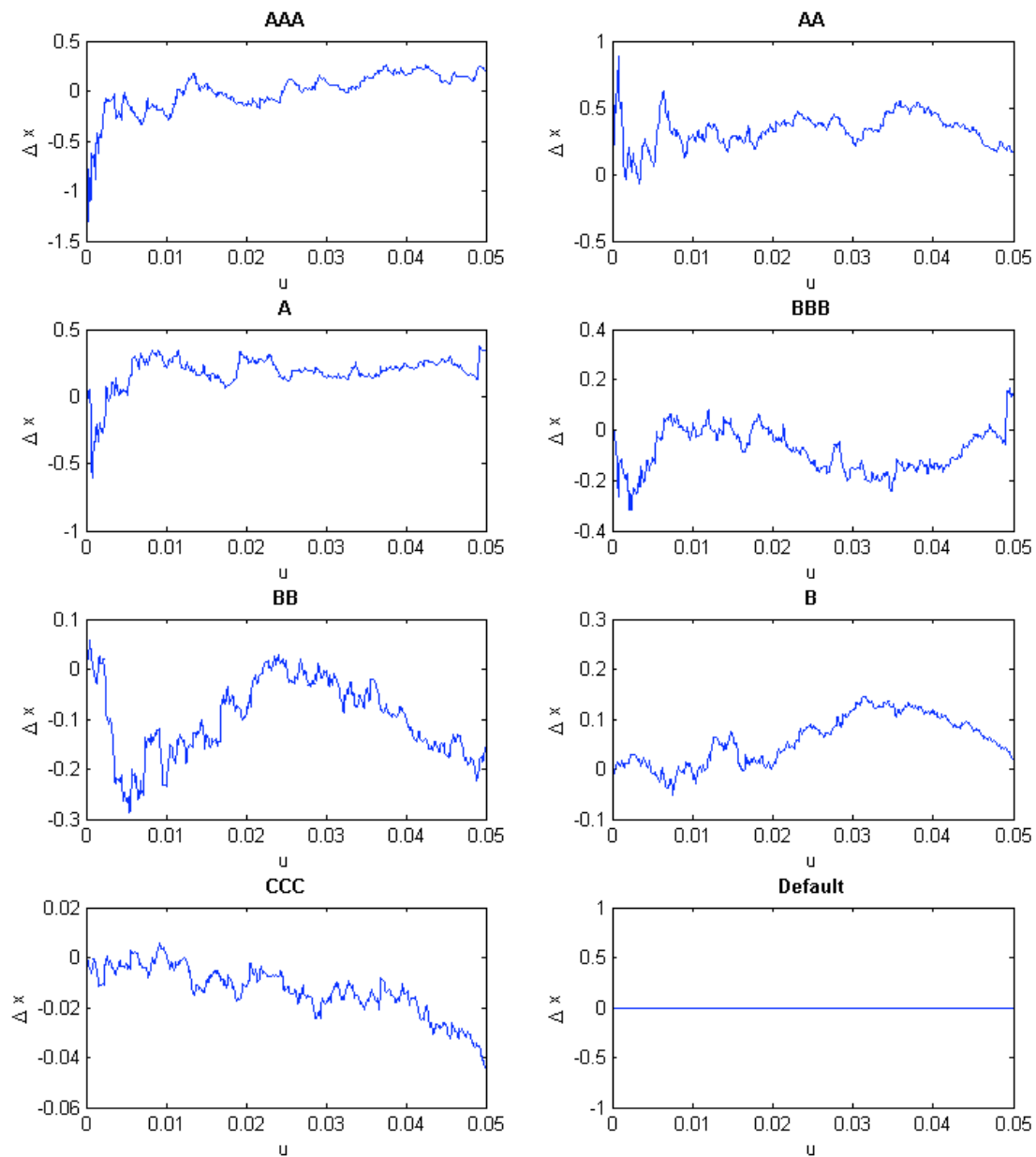


**Fig. 19: Hybride benadering op bereik [0, 50] jaar (Absolute fout)**



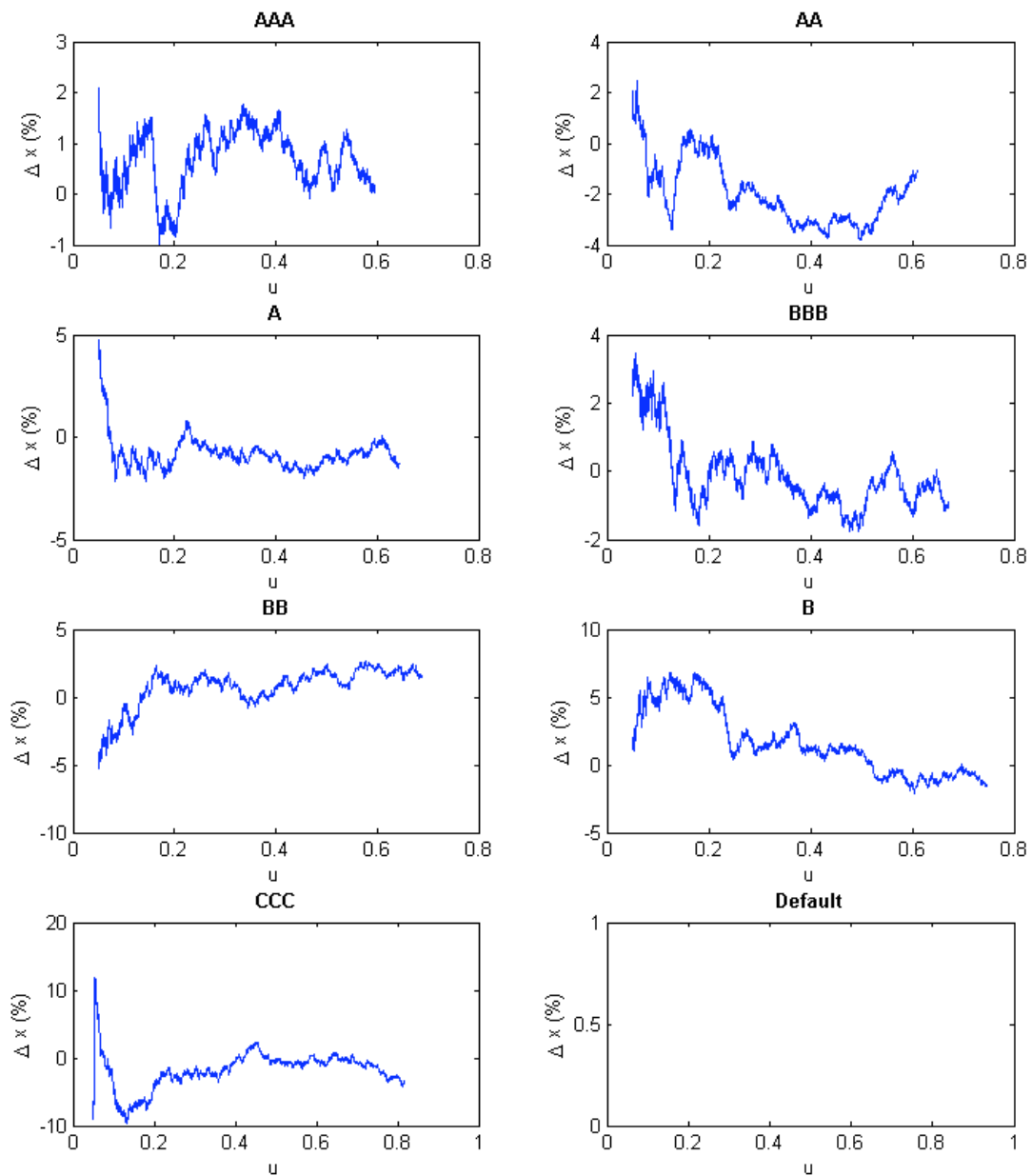
**Fig. 20: Hybride benadering op bereik [0, 50] jaar (Relatieve fout)**

De fout lijkt vanaf  $u=0.05$ , dus het door een functie benaderde gedeelte, mee te vallen, afgezien van het laatste stuk van rating CCC. Op het domein  $[0, 0.05]$  zou de relatieve fout groter kunnen zijn omdat het lage trekkingswaarden betreft. Dit laatste is in de getoonde figuren echter niet te zien. We zoomen apart in op deze twee domeinen, het geïnterpoleerde en het door functie benaderde, voor een beter beeld (Fig. 21 en Fig. 22).



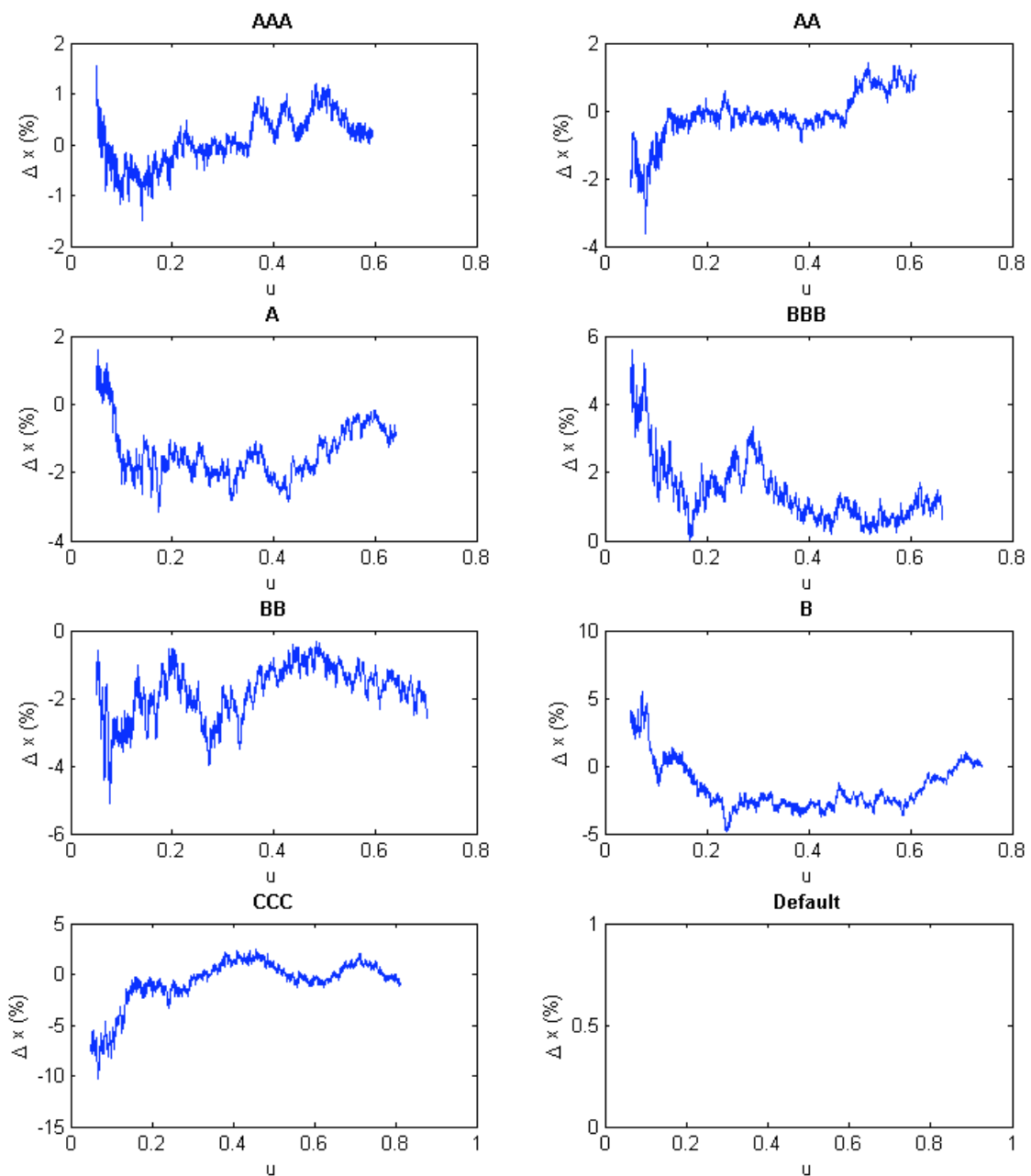
**Fig. 21: Hybride benadering op domein  $[0, 0.05]$  (Absolute fout)**

De fout valt mee, afgezien van het begin bij ratings AAA, AA en A. Op zoek naar een verklaring blijkt dat het komt doordat de ‘laagste waarden van de verdeling’ een zeer lage kans op voorkomen hebben. Dus de data waarmee is geleerd bevat in deze linkerrand al gauw andere getallen dan de data waarmee we approximatief uiteindelijk valideren.



**Fig. 22: Hybride benadering op domein vanaf  $u=0.05$  (Relatieve fout)**

De foutmarge blijft voor de ratings AAA tot en met BB binnen een acceptabele marge (0 tot 5%). Bij rating B wordt deze enigszins overschreden, maar de fout blijft acceptabel. Een foutpercentage van 5% zou bijvoorbeeld bij een waarde van 1 jaar een fout van ongeveer een halve maand betekenen. Rating CCC vertoont voor waarden van  $u$  dichtbij 0.05 een wat grotere fout van ongeveer maximaal 10%. Maar gezien het hier om een in absolute termen nihil fout gaat is het geen reden voor zorgen.



**Fig. 23: Interpolatie op domein vanaf  $u=0.05$  (Relatieve fout)**

Er is ook gekeken naar hoe de gefitte functies zich verhouden tot interpolatie (Fig. 23). Wanneer we interpolatie als referentie nemen zien we dat de functies die we hebben gefit in Matlab globaal eenzelfde foutmarge vertonen (in bepaalde gevallen zelfs beter).

Functie fitting is in dit geval dus een goed alternatief. De fout in de linkerrand van het domein  $[0, 0.05]$  komt zéér weinig voor, denk in de orde van  $10/10000$ , en ik neem aan dat het voor de verdere experimenten een verwaarloosbare invloed zal hebben.



### **3.4 Uitbreiding 2 (Theorie): Afhankelijkheid tussen defaults**

Een lastig vraagstuk waar nog relatief weinig onderzoek naar is verricht, is gecorreleerde huurders defaults. Huurders kunnen in default gaan, maar dit hoeft niet onafhankelijk van elkaar te gebeuren. Er kan een direct of indirect verband zijn dat huurders met elkaar bindt (Fig. 24). Huurder x kan bijvoorbeeld een bedrijf zijn dat afhankelijk is van grondstoffen die alleen huurder y, een ander bedrijf, kan leveren. Wanneer huurder y in default gaat, kan dit een verhogende impact hebben op de bedrijfsvoering en dus de defaultkans van huurder x. Ook kan het zo zijn dat wanneer een aantal huurders uit eenzelfde groep komen, een default van een huurder samenhangt met verhoogde kans op default van de andere huurders in de groep. Let op, dit is niet hetzelfde als het aan het begin van dit hoofdstuk uitgesloten één-op-één relaties (specifieke relaties tussen een specifiek bedrijf x en een specifiek bedrijf y), maar meer groepsgewijze relaties bijvoorbeeld tussen twee verbonden industrietakken (halfgeleiders en consumentenelektronica). Een groep kan bijvoorbeeld zijn: huurders uit eenzelfde land, eenzelfde regio, eenzelfde sector, etc. Een default van een huurder kan wijzen op een slechter economische toestand, waar de rest van de groep ook last van kan hebben.

Wanneer correlatie tussen huurders niet in een model wordt opgenomen zal dat een te positief beeld geven van de toekomst. Wanneer daadwerkelijk een doemscenario optreedt, zullen zelfs die investeerders die geïnvesteerd hebben in de minst risicovolle tranches (senior tranche) geraakt worden door verliezen. In de volgende paragraaf wordt een redelijk recente ontwikkeling binnen de vastgoed beschreven om nog een beter beeld te krijgen van het belang van gecorreleerde default modellering.

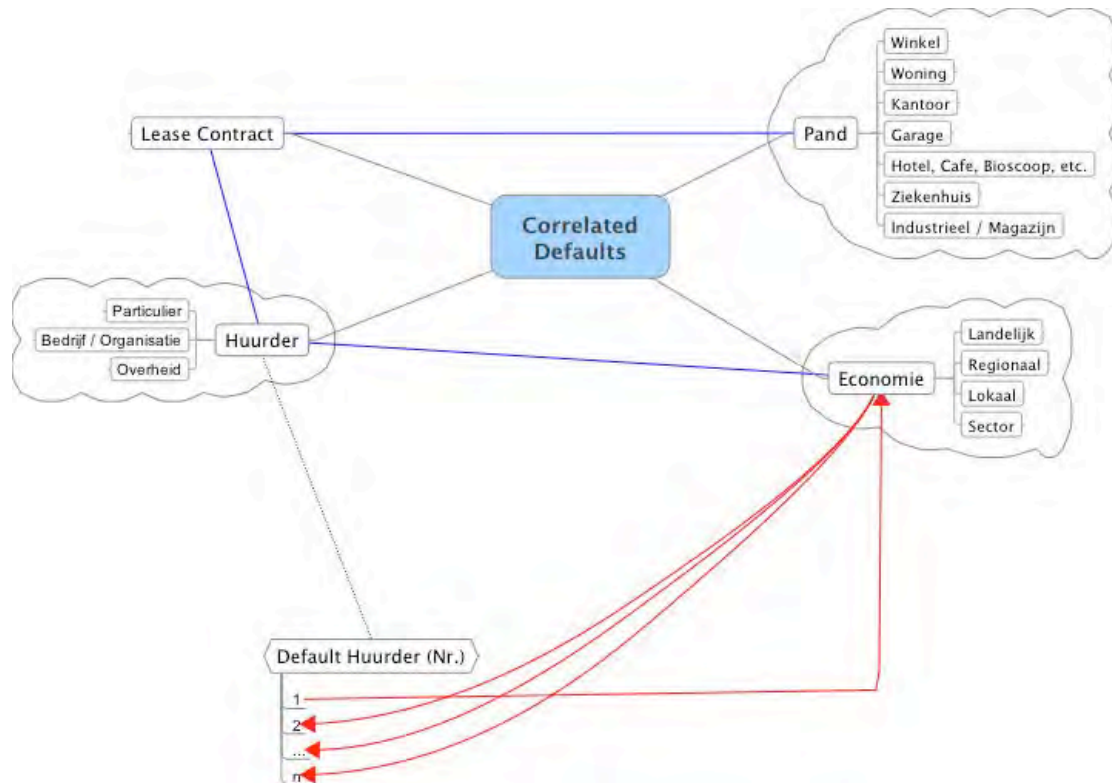


Fig. 24: Correlatie

In het model uit het vorige hoofdstuk is geen afhankelijkheid tussen de defaults opgenomen. Deze gecorreleerde default is een belangrijk component en zal uitgewerkt worden met behulp van copulas. Alvorens te beginnen met uitleg over wat copulas zijn en hoe ze in dit geval kunnen worden toegepast, wordt eerst uitgelegd waarom er niet is gekozen voor een andere veelgebruikte technieken, de correlatiecoëfficiënt en lineaire regressie.

### 3.4.1 Waarom geen lineaire regressie?

Bij het opgestelde model is gebruik gemaakt van copulas. Een betrekkelijk jonge techniek die voordelen kent ten opzichte van klassieke lineaire regressie. In deze paragraaf wordt beknopt lineaire regressie en haar beperkingen beschreven.

Bij lineaire regressie wordt een lineair model opgesteld dat wordt gefit aan de data. Een algemene formulering voor een lineair regressiemodel is:

$$\hat{y}_j = a_0 + \sum_{i=1}^n (a_i x_{i,j}) = a_0 + a_1 x_{1,j} + \dots + a_n x_{n,j}$$

$$\hat{y}_j - y_j = e_j,$$

( 17 )

waarbij de  $j$  de leersamples indexeert, en  $i$  de verklarende variabelen  $x_{i,j}$  en bijhorende coëfficiënten  $a_i$  indexeert.  $y_j$  is de waargenomen waarde bij leersample  $j$ . Het model

wordt gefit aan data door de waarden van de  $a_i$ 's zó worden gekozen opdat de som van de gekwadrateerde fouten ( $e_j$ ) geminimaliseerd wordt.

$$\min \sum_{i=1}^n (y_j - \hat{y}_j)^2 = \min(e_j^2) \quad (18)$$

Aan het gebruik van lineaire regressie kleven echter enkele assumpties waaraan moet worden voldaan:

1. De fout moet normaal verdeeld zijn met een gemiddelde van 0.

$$e_j \sim Norm(0, \sigma^2) \quad (19)$$

2. De fouten  $e_j$  en  $e_k$  moeten ongecorrleerd zijn.
3. De fout  $e_j$  moet ongecorrleerd zijn met de verklarende variabelen.

$$Cov(e_j, x_{i,j}) = 0 \quad (20)$$

4. De fout  $e_j$  moet ongecorrleerd zijn met de afhankelijke variabele.

$$Cov(e_j, \hat{y}_j) = 0 \quad (21)$$

Diverse papers geven aan dat bij de modellering van default niet aan de veronderstellingen van lineaire regressie wordt voldaan, maar dat deze methode alsnog wordt gebruikt. Deels uit onwetendheid, deels doordat men geen betere methode kent en deels doordat lineaire regressie een veelgebruikte techniek is die na jarenlang gebruik een acceptatie heeft in het bedrijfsleven. Typerend bij defaults is dat de staarten, en met name de economisch slechtere kant, een voortdurende bron van zorg zijn: ze zijn dikker en scheef. Dit effect is met lineaire regressie niet te reproduceren. De methode die in dit onderzoek wordt gebruikt is daarom niet lineaire regressie, maar de relatief jonge techniek van copulas. Copulas wordt onder andere veel door pensioenfondsen gebruikt.

De correlatie coëfficiënt wordt bij lineaire regressie vaak gebruikt als maat voor afhankelijkheid. Een uitgebreidere discussie over de beperkingen van de correlatie coëfficiënt als afhankelijkheidsmaat is te vinden in het laatste hoofdstuk, omwille hier niet te ver af te dwalen over een techniek die niet in de modellering is toegepast.

### 3.4.2 Introductie van copulas

Wanneer alle default tijden onafhankelijk van elkaar zouden zijn, is het eenvoudig om deze samen te modelleren. Iedere default tijd kan dan namelijk onafhankelijk van andere default tijden worden getrokken uit een bepaalde kansverdeling. Het wordt ingewikkelder wanneer er afhankelijkheid tussen de defaults bestaat, en dit blijkt in de praktijk ook zo te zijn bij default tijden. Dit is ook al in eerder genoemde assumpties gemeld. Wanneer de default tijd van een huurder laag is, is er een grotere kans dat de default tijd van een andere huurder ook laag uitvalt.

Een betrekkelijk jonge methode om dit te modelleren is met behulp van zogenaamde copulas. De grondslagen van de techniek liggen bij Sklar (1933), maar ongeveer sinds 15 jaar geleden is men pas stilaan begonnen om copulas toe te passen in de financiële sector. Dit komt deels door terughoudendheid, mensen zijn gewend om te werken met regressiemodellen. Ook al zijn deze soms niet geschikt vanwege assumpties van lineariteit en achterliggende normale verdelingen, het is stilaan door velen geaccepteerd als een standaardmethodiek door mensen, een norm. Deels komt het ook door onbekendheid en onervarenheid van toepassing van deze redelijk jonge techniek.

#### Wat is een copula?

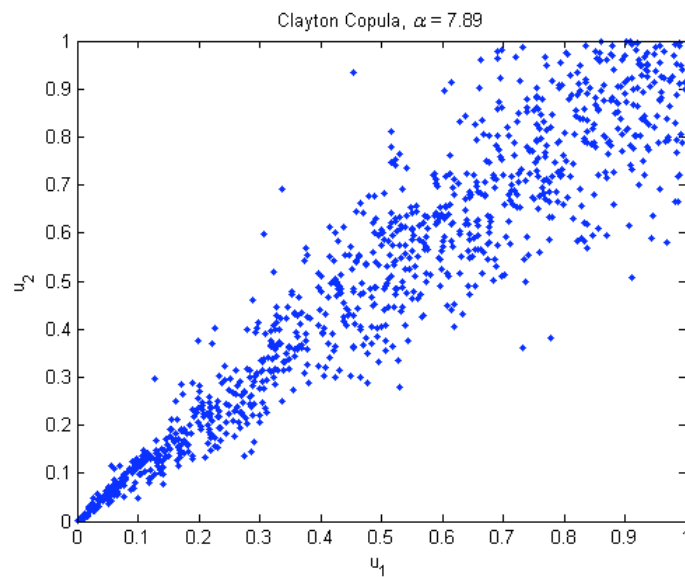
Een copula is een multivariate functie die marginale distributiefuncties bindt tot een nieuwe simultane distributiefunctie, waarbij de uitkomstwaarde een functie is van de marginale distributiefuncties.

$$\begin{aligned}
 F_{x_1, \dots, x_n}(x_1, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \\
 &= P(F(X_1) \leq F(x_1), \dots, F(X_n) \leq F(x_n)) \\
 &= P(U_1 \leq u_1, \dots, U_n \leq u_n) \\
 &= C(u_1, \dots, u_n)
 \end{aligned}
 \tag{22}$$

De copula wordt vaak als functie van  $u_i$ 's uitgedrukt. Dit omdat  $F(X_i)$  standaard uniform verdeeld is, wat eenvoudig is af te leiden uit het bewijs van de inverse transformatie methode dat in appendix A is gegeven.); Standaard uniforme variabelen worden vaak uitgedrukt als een hoofdletter U.

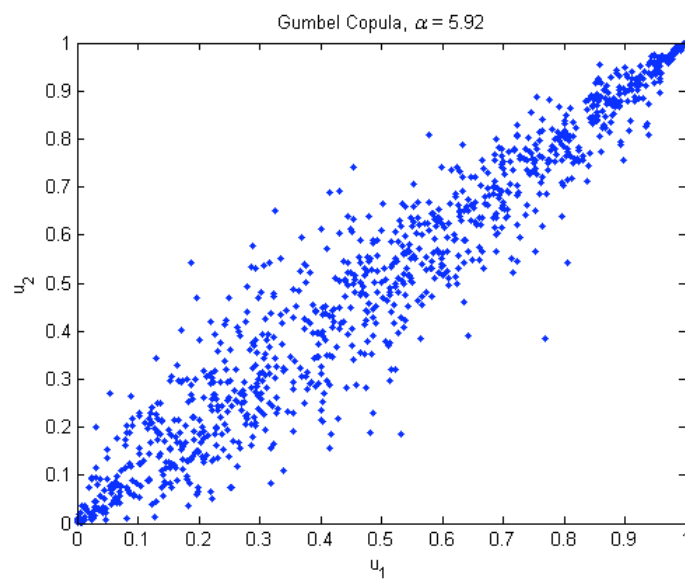
$$F_X^{-1}(U) = X \Rightarrow F_X(F_X^{-1}(U)) = U = F_X(X)
 \tag{23}$$

Door de copula uit te drukken op een dergelijke manier is ook duidelijk te zien dat de copula bestaat uit twee gedeelten. Een gedeelte waarbij marginale distributiefuncties worden gespecificeerd, en een copulafunctie die deze samenbindt ongeacht de onderliggende marginale distributies. Er zijn diverse populaire copulafuncties met verschillende eigenschappen. Zo legt de Clayton copula (Fig. 25), welke we later zullen gebruiken, een grotere mate van afhankelijkheid bij lagere waarden.



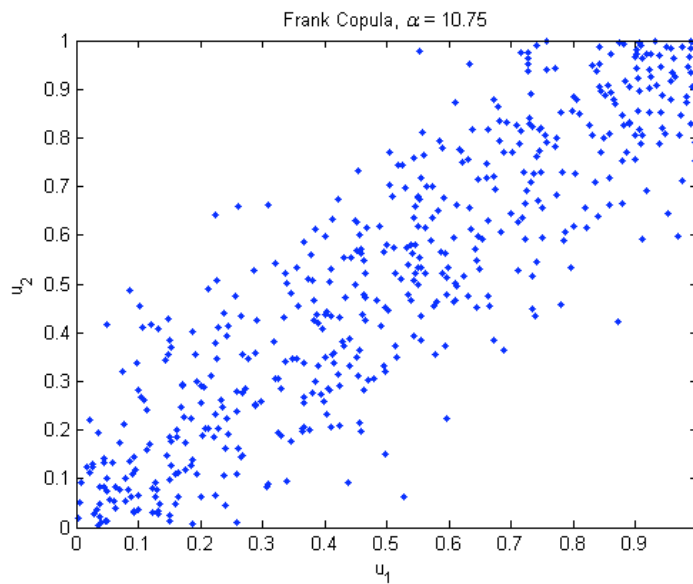
**Fig. 25: Plot van (bivariate) Clayton copula**

De Gumbel copula (Fig. 26) daarentegen vertoont juist hoge mate van afhankelijk bij hoge waarden en een wat lagere mate van afhankelijkheid bij lage waarden.



**Fig. 26: Plot van (bivariate) Gumbel copula**

De Frank copula (Fig. 27) is in tegenstelling tot de eerder genoemde copulas symmetrisch, met een hogere (doch niet zo sterke) mate van afhankelijkheid in de staarten. Dit laatste is te zien aan de iets dichtere dichtheid van punten in de staarten.



**Fig. 27: Plot van (bivariate) Frank copula**

En zo zijn er nog veel meer verschillende copulas, met hun eigen afhankelijkheidssignatuur. Afhankelijk van datgeen wat men wil modelleren kan dan een passende copula worden gevonden, of geprobeerd worden om zelf een passende copula op te stellen. Een copula bevat een of meerdere schaal parameters om de afhankelijkheid nog beter te laten fitten op datgeen dat gemodelleerd wordt. Een copula geeft een veel “vollere” omschrijving van de afhankelijkheidsstructuur dan een maat zoals de (Pearson) correlatie coëfficiënt kan geven.

Om de sterke punten van copulas op te sommen:

- Beschrijft afhankelijkheid beter dan een scalar maat als de (Pearson moment-product) correlatie coëfficiënt dat kan. Deze laatste geeft een kengetal dat afhankelijkheid over de hele linie moet omschrijven en mist daardoor veel detail.
- Een manier om marginale verdelingen en afhankelijkheid apart te modelleren, wat het opstellen van multivariabel distributie functies makkelijker maakt. Met deze laatste verdeling is het vervolgens mogelijk om bijvoorbeeld simulaties uit te voeren.
- Doordat de gekozen marginale verdelingen onafhankelijk zijn van de copula, kan er eenvoudig andere marginale verdelingen in worden gestopt. Niet gelimiteerd tot normale verdelingen zoals standaard veel (soms te onpas) wordt gebruikt. De marginale verdelingen hoeven niet dezelfde parameters te delen, of uit dezelfde familie te komen, wat flexibiliteit verder vergroot.
- Verschillende bestaande copulas in de literatuur die tegemoet komen aan verschillende afhankelijkheidssignalen.

- Geen normale verdelingen vereist. Iedere strict stijgende functie kan gebruikt worden.

### 3.4.3 Modelling en simulatie met de Clayton Copula

De Clayton copula is een copula die in de financiële wereld onder andere wordt gebruikt voor levensverzekeringen. Voor oudere stellen is bekend dat er sprake is van het “broken heart syndrome”, waarbij het vaak voorkomt dat wanneer een van de twee sterft, de ander na niet al te lange tijd ook sterft. Wanneer deze afhankelijkheid sterk is bij lage resterende levensjeden, dan is de Clayton copula een goede keus<sup>3</sup>. De Clayton copula is mede in trek doordat er slechts een parameter geschat hoeft te worden uit de data, wat het gebruik ervan relatief vereenvoudigt ten opzichte van copulas met meerdere parameters.

De Clayton copula is bij de modellering van default tijden een geschikte keuze omdat de linkerstaart een hogere mate van afhankelijkheid kent. Deze gaan we dan ook uitwerken in deze paragraaf. De formule van de Clayton copula is:

$$C(u_1, \dots, u_n) = \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-1/\theta} \quad (24)$$

#### Afhankelijkheidsparameter schatten

Een eerste stap naar simulatie is om de juiste parameter te vinden. We gebruiken hiervoor de methode van de meest waarschijnlijk schatter. Ofwel, we zoeken die waarde van de parameter die volgens de gegeven data het meest waarschijnlijk is, dus waar de kans het hoogst is. In dit geval gaan we kijken naar de kansdichtheidsfunctie van de copula als functie van de afhankelijkheidsparameter  $\theta$  (spreek uit: Theta). Net zoals we bij reguliere distributiefuncties de kansdichtheidsfunctie verkrijgen uit differentiatie, verkrijgen we ook bij copulas de kansdichtheidsfunctie uit differentiatie. De geïnteresseerde lezer verwijs ik naar appendix B voor een uitgebreide uitwerking.

$$\begin{aligned} c(u_1, \dots, u_n) &= \frac{d^n}{du_1 \dots du_n} C(u_1, \dots, u_n) \\ &= \frac{d^n}{du_1 \dots du_n} \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-1/\theta} \\ &= \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-1/\theta - n} \prod_{i=1}^n u_i^{-\theta - 1} (1 + \theta i - \theta) \end{aligned} \quad (25)$$

Hierbij zijn de  $u_i$ 's geen variabelen. De  $u_i$ 's worden vanuit de  $x_i$ 's van de dataset berekend uit de distributiefuncties (van de respectievelijke marginale

kansverdelingen). Een omgekeerd proces dus van eerder behandelde simulatie met behulp van inverse transformatie methode. Bij eerder genoemde werd vanuit een standaard uniforme verdeling, aan de hand van de inverse van de distributiefunctie een waarde voor de variabele bepaald. Nu wordt tegenovergesteld aan de hand van een gerealiseerde waarde van een variabele een bijhorende (cumulatieve) kans bepaald. De kansdichtheidsfunctie van de copula reduceert in dit geval dus tot een functie van één variabele  $\theta$ . Wiskundige programma's als Matlab bevatten numerieke methoden om de waarde van  $\theta$  te berekenen.

### Simuleren

Nu de afhankelijkheidsparameter is geschat, kan daadwerkelijk met behulp van de copula random getallen gesimuleerd worden. Dit is een iteratief proces waarvan de volledige uitwerking in de appendix staat, maar waarvan de kerngedachten hier worden besproken:

We weten dat voor iedere strikt stijgende continue kansdeling geldt dat

$$F_X(X) = U, \tag{26}$$

waarbij  $U$  standaard uniform verdeeld is. Stel dat we een bivariate kansverdeling hebben, waarbij de distributiefunctie naar alle dimensies strikt stijgend is. Dan geldt er ook voor de voorwaardelijke distributiefunctie:

$$F_{X|Y}(X | y) = U \tag{27}$$

En in het algemenere multivariabel geval:

$$F_{X|Y}(X_n | x_1, \dots, x_{n-1}) = U \tag{28}$$

Een op het oog eenvoudige constatering, maar wel een die ons in staat stelt simulaties op een iteratieve manier uit te voeren. NB. We simuleren waarden voor de  $u_i$ 's, de waarden van de  $x_i$ 's kunnen direct met inverses van de marginale distributiefuncties worden berekend.

1. We simuleren een vector  $r$ , ter lengte  $n$ , met standaard uniform getrokken getallen. Dit zijn de waarden voor de voorwaardelijke copulas.

$$\begin{aligned} r(1) &= C(u_1) \\ r(i) &= C(u_i | u_1, \dots, u_{i-1}) \end{aligned} \tag{29}$$



2. Beginnende met  $r(1)$  volgt:

$$\begin{aligned} r_1 &= C(u_1) = \left(u_1^{-\theta}\right)^{-1/\theta} = u_1 \\ \Rightarrow u_1 &= r_1 \end{aligned} \tag{30}$$

3. Nu  $u_1$  bekend is kan  $u_2$  berekend worden:

$$\begin{aligned} r_2 &= \frac{C(u_1, u_2)}{C(u_1)} = \frac{\left(-1 + u_1^{-\theta} + u_2^{-\theta}\right)^{-1/\theta}}{\left(u_1^{-\theta}\right)^{-1/\theta}} \\ \Rightarrow u_2 &= \left((r_2 u_1)^{-\theta} + 1 - u_1^{-\theta}\right)^{-1/\theta} \end{aligned} \tag{31}$$

4. Dit proces herhalende komen we aan alle waarden voor de  $u_i$ 's, en vinden we de algemene formule (appendix C):

$$\begin{aligned} u_n &= \left( (n-1) + r_n^{-\theta} (2-n) + \sum_{i=1}^{n-1} r_n^{-\theta} u_i^{-\theta} - \sum_{i=1}^{n-1} u_i^{-\theta} \right)^{-1/\theta} \\ &\text{voor } n \geq 3 \end{aligned} \tag{32}$$

5. Met de  $u_i$ 's kunnen we de  $x_i$ 's berekenen:

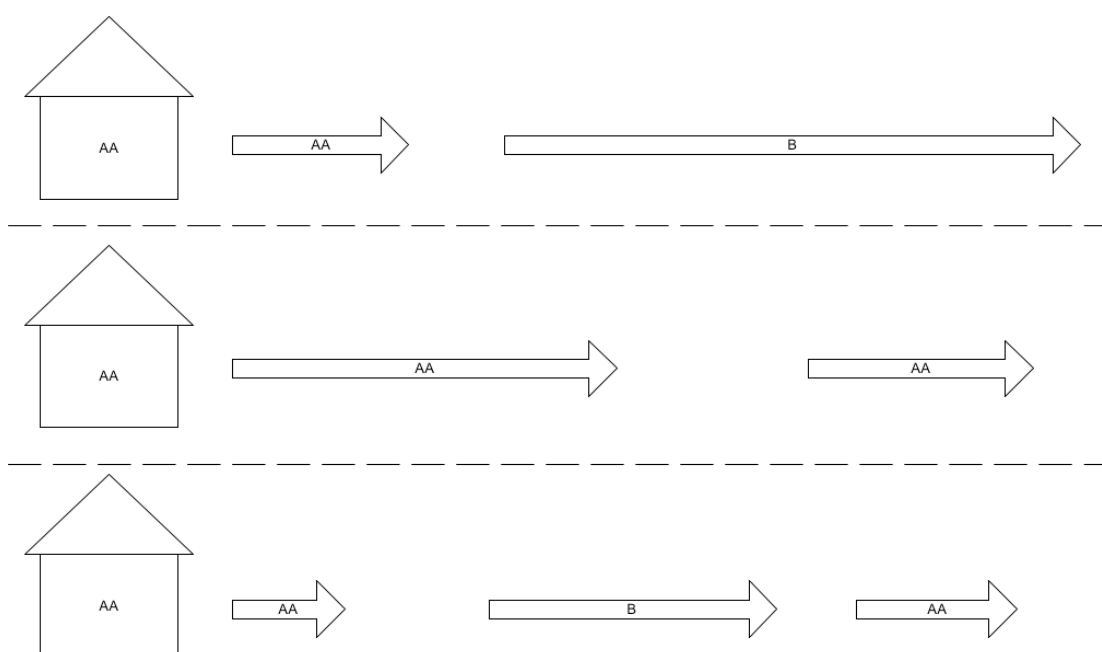
$$F_{X_i}^{-1}(u_i) = x_i \tag{33}$$

Programma's als Matlab bevatten routines om deze simulaties uit te voeren voor ingeprogrammeerde copulas, waaronder de Clayton copula. Maar aan de hand van bovenstaand stappenplan is het ook mogelijk om zelf de routines in te programmeren.

### 3.5 Uitbreiding 2 (Toegepast): Verwerking van copula's in model

#### 3.5.1 Groepering per Huurdersklasse

De uitwerking van copulas in de vorige paragraaf is in haar huidige vorm niet direct verenigbaar met het basismodel uit het vorige hoofdstuk. In het vorige hoofdstuk geschiedde simulatie van defaulttijden onafhankelijk van elkaar en was het mogelijk om de simulatie contractsgewijs, dus per groep contractgerelateerde rentable units, te verwerken.



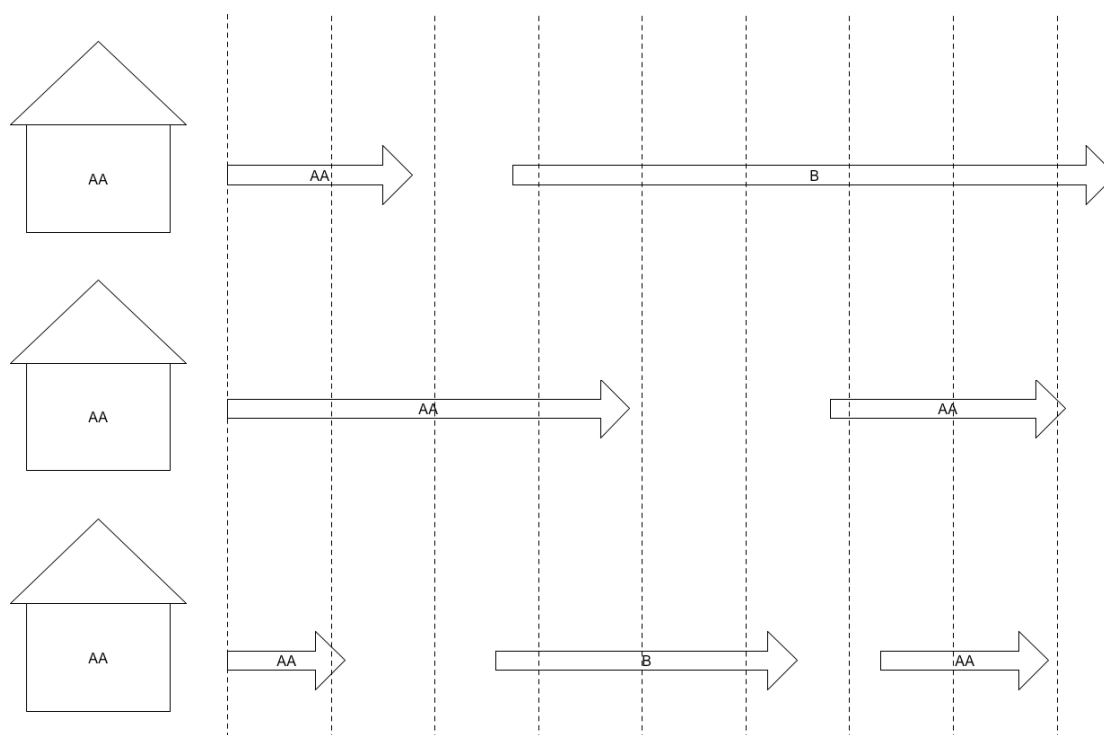
**Fig. 28: Onafhankelijke defaulttijden**

De stippellijnen in de bovenstaande figuur (Fig. 28) benadrukken de onafhankelijke generatie van defaulttijden. Voor iedere rentable unit zijn met pijlen de verhuurde periodes weergegeven, met in de pijl de rating van de huurder. De uitbreiding naar afhankelijke defaulttijden voor het begin van de simulatie kan geschieden zoals in de vorige paragraaf is beschreven. Dit omdat al deze tijden hetzelfde beginmoment kennen. Voor het vervolg van de simulatie is het echter zo dat ratings kunnen veranderen en aanvangsmomenten van verhuurde periodes uiteenlopen. Om deze twee punten te verwerken volgen we de volgende aanpak. We stellen een tabel ter grootte van het aantal klassen en het aantal te simuleren jaren. Iedere cel van de tabel bevat een vector die bijhoudt welke getallen er zijn getrokken voor de copula-gebaseerde simulatie (Fig. 29).

<i>Klasse / jaar</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>...</i>
<i>AAA</i>	$\underline{U}_{AAA,1}$	$\underline{U}_{AAA,2}$	$\underline{U}_{AAA,3}$	$\underline{U}_{AAA,4}$	...
<i>AA</i>	$\underline{U}_{AA,1}$	$\underline{U}_{AA,2}$	$\underline{U}_{AA,3}$	$\underline{U}_{AA,4}$	...
<i>A</i>	$\underline{U}_{A,1}$	$\underline{U}_{A,2}$	$\underline{U}_{A,3}$	$\underline{U}_{A,4}$	...
<i>BBB</i>	$\underline{U}_{BBB,1}$	$\underline{U}_{BBB,2}$	$\underline{U}_{BBB,3}$	$\underline{U}_{BBB,4}$	...
<i>BB</i>	$\underline{U}_{BB,1}$	$\underline{U}_{BB,2}$	$\underline{U}_{BB,3}$	$\underline{U}_{BB,4}$	...
<i>B</i>	$\underline{U}_{B,1}$	$\underline{U}_{B,2}$	$\underline{U}_{B,3}$	$\underline{U}_{B,4}$	...
<i>CCC</i>	$\underline{U}_{CCC,1}$	$\underline{U}_{CCC,2}$	$\underline{U}_{CCC,3}$	$\underline{U}_{CCC,4}$	...

**Fig. 29: Opslag voor copula-gebaseerde simulatie**

Voor het eerder getoonde plaatje zou de situatie dan worden (Fig. 30):



**Fig. 30: Afhankelijke defaulttijden**

- Het eerste getal van vector  $\underline{U}_{AA,1}$ , welke we noteren als  $\underline{U}_{AA,1}(1)$ , is een standaard uniform getrokken random getal.
- Voor het tweede contract met een AA huurder aan het begin van de simulatie wordt met behulp van copula-gebaseerde simulatie een tweede getal gesimuleerd voor de vector  $\underline{U}_{AA,1}$ , de notatie hiervoor is  $\underline{U}_{AA,1}(2)$ .
- Met de twee al getrokken getallen wordt voor het derde contract met een AA huurder  $\underline{U}_{AA,1}(3)$  gesimuleerd.
- De eerstvolgende gebeurtenis hierna is dat voor het derde contract de volgende huurder van klasse B is. We voegen aan onze tabel  $\underline{U}_{B,3}(1)$  toe, een standaard

uniform getal, omdat er nog geen getal is gesimuleerd in voor deze klasse en jaar combinatie.

- Met behulp van  $\underline{U}_{B,3}(1)$  kunnen we voor het eerste contract, wat toevallig ook een huurder van klasse B kent in hetzelfde jaar,  $\underline{U}_{B,3}(2)$  simuleren.
- Het tweede contract krijgt vervolgens een AA huurder. Voor deze klasse en jaar combinatie zijn er nog geen getallen gesimuleerd. Derhalve trekken we voor  $\underline{U}_{AA,6}(1)$  een standaard uniform random getal.
- Laatste gebeurtenis in het voorbeeld is dat voor het derde contract de volgende huurder een AA huurder is. Net zoals bij de vorige stap bevat de relevante vector nog geen getallen.  $\underline{U}_{AA,7}(1)$  vullen we met een random standaard uniform getal.
- Etcetera. Vervolg is niet uitgetekend maar het idee is duidelijk.

Merk op dat bovenstaand voorbeeld versimpeld kan worden. Het voorbeeld gaat er namelijk van uit dat steeds de eerstvolgende gebeurtenis in de tijdlijn verwerkt, wat er dus impliciet van uit gaat dat er een algoritme ingebouwd moet zijn dat de eerstvolgende gebeurtenistijd kan opzoeken. Doordat we echter alle gesimuleerde copula getallen opslaan, is het mogelijk om de simulatie per contract uit te voeren. Hierbij moet wel bij ieder te simuleren contract de voor copula-gebaseerde simulatie opgeslagen getallen in acht worden genomen. Dit bespaart het uitwerken van een algoritme dat de eerstvolgende event-tijdstip opzoekt en levert naast een eenvoudiger simulatie algoritme ook een snellere simulatie op.

### 3.5.2 Verdere groepering

In de vorige paragraaf is er per huurderklasse een afhankelijkheid. Dus AAA huurders zijn met elkaar gerelateerd, AA huurders zijn met elkaar gerelateerd, etc. Dit doen we om per klasse een afwijkende afhankelijkheidsparameter in te voeren. Immers, alle klassen volgen dezelfde algemene relatie tussen defaulttijden, (namelijk grotere afhankelijkheid naarmate de defaulttijden laag zijn), maar niet alle klassen zijn even gevoelig voor dit effect. Zoals eerder al eens is vermeld hebben huurders met een lagere klasse hebben meer last van dit effect, omdat ze een financieel slechtere positie hebben om financiële klappen op te vangen.

Het is echter mogelijk dat verschillende economische sectoren een verschillende relatie hebben. De olie-industrie zal bijvoorbeeld sterkere correlatie vertonen dan een halfgeleider industrie. Dus naast groepering per huurderklasse kan er nog verdere groepering zijn op bijvoorbeeld industriële sector. Een mogelijke modellering hiervoor is om de eerdere voor copula-gebaseerde simulatie gebruikte tabel uit te breiden met een derde dimensie welke de industriële sector representeert. De vectoren nemen dan de vorm aan:

$$\underline{U}_{\text{Huurdersklasse, industrie, jaar}}$$

( 34 )

Simulatie kan nog steeds per contract geschieden, mits de getallen voor copula-gebaseerde simulatie in de juiste vector worden opgeslagen en gehaald.



## 4. Simulatie

### 4.1 Inleiding

Om de in het vorige hoofdstuk geschetste model vorm te geven en verder te onderzoeken is er besloten een simulatie te bouwen om verdere tests uit te voeren. Dit hoofdstuk staat in het teken van een beschrijving van deze simulatie. Analyses die zijn uitgevoerd met de simulatie zijn bewaard voor het volgende hoofdstuk.

### 4.2 Data

#### 4.2.1 Fictieve data

Data over de één-jaar transitie kansen van huurderklassen is beschikbaar, maar gedetailleerde data over huurders, contracten, leegstand/verhuur panden, etc. is niet beschikbaar. Om toch de modellering te kunnen analyseren is besloten om zelf een fictieve dataset te genereren. Hiermee is het in ieder geval mogelijk om te onderzoeken hoe het model zich gedraagt.

Het juist kalibreren van het model is een taak voor degenen die het uiteindelijk kunnen gebruiken. Zoals woningcorporaties en bedrijven die werken met vastgoedfondsen. De data waarover deze samen beschikken zou genoeg moeten zijn om de gewenste parameters te vinden en het model naar inzicht te verfijnen. Voor een onafhankelijk onderzoeker is deze data echter niet beschikbaar. Ratingagencies claimen over uitgebreide databases te beschikken rondom default onderzoek, maar het kost veel geld om hier toegang toe te verkrijgen.

#### 4.2.2 De dataset

De dataset bevat allereerst een één-jaars transitie matrix van een rating agency, die vrijelijk via het internet beschikbaar is. Dezelfde transitie matrix is ook gebruikt in het vorige hoofdstuk. Voor de volledigheid:

→	AAA	AA	A	BBB	BB	B	CCC	Default	$\Sigma$
AAA	66.26	22.22	7.37	2.45	0.86	0.67	0.14	0.02	100
AA	21.66	43.04	25.83	6.56	1.99	0.68	0.2	0.04	100
A	2.76	20.34	44.19	22.94	7.42	1.97	0.28	0.1	100
BBB	0.3	2.8	22.63	42.54	23.52	6.95	1	0.26	100
BB	0.08	0.24	3.69	22.93	44.41	24.53	3.41	0.71	100
B	0.01	0.05	0.39	3.48	20.47	53	20.58	2.01	100
CCC	0	0.01	0.09	0.26	1.79	17.77	69.94	10.13	100
Default	0	0	0	0	0	0	0	100	100

Tabel 3: Eén-jaar transitiematrix gebaseerd op periode 1990-1995 (in %)

De dataset bevat de volgende componenten:

- **Pand ID**  
Een uniek getal dat gebruikt wordt om panden te onderscheiden.
- **Contract ID**  
Een uniek getal dat gebruikt wordt om contracten te onderscheiden. De contract ID wordt zó gegenereerd dat een contract (geometrisch verdeeld) gemiddeld twee panden bevat.
- **Init. Huurder ID**  
Geeft de ID van de eerste huurder van het pand weer. De huurder ID wordt zó gegenereerd dat een huurder (geometrisch verdeeld) gemiddeld 4/3 contracten heeft.
- **Init. Huurder rating**  
Geeft de klasse van de eerste huurder van het pand weer. De rating van de huurder is gebaseerd op de huur van het pand. De huur is uniform verdeeld, en ook de klassen zijn uniform verdeeld. Indien een huurder meerdere panden huurt wordt de klasse bepaald aan de hand van de huur van een willekeurig (conventie: eerst tegengekomen in dataset) pand.
- **Begin contract**  
Er zijn eigenlijk twee “simulatie starttijdstippen” in de huidige simulatie. Het tijdstip wat  $T = 0$  wordt gedoopt is het moment vanwaar statistieken worden bijgehouden. Maar daarnaast is het zo dat we vóór dit tijdstip al moeten beginnen met simuleren vanaf het moment dat het eerste contract ingaat. Dit laatste verschilt per contract en noemen we “Begin contract”, uitgedrukt in jaren vanaf  $T = 0$ . De begindatum is uniform getrokken uit het interval  $[-1, 0]$ .
- **Huur (maand)**  
De maandelijkse huur van een pand. Deze wordt uniform getrokken vanuit de range [“Min. Huur”, “Min. Huur” + Huur spread]



- **Leegstandskosten (maand)**  
De kosten voor leegstand van het pand, per maand. Dit zijn o.a. kosten voor service en onderhoud die niet meer op de huurder verhaald kunnen worden. Dit wordt gemodelleerd als een bepaald percentage van de maandhuur.
- **Leegstandskosten (per keer)**  
Per leegstand, en ongeacht de duur van de leegstand, zijn er kosten. Bijvoorbeeld kosten voor het zoeken naar een nieuwe huurder en de kosten voor contracten opstellen. Dit wordt gemodelleerd als een bepaald percentage van de maandhuur.
- **# Huurdersklassen**  
Het aantal huurderklassen (exclusief default als klasse). Voor de transitie matrix in dit onderzoek bedraagt dit zeven klassen.
- **Looptijd (jaar)**  
Looptijd van het contract in jaren. In de praktijk is dit meestal 5 jaar. Er is gekozen om alle contracten een looptijd van 5 jaar te laten hebben.
- **Gemiddelde leegstand (jaren)**  
De gemiddelde leegstandsperiode van een pand in jaren. Deze modelleren we als 6 maanden voor alle panden.
- **Minimale leegstand (jaren)**  
De minimale leegstandsperiode van een pand in jaren. Het vinden van een nieuwe huurder kost minimaal 3 maanden, dus 0,25 jaar.
- **Kans op contract verlenging**  
De kans dat een huurder aan het eind van een contract periode zijn of haar contract verlengt met nog eens vijf jaar. Voor de simulatie is een kans gekozen van 0,5 jaar.

Een screenshot van de dataset staat op de volgende pagina (Fig. 31). De dataset bevat 200 panden.

## Huurders Default – Risico modellering

Data (Frozen).xls											
	A	B	C	D	E	F	G	H			
1	Pand ID	Contract ID	Init. Huurder ID	Init. Huurder Rating	Begin contract	Huur (maand)	Leegstandskosten (maand)	Leegstandskosten (per keer)			
2	1	1	1	7	-1.27	€ 14,367	€ 4,310	€ 11,493			
3	2	1	1	7	-1.26	€ 13,291	€ 3,987	€ 10,633			
4	3	2	1	5	-0.11	€ 11,219	€ 3,366	€ 8,975			
5	4	3	2	7	-0.20	€ 14,554	€ 4,366	€ 11,643			
6	5	4	3	1	-0.06	€ 6,235	€ 1,871	€ 4,988			
7	6	5	4	5	-1.33	€ 11,195	€ 3,358	€ 8,956			
8	7	5	4	5	-0.01	€ 13,802	€ 4,140	€ 11,041			
9	8	5	4	5	-0.37	€ 11,928	€ 3,578	€ 9,542			
10	9	6	4	4	-1.03	€ 10,493	€ 3,148	€ 8,394			
11	10	7	5	1	-1.68	€ 5,316	€ 1,595	€ 4,253			
12	11	8	6	2	-0.09	€ 7,411	€ 2,223	€ 5,929			
13	12	9	6	6	-1.47	€ 12,823	€ 3,847	€ 10,258			
14	13	9	6	6	-1.41	€ 7,353	€ 2,206	€ 5,883			
15	14	10	7	5	-1.26	€ 10,980	€ 3,294	€ 8,784			
16	15	10	7	5	-0.21	€ 11,968	€ 3,590	€ 9,575			
17	16	11	7	5	-1.14	€ 10,779	€ 3,234	€ 8,623			
18	17	11	7	5	-0.38	€ 9,899	€ 2,970	€ 7,919			
19	18	11	7	5	-0.34	€ 8,060	€ 2,418	€ 6,448			
20	19	11	7	5	-1.28	€ 13,151	€ 3,945	€ 10,520			
21	20	12	7	1	-0.60	€ 5,451	€ 1,635	€ 4,361			
22	21	13	7	5	-0.92	€ 11,712	€ 3,514	€ 9,370			
23	22	13	7	5	-1.83	€ 9,770	€ 2,931	€ 7,816			
24	23	13	7	5	-0.56	€ 9,479	€ 2,844	€ 7,583			
25	24	14	8	2	-1.70	€ 7,771	€ 2,331	€ 6,217			
26	25	14	8	2	-1.11	€ 10,128	€ 3,039	€ 8,103			
27	26	14	8	2	-1.12	€ 12,405	€ 3,722	€ 9,924			
28	27	15	8	6	-0.80	€ 12,723	€ 3,817	€ 10,179			
29	28	16	9	3	-1.15	€ 8,748	€ 2,624	€ 6,998			
30	29	17	10	5	-1.52	€ 12,126	€ 3,638	€ 9,700			
31	30	17	10	5	-1.74	€ 6,024	€ 1,807	€ 4,819			
32	31	18	10	3	-0.54	€ 8,336	€ 2,501	€ 6,669			
33	32	19	10	6	-0.27	€ 13,249	€ 3,975	€ 10,600			
34	33	20	11	2	-0.91	€ 7,853	€ 2,356	€ 6,283			
35	34	21	12	5	-0.01	€ 11,173	€ 3,352	€ 8,938			
36	35	22	13	1	-1.47	€ 5,725	€ 1,717	€ 4,580			
37	36	22	13	1	-1.63	€ 14,063	€ 4,219	€ 11,250			
38	37	23	14	4	-1.09	€ 9,894	€ 2,968	€ 7,915			
39	38	23	14	4	-1.64	€ 9,010	€ 2,703	€ 7,208			
40	39	24	15	7	-0.05	€ 13,638	€ 4,091	€ 10,910			
41	40	24	15	7	-1.59	€ 11,460	€ 3,438	€ 9,168			
42	41	25	16	2	-0.13	€ 7,693	€ 2,308	€ 6,155			
43	42	25	16	2	-0.88	€ 7,788	€ 2,328	€ 6,208			

# Huurdersklassen	7
Looptijd (jaar)	5
Vroegst mogelijke "Begin contract"	-2
Min. Huur (maand)	€ 5,000
Huur spread	€ 10,000
Leegstandskosten (maand, % maandhuur)	30%
Leegstandskosten (per keer, % maandhuur)	80%
Gemiddelde leegstand (jaren)	0.5
Minimale leegstand (jaren)	0.25
Kans op contract verlenging	50%
Gemiddeld #panden per contract	2
Gemiddeld #contracten per huurder	1.333333333

AAA	1
AA	2
A	3
BBB	4
BB	5
B	6
CCC	7

Fig. 31: Screenshot simulatie invoer data

## 4.3 Assumpties

Een deel van de assumpties zijn al in de paragraaf over de data vermeld. De resterende assumpties zullen hier gepresenteerd worden.

### **Pand - Rentable Units**

In de praktijk is er per contract een aantal rentable units. In de simulatie is dit versimpeld naar hele panden per contract. Hierbij wordt de rentable unit laag overgeslagen en als het ware geaggregeerd. Dit versimpelt het programmeerwerk voor de simulatie. We kunnen datgeen wat in de data “Pand (ID)” wordt genoemd interpreteren als “Rentable Unit groepering”. De versimpeling naar “Pand” heeft dus geen invloed op de generiekheid van het model; ofwel, niet perse voor contracten met gehele aantallen fysieke panden toepasbaar, maar ook voor contracten met de gebruikelijke verzameling rentable units.

### **Initieel verhuurd**

Alle panden in de dataset zijn initieel verhuurd. Dit is gedaan om niet apart routines in te hoeven programmeren die rekening houden met panden die aan het begin van de simulatie leeg staan. Voor de bestudering van de modellering zal dit waarschijnlijk weinig effect hebben. Het bespaart echter een redelijke hoeveelheid programmeren.

### **Leegstand multi-pand contract**

Wanneer een contract slechts een pand bevat is eenvoudig om met de minimale leegstand van dat pand en de gemiddelde leegstand van dat pand een leegstandperiode te genereren. Wanneer een contract meerdere panden bevat is het minder eenduidig hoe de leegstandperiode berekend moet worden. Moet men bijvoorbeeld voor alle panden een trekking verrichten en de hoogste leegstandperiode kiezen? Dit is mogelijk maar leidt bij een groot aantal panden binnen een contract mogelijk te vaak tot extreme waarden. Het is ook mogelijk om de grootste gemiddelde leegstand van alle panden te nemen en die te gebruiken voor het genereren van een leegstandperiode. Dit kan echter weer onrealistisch (in dit geval te hoog) uitpakken wanneer bijvoorbeeld vijf panden een gemiddelde leegstand kennen van zes maanden en een pand een gemiddelde leegstand kent van twaalf maanden. We gaan daarom uit van de gemiddelde leegstand, middelend over alle panden. In onze dataset hebben we aangenomen dat voor elk pand een zelfde minimale leegstand en gemiddelde leegstand geld.

## **4.4 Implementatie**

### **4.4.1 Omgeving / programmeertaal**

Voor de implementatie is gekozen voor het in wetenschappelijke kringen veelgebruikte programma Matlab. Dit omdat het sneller is dan van de grond af aan alles programmeren, inclusief opslagstructuren, data importeren en exporteren, etc. In Matlab zijn bepaalde routines voor random trekken van getallen uit bekende verdelingen al ingebouwd. Ook zijn er mogelijkheden tot experimenteel onderzoek, statistisch onderzoek en presentatie mogelijk in Matlab.

Een nadeel is dat Matlab beperkte mogelijkheden tot object georiënteerd programmeren heeft. Maar eerder genoemde voordelen wegen hier tegenop.

### **4.4.2 Beschouwing**

Het is oninteressant om de geschreven code door te nemen. In plaats daarvan wil ik op een abstracter niveau kort de implementatie beschouwen.

#### **Object georiënteerd versus procedureel**

Vanwege het gemak waarmee in Matlab experimenten zijn uit te voeren en relatief snel te programmeren valt, is gekozen voor Matlab. Het doel is hier een eindig onderzoek. Wanneer er een adviestool wordt gemaakt zou een object georiënteerde omgeving de voorkeur verdienen omdat deze makkelijker uit te breiden en te wijzigen is. Het ontbreken van “echte” objecten leidt in Matlab soms tot onnodig ingewikkelde structuren.

#### **Caching**

De simulaties vereisen een grote hoeveelheid rekentijd. De hoeveelheid gebruikt geheugen valt relatief mee. Om rekentijd te besparen kan het zeer lonend zijn om bepaalde bewerkingen, waarvan de resultaten meermaals hergebruikt worden, zo min mogelijk onnodig te herhalen.

#### **Model op papier omzetten naar implementatie en debugging**

De vertaalslag van een model op papier naar implementatie moet niet onderschat worden. Concepten die op papier eenvoudig lijken kunnen bij implementatie moeilijk of onhaalbaar blijken. Niet alles kan vooraf voorzien worden, maar een computerprogramma is zeer exact en alles moet kloppen. En wanneer de implementatie voltooid is, blijkt het opsporen van niet syntactische fouten nog een flinke hoeveelheid tijd in beslag te nemen. Mijn advies is om voor implementatie en debugging voldoende tijd in te plannen bij het bouwen van een eventueel adviestool.





## 5. Analyse

### 5.1 Inleiding

Aan de hand van de in Matlab geschreven simulatie is het mogelijk om de opgestelde modellering nader te bestuderen. De uitgevoerde analyses en de verkregen resultaten worden in dit hoofdstuk beschreven. In het bijzonder zijn we geïnteresseerd hoe het model reageert op veranderingen in invoerparameters, zogenaamde gevoeligheidsanalyse.

### 5.2 Basisparameters

De basisparameters zijn beschreven in het vorige hoofdstuk in de paragraaf over de invoerdata. Er zijn 1000 simulaties uitgevoerd. En de data die hieruit is verkregen, is samengevat in plots. Voor ieder van de 20 simulatiejaren zijn acht verschillende soorten output bijgehouden:

1. **Percentage rented**  
Gedeelte van de tijd dat verhuur plaatsvond. Gemiddeld over alle panden.
2. **Percentage void**  
Gedeelte van de tijd dat leegstand plaatsvond. Gemiddeld over alle panden.
3. **#Defaults**  
Aantal contract default gevallen (huurders default + regulier contract afloop).  
Totaal over alle panden.
4. **Fixed Costs**  
Vaste kosten door leegstand Totaal over alle panden.
5. **Variable Costs**  
Variabele kosten door leegstand. Totaal over alle panden.
6. **Total Costs**  
Vaste kosten + variabele kosten. Totaal over alle panden.
7. **Rent**  
Huurinkomsten. Totaal over alle panden.

## 8. Netto

Huurinkomsten – kosten door leegstand. Totaal over alle panden.

De rode lijn in de volgende plots stelt de gemiddelde waarde voor. Om een inzicht te krijgen in de spreiding van de data zijn ook percentielen (per tien procent: 0.1%, 10%, 20%, 30%, 40%, 60%, 70%, 80%, 90%, 100%) met een stippellijn weergegeven (Fig. 32).

### Cyclisch patroon

Wat opvalt in iedere outputcategorie is een cyclisch patroon dat zich iedere vijf jaar herhaalt. Dit is te verklaren wanneer we naar het aantal defaults (Let op: het gaat hier over contract defaults = huurders defaults + reguliere contract beëindigingen) kijken en bedenken dat de contractperiode vijf jaar bedraagt. We zien dat gedurende de eerste paar jaren het aantal defaults per jaar vrij laag is, licht toenemend naarmate de tijd vordert; de defaultkans per jaar is immers vrij laag en de kans op default is hoger naarmate de tijdspanne toeneemt. De piek die iedere vijf jaar te zien is wordt veroorzaakt door contracten die massaal aflopen en al dan niet worden verlengd. Omdat contracten in default kunnen gaan neemt het aantal contracten dat iedere vijf jaar verlengd kan worden af naarmate de tijd vordert. Dit is terug te zien in een gestage demping van de pieken en ook een verbreding van deze pieken. Het aantal defaults is bepalend voor de andere outputcategorieën, waardoor overal een cyclisch patroon te zien is.

### Buitenste 0%-10% / 90%-100%

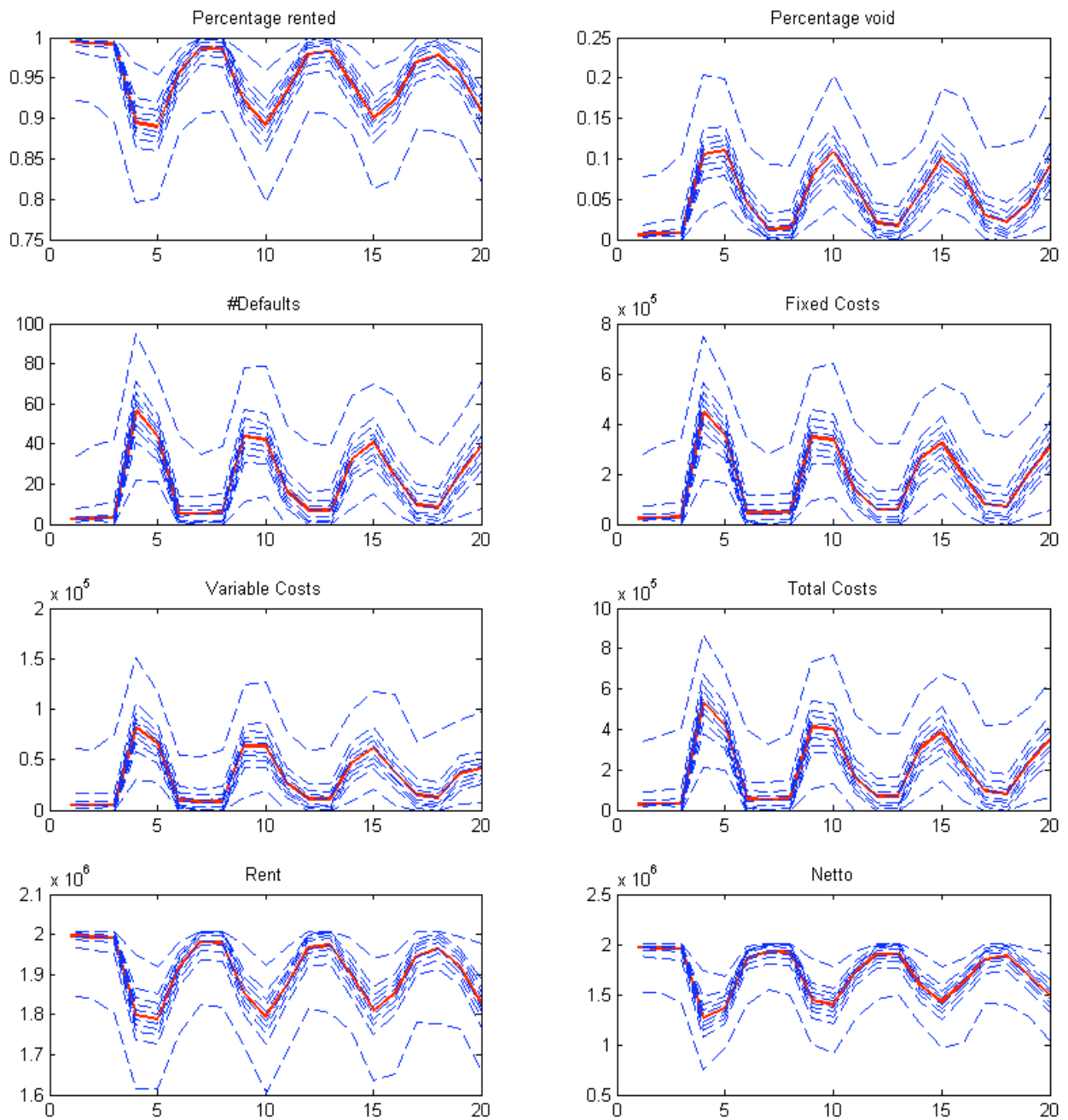
Wanneer er naar het aantal defaults wordt gekeken, zien we dat de bandbreedte 90% tot 100% aanzienlijk breder is dan andere bandbreedtes van 10%. Dit is te verklaren vanuit de modellering met behulp van de Clayton Copula. Lage defaultduraties (tijd tot de volgende default) zijn hierbij gecorreleerd, en dit valt samen met de meest negatieve situatie die in de bandbreedte behorende bij 90% tot 100% is te zien. De Clayton copula heeft aan de andere kant van het spectrum, hoge defaultduraties, veel lagere geborneerdheid. Wat te merken is aan een kleinere bandbreedte behorende bij 0% tot 10%. Het aantal defaults is zeer bepalend voor andere outputcategorieën waardoor ook daarin aan de boven of onderzijde een brede categorie is te zien.

### Standaard deviatie

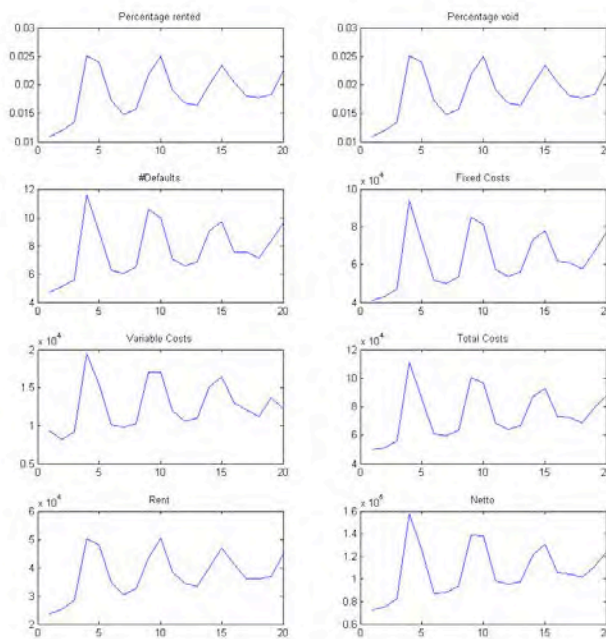
De standaard deviatie vertoont ook een cyclisch beeld (Fig. 33). Verklaring hiervoor is in lijn met eerder al uiteengezette gedrag van het aantal defaults. Op het moment dat massaal contracten aflopen is er de grootste deviatie mogelijk in het aantal defaults. De rest van de tijd is het aantal defaults laag, resulterend in een lage standaard deviatie. Zoals al meerdere malen is gezegd is het aantal defaults een belangrijke factor voor andere outputcategorieën, waardoor deze ook eenzelfde cyclisch patroon vertonen in hun standaard deviatie. Merk ook op dat pieken niet samenvallen met dalen, maar met pieken (anders dan bij eerdere grafieken van de waardes), dit omdat de standaard deviatie geen onderscheid maakt in positieve en negatieve afwijking.



## Huurders Default – Risico modellering



**Fig. 32: Resultaten met basisparameters**



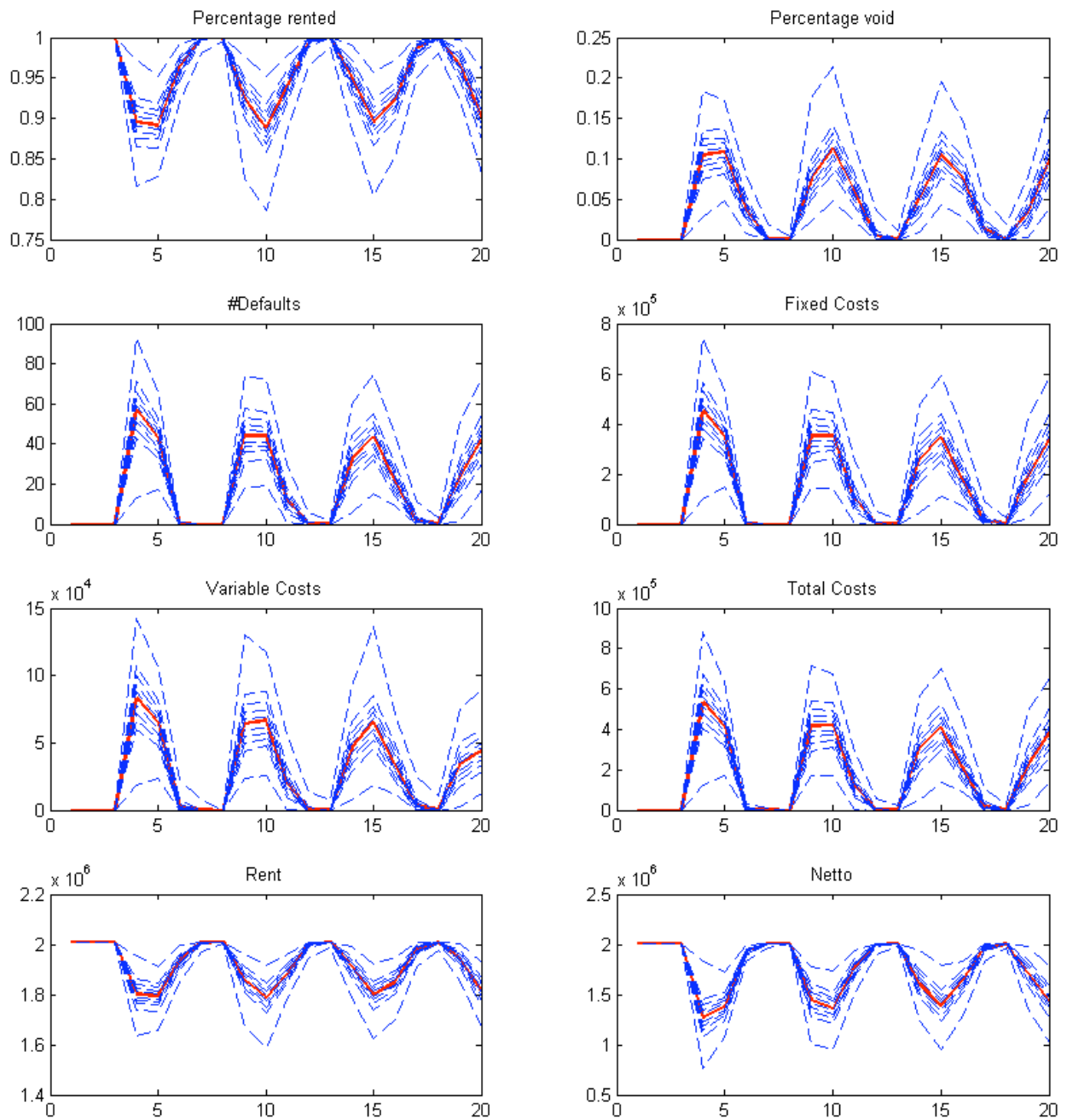
**Fig. 33: Standaard deviatie met basisparameters**

### 5.3 Gecorreleerde vs. ongecorreleerde defaulttijden

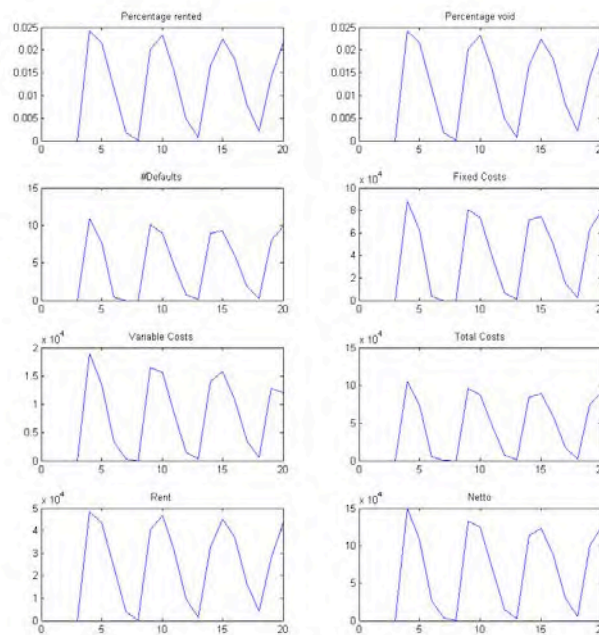
In het geïmplementeerde model is gewerkt met gecorreleerde defaulttijden. Om te zien wat voor effect deze correlatie heeft zijn er ook simulaties gedraaid met ongecorreleerde defaulttijden.

Op het eerste gezicht vertoont het resultaat in geval van ongecorreleerdheid van defaulttijden hetzelfde beeld als het resultaat in geval van gecorreleerde defaulttijden; dus een cyclisch patroon en ook een cyclisch patroon voor de standaard deviaties (Fig. 34 en Fig. 35). Meest in oog springend verschil is de grootte van de standaard deviaties. In het geval van gecorreleerde defaults is de standaard deviatie groter dan in het ongecorreleerde geval. Vooral in het ergste geval, dat correspondeert met, in het geval van aantal defaults, de buitenste 90%-100%. Dit is vooral te merken tussen de “piekmomenten” (waar massaal contracten aflopen) door. Bij ongecorreleerde default is de standaard deviatie zeer klein, maar in het geval van gecorreleerde defaults zien we dat de uiterste 90%-100% een grote spreiding kent. Dit laatste wordt veroorzaakt door de geborneerdheid van het negatieve effect (korte defaulttijden kennen een grote correlatie).

Modellering van gecorreleerde defaulttijden geeft een ander beeld dan het geval van ongecorreleerde defaulttijden, en correlatie zou in iedere vastgoed modellering verwerkt moeten worden voor een realistischer en accurater model.



**Fig. 34: Resultaten in het geval van ongecorreleerde defaulttijden**



**Fig. 35: Standaard deviatie in het geval van ongecorreleerde defaulttijden**

Door de opzet van de simulatie, namelijk dat alle begincontracten beginnen tussen jaar -1 en jaar 0, zien we een redelijk scherp (smal en steil) piekpatroon iedere vijf jaar. Wanneer de begintijd uit een groter interval wordt getrokken, bijvoorbeeld van jaar -3 tot jaar 0, dan zullen de pieken breder en vlakker zijn.

## 5.4 Gevoeligheidsanalyse

### 5.4.1 Output

Voor de komende sensitiviteit analyses aggregeren we de resultaten over de gehele simulatieperiode om de analyse handelbaar te maken. Een kleine aanpassing in de definities van de outputcategorieën is vereist:

**1. Percentage rented**

Gedeelte van de tijd dat verhuur plaatsvond. Gemiddeld over alle panden en gemiddeld over alle simulatiejaren.

**2. Percentage void**

Gedeelte van de tijd dat leegstand plaatsvond. Gemiddeld over alle panden en gemiddeld over alle simulatiejaren.

**3. #Defaults**

Aantal contract default gevallen (huurders default + regulier contract afloop).  
Totaal over alle panden, gemiddeld over alle simulatiejaren.

**4. Fixed Costs**

Vaste kosten door leegstand Totaal over alle panden en totaal over de gehele simulatieperiode.

**5. Variable Costs**

Variabele kosten door leegstand. Totaal over alle panden en totaal over de gehele simulatieperiode.

**6. Total Costs**

Vaste kosten + variabele kosten. Totaal over alle panden en totaal over de gehele simulatieperiode.

**7. Rent**

Huurinkomsten. Totaal over alle panden en totaal over de gehele simulatieperiode.

**8. Netto**

Huurinkomsten – kosten door leegstand. Totaal over alle panden en totaal over de gehele simulatieperiode.

### 5.4.2 Contract periode

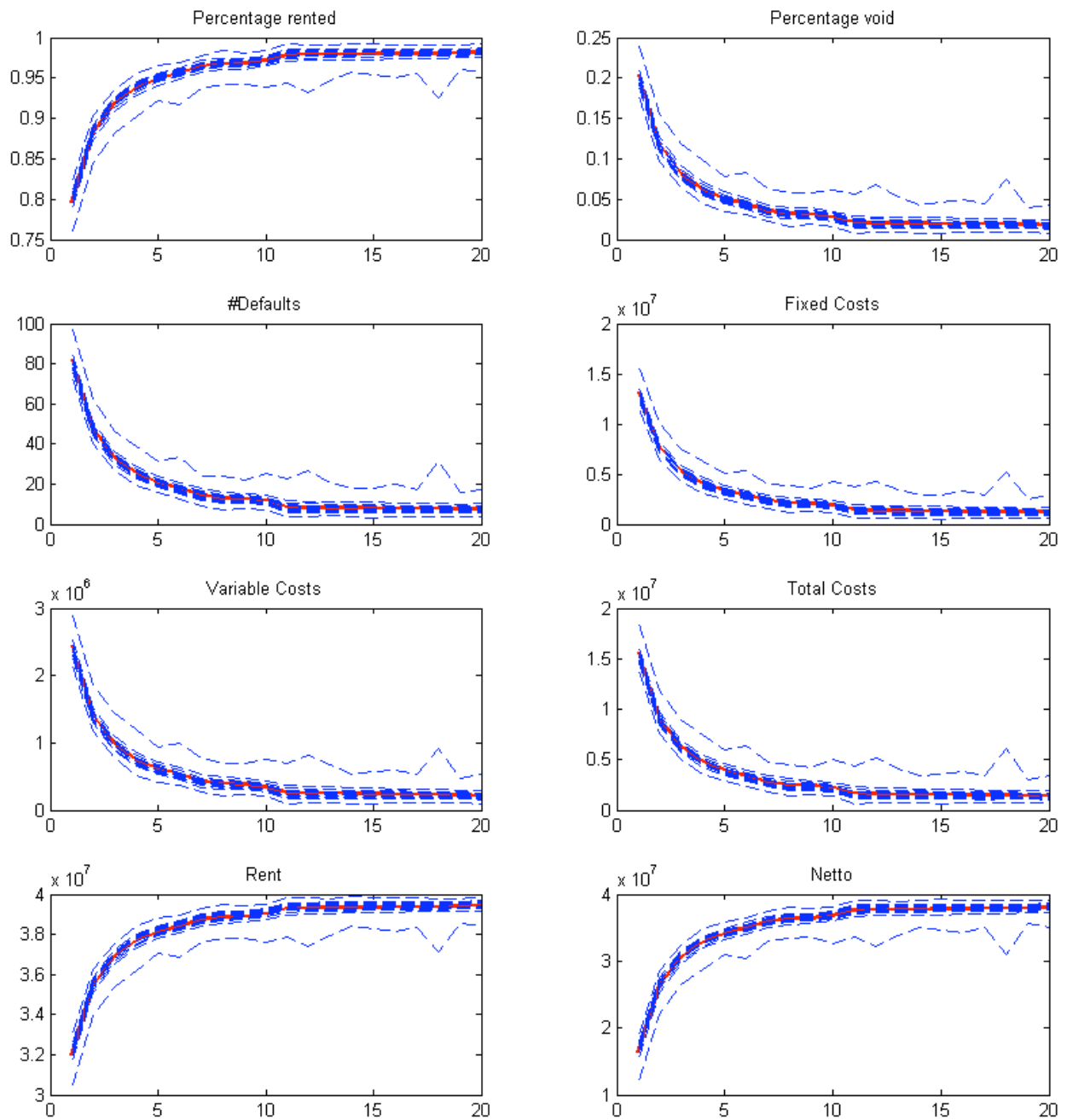
De standaardwaarde voor de contractperiode in de simulatie bedraagt vijf jaar. Er zijn voor verschillende waarden van de contractperiode simulaties gedraaid: 1 jaar, 2 jaar, ..., 20 jaar. De resultaten zijn in de plots te zien (Fig. 36). Voor het gemiddelde is wederom een rode lijn gebruikt en om de 10% zijn de percentielen weergegeven als stippellijn.

Er is te zien dat voor 90% van de simulaties de waarden zich weinig van de gemiddelde waarde bevinden. De buitenste 10% van de simulaties (aan één zijde) vertonen een grotere spreiding, wat komt door het al meerdere malen besproken effect van de gecorreleerde defaults door de Clayton copula modellering.

Het valt op dat voor de eerste twee à drie jaar de contractperiode een zeer sterk effect heeft op de verschillende outputcategorieën. Dit effect vlakt sterk af tussen 3 en 5 jaar, en boven de 6 is het effect zeer klein.

Wat we hieruit kunnen leren is dat een contractperiode van 5 jaar, wat zeer gebruikelijk is in de vastgoedwereld, een goede waarde is. Wanneer men neigt naar kortere contractperiodes dan zal dit heel snel een negatief effect hebben op investeringen. Er is nog winst te behalen door leningen te kiezen waarbij de contractperiodes hoger dan 5 jaar zijn. Bij dit laatste moet ook in acht worden genomen dat boven de 8 jaar het positieve effect marginaal wordt.

## Huurders Default – Risico modellering



**Fig. 36: Gevoeligheidsanalyse – Contract periode vs. in 5.4.1 beschreven outputcategorieën.**

### 5.4.3 Gemiddelde leegstandsperiode

Er zijn simulaties gedraaid met verschillende waarden voor de gemiddelde leegstandsperiode. In de plots zijn de resultaten te zien voor een gemiddelde leegstand van 3 maanden, 4 maanden, ..., 12 maanden. Een gebruikelijke aanname is een leegstand van 6 maanden. De rode lijn geeft het gemiddelde aan en per 10% zijn er met stippellijnen de percentielen weergegeven (Fig. 37).

Er is een duidelijk lineair verband te zien. Verder zien we dat de spreiding groeit naarmate de gemiddelde leegstandsperiode toeneemt, behalve bij het aantal defaults en bij de vaste leegstandskosten. Net als bij alle eerdere resultaten herkennen we ook hier weer een aan een zijde het typerende effect van gecorreleerde defaults.

Het is interessant om te zien dat het gemiddelde aantal defaults slechts zeer gering varieert, en ook dat dit toch een groot effect heeft op bepaalde outputcategorieën zoals het percentage tijd dat leegstand/verhuur is en de variabele kosten. De variabele kosten nemen redelijk lineair toe bij ongeveer gelijk blijvend aantal defaults, omdat variabele kosten direct worden beïnvloedt door leegstandsduratie. Iedere eenheid toename in gemiddelde leegstandsperiode, gaat gepaard met een vaste toename in de variabele kosten:

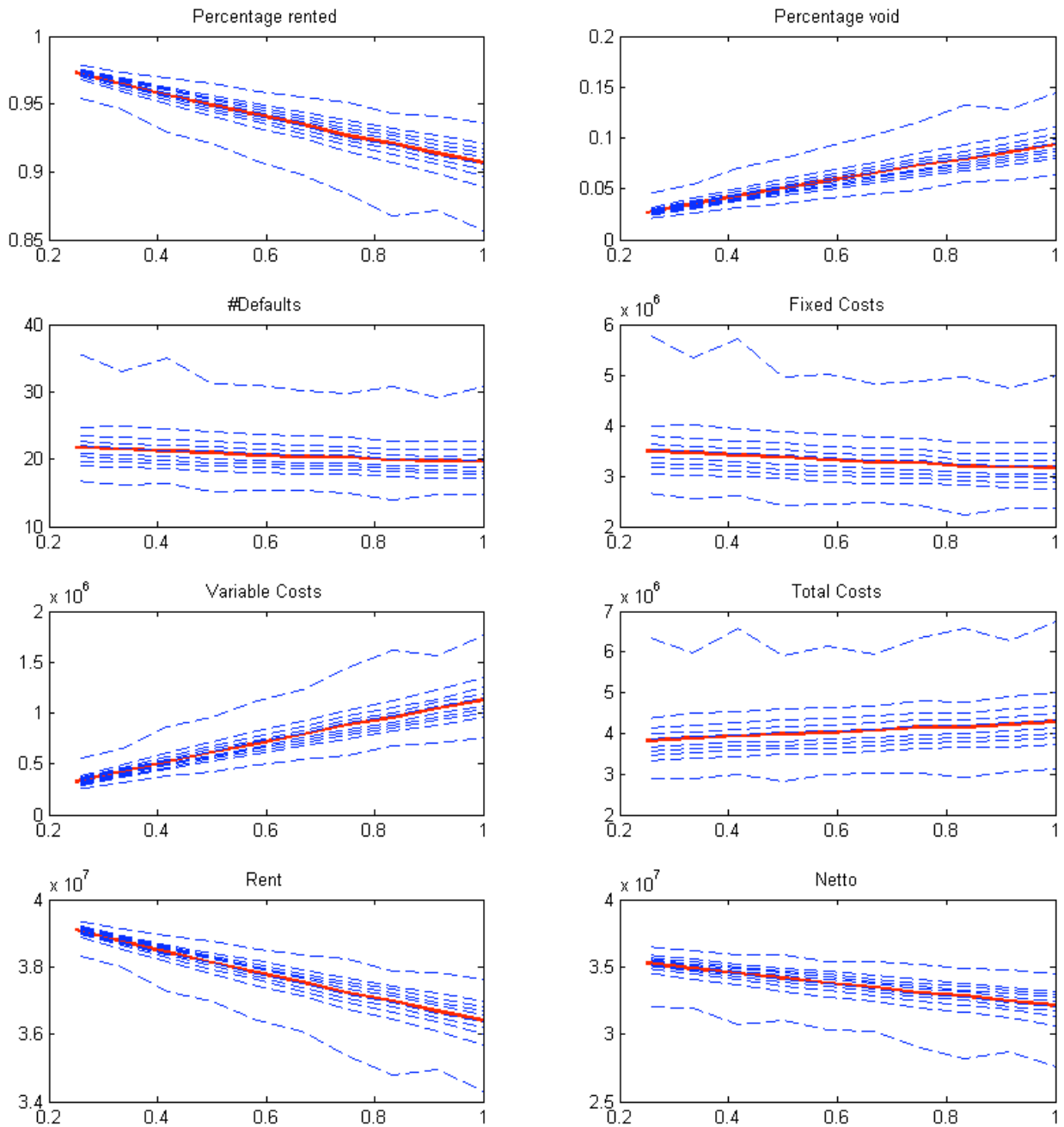
$$\text{Leegstandsperiode} * \text{variabele leegstandskosten per eenheid tijd} = \text{variabele leegstandskosten}$$

De lering die we hieruit kunnen trekken is dat ondanks het aantal defaults een van de belangrijkste factoren is, het altijd samen bekeken moet worden met andere metrieken om een goed beeld van de situatie te krijgen.

Verder is het belangrijk om de gemiddelde leegstand vanuit historische data zo nauwkeurig mogelijk te schatten. Een verschil tussen bijvoorbeeld zes maanden en negen maanden leidt al gauw tot een redelijk verschil in de kosten. In de vastgoedwereld wordt soms te eenvoudig uitgegaan van ongeveer een half jaar.



## Huurders Default – Risico modellering



**Fig. 37: Gevoeligheidsanalyse – Gemiddelde leegstandsperiode vs. in 5.4.1 beschreven outputcategorieën.**

#### **5.4.4 Contract voortzettingkans**

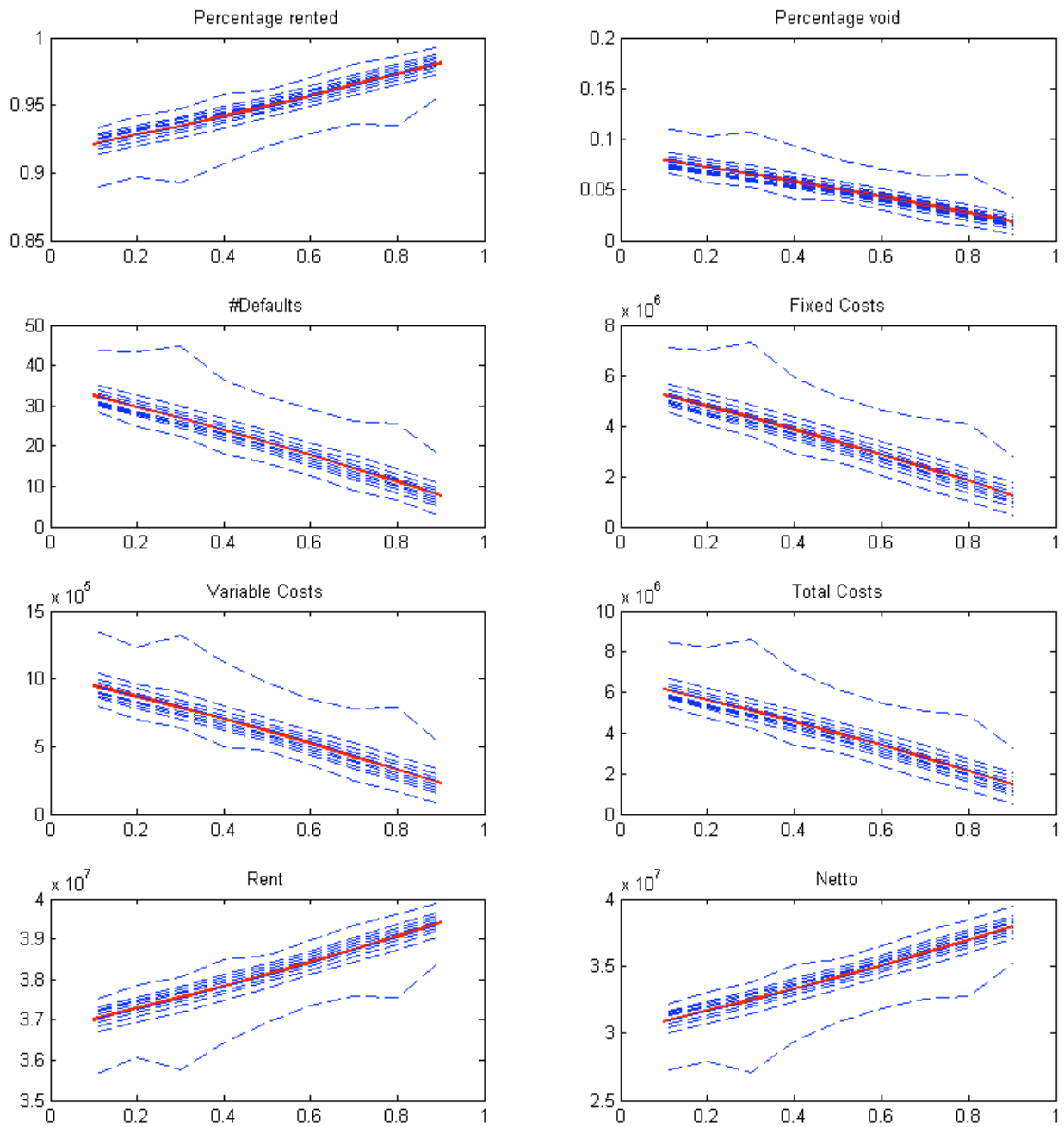
Er zijn simulaties gedraaid met verschillende waarden voor de contract voortzettingkans. In de plots zijn de resultaten te zien voor kansen 0.1, 0.2, ..., 0.9 maanden. Een aanname die in de vastgoedwereld wordt gehanteerd is een kans van 0.5. De rode lijn geeft het gemiddelde aan en per 10% zijn er met stippellijnen de percentielen weergegeven (Fig. 38).

Ook hier is weer een duidelijk lineair verband te zien. Verder zien we dat de spreiding zeer constant blijft. Net als bij alle eerdere resultaten herkennen we ook hier weer een aan een zijde het typerende effect van gecorreleerde defaults.

In tegenstelling tot variatie van de gemiddelde leegstand, heeft variatie van de contract voortzettingkans grote invloed op alle categorieën (dus ook op het aantal defaults en de vaste kosten voor leegstand).

Hoewel de contract voortzettingkans van grote invloed is op het model, wordt er in de praktijk weinig speciale aandacht aan besteed. Er zijn gevallen waar simpelweg met een kans van 0.5 wordt gewerkt. Een goede schatting van de waarde voor de contract voortzettingkans is vereist.

## Huurders Default – Risico modellering



**Fig. 38: Gevoeligheidsanalyse – Contract voortzettingkans vs. in 5.4.1 beschreven outputcategorieën.**

## 5.5 Conclusie

In dit hoofdstuk zijn de resultaten getoond van de vele simulaties die zijn gedraaid. De observaties die hierbij zijn gemaakt, zijn in de voorgaande paragrafen uiteengezet. De belangrijkste lessen die geleerd zijn, zijn onder andere:

- Afhankelijke defaults geven een wezenlijk verschil met het geval van onafhankelijke defaults. Afhankelijkheid zou in ieder model inbegrepen moeten worden.
- De afhankelijke defaults leiden tot een grote spreiding, vooral merkbaar aan de laatste 10% aan waarnemingen aan de “negatieve” (negatief vanuit het oogpunt van inkomsten) zijde. Worst case scenario’s worden versterkt door afhankelijke defaults. Dit is in lijn met de gekozen modellering met behulp van de Clayton copula; korte tijden tot default zijn sterk gecorreleerd.
- Contract periodes van vijf jaar of hoger zijn gewenst. Vanaf acht jaar maakt dit echter weinig uit. Vier jaar en lager leidt rap tot verslechterende situaties.
- Gemiddelde leegstandsperiode is van weinig invloed op het aantal defaults. Omdat de vaste kosten direct gekoppeld zijn aan het aantal defaults is ook hierop weinig invloed. De variabele kosten voor leegstand vertonen een duidelijk positief lineair verband met de gemiddelde leegstandsperiode.
- De contract voortzettingkans is van invloed op het aantal defaults, de vaste kosten voor leegstand en de variabele kosten van leegstand. Er is een negatief lineair verband tussen de contract voortzettingkans en eerder genoemde output categorieën.
- Het is van groot belang om de gemiddelde leegstandsperiode en de contract voortzettingkans zo goed mogelijk te schatten vanuit beschikbare data.





## 6. Conclusie

### 6.1 Inleiding

Modellering is geen objectieve bezigheid. Verschillende personen kunnen bij de modellering van eenzelfde systeem op verschillende wijze te werk gaan en op verschillende resultaten eindigen. Zo is het ook met de modellering in dit verslag. In dit laatste hoofdstuk wil ik kritisch ingaan op de keuzes die zijn gemaakt bij de modellering. Tevens komen onderwerpen aan bod die van belang zijn, maar niet in eerdere hoofdstukken zijn geplaatst omwille het verloop daarvan niet te verstoren.

### 6.2 Achtergronden

#### 6.2.1 De correlatiecoëfficiënt

De correlatiecoëfficiënt geeft de mate weer waarin twee variabelen aan elkaar gerelateerd zijn. Wat sommigen niet weten is dat de correlatiecoëfficiënt alleen voor lineaire verbanden zinnig is. Twee variabelen kunnen wel degelijk een verband vertonen terwijl de correlatiecoëfficiënt een waarde dichtbij 0 heeft. Een waarde van de correlatiecoëfficiënt dichtbij 1 wijst op een lineair verband tussen de variabelen. Andersom geldt de logica niet, een verband tussen variabelen hoeft niet te leiden tot een waarde van de correlatiecoëfficiënt die dichtbij 1 ligt (Fig. 39).

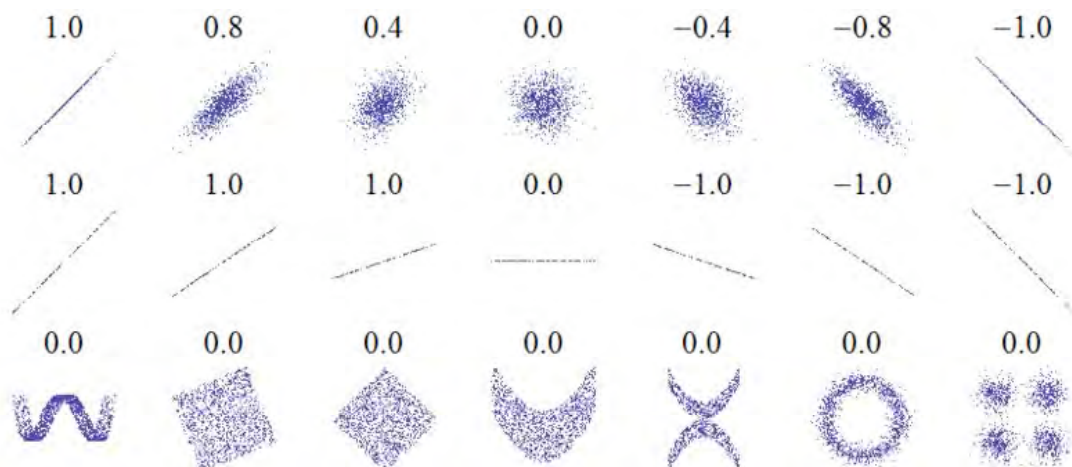
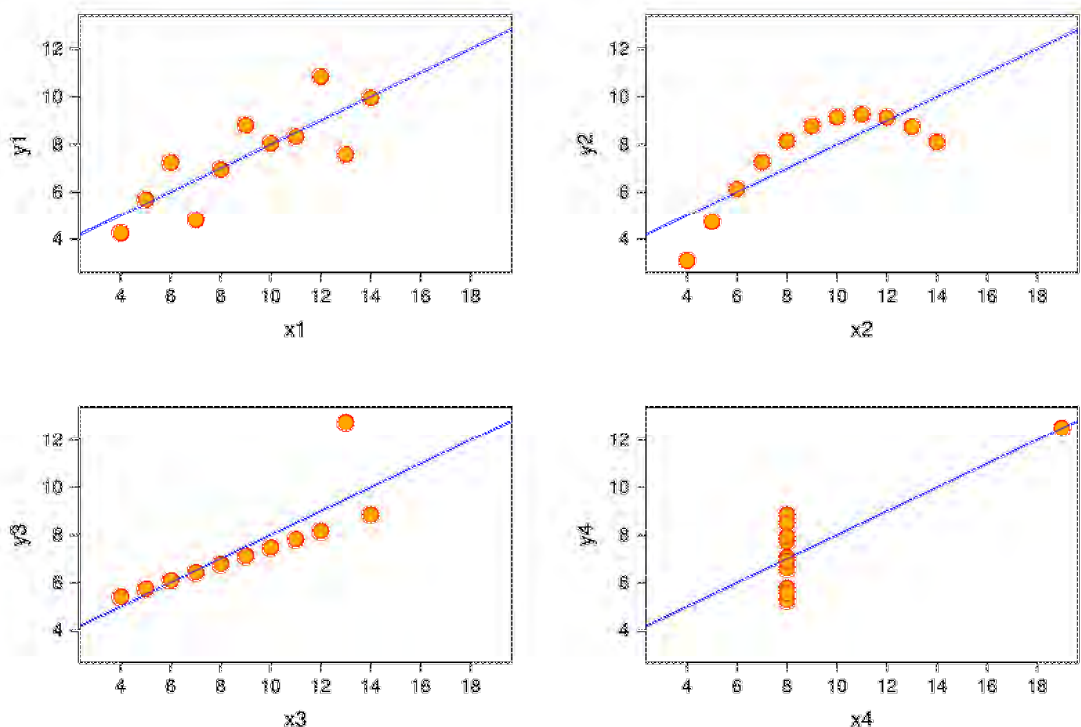


Fig. 39: Correlatiecoëfficiënten en beperkte afhankelijkheidsinformatie (1)<sup>4</sup>

Belangrijk is het dat men beseft dat het hier gaat om statistisch verband. Een “echt” verband, of een causaliteitsrelatie kan men uit de correlatiecoëfficiënt, of wat betreft in de statistiek, niet afleiden. Een voorbeeld hiervan is onderzoek naar agressie bij kinderen en de hoeveelheid gewelddadige computerspellen die deze kinderen spelen. Sommige onderzoekers menen dat hun statistisch onderzoek aantoon(!)t dat er een positief verband tussen de twee is en dat computerspellen agressiviteit veroorzaakt. Dit kan men niet zo stellen. Ten eerste kan het een de oorzaak zijn van de ander maar ook andersom. Maar belangrijker, misschien is er helemaal geen verband en is het geobserveerde puur toeval, of misschien wordt deze veroorzaakt door eenzelfde achterliggende oorzaak. Statistisch verband  $\neq$  oorzakelijk verband.

Voor eenzelfde waarde van de correlatiecoëfficiënt, bijvoorbeeld 0.81, zijn verschillende achterliggende verklaringen mogelijk zoals in het volgende figuur te zien is (Fig. 40). De grafiek linksboven geeft ongeveer aan wat iemand zou verwachten bij een correlatie van 0.81, namelijk een redelijk lineair verband met nog enige ruis. De grafiek rechtsboven vertoont een stijgende trend maar visueel is duidelijk dat het verband niet lineair is (bij meer punten aan de uiteindes zou een parabolachtig figuur ontstaan). De grafiek linksonder toont een perfect lineair verband aan, ware het niet dat een uitbijter de waarde van de correlatiecoëfficiënt significant doet afwijken van 1. In de grafiek rechtsonder is te zien dat de waarde van de dimensie op de verticale as geen invloed heeft op de waarde behorende bij de dimensie op de horizontale as, ware het niet dat er ook hier weer een uitbijter is. Deze is zelfs zo extreem dat het de waarde van de correlatiecoëfficiënt omhoog weet te trekken naar 0.81, terwijl het lineaire verband in dit geval gezien de data praktisch nul is.



**Fig. 40: Correlatiecoëfficiënten en beperkte afhankelijkheidsinformatie (2)**



De laatste twee “fouten” kunnen voorkomen worden wanneer uitbijters uit de data gefilterd worden alvorens deze wordt gebruikt. De “fout” in de figuur rechtsboven kan voorkomen worden door een grotere dataset te gebruiken waaruit de volledige vorm van het verband naar voren komt. Het woord “fout” is tussen haakjes geschreven omdat het niet een fout in de correlatiecoëfficiënt metriek betreft, deze doet zijn werk volgens de specificatie. De “fout” ligt in de foute interpretatie van de metriek waarin men kan vallen wanneer niet wordt gelet op karakteristieke gedragingen van de metriek.

### **6.2.2 Model versus werkelijkheid**

Modellen zijn vereenvoudigingen van de werkelijkheid en dit feit moet altijd in het achterhoofd gehouden worden. Als er uitkomsten uit een model komen waarvan men zeker weet dat die verre van de werkelijkheid liggen is het model, in ieder geval voor de geteste gevallen, niet goed genoeg.

Vooraf in een context waarbij veel variabelen en afhankelijkheden een rol spelen, en een modelresultaat volgens “buikgevoel” geen juiste waarde geeft maar dit niet met zekerheid kan worden bepaald, wordt de plaatsing van een model binnen het businessproces kritiek. Want moet men aan een model een adviesrol toekennen of een beslissende rol? Aan de ene kant is het probleem te complex zodat het alleen kwantitatief is in te schatten middels een model. Aan de andere kant kan het model goed zijn maar nog fouten of niet in bevredigende mate gemodelleerde zaken zitten. Er is geen eenduidig antwoord te geven, maar ik zou persoonlijk een model in de eerste instantie als adviesmiddel gebruiken. En wanneer er geen fouten of tekortkomingen in de modellering te vinden zijn het advies dat het model geeft accepteren als het best onderbouwde antwoord dat op dat moment mogelijk is. Het is altijd dan de alternatieve optie, keuzes nemen op “buikgevoel”. Immers, wie zegt dat een onverwachte uitkomst die niet strookt met het gevoel perse fout is. Een klassiek voorbeeld hiervan is bijvoorbeeld wanneer iemand de vraag wordt voorgelegd hoe groot de kans is dat twee personen op eenzelfde dag jarig zijn in een groep van 30 personen. Iemand die niet bekend is met dit vraagstuk zal de kans gevoelsmatig laag inschatten en verrast zijn met de voor het gevoel hoge kans.

Het is belangrijk om ook relatief jonge modellen hun kans te geven om zich te laten bewijzen, en om niet direct bij de eerste proefrondes in de business waarbij het voorspelde resultaat afwijkt van het werkelijke resultaat, te verwerpen. Uiteraard moet er bij een lange trackrecord van fouten vraagtekens bij het model worden gezet en uitgebreidere modelcontroles worden ingezet.

## **6.3 Suggesties voor vervolgonderzoek**

Projecten zijn per definitie gelimiteerd in tijd en er moet een keuze worden gemaakt wat men wil onderzoeken. Zo ook bij dit afstudeerwerk. Bij ieder onderzoek of model zijn er ook altijd verbeteringen mogelijk. In de volgende paragrafen wil ik suggesties doen voor mogelijk vervolgonderzoek.

### **6.3.1 Correlatie / regressie zijn niet de enige technieken**

In dit onderzoek is gewerkt met simulatie van kansverdelingen en copulas. Dit zijn echter niet de enige technieken voor modelvorming. Naast deze statistisch georiënteerde technieken zijn er ook technieken uit de machine learning (bijvoorbeeld neurale netwerken) en data mining (bijvoorbeeld decision trees) die bruikbaar kunnen zijn om bruikbare structuren uit de data te halen. Deze kunnen dan weer vertaald worden voor modelverfijning en kunnen mogelijk bijdragen aan een sterker model.

### **6.3.2 Contractstructuren**

In de modellering is er uitgegaan van starre contract – verhuurbare eenheid structuren. Hierbij wordt er dus vanuit gegaan dat verhuurbare eenheden altijd in eenzelfde groepering via een contract worden verhuurd, wat in de praktijk niet altijd zo is. Verder onderzoek is vereist om dit realistischer te kunnen modelleren.

### **6.3.3 Opname van sociaal-economische informatie**

In dit onderzoek is een fictieve dataset gebruikt. In de echte wereld zijn er (tegen betaling) databases vol met uitgebreide demografische en economische data. Met deze data kan het model eventueel verder verfijnd worden. Wellicht is het van belang om bijvoorbeeld de industrie van de huurder in de modellering te betrekken.

## **6.4 Slotconclusie / Eindwoord**

In dit afstudeeronderzoek is een model opgesteld voor het verhuurproces in een vastgoedportefeuille. Hierbij is onder andere afhankelijkheid van defaults opgenomen met behulp van modellering met behulp van copulas, in het bijzonder de Clayton copula.

Aan de hand van een geïmplementeerde simulatie zijn verschillende tests uitgevoerd. Hieruit blijkt dat de opname van gecorreleerde defaults tot wezenlijk verschil leidt in vergelijking met de situatie met ongecorreleerde defaults. Verder bleek de simulatie een handige tool om gevoeligheidsanalyse uit te voeren naar verschillende input parameters (voor uitgebreide bespreking verwijs ik naar hoofdstuk 6).

Hoewel de modellering in dit afstudeerwerk een goede basis vormt, is er nog voldoende mogelijkheid tot vervolgonderzoek om tot een nog betere modellering te komen.



## Appendix A – Inverse Transformatie Methode

De inverse transformatie methode stelt ons in staat random trekkingen te verrichten van een stochast  $X$ , waarvan de distributiefunctie strikt stijgend is, met behulp van een random trekkingen uit de standaard uniforme verdeling.

De stelling is dat:

$$U \sim \text{Uniform}(0,1)$$

$$F_X(X) \sim \text{Uniform}(0,1)$$

$$F_X(X) = U$$

$$F_X^{-1}(U) = X$$

De laatste vergelijking stelt ons in staat om een random  $X$  trekking te berekenen door een random getrokken standaard uniform getal in de inverse van de distributiefunctie van  $X$  te stoppen.

Hieronder volgt het bewijs:

$$F_X^{-1}(U) = C$$

$$\Rightarrow F_C(c) = P(C \leq c) = P(F_X^{-1}(U) \leq c) = P(F_X(F_X^{-1}(U)) \leq F_X(c)) = P(U \leq F_X(c)) = F_X(c)$$

$$\Rightarrow F_X^{-1}(U) \sim F_C(c) = F_X(c) \sim X$$

Zie (Ross, 2002)<sup>5</sup> voor een alternatief bewijs.

Noot: De eis dat de distributiefunctie strikt stijgend moet zijn komt voort uit het feit dat er een één-op-één relatie moet zijn tussen invoer en uitkomst (injectie), immers, het kan niet zo zijn dat bij een bepaalde waarde van  $U$  meerdere mogelijke waarden voor  $X$  uit een simulatie trekking komen. Verder geldt er dat voor iedere uitkomst een invoer moet zijn (surjectie). De distributiefunctie is dus bijjectief (injectie + surjectie).

## Appendix B – Clayton Copula (Kansdichtheid)

De kansdichtheidsfunctie van de Clayton copula wordt gevonden door de Clayton copula te differentiëren:

$$\begin{aligned}
 \frac{d^n}{du_1 \dots du_n} C(u_1, \dots, u_n) &= \frac{d^n}{du_1 \dots du_n} \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-1/\theta} = \frac{d^{n-1}}{du_2 \dots du_n} \left[ \frac{d}{du_1} \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-1/\theta} \right] \\
 &= \frac{d^{n-1}}{du_2 \dots du_n} \left[ -\frac{1}{\theta} \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-1/\theta-1} \cdot -\theta u_1^{-\theta-1} \right] = \frac{d^{n-1}}{du_2 \dots du_n} \left[ u_1^{-\theta-1} \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-(1/\theta+1)} \right] \\
 &= \frac{d^{n-2}}{du_3 \dots du_n} \left[ \frac{d}{du_2} u_1^{-\theta-1} \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-(1/\theta+1)} \right] \\
 &= \frac{d^{n-2}}{du_3 \dots du_n} \left[ u_1^{-\theta-1} \cdot -(1/\theta+1) \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-(1/\theta+1)-1} \cdot -\theta u_2^{-\theta-1} \right] = \frac{d^{n-2}}{du_3 \dots du_n} \left[ u_1^{-\theta-1} u_2^{-\theta-1} \theta (1/\theta+1) \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-(1/\theta+2)} \right] \\
 &= \frac{d^{n-3}}{du_4 \dots du_n} \left[ \frac{d}{du_3} u_1^{-\theta-1} u_2^{-\theta-1} \theta (1/\theta+1) \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-(1/\theta+2)} \right] \\
 &= \frac{d^{n-3}}{du_4 \dots du_n} \left[ u_1^{-\theta-1} u_2^{-\theta-1} \theta (1/\theta+1) \cdot -(1/\theta+2) \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-(1/\theta+2)-1} \cdot -\theta u_3^{-\theta-1} \right] = \frac{d^{n-3}}{du_4 \dots du_n} \left[ u_1^{-\theta-1} u_2^{-\theta-1} u_3^{-\theta-1} \theta^2 (1/\theta+1)(1/\theta+2) \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-(1/\theta+3)} \right] \\
 &= \dots \\
 &= \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-(1/\theta+n)} \prod_{i=1}^n u_i^{-\theta-1} (\theta)(1/\theta+i-1) = \left( 1 - n + \sum_{i=1}^n u_i^{-\theta} \right)^{-1/\theta-n} \prod_{i=1}^n u_i^{-\theta-1} (1+\theta i - \theta)
 \end{aligned}$$

## Appendix C – Clayton Copula (Simulatie)<sup>6</sup>

$$C(u_1, \dots, u_n) = \left(1 - n + \sum_{i=1}^n u_i^{-\theta}\right)^{-1/\theta}$$

$$r_1, \dots, r_n \sim \text{Uniform}(0,1)$$

$$r_1 = C(u_1) = (u_1^{-\theta})^{-1/\theta} = u_1$$

$$\Rightarrow u_1 = r_1$$

$$r_2 = \frac{C(u_1, u_2)}{C(u_1)} = \frac{(-1 + u_1^{-\theta} + u_2^{-\theta})^{-1/\theta}}{(u_1^{-\theta})^{-1/\theta}} = \left(\frac{-1 + u_1^{-\theta} + u_2^{-\theta}}{u_1^{-\theta}}\right)^{-1/\theta}$$

$$\Rightarrow r_2^{-\theta} = \left(\frac{-1 + u_1^{-\theta} + u_2^{-\theta}}{u_1^{-\theta}}\right) \Rightarrow r_2^{-\theta} u_1^{-\theta} = -1 + u_1^{-\theta} + u_2^{-\theta} \Rightarrow r_2^{-\theta} u_1^{-\theta} + 1 - u_1^{-\theta} = u_2^{-\theta}$$

$$\Rightarrow u_2 = (r_2^{-\theta} u_1^{-\theta} + 1 - u_1^{-\theta})^{-1/\theta} = ((r_2 u_1)^{-\theta} + 1 - u_1^{-\theta})^{-1/\theta}$$

$$r_3 = C(u_3 | u_1, u_2) = \frac{C(u_1, u_2, u_3)}{C(u_1, u_2)} = \frac{(-2 + u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta})^{-1/\theta}}{(-1 + u_1^{-\theta} + u_2^{-\theta})^{-1/\theta}} = \left(\frac{-2 + u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta}}{-1 + u_1^{-\theta} + u_2^{-\theta}}\right)^{-1/\theta}$$

$$\Rightarrow r_3^{-\theta} = \left(\frac{-2 + u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta}}{-1 + u_1^{-\theta} + u_2^{-\theta}}\right) \Rightarrow r_3^{-\theta} (-1 + u_1^{-\theta} + u_2^{-\theta}) = -2 + u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta}$$

$$\Rightarrow u_3^{-\theta} = r_3^{-\theta} (-1 + u_1^{-\theta} + u_2^{-\theta}) + 2 - u_1^{-\theta} - u_2^{-\theta} = -r_3^{-\theta} + r_3^{-\theta} u_1^{-\theta} + r_3^{-\theta} u_2^{-\theta} + 2 - u_1^{-\theta} - u_2^{-\theta}$$

$$\Rightarrow u_3 = (-r_3^{-\theta} + r_3^{-\theta} u_1^{-\theta} + r_3^{-\theta} u_2^{-\theta} + 2 - u_1^{-\theta} - u_2^{-\theta})^{-1/\theta} = (-r_3^{-\theta} + (r_3 u_1)^{-\theta} + (r_3 u_2)^{-\theta} + 2 - u_1^{-\theta} - u_2^{-\theta})^{-1/\theta}$$

... voor  $n \geq 3$

$$r_n = C(u_n | u_1, \dots, u_{n-1}) = \frac{C(u_1, \dots, u_n)}{C(u_1, \dots, u_{n-1})} = \frac{(1 - n + u_1^{-\theta} + \dots + u_n^{-\theta})^{-1/\theta}}{(2 - n + u_1^{-\theta} + \dots + u_{n-1}^{-\theta})^{-1/\theta}} = \left(\frac{1 - n + u_1^{-\theta} + \dots + u_n^{-\theta}}{2 - n + u_1^{-\theta} + \dots + u_{n-1}^{-\theta}}\right)^{-1/\theta}$$

$$\Rightarrow r_n^{-\theta} = \left(\frac{1 - n + u_1^{-\theta} + \dots + u_n^{-\theta}}{2 - n + u_1^{-\theta} + \dots + u_{n-1}^{-\theta}}\right) \Rightarrow r_n^{-\theta} (2 - n + u_1^{-\theta} + \dots + u_{n-1}^{-\theta}) = 1 - n + u_1^{-\theta} + \dots + u_n^{-\theta}$$

$$\Rightarrow u_n^{-\theta} = r_n^{-\theta} (2 - n + u_1^{-\theta} + \dots + u_{n-1}^{-\theta}) + (n - 1) - u_1^{-\theta} - \dots - u_{n-1}^{-\theta} =$$

$$= r_n^{-\theta} (2 - n) + r_n^{-\theta} u_1^{-\theta} + \dots + r_n^{-\theta} u_{n-1}^{-\theta} + n - 1 - u_1^{-\theta} - \dots - u_{n-1}^{-\theta} = (n - 1) + r_n^{-\theta} (2 - n) + \sum_{i=1}^{n-1} r_n^{-\theta} u_i^{-\theta} - \sum_{i=1}^{n-1} u_i^{-\theta}$$

$$\Rightarrow u_n = \left( (n - 1) + r_n^{-\theta} (2 - n) + \sum_{i=1}^{n-1} r_n^{-\theta} u_i^{-\theta} - \sum_{i=1}^{n-1} u_i^{-\theta} \right)^{-1/\theta}$$

$u_i$ 's gebruiken met inverse transformatie methode:  $F_{X_i}^{-1}(u_i) = x_i$

## Bronnenlijst

---

1

A.M. Cowan, C.D. Cowan

*Default correlation: An empirical investigation of a subprime lender*

Journal of Banking & Finance 28 (2004)

2

KMV Corporation

*Default and Credit Quality Migration*

CREDITMETRICS - Technical Document, April 2, 1997, pp. 70.

<http://www.efalken.com/banking/html/s/matrices.htm>

3

P.K. Trivedi, D.M. Zimmer

*Copula Modeling: An Introduction for Practitioners*

Foundations and Trends in Econometrics Vol 1, No 1 (2005)

4

*Correlation*

<http://en.wikipedia.org/wiki/Correlation>

5

S. Ross

*A First Course in Probability*

Sixth Edition

6

W. Vecchiato

*“Copula Methods in Finance” Special focus to Credit Derivatives*