

Operationeel Risico Management:
Zwaarteverdelingen en data kwaliteit in verliesdatabases
En het gevolg op economisch kapitaal

Master Project BA

Loey Wittink

Juli 2012



vrije Universiteit amsterdam

Operationeel Risico Management: Zwaarteverdelingen en data kwaliteit in verliesdatabases En het gevolg op economisch kapitaal

MSc Scriptie Business Analytics

Loey Wittink

Juli 2012

Begeleider: Prof. Dr. Aad van der Vaart

Tweede lezer: Dr. René Bekker

Vrije Universiteit Amsterdam

Faculteit Exacte Wetenschappen

De Boelelaan 1081a

1081 HV Amsterdam

Begeleider: Rob Voster M.Eng

Tweede begeleider: ir. Satiesh Bajnath

KPMG

IT Advisory – Financial Services

Laan van Langerhuize 1

1186 DS Amstelveen

Voorwoord

De studie Business Analytics – voorheen Business Mathematics & Informatics – wordt afgesloten met het lopen van een afstudeerstage en het schrijven van een scriptie. Deze scriptie is het sluitstuk van een stage die ik heb gevolgd bij KPMG, afdeling IT Advisory Financial Services in Amstelveen. Ik heb onderzoek gedaan naar geschikte zwaarteverdelingen binnen Operationeel Risicomanagement en naar de gevolgen van variërende data kwaliteit binnen dit vakgebied.

Tijdens de stage ben ik begeleid door Aad van der Vaart (Vrije Universiteit), René Bekker (Vrije Universiteit), Rob Voster (KPMG) en Satiesh Bajnath (KPMG). Ik wil alle begeleiders van harte bedanken voor hun hulp bij het geven van feedback, geven van nieuwe inzichten en verdere ondersteuning. In het bijzonder wil ik Aad van der Vaart bedanken voor zijn begrip en geduld bij de lastigheden rondom het vinden van een geschikte dataset. Verder wil ik ook Rob Voster bedanken voor zijn enorme inzet, zijn tijd en moeite bij het vinden van een geschikte dataset. Het was een flinke zoektocht, maar uiteindelijk is het gelukt.

Ook wil ik graag prof. dr. Paul Embrechts (ETH Zürich), dr. Pavel Shevchenko (CSIRO Australia) en Michael McAssey, PhD bedanken voor hun hulp en/of aanbevelingen bij respectievelijk berekeningsmethodiek, verkrijgen van data en het schatten van parameters voor verdelingen zonder analytische uitdrukking.

Verder wil ik de gehele afdeling IT Advisory bedanken voor de leuke en leerzame tijd. Ik heb in deze tijd veel kunnen leren over hoe een organisatie als KPMG in elkaar steekt en hoe het is om bij zo'n type organisatie te werken. Ik ben KPMG erg dankbaar voor de mogelijkheid om hier mijn scriptie te schrijven en op deze manier bij te dragen aan de afdeling.

Loey Wittink

Juli 2012

Samenvatting

Sinds de invoering van Basel I in 1988 moeten banken kapitaal opzij zetten voor het afdekken van krediet- en marktrisico's. Door de jaren heen zijn echter grote verliezen geleden die in geen van beide categorieën vallen. Soms zijn deze verliezen voor banken zelfs catastrofaal gebleken met faillissement of overname als gevolg. Verliezen als gevolg van fraude, terrorisme, natuurrampen of, meer recent, cybercrime kunnen grote vormen aannemen en worden niet afgedekt door kapitaal voor krediet- of marktrisico. Sinds de invoering van Basel II moet ook rekening gehouden worden met operationeel risico, wat wil zeggen het risico op verlies resulterende uit inadequate of gefaalde interne processen, mensen en systemen of van externe gebeurtenissen.

Voor het berekenen van het economisch kapitaal voor operationeel risico kan gebruik gemaakt worden van één van drie benaderingen. De meest eenvoudige benadering is de Basic Indicator Approach (BIA), waarbij een vastgesteld percentage van het gemiddelde inkomen over de afgelopen drie tot vijf jaar als economisch kapitaal wordt gerekend. Dit percentage is op dit moment 15%. Deze methode wordt vaak door kleinere banken gebruikt en door banken waarbij het onvoordelig is om de Standardized Approach (SA) te gebruiken. Bij SA moet de instelling haar handelsactiviteiten opdelen in Business Lines. Vervolgens wordt per Business Line het economisch kapitaal berekend door het gemiddelde inkomen van de afgelopen drie tot vijf jaar per Business Line te bepalen en vervolgens het kapitaal voor alle Business Lines bij elkaar op te tellen. Het vastgestelde percentage verschilt per Business Line en ligt tussen 12% en 18%. Het kan dus zo zijn dat een instelling actiever is in Business Lines met een hoger percentage en hierdoor gebruik van BIA voordeliger is. De laatste methode voor het berekenen van kapitaal voor operationeel risico is de Advanced Measurement Approach (AMA). Waar BIA en SA zogenaamde 'top-down' benaderingen zijn, daar is AMA een 'bottom-up' benadering. Er is voor deze methode geen vastgesteld percentage over inkomen dat als kapitaal gerekend moet worden. Bij de invoering van Basel II werd veel vrijheid gegeven aan banken om zelf met methodieken te komen en werd vanuit BIS de Loss Distribution Approach (LDA) geopperd. Voor de LDA geldt dat een bank minimaal vijf jaar historische data – bij de invoering werd drie jaar ook toegestaan – van operationele verliezen aanwezig moet zijn in een operationele verliesdatabase. Vanuit deze database worden verdelingen geschat voor frequentie van verliesincidenten en zwaarte van verliesincidenten. Hier worden ook scenario's en externe data toegevoegd in combinatie met expert opinies. Uiteindelijk wordt een Monte Carlo simulatie gebruikt om een groot aantal verliesjaren te simuleren vanuit de zwaarte- en frequentieverdelingen, omdat een analytische oplossing vaak lastig is wanneer een verdeling geen gesloten vorm heeft.

In dit onderzoek is gebruik gemaakt van een dataset van [REDACTED] waarin [REDACTED] verliesincidenten zijn opgenomen over de afgelopen [REDACTED] jaar. Data van de laatste tien jaar is gebruikt en alleen Business Lines, Event Types en risico cellen met genoeg data om goede conclusies uit te trekken. Uiteindelijk zijn 2532 datapunten gebruikt, verdeeld over vijf Event Types, zes Business Lines en vijf risico cellen. Aan de hand van deze dataset is onderzocht hoe goed de exponentiële verdeling, de lognormale verdeling, de g-and-h verdeling, Extreme Value Theory, de alpha-stable verdeling en de gegeneraliseerde Pareto verdeling (GPD) passen op de data. Na het schatten van parameters zijn Goodness-of-Fit tests gedaan waaruit is gebleken dat de g-and-h verdeling en Extreme Value Theory aan de hand van methode Peaks Over Thresholds (EVT-POT) beter presteren dan de andere verdelingen. De exponentiële verdeling en de GPD hebben in geen enkel geval een goede fit. De exponentiële verdeling geeft veel te lage schattingen en de GPD geeft veel te hoge schattingen. In dit onderzoek komt ook naar voren dat de alpha-stable in veel gevallen geen goede schatting geeft, mede door het te hoog inschatten van de skewness parameter. Aan de hand van de Block Maxima methode presteert EVT wisselvallig.

Data kwaliteit in databases is moeilijk te kwantificeren. Er kan moeilijk van een database gesteld worden waar deze exact scoort op een schaal van één tot tien. Toch kan een onvolledige of onnauwkeurige database een verkeerd beeld geven van de daadwerkelijke situatie. Daarnaast gaat niet elke verdeling goed om met variërende data kwaliteit. Daarom is in dit onderzoek een scenario analyse uitgevoerd, waarbij gekeken is wat de gevolgen zijn van gebrekkige kwaliteit in de dimensies accuraatheid, compleetheid en punctualiteit. Voor accuraatheid is gekeken naar de effecten van afronden. In de gebruikte dataset, waarbij alle waarden groter zijn dan één miljoen dollar, is afgerond op tienduizend dollar. In de analyse is onderzocht wat het effect is van afronden op honderdduizendtallen en op miljoenen. Hier blijkt dat EVT-POT minder variatie ervaart dan de g-and-h verdeling, waar de nieuwe schattingen gemiddeld bijna anderhalf keer zo hoog liggen als de oorspronkelijke schattingen. Toch geldt ook voor EVT-POT dat er schommelingen plaatsvinden, omdat de nieuwe parameters vaak net anders zijn. Voor compleetheid is gekeken naar de gevolgen van het weghalen van 5% van de metingen. Hier zijn twee scenario's geanalyseerd: één waarbij de weggehaalde metingen willekeurige punten in de dataset zijn en één waarbij de grootste punten zijn weggehaald. Hier blijkt dat bij het weghalen van de grootste metingen vaak 20% of minder van de oorspronkelijke schatting overblijft. Alleen EVT-BMM presteert hier wisselvallig. In het geval dat willekeurige metingen worden weggehaald, zijn de nieuwe schattingen gemiddeld gezien niet veel anders dan de oorspronkelijke schattingen. Toch komt het ook hier voor dat parameters net anders worden geschat naar aanleiding van het weghalen van punten en de schatting anders uitvalt. Als laatste scenario is punctualiteit beschouwd. Bij het registreren van verliesincidenten in de database wordt de waarde van verlies op het moment van verlies verwerkt. Bij de berekening is het echter van belang dat de actuele waarde wordt gebruikt, omdat de verliesincidenten representatief moeten zijn voor de verliezen in het komende jaar. Bij de analyse is bekeken hoezeer de schatting voor kapitaal verandert wanneer niet de actuele waarde wordt gebruikt maar de waarde op het moment van verlies. Zoals verwacht zijn de schattingen lager dan de oorspronkelijke schattingen.

Het is voor banken dus belangrijk om goede controle te hebben over de processen rondom de database zodat goede data kwaliteit zo goed mogelijk gegarandeerd wordt, zodat ook een nauwkeurige schatting van het economisch kapitaal gemaakt kan worden. Ook is het belangrijk om rekening te houden met hoe een verdeling presteert bij variërende data kwaliteit, omdat niet elke verdeling even goed omgaat met variërende data kwaliteit. Dit is zeker het geval voor organisaties die onzeker zijn over hun proces rondom data kwaliteit of geen goede data kwaliteit kunnen garanderen.

Inhoudsopgave

Voorwoord	v
Samenvatting	vi
Inhoudsopgave	viii
Lijst van tabellen.....	x
Lijst van figuren	xi
1 – Inleiding	1
1.1 – Probleemstelling.....	2
1.2 – Over KPMG en KPMG IT Advisory	3
2 – Operationeel Risico	5
2.1 – Definitie en omschrijving Risico.....	5
2.2 – Definitie en omschrijving Operationeel Risico.....	5
2.3 – Regelgeving.....	7
2.3.1 – Loss Databases.....	7
2.3.2 – Basel II.....	8
2.4 – Meten van operationele risico’s	9
2.4.1 – Basic Indicator Approach	10
2.4.2 – Standardized Approach.....	10
2.4.3 – Advanced Measurement Approach	11
3 – Verdelingen voor zwaarte van verliesincidenten.....	13
3.1 – Parametrische en non-parametrische verdelingen	13
3.2 – De α -stable verdeling.....	16
3.2.1 – Definitie en eigenschappen.....	16
3.2.2 – Schatten van parameters	18
3.3 – Extreme Value Theory.....	19
3.3.1 – Het Block Maxima Model	19
3.3.2 – Het Peak-Over-Threshold model.....	21
3.3.2 – Schatten van parameters.....	23
3.4 – De g-and-h verdeling.....	24
3.4.1 – Schatten van parameters.....	26
4 – Goodness of Fit.....	27
4.1 – Grafische tests.....	27
4.1.1 – QQ-plot	27
4.1.2 – Mean Excess Plot.....	28
4.2 – Formele tests.....	29
4.2.1 – Chi-kwadraat tests.....	29
4.2.2 – Empirische verdelingsfunctie (EDF) tests	30

5 – Value-at-Risk.....	32
5.1 – Berekenen van Value-at-Risk	32
5.2 – Monte Carlo Simulatie	34
6 – Data Kwaliteit	36
6.1 – Data Kwaliteit strategie	36
6.2 – Dimensies van Data Kwaliteit	37
7 – Methodologie	39
7.1 – Literatuuronderzoek.....	39
7.2 - Praktijkonderzoek	39
7.2.1 – Schatten van parameters	40
7.2.2 – Bepalen van Goodness of Fit	41
7.2.3 –Onderzoeken invloed data kwaliteit.....	41
8 – Resultaten	43
8.1 - Literatuuronderzoek.....	43
8.2 - Praktijkonderzoek	44
8.2.1 – Verkennende data analyse	44
8.2.2 – Analyse Event Types.....	48
8.2.3 – Analyse Business Lines.....	51
8.2.4 – Analyse Risico Cellen.....	54
8.2.5 – Analyse Data Kwaliteit.....	57
9 – Conclusie en discussie.....	61
9.1 – Conclusie	61
9.1.1 – Zwaarteverdelingen.....	61
9.1.2 – Data Kwaliteit.....	61
9.2 – Verdere opmerkingen en aanbevelingen	62
Referenties	64
Bijlagen	66
Bijlage 1 Classificatie van Event Type	66
Bijlage 2.....	67
Bijlage 3: Specificatie Business Lines en Event Types	69
Bijlage 4: Gebruikte en geschreven functies in R.....	71
Bijlage 4.2 – Monte Carlo Simulatie	74
Bijlage 5: QQ-plots voor GoF.....	78
Bijlage 5.1 – QQ-plots voor GoF: Event Types	78
Bijlage 5.2 – QQ-plots voor GoF: Business Lines	79
Bijlage 5.3 – QQ-plots voor GoF: Risico Cellen	81
Bijlage 6 – Gebruikte packages in R.....	82

Lijst van tabellen

Tabel 1: <i>Operationeel risico en voornaamste factoren</i>	6
Tabel 2: <i>Eigenaardigheden operationeel risico</i>	7
Tabel 3: <i>Overzicht Business Lines en Event Types</i>	8
Tabel 4: <i>β voor elke Business Lines. Bron: ERD 2006 [1]</i>	11
Tabel 5: <i>Data statistieken van Rippel & Teplý in USD</i>	43
Tabel 6: <i>Incidenten per jaar en per regio</i>	45
Tabel 7: <i>p-waarden Kolmogorov-Smirnov test voor onderliggende verdelingen</i>	45
Tabel 8: <i>Frequentietabel voor alle risicocellen</i>	46
Tabel 9: <i>Totale verlies voor alle risicocellen (in miljoen USD)</i>	46
Tabel 10: <i>beschrijvende numerieke statistieken per Business Line</i>	47
Tabel 11: <i>beschrijvende numerieke statistieken per Event Type</i>	47
Tabel 12: <i>te onderzoeken vectoren</i>	48
Tabel 13: <i>parameters van Event Types</i>	48
Tabel 14: <i>p-waarden uit de Anderson Darling test voor Event Types</i>	50
Tabel 15: <i>p-waarden uit de Kolmogorov-Smirnov test voor Event Types</i>	50
Tabel 16: <i>Value-at-Risk schattingen per Event Type</i>	51
Tabel 17: <i>parameters van de verdelingen die horen bij Event Types</i>	51
Tabel 18: <i>p-waarden uit de Anderson Darling test voor Business Lines</i>	53
Tabel 19: <i>p-waarden uit de Kolmogorov-Smirnov test voor Business Lines</i>	53
Tabel 20: <i>Value-at-Risk schattingen per Business Line</i>	53
Tabel 21: <i>parameters voor de verdelingen van risico cellen</i>	54
Tabel 22: <i>p-waarden uit de Anderson Darling test voor risico cellen</i>	55
Tabel 23: <i>p-waarden uit de Kolmogorov-Smirnov test voor risico cellen</i>	56
Tabel 24: <i>Value-at-Risk schattingen per risico cel</i>	56
Tabel 25: <i>verhoudingen originele Value-at-Risk ten opzichte van de nieuwe schatting voor accuraatheid</i>	57
Tabel 26: <i>verhoudingen originele Value-at-Risk ten opzichte van de nieuwe schatting voor compleetheid</i>	58
Tabel 27: <i>verhoudingen originele Value-at-Risk ten opzichte van de nieuwe schatting voor punctualiteit</i>	60

Lijst van figuren

Figuur 1: <i>Organigram van KPMG Nederland</i>	4
Figuur 2: <i>De drie pijlers van Basel II. Bron: NOREA rapport Basel II en Solvency II (2010)</i>	8
Figuur 3: <i>Verdeling vereist kapitaal en economisch kapitaal. Bron: Chalupka & Teplý (2008) [18]</i>	9
Figuur 4: <i>Histogram van de frequentie van verliezen gecategoriseerd per verliesgrootte (links) en de procentuele grootte van verlies per categorie (rechts). Bron: ORX Operational Report June 2010</i>	12
Figuur 5: <i>voorbeeld van een empirische verdelingsfunctie [7]</i>	14
Figuur 6: <i>mixed lognormale verdeling</i>	14
Figuur 7: <i>verdelingsfuncties van de exponentiële verdeling, lognormale verdeling, Weibull verdeling en de Paretoverdeling</i>	15
Figuur 8: <i>vergelijking van zwaarte van staarten voor de verdelingen</i>	15
Figuur 9: <i>alpha-stable verdeling met vaste parameters voor skewness, schaal en locatie en variërende parameter voor shape</i>	18
Figuur 10: <i>alpha-stable verdeling met vaste parameters voor shape, scale en locatie en variërende parameter voor skewness</i>	18
Figuur 11: <i>Extreme waarden volgens de BMM methode</i>	19
Figuur 12: <i>De drie gevallen van GEV: dichtheidsfuncties van de Gumbel familie, Fréchet familie en Reversed Weibull familie</i>	21
Figuur 13: <i>Extreme waarden volgens de POT methode</i>	22
Figuur 14: <i>mean excess Plot voor een GPD verdeling</i>	23
Figuur 15: <i>Hill Plot voor GPD verdeling</i>	24
Figuur 16: <i>g-and-h verdelingsfuncties met $a = 1$ en $b = 1$</i> Figuur 17: <i>g-and-h verdelingsfuncties met $a = 10$</i>	25
Figuur 18: <i>G-and-h verdeling met parameters gebruikt door Dutta & Perry</i>	25
Figuur 19: <i>QQ-plots van operationele verliezen tegen een normale verdeling (links) en een GPD (rechts)</i>	27
Figuur 20: <i>Mean Excess Plot van dataset als in figuur 19</i>	28
Figuur 21: <i>Verdeling vereist kapitaal en economisch kapitaal. Bron: Chalupka & Teplý (2008) [18]</i>	32
Figuur 22: <i>de werking van het actuariële model: samengestelde verliesverdeling (Bron: [7])</i>	33
Figuur 23: <i>histogram van de verdeling van de combinatie van de frequentie en zwaarte verdeling</i>	35
Figuur 24: <i>QQ-plots kwantilen van verschillende geografische gebieden</i>	45
Figuur 25: <i>QQ-plot met de onderzochte verdelingen tegenover de sample verliezen voor EDPM</i>	49
Figuur 26: <i>QQ-plot met de onderzochte verdelingen tegenover de sample verliezen voor Corporate Finance</i>	52
Figuur 27: <i>QQ-plot met de onderzochte verdelingen tegenover de sample verliezen voor Corporate Finance</i>	55

1 - Inleiding

Sinds eind jaren '80 wordt de nodige aandacht besteed aan risicomanagement, zowel bij overheden als financiële instellingen en andere commerciële bedrijven. Door globalisatie en deregulatie groeide de blootstelling aan verschillende risico's. Ook door innovatie, grote vorderingen op technologisch vlak, automatisering en door een groei in omvang van aangeboden producten is de complexiteit van het bancaire systeem enorm toegenomen [7]. Een simpel voorbeeld van de toegenomen gevoeligheid voor risico's door innovatie is het risico op een computervirus of creditcard fraude sinds het grootschalige gebruik van computers en de opkomst van het internet.

Er werd duidelijk dat om financiële en economische stabiliteit te stimuleren, er meer regulatie nodig was voor grote banken, mede om een 'too-big-to-fail' houding te voorkomen in de toekomst. In navolging van het sluiten van de Duitse Herstatt Bank als gevolg van settlement risico, werd vanuit de centrale banken van de G10 de Bank for International Settlements (BIS) gesticht. Hieruit kwam de Basel Committee on Banking Supervision (BCBS) in 1974 voort. De BCBS heeft geen officiële autoriteit aangaande de door hun gestelde regelgeving, maar toch wordt de door hun gestelde regelgeving en aanbevelingen gevolgd door banken, centrale banken en overheden wereldwijd. Deze regelgeving staat bekend als Basel I (voorheen Basel I en later Basel II) en zal later behandeld worden. Hierin staat aangegeven dat financiële instellingen voor krediet-, markt- en operationeel risico economisch kapitaal in kas moeten houden om risico's af te dekken. Mejtrik, Pecena en Teplý (2008) definiëren dit kapitaal als "een buffer tegen toekomstige, onverwachte verliezen toegebracht door krediet, markt en operationele risico's die inherent zijn aan de ondernemingsactiviteiten in het lenen van geld".

Bij het eerste Basel akkoord in 1988 werd voornamelijk aandacht geschonken aan krediet en marktrisico. Risico's die niet in één van beide categorieën hoorden werden "andere risico's" genoemd. Er werd in die tijd nog geen apart label gehangen aan operationele risico's. Toch waren de grootste verliezen in de jaren '90 vaak een gevolg van gebrekkig operationeel risicomanagement (ORM), soms oplopend tot meer dan één miljard dollar. In het hoofdstuk 'Operationeel Risico' zal er dieper in worden gegaan op wat precies operationeel risico is en zullen enkele voorbeelden gegeven worden van grootschalige verliezen als gevolg van gebrekkig ORM de afgelopen twintig jaar.

Nadat duidelijk werd dat de regelgeving aangegeven in Basel I ontoereikend was, presenteerde de BCBS in 2004 de nieuwe versie van het akkoord, Basel II. Al eerder, in 2001 werd een document uitgebracht door BIS waarin werd uitgeweid over operationele risico's en een beschrijving werd gegeven hoe het kapitaalbeslag berekend kan worden aan de hand van ofwel het bruto inkomen van het bedrijf of specifieke Business Lines (BLs) of aan de hand van interne risicometingen. Bij deze laatste methode is niet specifiek vastgesteld welk percentage van het bruto inkomen in voorziening gehouden moet worden als economisch kapitaal. Echter om in aanmerking te komen om deze methode, de Advanced Measurement Approach (AMA) genaamd, moet de informatievoorziening van de bank aan bepaalde voorwaarden voldoen [1]. Een van deze voorwaarden is dat de bank de afgelopen vijf jaar historische data moet hebben verzameld als het gaat om operationele verliezen. Bij de invoering van de AMA is drie jaar historische data toegestaan, mits aan de andere kwalitatieve eisen is voldaan (zie [1] voor een volledige lijst van kwalificaties).

Deze historische data wordt opgeslagen in een verliesdatabase (Loss Database, LDB), die op zijn beurt gebruikt wordt voor het berekenen van het economisch kapitaal. Dit gebeurt doordat vanuit de historische data verdelingen, aangevuld met externe data en scenario's, voor zwaarte en frequentie worden geschat. Van

hieruit worden deze verdelingen samengevoegd en wordt het economisch kapitaal berekend. Omdat de LDB dus als referentiepunt wordt gebruikt voor de berekening, is de kwaliteit van de data van kritiek belang. Over de verdeling van de frequentie is in literatuur [6, 7, 8, 10, 18, 22] overeenstemming dat een Poisson verdeling een bevredigende weergave is van de frequentie. In diezelfde literatuur wordt tegelijkertijd duidelijk dat het vinden van de juiste verdeling voor de zwaarte van verliesincidenten een groter obstakel vormt. In hoofdstuk 3 zal er verder ingegaan worden op zwaarte verdelingen voor verliesincidenten, waarna in het hoofdstuk erna besproken wordt hoe economisch kapitaal berekend kan worden vanuit de verdelingen voor frequentie en zwaarte.

Omdat deze verdelingen geschat worden aan de hand van historische data, aangevuld met externe data en scenario's, is het belangrijk dat de verliesdata daadwerkelijk representatief is voor de werkelijkheid. Incomplete, incorrecte of incompleete data geeft een verkeerd beeld van het kapitaalbeslag. Mede doordat elke bank vrij is geweest een eigen loss database te ontwikkelen, zal het moeilijk zal een generiek data kwaliteitsmodel te ontwikkelen voor LDBs. Echter, er kan wel aangegeven worden waar knelpunten liggen en wat belangrijk is voor banken. Ook is het belangrijk om een indicatie te krijgen van wat de invloed is van de aanwezige data kwaliteit in huidige LDBs op het economisch kapitaal en wellicht het risicobewustzijn van financiële instellingen. In hoofdstuk 6 wordt verder ingegaan op dimensies van data kwaliteit.

In het daaropvolgende hoofdstuk wordt de methodologie die is gebruikt bij het uitvoeren van de berekeningen besproken. Het schatten van de verdelingen en het bepalen van de goodness-of-fit wordt gedaan met statistisch pakket R (www.r-project.org). In dit hoofdstuk zal praktisch ingegaan worden op de genomen stappen bij het bepalen van de meest geschikte verdelingen, het uitvoeren van de Monte Carlo simulatie en het onderzoeken van de invloed van data kwaliteit.

In de laatste hoofdstukken worden de resultaten besproken en worden deze resultaten geëvalueerd. Ook worden aanbevelingen gedaan aangaande omgaan met data kwaliteit en het kiezen van de juiste verdeling en worden conclusies getrokken over de invloed van beide factoren op het economisch kapitaal.

1.1 – Probleemstelling

Zoals in de inleiding is aangegeven, komen er veel verschillende facetten kijken bij het berekenen van het kapitaalbeslag voor operationele risico's. Allereerst moet bepaald worden welke berekeningsmethoden gebruikt gaat worden. Het hoofdstuk 'Methodologie' zal meer inzicht geven in deze berekeningsmethoden. Nadat bepaald is welke methode gebruikt gaat worden moet data worden opgeslagen en later verzameld in een LDB. Vanuit deze database wordt vervolgens aan de hand van een van de berekeningsmethoden het kapitaalbeslag berekend. In literatuur [1, 4, 5, 8, 10] wordt gesteld dat de meest complexe berekeningsmethode, de AMA, ervoor kan zorgen dat het economisch kapitaal een lager percentage van het bruto inkomen is dan bij de andere berekeningsmethoden. Echter, om in aanmerking te komen voor deze methode moet de LDB aan een aantal eisen voldoen (zie [1, 2, 3]). Ook is bij een aantal van de berekeningsmethoden de kwaliteit van de data in de LDB van groot belang, omdat deze in direct verband staat met het berekenen van het economisch kapitaal. Hiermee zijn we aangekomen bij de twee hoofdvragen van deze scriptie:

- *Wat is de best passende verdeling voor zwaarte van verliesincidenten in verliesdatabases en wat is de invloed hiervan op het economisch kapitaal volgens Basel regelgeving?*
- *Wat is de invloed van variërende data kwaliteit op het economisch kapitaal in verliesdatabases?*

Om uiteindelijk antwoord te kunnen geven op deze vraag, zijn hieronder deelvragen gespecificeerd die in dit onderzoek beantwoord gaan worden. Door antwoord te geven op de volgende subvragen zal er een duidelijke situatie geschetst worden van waaruit de hoofdvragen beantwoord kunnen worden.

- Wat houdt operationeel risico precies in?
- Hoe wordt economisch kapitaal berekend?
- Wat zijn veelbesproken verdelingen in recente literatuur?
- Wat houdt data kwaliteit in en welke dimensies heeft data kwaliteit in verliesdatabases?

In de inleiding is kort besproken wat de structuur van deze scriptie is. Naast het beantwoorden van deze vragen zullen ook enkele aanbevelingen gedaan worden aan banken die gebruik maken van AMA.

1.2 – Over KPMG en KPMG IT Advisory

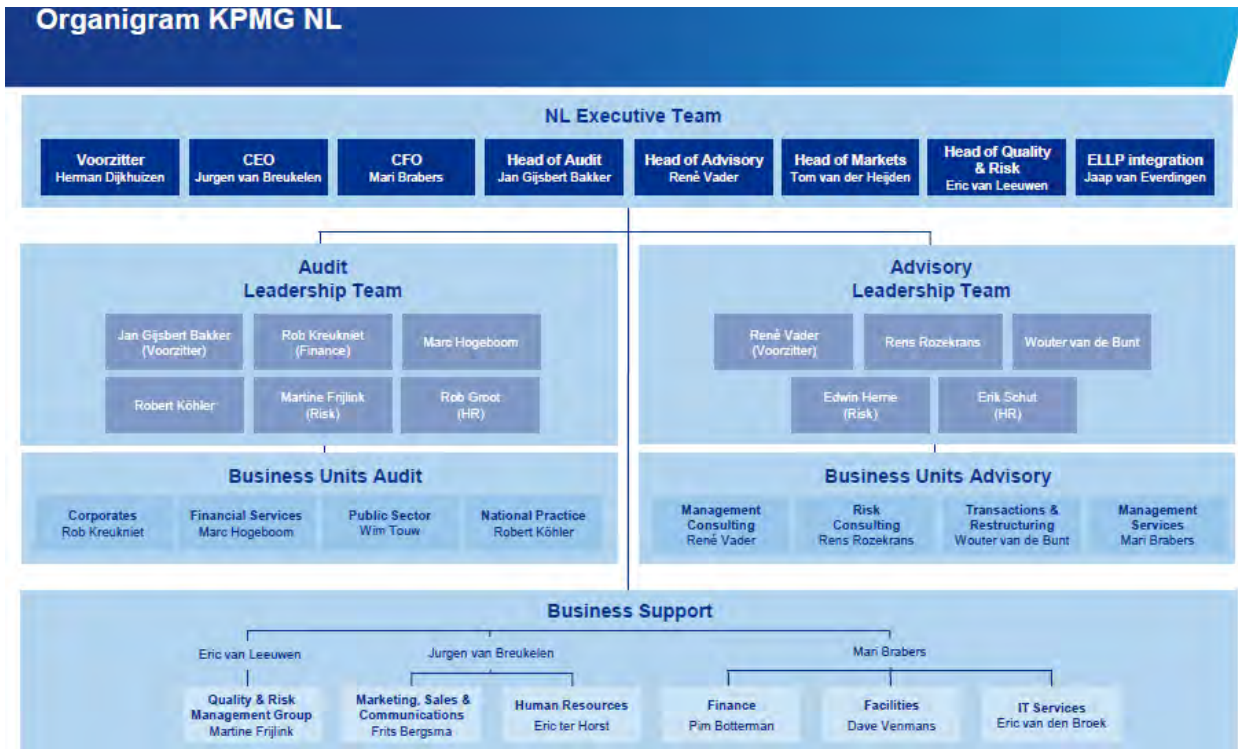
KPMG biedt dienstverlening op het gebied van audit, tax en advies en werkt voor een brede groep opdrachtgevers op bovenstaande gebieden.

KPMG heeft kantoren in meer dan 152 landen met in totaal meer dan 145.000 werknemers. In Europa is KPMG Nederland aangesloten bij KPMG Europe LLP waar de krachten gebundeld wordt met de Britse, Duitse, Luxemburgse, Zwitserse, Spaanse, Russische en Turkse KPMG praktijken met gezamenlijk 30.000 partners en medewerkers. In Nederland waren in het boekjaar 2010/2011 gemiddeld 3.872 FTE's geregistreerd.

KPMG is dus verdeeld in KPMG Audit, KPMG Meijburg & Co (tax) en KPMG Advisory. KPMG Advisory is gebundeld in drie Service Groups: Transactions & Restructuring, Risk Consulting en Management Consulting.

Risk Consulting draait om het beheersen van risico's, de wijze van besturen, het inrichten van duidelijke controlesystemen en heldere communicatie over de organisatie naar de kapitaalmarkten. Dit heeft betrekking op alle processen in een organisatie en de daarbij gebruikte IT-systemen. Dit hangt bovendien nauw samen met het afleggen van verantwoording, het voorkomen van fraude en het naleven van wet- en regelgeving.

IT Advisory – Financial Services is onderdeel van Risk Consulting. Bij deze afdeling ligt de focus voornamelijk op risicobeheer en IT-systemen bij financiële instellingen. Betreffende kennis op het gebied van data kwaliteit is er begeleiding vanuit deze afdeling. Het organogram van KPMG wordt hieronder weergegeven.



Figuur 1: Organigram van KPMG Nederland

2 – Operationeel Risico

Zoals gemeld in de inleiding, werd in eerste instantie door BCBS vooral focus gelegd op kredietrisico en marktrisico. Sinds de invoering van Basel I zijn er duidelijke definities wat betreft risico en sinds Basel II ook wat betreft operationeel risico. Deze definities zullen hieronder besproken worden.

2.1 – Definitie en omschrijving Risico

Risicomanagement komt voort uit de verzekeringsbranche [7]. In productie werd in de jaren '80 total quality management ingevoerd door risico managers. Pas in de jaren '90 werd door een breder publiek het belang van risicomanagement in zowel de financiële als non-financiële sector erkend. Begin jaren '90 werd voor het eerst bij een aantal grote bedrijven naast een CEO, CFO en COO ook een CRO (Chief Risk Officer) aangesteld. Mede door het opkomen van risicomangers bij grote bedrijven (bijvoorbeeld James Lam in 1993 bij GE Capital) en het oprichten van de Global Association of Risk Professionals (GARP) in 1996 werd het nodig dat duidelijk gedefinieerd werd wat risicomanagement precies omhelst. Volgens de International Organization for Standardization (ISO) in Guide 7, uitgave 31000 behelst risicomanagement zich met "*identificatie, inschatting en prioriteiten van risico's*", waarbij risico's gedefinieerd worden als het effect van onzekerheid op doelstellingen, zowel positief als negatief. Vervolgens hoort "*economisch kapitaal toegepast te worden zodat de kans en/of impact op ongewenste gebeurtenissen geminimaliseerd en gecontroleerd kan worden en de totstandkoming van mogelijkheden gemaximaliseerd kan worden*". [15, 16].

Op het moment wordt voornamelijk onderscheid gemaakt tussen kredietrisico, marktrisico en operationeel risico. Kredietrisico is van deze drie het grootst in omvang. De definitie van kredietrisico is het risico van een investeerder op verlies doordat een lener niet de beloofde betalingen maakt [1, 2]. Marktrisico werd net als kredietrisico al omschreven in het eerste Basel akkoord. Marktrisico is, verschillend per bedrijf, groter of kleiner dan operationeel risico, maar altijd kleiner dan kredietrisico. Marktrisico is het risico dat de waarde van een portfolio daalt door een waardeverandering in marktrisico factoren (aandeelprijs, rente, wisselkoers en de prijs van grondstoffen) [2]. Het begrip operationeel risico wordt in detail besproken in de volgende paragraaf.

2.2 – Definitie en omschrijving Operationeel Risico

Sinds de invoering van Basel II is operationeel risico één van de meest besproken onderwerpen bij wetenschappers en professionals in de financiële sector. Dit komt niet alleen door de regelgeving, maar eerder ook door de hoge investeringen in nieuwe informatiesystemen en technologie, toename van nieuwe financiële instrumenten en de groei van elektronisch handelen op de beurs [17]. Een combinatie van beiden zorgt ervoor dat financiële instellingen sterk gemotiveerd zijn om preciezer te zijn aangaande meten en managen van operationeel risico. Echter, voor de invoering van Basel II kwamen vooral banken erachter dat ze nog niet voorbereid genoeg waren voor het omgaan met operationeel risico [7]. BCBS omschrijft operationeel risico als: "*het risico op verlies resulterende uit inadequate of falende interne processen, mensen en systemen of van externe gebeurtenissen*".

Sinds eind 1980 zijn meer dan honderd operationele verliezen geleden in de financiële sector van een omvang van groter dan 100 miljoen dollar [18]. Enkele voorbeelden van zulke verliezen zijn 7.3 miljard dollar verlies in trading bij Société Générale. Société Générale claimt dat frauduleuze transacties de oorzaak waren

van dit verlies. Een van de bekendste voorbeelden is het Enron schandaal in 2001, dat een direct verlies van 1 miljard dollar tot gevolg had en daarmee het failliet gaan van Enron zelf. Indirect was de schade nog een stuk groter, ondermeer door het ten onder gaan van Arthur Anderson, op dat moment één van de Big Five. Andere grote verliezen ten gevolge van gebrekkig operationeel risicomanagement zijn de interne fraude en rogue trading bij Allied Irish Banks met meer dan 700 miljoen dollar aan directe verliezen en het failliet gaan van het district Orange County in Californië, Verenigde Staten na een verlies van 1.7 miljard dollar door een gebrek aan controle en een gebrek aan kennis [7]. Ook de fysieke schade na de terroristische aanval op New York 9 september 2001 leverde een geschatte totale schade op van tussen de 40 en 70 miljard dollar, met bijvoorbeeld voor The Bank NY 140 miljoen schade en de tsunami in Zuidoost Azië in 2004, worden gezien als operationele verliezen [19]. Door de technologische vooruitgang en wijdverspreide toegang tot (vergevoerde) ICT mogelijkheden, is cybercrime in opkomst. Ook dit is een operationeel risico, en in Nederland blijkt uit een onderzoek van KPMG Advisory Nederland dat een flink deel van organisaties in de financiële sector verliezen van jaarlijks meer dan 1.5 miljoen euro lijden door cybercrime [20], maar dit getal kan voor grotere banken hoger willen.

Nu is het niet zo dat operationeel risico alleen grote, externe risico's omvat. Zoals de definitie omvat, het draait om inadequate of falende interne processen, mensen en systemen en ook externe gebeurtenissen. Dit omvat bijvoorbeeld ook dagelijkse (kleine) verliezen als administratieve fouten van het eigen personeel, uitvallen van machines of netwerken of verliezen door bijvoorbeeld ziekte-uitval. De operationele risico's lopen dus uiteen van kleine, vaak voorkomende verliezen (High Frequency Low Impact, HFLI) tot grote, weinig voorkomende verliezende (Low Frequency High Impact, LFHI). De andere twee scenario's, HFHI en LFLI zijn ofwel niet realistisch ofwel niet interessant, omdat ofwel de frequentie van verliezen en grootte van verliezen te groot zijn en hierdoor het bestaan van de organisatie in gevaar komt, waardoor de financiële instelling zich het best kan beraden over positionering, ofwel de verliezen zijn klein in grootte en frequentie en daardoor niet de moeite waard om veel tijd of middelen in te steken. Sironi en Resti (2007) [17] geven de meest voorkomende operationele risico's als volgt weer:

Tabel 1: Operationeel risico en voornaamste factoren

Mensen	Systemen	Processen	Externe Gebeurtenissen
Fraude, collusie en andere criminele activiteiten	IT problemen (hardware, software, hackers of virussen)	Uitvoering, registratie, settlement en documentatie fouten (transactie risico)	Criminele activiteiten als diefstal, terrorisme of vandalisme
Overtreding van interne of externe regelgeving (ongeautoriseerde trading, insider dealing)	Ongesautoriseerde toegang tot informatie en de veiligheid van systemen	Fouten in modellen, methodologie en model risico	Politieke en militaire gebeurtenissen (oorlogen of internationaal gerecht)
Fouten aangaande management tekortkomingen of nalatigheid	Onbeschikbaarheid en twijfelachtige integriteit van de data	Accounting en taxatie fouten, inadequate formalisering van interne procedures	Verandering in de politieke, wetmatige en belasting omgeving (strategisch risico)
Verlies van belangrijk personeel (ziekte, blessures, problemen met behouden van stafleden)	Storingen in telecommunicatie	Regelgeving problemen, schending van bevelschriften	Verandering in juridische omgeving (juridisch risico) en natuurlijke gebeurtenissen (brand, aardbeving, overstroming)
Overtredingen van veiligheid van het systeem	Uitval van hulpprogramma's	Inadequate definitie en verdeling van verantwoordelijkheden	Operationele storingen bij leveranciers of uitbestede operaties

De vraag waarin operationeel risico dan verschilt van kredietrisico en marktrisico zou gesteld kunnen worden. Over het inschatten, meten en omgaan met die risico's is al langer veel over bekend, omdat er simpelweg al langer onderzoek in wordt gedaan. Het grootste verschil is dat er geen keuze wordt gemaakt over de grootte van het operationeel risico, als wel het geval kan zijn bij marktrisico en kredietrisico, maar dat het een natuurlijk gevolg is uit de activiteiten van een financiële instelling [17]. Onderstaande tabel uit [18] geeft een aantal belangrijke verschillen aan:

Tabel 2: *Eigenaardigheden operationeel risico*

Marktrisico en kredietrisico	Operationeel risico
Bewust gekozen in kwantiteit	Onontkoombaar
"Speculatief" risico, impliceert verliezen en opbrengsten	Pure risico's, impliceert alleen verliezen
Consistent met een toenemende relatie tussen risico en verwachte opbrengsten	Niet consistent met een toenemende relatie tussen risico en verwachte opbrengsten
Makkelijk te identificeren en begrijpen	Moeilijk te identificeren en te begrijpen
Relatief makkelijk te meten en identificeren	Moeilijk te meten en te identificeren
Grote beschikbaarheid van hedging instrumenten	Gebrek aan effectieve hedging instrumenten
Relatief makkelijk een prijs aan te koppelen en over te dragen	Moeilijk een prijs aan te koppelen en over te dragen

De verschillen in de bovenstaande tabel geven aan dat de methode waarmee operationele risico's moeten worden bepaald en berekend van een fundamenteel andere aard moeten zijn dan voor marktrisico en kredietrisico, mede door grote onzekerheid en door de breedte van het begrip: operationeel risico omhelst heel erg veel verschillende type incidenten.

2.3 – Regelgeving

2.3.1 – Loss Databases

In Basel I en in Basel III komt het onderwerp operationeel risico respectievelijk niet en niet veel aan bod. In Basel I niet doordat operationele risico's nog niet bestonden, maar de meeste instellingen en risicomangers er geen specifieke aandacht aan besteedden. In Basel III doordat de financiële crisis in 2008 vooral zwakte bij financiële instellingen aan bod bracht op het gebied van kredietrisico en marktrisico. In Basel III wordt voor operationeel risico verwezen naar Basel II [2, 3, 4].

In deze Basel richtlijnen wordt een raamwerk aangegeven voor het identificeren, meten, controleren, rapporteren en omgaan met operationele risico's. BCBS eist procedures voor het uitvoerig meten van operationele risico verliezen en ook voor actief management van operationeel risico. Hiervoor moeten incidenten op het gebied van operationeel risico in een LDB opgenomen worden. Behalve de gebeurtenis en de grootte van verlies geeft de BCBS ook aan dat het incident geassocieerd moet worden. Er wordt aangegeven er onderscheid gemaakt kan worden tussen zeven categorieën (BL, Business Lines) en acht type gebeurtenissen (ET, Event Types) [2], welke zijn aangegeven in tabel 3. Hier moet de opmerking geplaatst worden dat de BLs en ETs die naast elkaar staan niet per definitie bij elkaar horen. Een loss gebeurtenis zou bijvoorbeeld bij de BL Corporate Finance kunnen horen en bij ET Fysieke Schade.

Tabel 3: Overzicht Business Lines en Event Types

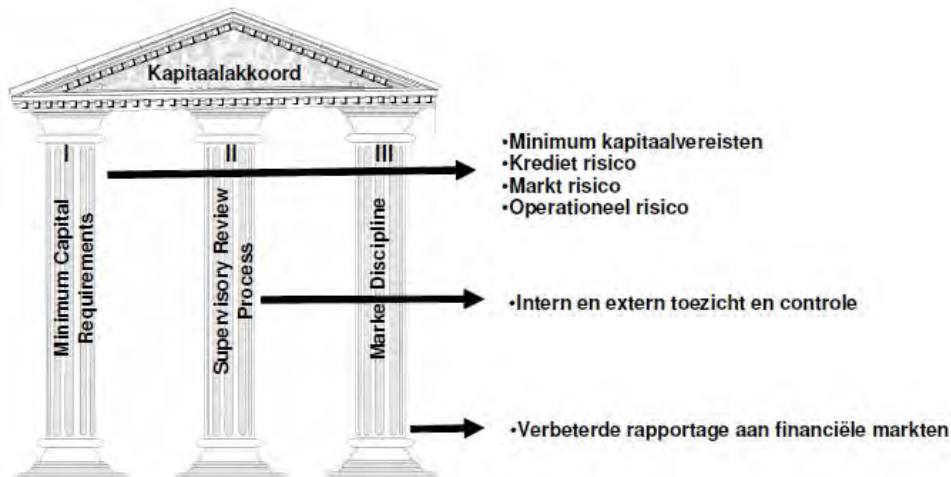
Business Lines	Event Types
<i>Corporate Finance</i>	<i>Interne Fraude</i>
<i>Trading & Sales</i>	<i>Externe Fraude</i>
<i>Retail Banking</i>	<i>Werkgeverszaken en Veiligheid op de werkplaats</i>
<i>Commercial Banking</i>	<i>Klanten, producten en diensten</i>
<i>Payment & Settlement</i>	<i>Fysieke schade</i>
<i>Agency Services</i>	<i>Onderbreking van processen en storingen van het systeem</i>
<i>Asset Management</i>	<i>Uitvoering, bezorging en proces management</i>
<i>Retail Brokerage</i>	

In [21] wordt uitgebreid beschreven wat voor activiteiten bij een bepaalde BL horen (bijvoorbeeld Fusies en Overnames bij Corporate Finance, Privé bankieren bij Retail Banking en Onroerend Goed bij Commercial Banking). In bijlage 1 staat een overzicht waarmee bepaald kan worden bij welke ET een incident ingedeeld zou moeten worden en in bijlage 3 staan meer specificaties over Business Lines en Event Types.

2.3.2 - Basel II

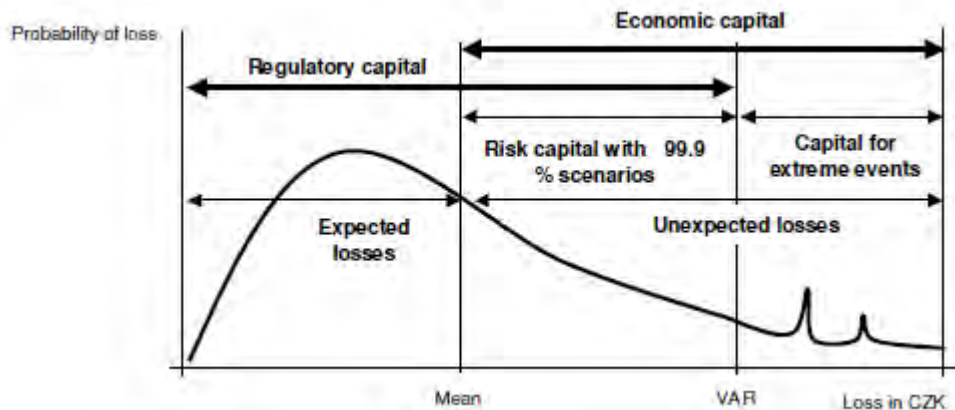
Wat betreft regelgeving, kan er onderscheid gemaakt worden tussen wereldwijde regelgeving, Europese regelgeving en nationale regelgeving. Wat betreft operationeel risico, wordt wereldwijd de regelgeving van de BCBS gehanteerd, die in Europa grotendeels is overgenomen in de Capital Requirements Directive (CRD [1]) en in Nederland in de Wet Financieel Toezicht (WFT). Wat betreft de Basel II wetgeving, stelt het BCBS drie Pijlers op (zie figuur 1, onderstaand):

1. Minimum voor Kapitaalvereisten
2. Proces voor toezichthouding en beoordeling
3. Marktdiscipline



Figuur 2: De drie pijlers van Basel II. Bron: NOREA rapport Basel II en Solvency II (2010)

Pijler I is wat betreft operationeel risico de voornaamste pijler. Hier wordt het minimum aan kapitaalvereisten berekend aan de hand van één van de door BCBS voorgestelde berekeningsmethoden, die in paragraaf 2.3.3 besproken zullen worden. Er is echter een verschil tussen de twee termen “kapitaalvereisten” of “kapitaalbeslag” en de term “economisch kapitaal”. Kapitaalvereisten vallen onder de eerste pijler. Economisch kapitaal valt onder de tweede pijl, dit wordt namelijk bepaald door interne controle. Er is geen unieke of door BCBS of CRD vastgestelde definitie van economisch kapitaal. Mejstřík, Pecená en Teplý (2007) stellen dat “*economisch kapitaal een buffer is tegen toekomstige, onverwachte verliezen toegebracht door krediet, markt en operationele risico’s inherent aan de bedrijvigheid van het geld lenen.*” De kapitaalvereisten zijn nodig om aan de BCBS regelgeving te voldoen, economisch kapitaal wordt gebruikt om de zwaarste verliezen (LFHI) te dekken. Zie figuur 3 voor een grafische weergave van het verschil tussen kapitaalvereisten en economisch kapitaal.



Figuur 3: Verdeling vereist kapitaal en economisch kapitaal. Bron: Chalupka & Teplý (2008) [18]

Evenals in de BCBS richtlijnen, staan in de CRD richtlijnen de drie berekeningsmethoden uitgewerkt met daarbij gemeld de kwantitatieve vereisten en de kwalitatieve vereisten die daarbij horen [1].

2.4 – Meten van operationele risico’s

De BCBS en CRD stellen in hun richtlijnen drie verschillende methoden voor om de kapitaalvereisten te berekenen: de Basic Indicator Approach (BIA), de Standardized Approach (SA) en de Advanced Measurement Approach (AMA). Hierbij zijn BIA en SA top-down benaderingen en de AMA is een bottom-up benadering. Het voornaamste verschil tussen beide is dat in het geval van een top-down benadering geen poging wordt gedaan de daadwerkelijke risico’s te kwantificeren. De verliezen worden aan de hand van een macro-benadering ingeschat. Het voordeel van top-down benaderingen is dat geen middelen ingezet hoeven worden om de risico’s te verklaren en informatie in te zamelen en hierdoor goedkoper is en veel minder tijd kost. Het voordeel van bottom-up benaderingen is dat het hoe en waarom achter operationele risico’s duidelijker wordt, omdat de risico’s wel gekwantificeerd worden.

Voorbeelden van top-down methodes zijn scenario-analyse, risico indicator methoden (zoals SA) en inkomensgebaseerde methoden (zoals BIA en SA) en stress tests. Bij stress tests en scenario-analyse wordt gekeken naar de robuustheid van de gekozen methode om het vereiste kapitaal te berekenen. Bottom-up methoden zijn grofweg onder te verdelen in op proces gebaseerde methoden, actuariële modellen en eigendomsmodellen (proprietary models). De bekendste op proces gebaseerde methode is de scorekaart

methode. Aan actuariële (statistische) methoden is in literatuur [5, 7, 8, 9, 10, 18, 21] de laatste jaren de meeste aandacht besteed. De meest gebruikte methode is de Loss Distribution Approach (LDA), hierover meer in 2.4.3. Proprietary models maken vaak gebruik van een combinatie van top-down en bottom-up methodes. Een bekend voorbeeld is six-sigma.

2.4.1 – Basic Indicator Approach

Voor de BIA geldt dat een percentage α van het netto inkomen, voor het aftrekken van voorzieningen en bedrijfskosten, genomen wordt als kapitaalvereiste [1]. Op dit moment is α door BCBS gesteld op 15%. Dit wordt gedaan als gemiddelde van de afgelopen drie jaar. De kapitaalvereiste K wordt berekend als:

$$K_{BIA} = \alpha \frac{\sum_{t=1}^n NI_t}{n}$$

Hier geldt dat α het door BCBS vastgestelde percentage is, NI is het netto inkomen en n is het aantal jaren, vaak op drie gesteld. Hier wordt het netto inkomen genomen omdat dit is vastgesteld in de Europese CRD. In Basel II van BCBS wordt het bruto inkomen genomen. Nu geldt dus bijvoorbeeld dat als het netto inkomen €1000 is, dat het vereiste kapitaal €150 is.

Er zijn een aantal eisen waar een financiële instelling aan dient te voldoen bij het gebruiken van deze methode. Dit zijn bijvoorbeeld de eis dat de indicator berekend wordt voor het aftrekken van voorzieningen en operationele uitgaven, en de eis dat elementen als gerealiseerde winsten of verliezen van verkopen die niet deel zijn van het handelsportefeuille of de inkomsten uit verzekeringen niet worden meegenomen in de berekening [1].

2.4.2 – Standardized Approach

Voor de SA geldt dat er onderscheid gemaakt wordt tussen de verschillende BLs die waren gedefinieerd in 2.3.1. Opnieuw geldt dat een vastgesteld percentage β_i van het netto inkomen, voor het aftrekken van voorzieningen en bedrijfskosten, wordt genomen als kapitaalvereiste. In dit geval is β_i het percentage dat geldt voor BL i . De kapitaalvereiste voor SA kan als volgt worden berekend:

$$K_{SA} = \frac{\sum_{t=1}^3 \max \left\{ \sum_{k=1}^8 NI_{tk} \cdot \beta_k, 0 \right\}}{3}$$

Hier staat NI opnieuw voor het netto inkomen, k voor de verschillende BLs en t voor het aantal jaren, dat wordt vastgesteld op drie jaar. Zie tabel 4 voor een lijst van β 's voor de verschillende BLs. Nu geldt dus dat het vereiste kapitaal afhankelijk is van het netto inkomen van de verschillende BLs. De instelling moet dus haar activiteiten indelen in de verschillende BLs. Geldt nu dat alle BLs hetzelfde inkomen hebben, dan is het vereiste kapitaal gelijk aan het vereiste kapitaal als de BIA gebruikt zou worden. Geldt dat het inkomen voor de BLs Corporate Finance, Trading & Sales en Payment & Settlement hoger is dan het inkomen voor de BLs Retail Banking, Asset Management en Retail Brokerage, dan is het vereiste kapitaal hoger dan wanneer de BIA

gebruikt zou worden. In het geval dat het inkomen voor Retail Banking, Asset Management en Retail Brokerage hoger is dan het inkomen voor Corporate Finance, Trading & Sales en Payment & Settlement, dan is het vereiste kapitaal lager dan wanneer de BIA als berekeningsmethode gebruikt zou worden. Stel bijvoorbeeld dat het inkomen voor de BLs in onderstaande tabel in het eerste jaar voor de eerste vier BLs €1200 zou zijn en voor de laatste vier BLs €800, in het tweede jaar €1300 en €950 en in het derde jaar €1250 en €1000, dan zou het vereiste kapitaal zijn:

$$K_{SA} = \frac{(0.18 + 0.18 + 0.12 + 0.15) * (1200 + 1300 + 1250) + (0.18 + 0.15 + 0.12 + 0.12) * (800 + 950 + 1000)}{3} = 1310$$

Hier zou bij het gebruik van de BIA het vereist kapitaal €1300 zijn. Het hoeft dus niet zo te zijn dat bij gebruik van SA het vereist kapitaal lager is dan wanneer BIA als berekeningsmethode wordt gebruikt. Ook is het zo dat er extra kosten komen kijken bij het invoeren van SA, omdat de BLs geïdentificeerd moeten worden en er strengere eisen gesteld worden, wel kwalitatief als kwantitatief.

Tabel 4: β voor elke Business Lines. Bron: ERD 2006 [1]

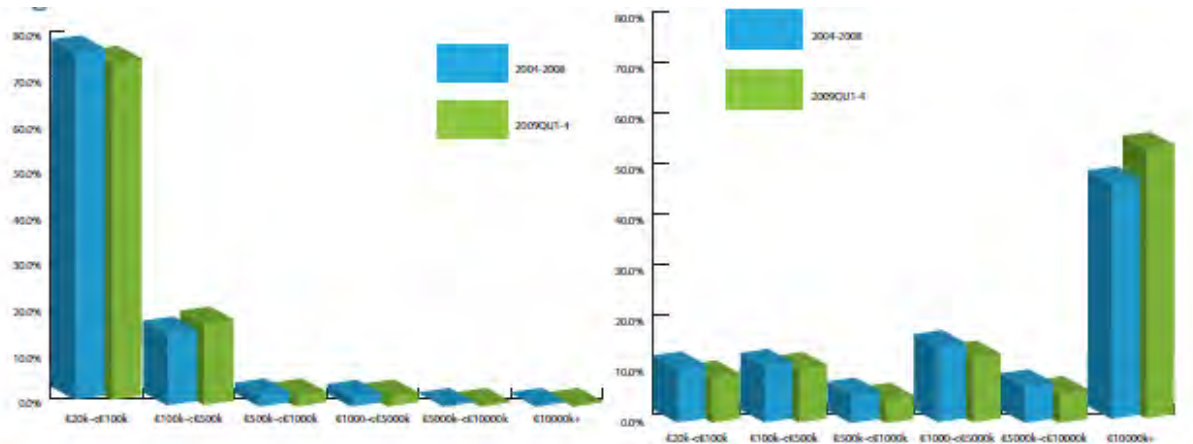
Business Line	β
Corporate Finance	18%
Trading & Sales	18%
Retail Banking	12%
Commercial Banking	15%
Payment & Settlement	18%
Agency Services	15%
Asset Management	12%
Retail Brokerage	12%

Deze eisen staan in detail vermeld in CRD [1] en Basel II [3]. Zo moeten de activiteiten van de financiële instelling ingedeeld worden in BLs beschreven in bovenstaande tabel, zodat operationele verliezen ingedeeld kunnen worden. Ook staat hier in detail beschreven welke principes gevolgd moeten worden bij het indelen van activiteiten in BLs. Verdere eisen die worden gesteld zijn (1) dat de instelling een goed gedocumenteerd beoordeling- en managementsysteem voor operationeel risico, met duidelijke verantwoordelijken voor het systeem, (2) dat het operationeel risico beoordelingssysteem geïntegreerd is in het risicomanagement proces van de instelling en (3) dat de instelling een systeem implementeert voor rapportage aan relevante functies binnen de instelling [1].

2.4.3 – Advanced Measurement Approach

De laatste methode is de meest geavanceerde methode. Waar BIA en SA zogenaamde top-down benaderingen zijn voor het modelleren van operationeel risico, daar is AMA een bottom-up benadering. Dit ligt in het feit dat niet van bovenaf (BCBS, CRD, WFT) percentages worden vastgesteld die worden uitgewerkt in de lagen eronder (de centrale bank, reguliere banken, BLs), maar de bank via een interne meetmethode het operationeel risico vast kan stellen. Voor AMA zijn daarom geen door het BCBS vastgestelde formules beschikbaar. Ook zijn er verschillende vormen van AMA bekend. De meest gangbare methode is echter de LDA, waarbij gebruik wordt gemaakt van een Loss Database. Bij deze methode worden de verliezen

opgeslagen in de database zoals beschreven in 2.3.1. Voor elke risico cel, dat is zijn de verliezen die horen bij een specifieke BL – ET combinatie, wordt vervolgens een verdeling geschat van de frequentie van verliezen en een verdeling van de zwaarte per verlies. Vervolgens worden beide verdelingen geaggregeerd en wordt op basis van een Monte Carlo simulatie de Value-at-Risk berekend. Hierdoor kunnen de bijvoorbeeld 99.9% meest waarschijnlijke risico's worden afgedekt door het economisch kapitaal. Hierdoor kan ook in één keer het vereiste kapitaal en het economisch kapitaal worden berekend, hoewel het volgens het CRD ook nodig is om onderscheid te maken tussen de verwachte en onverwachte verliezen.



Figuur 4: Histogram van de frequentie van verliezen gecategoriseerd per verliesgrootte (links) en de procentuele grootte van verlies per categorie (rechts). Bron: ORX Operational Report June 2010

Voor de verdeling wordt onderscheid gemaakt tussen frequentie en zwaarte van verliesincidenten, zoals in figuur 4 is aangegeven. De linkerfiguur is een histogram met de frequentie per verliesgrootte categorie en de rechterfiguur is het totale verlies per verliesgrootte categorie.

De verdeling voor de frequentie is over het algemeen geen knelpunt, onderzoek heeft al snel uitgewezen dat de Poisson verdeling geschikt is [7] voor het modelleren van frequentie van verliesincidenten. Dit blijkt vooral uit empirische onderzoeken van Cruz (2002) en Chapelle et. al. (2005) [22, 23]. Uit een onderzoek van Chernobai et. al. (2005) blijkt dat ook non-homogene Poisson processen gebruikt kunnen worden voor het modelleren van deze data [24]. Vanwege de simpliciteit van de Poisson verdeling wordt is deze echter gangbaarder dan een non-homogeen Poisson proces. De parameter wordt geschat aan de hand van het aantal verliesincidenten. Dit wordt gedaan voor elke BL-ET combinatie. Vervolgens wordt een trekking gedaan uit de Poisson verdeling met als parameter het gemiddelde aantal incidenten voor elke risk cel. Een idee over frequentie per risico cel wordt gegeven in bijlage 2. Hier zijn het totaal aantal verliesincidenten van bij het ORX aangesloten banken te zien (www.orx.org/orx-members). De Poisson verdeling heeft de onderstaande dichtheidsfunctie.

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Andere frequentieverdelingen die overwogen kunnen worden zijn de Binomiale verdeling, de geometrische verdeling, een cox proces of een empirische verdeling. Een Poisson verdeling wordt echter het meest gehanteerd. Naast de frequentie per risicocel wordt ook de relatieve zwaarte per risk cel wordt weergegeven in bijlage 2. De verdelingen voor zwaarte van verliesincidenten worden in het volgende hoofdstuk besproken.

3 – Verdelingen voor zwaarte van verliesincidenten

De verdeling voor zwaarte van verliesincidenten is moeilijker vast te stellen dan de verdeling voor frequentie. Dit vooral door de vorm van de data. Zoals in figuur 4 valt te zien zitten verreweg de meeste verliezen aan de linkerkant van de figuur, terwijl ook te zien valt dat hoewel de frequentie voor de grotere verliezen klein is, deze de grootste bijdrage hebben aan het totale operationele verlies. Dit is waarom enkele grote banken hele grote drempels gebruiken bij het bepalen van het economisch kapitaal voor operationeel risico. ING Groep gebruikt bijvoorbeeld een drempel van €1 miljoen bij het berekenen van het economisch kapitaal. Of deze hoge drempels ook terecht zijn, wordt door sommige wetenschappers betwijfeld [6, 7].

Er kan grofweg gebruik gemaakt worden van twee verschillende groepen van benaderingen. Aan de ene kant kan gebruik gemaakt worden van een non-parametrische aanpak of aan de andere kant een parametrische aanpak. Hieronder zullen kort parametrische en non-parametrische verdelingen besproken, waarna er stilgestaan wordt bij Extreme Value Theory (EVT), de α -stable verdeling, de g-and-h verdeling en de empirische verdeling.

3.1 – Parametrische en non-parametrische verdelingen

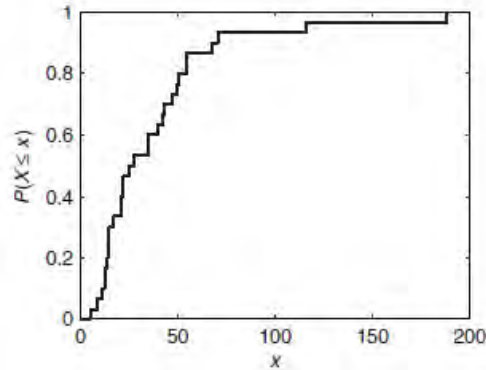
Het doel bij het fitten van verdelingen is het vinden van een model dat de data zo goed mogelijk weergeeft, zodat dit gebruikt kan worden bij het inschatten van toekomstige (onzekere) operationele verliezen. Bij non-parametrische verdelingen worden, zoals de naam doet vermoeden, geen parameters geschat en wordt een verdeling direct bepaald uit de verliesdata. Twee belangrijke aannames moeten gemaakt worden als deze methode gebruikt wordt:

1. Historische verliesdata zijn uitgebreid genoeg, dat wil zeggen de dataset is groot genoeg om een representatieve verdeling op te zetten.
2. Het is even waarschijnlijk dat de historische verliezen in de toekomst opnieuw plaatsvinden, en verliezen van andere groottes, zoals potentiële extreme gebeurtenissen die niet zijn opgenomen in de huidige database) kunnen niet plaatsvinden.

De tweede aanname kan vervangen worden door het gebruik van uitvoerige scenario analyse. Twee redenen dat empirische of non-parametrische verdelingen (nog) niet vaak gebruikt worden, zijn dat LDBs op het moment niet gezien worden als uitgebreid genoeg, dat wil zeggen de data is nog niet zo uitgebreid dat de verdeling alleen op deze data gebaseerd kan worden. De tweede reden is dat het twijfelachtig is om te stellen dat historische verliezen met dezelfde kans op gebeuren in de toekomst zullen plaatsvinden, zeker door technologische ontwikkelingen. De empirische verdelingsfunctie wordt gevonden aan de hand van de volgende formule:

$$P(X \leq x) = \frac{\text{aantal verliezen} \leq x}{\text{totaal aantal verliezen}}$$

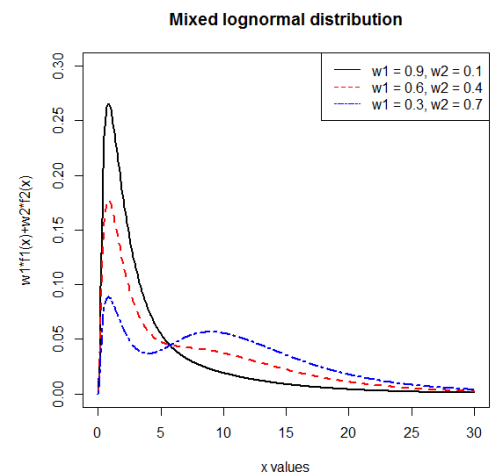
De empirische verdelingsfunctie is trapsgewijs, voor elke observatie x geldt dat er een stap gemaakt wordt. Zie Figuur 5 hieronder voor een voorbeeld.



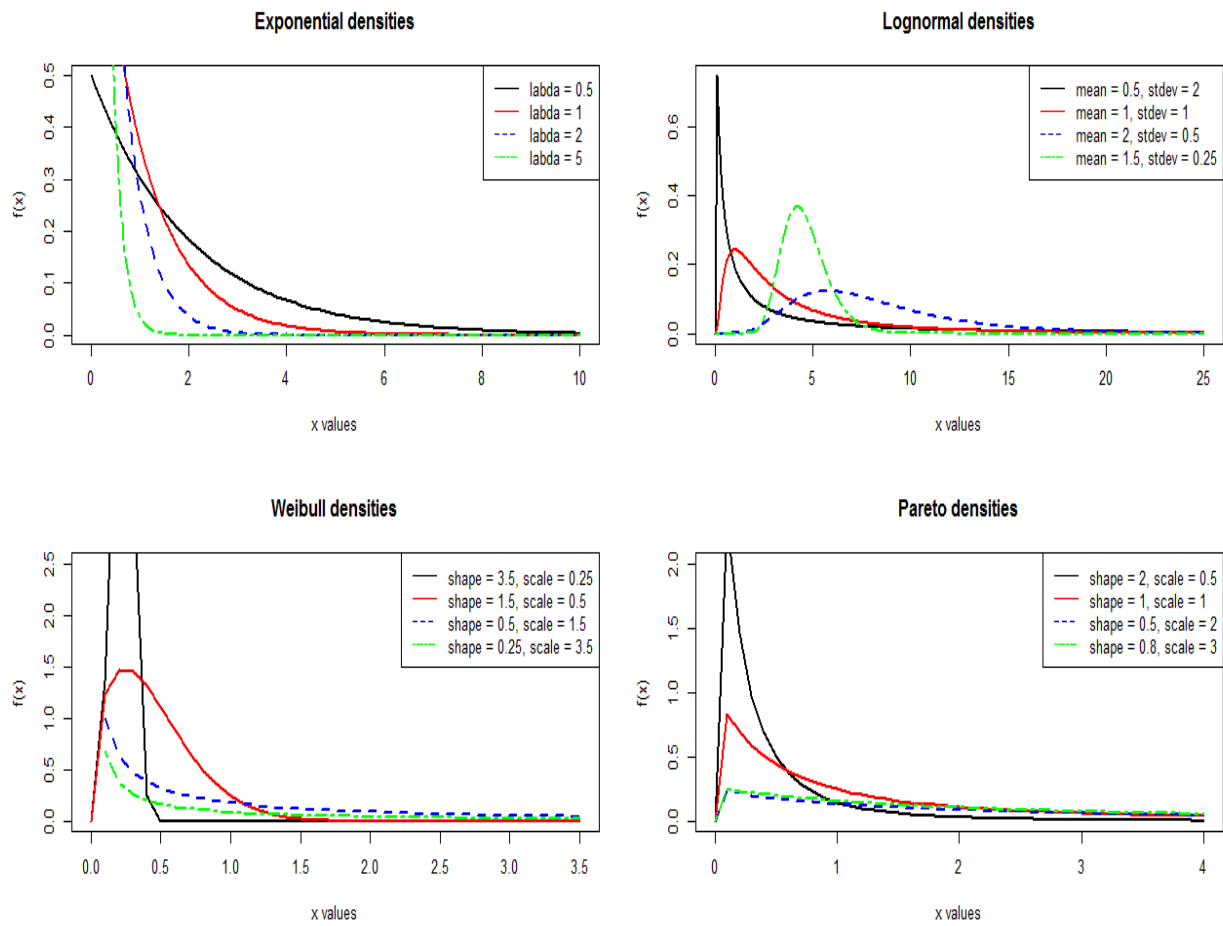
Figuur 5: voorbeeld van een empirische verdelingsfunctie [7]

In het geval van parametrische verdelingen wordt een poging gedaan een curve of een analytische functie te fitten. Het doel is bij parametrische verdelingen is het vinden van een verdeling die zo dicht mogelijk ligt bij de beschikbare data [7]. Onder anderen worden de exponentiële verdeling, de lognormale verdeling, de Weibull verdeling en de Pareto verdeling gebruikt als zwaarte verdeling. Echter, de kritiek op veel van deze verdelingen is dat ze het karakter van operationele verliezen niet goed beschrijven. In de Figuur 7 en Figuur 8 worden deze verdelingen weergegeven. In Figuur 6 zijn deze vier verdelingen uitgezet met verschillende parameters. Hier is duidelijk dat zelfs voor dezelfde verdelingen de vorm verschillend kan zijn bij verschillende parameters. Dit is vooral het geval bij de verdelingen met twee parameters, omdat hier de schaal en de vorm mee bepaald kan worden. In Figuur 8 valt te zien dat de Pareto en de lognormale verdeling zwaardere staarten hebben dan de Weibull en de exponentiële verdeling. Dit kan erop duiden dat deze verdelingen meer geschikt zijn voor het modelleren van operationele verliezen.

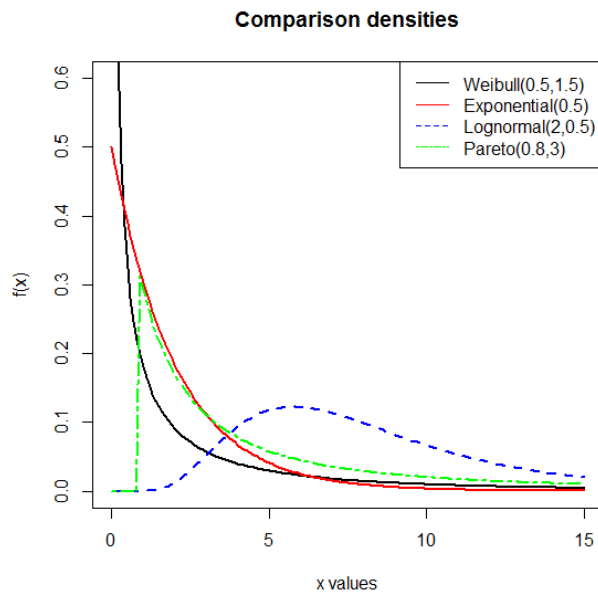
Aan de andere kant hebben deze verdelingen aan de linkerkant een veel minder groot lichaam, hetgeen vaak juist het geval is bij operationele verliezen. Het combineren van twee verdelingen zou een uitkomst kunnen bieden. In Figuur 6 hiernaast is een aantal mixed verdelingen te zien van de lognormale verdeling met aan de ene kant parameters 0.8 en 1 en aan de andere kant parameters 2.5 en 0.5. De mixed verdeling wordt gevonden door de vergelijking:



Figuur 6: mixed lognormale verdeling



Figuur 7: verdelingsfuncties van de exponentiële verdeling, lognormale verdeling, Weibull verdeling en de Paretoverdeling



Figuur 8: vergelijking van zwaarte van staarten voor de verdelingen

$$f_{mixed}(x) = \sum_{i=1}^n w_i \cdot f_i(x)$$

Er valt in Figuur 6 te zien dat bijvoorbeeld bij de gewichtsverdeling $w_1 = 0.3$, $w_2 = 0.7$ er wel een vorm ontstaat met een groot lichaam aan de linkerkant en stevige staart aan de rechterkant. Een dergelijke constructie zou, bij het kiezen van de juiste verdelingen, eventueel een goede kandidaat zijn voor het modelleren van de zwaarte van verliesincidenten. In de richtlijnen van BIS aangaande AMA [28, par. 202], staat echter wel aangegeven dat wanneer twee verdelingen gebruikt worden, bijvoorbeeld één voor het lichaam en een ander voor de staart, dat de keuze voor gewichten of drempel aangegeven moet worden aan de hand van statistische onderbouwing. Technieken als de Hill plot en de mean excess plot (besproken in hoofdstuk 3.3.2 respectievelijk 4.1.2) kunnen bijvoorbeeld gebruikt worden om een drempel te bepalen.

3.2 – De α -stable verdeling

In de vorige paragraaf zijn verdelingen met één of twee parameters besproken. In deze paragraaf wordt ingegaan op de alpha-stable verdeling, een verdeling met vier parameters en mede daarom kan deze verdeling een zware staart bezitten. De vier parameters (voor shape, skewness, scale en locatie) zorgen ervoor dat de alpha-stable verdeling flexibel is. Het grootste obstakel van deze verdeling is dat voor de dichtheidsfunctie geen gesloten vorm bestaat en daarom de oplossing vaak numeriek benaderd moet worden. De enige uitzonderingen hiervoor zijn drie bijzondere gevallen, namelijk bij bepaalde waarden voor de shape of skewness parameters, neemt de dichtheidsfunctie een Gaussian, Cauchy of Lévy vorm aan.

3.2.1 – Definitie en eigenschappen

Neem aan dat stochasten X_1, X_2, \dots, X_n iid (gelijk en onafhankelijk) verdeeld zijn, allen kopieën van X . De stochast X is dan alpha-stable verdeeld als geldt, met constante $C_n > 0$ en $D_n \in \mathbb{R}$, zodat geldt dat:

$$\sum_{i=1}^n X_n \stackrel{d}{=} C_n X + D_n$$

Hier staat de notatie $\stackrel{d}{=}$ voor gelijkheid in verdeling, zoals wordt gebruikt in [7]. De constante $C_n = n^{1/\alpha}$, zorgt ervoor dat voldaan wordt aan de stabiliteitseigenschap, die ervoor zorgt dat de vorm van de verdeling van een stochast behouden blijft bij lineaire transformaties. Dit is één van de vier belangrijke eigenschappen van de alpha-stable verdeling. Deze eigenschap zorgt er ook voor dat de Centrale Limiet Stelling hier niet van toepassing is: een grote hoeveelheid iid stochasten convergeren niet naar een normale verdeling, maar naar een alpha-stable stochast. De andere drie belangrijke eigenschappen van de alpha-stable verdeling:

1. De “power tail decay” eigenschap houdt in dat de staart van de dichtheidsfunctie afloopt als een machtfunctie in plaats van als een exponentiële functie. Hierdoor kan de verdeling extreme gebeurtenissen in de staart modelleren:

$$P(|X| > x) \propto \text{const} \cdot x^{-\alpha}, \quad x \rightarrow \infty$$

2. Ruwe momenten voldoen aan:

$$\begin{aligned} \mathbb{E}|X|^p < \infty, & \quad 0 < p < \alpha \\ \mathbb{E}|X|^p = \infty & \quad p \geq \alpha \end{aligned}$$

3. Door eigenschap 2 geldt ook dat het gemiddelde alleen eindig is als $\alpha > 1$.

Zoals eerder gesteld, heeft de dichtheid over het algemeen geen gesloten vorm. De verdeling wordt uitgedrukt door zijn karakteristieke functie:

$$\mathbb{E} \left| e^{itX} \right| = \begin{cases} \exp(-|\sigma|^\alpha (1 - i\beta(\text{sign } t) \tan \frac{\pi\alpha}{2}) + i\mu t), & \alpha \neq 1 \\ \exp(-\sigma|t|(1 + i\beta \frac{2}{\pi} (\text{sign } t) \ln|t|) + i\mu t), & \alpha = 1 \end{cases}$$

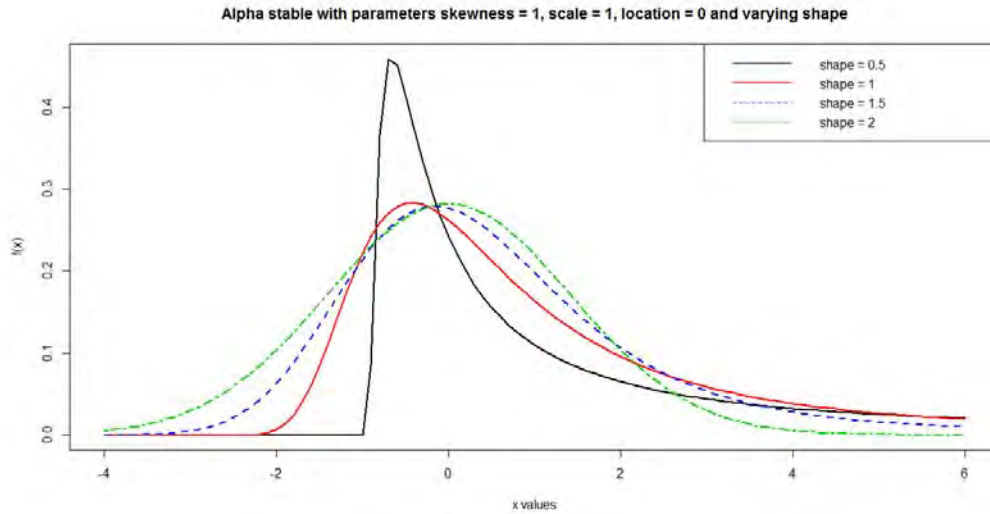
De vier parameters die deze verdeling karakteriseren zijn:

- α : de stabiliteitsindex, of de shape parameter. $\alpha \in (0, 2)$
- β : de parameter voor skewness, $\beta \in [-1, 1]$
- σ : de parameter voor schaal, $\sigma \in (0, \infty)$
- μ : de parameter voor locatie, $\mu \in (-\infty, \infty)$

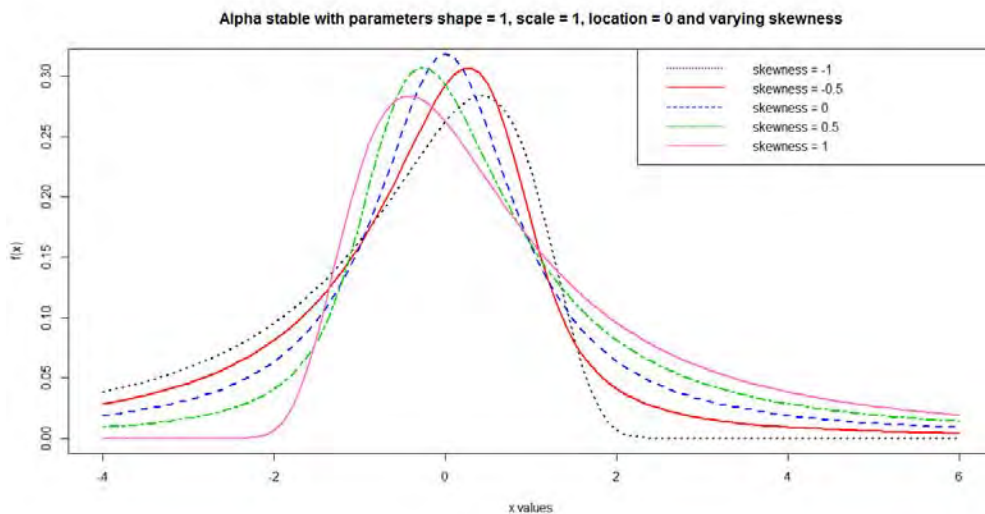
De uitzonderingen waarbij geldt dat de dichtheid wel een gesloten vorm heeft, gelden bij:

- $\alpha = 2$, Gaussian: $f(x) = \frac{1}{2\sigma\sqrt{\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $\alpha = 1$ en $\beta = 0$, Cauchy: $f(x) = \frac{\sigma}{\pi((x-\mu)^2 + \sigma^2)}$
- $\alpha = 1/2$ en $\beta = \pm 1$, Lévy: $f(x) = \frac{\sqrt{\sigma}}{\sqrt{2\pi}(x-\mu)^{3/2}} e^{-\frac{\sigma}{2(x-\mu)}}$

Voor de Gaussian en Cauchy gevallen geldt $-\infty < x < \infty$ en voor Lévy geldt $\mu < x < \infty$. Zoals eerder gezegd, zorgt het feit dat de verdeling vier parameters heeft voor flexibiliteit en maakt de verdeling geschikt voor het modelleren van asymmetrische data met zware staarten. In Figuur 9 en Figuur 10 zijn alpha-stable verdelingen weergegeven voor verschillende waarden van α en β . In Figuur 9 valt te zien dat een lage waarde van α zorgt voor een zware staart, terwijl in figuur 10 te zien valt dat β richting 1 een zwaardere rechterstaart geeft en een β richting -1 een zwaardere linkerstaart.



Figuur 9: *alpha-stable verdeling met vaste parameters voor skewness, schaal en locatie en variërende parameter voor shape*



Figuur 10: *alpha-stable verdeling met vaste parameters voor shape, scale en locatie en variërende parameter voor skewness*

3.2.2 - Schatten van parameters

Omdat er in het algemeen geen gesloten vorm voor de dichtheid van een alpha-stable verdeling van een stochast bestaat, kan voor het schatten van de parameters voor de verdeling geen analytische vorm van MLE (Maximum Likelihood Estimation) gebruikt worden. Er zijn twee methodes die over het algemeen worden gebruikt om de vier parameters te schatten.

De eerste van de twee is de “sample characteristic function approach”. Bij deze methode worden de observaties gebruikt om de sample karakteristieke functie te evalueren en de parameters zo te bepalen dat de afstand tussen de observaties en de theoretische functie zo klein mogelijk is. In de eerste stap worden observaties x_1, x_2, \dots, x_n gebruikt om de sample karakteristieke functie te berekenen:

$$\hat{\varphi}(t) = \frac{1}{n} \sum_{k=1}^n e^{itx_k}$$

Vervolgens wordt aan de hand van wiskundige software (in dit geval R) de theoretische karakteristieke functie gefit naar de sample karakteristieke functie. Dit wordt gedaan door schatters $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$, $\hat{\mu}$ te vinden zodat de afstand tussen de sample en theoretische karakteristieke functie zo klein mogelijk is.

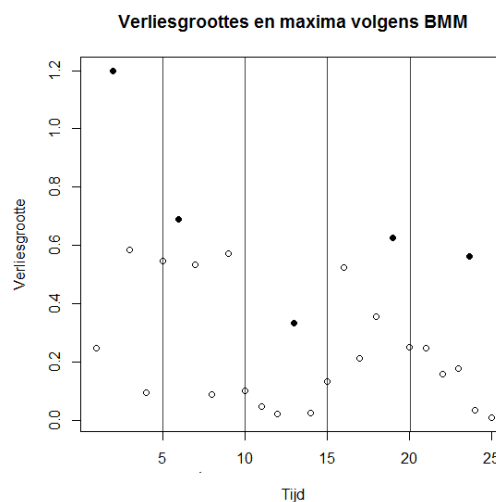
De andere benadering voor het schatten van de parameters is een numerieke benadering. Hier wordt in de eerste stap de dichtheidsfunctie berekend aan de hand van de karakteristieke functie met het FFT (Fast Fourier Transform) algoritme. Vervolgens kan een MLE voor de onbekende parameters uitgevoerd worden aan de hand van numerieke optimalisatie en worden schatters $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$, $\hat{\mu}$ gekozen die de likelihood functie maximaliseren.

3.3 – Extreme Value Theory

Extreme Value Theory (EVT) is een veld waarin specifieke aandacht geschonken wordt aan LFHI verliezen. Hier zullen twee EVT modellen besproken worden: het Block Maxima Model (BMM) en het Peak-Over-Threshold (POT) model. Het grootste verschil tussen EVT en de voorgaande modellen is dat er een duidelijk onderscheid gemaakt wordt tussen het 'lichaam' van de dichtheidsfunctie voor zwaarte van verliesincidenten en de staart van verliesincidenten. EVT heeft wel raakvlakken met mixed verdelingen, waar principe ook terugkomt, maar minder expliciet dan bij EVT, omdat daar geen harde scheiding gemaakt wordt zoals bij EVT.

3.3.1 – Het Block Maxima Model

Bij deze methode wordt de dataset onderverdeeld in onafhankelijke blokken van gelijke grootte en wordt er gekeken naar het maximum uit elk blok. Vervolgens wordt gekeken naar de verdeling van alleen deze specifieke datapunten.



Figuur 11: *Extreme waarden volgens de BMM methode*

Voor extreme verliesobservaties x , is de limiting verdeling de GEV (Generalized Extreme Value) verdeling:

$$f(x) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{(-1/\xi)-1} \cdot \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad 1 + \xi \left(\frac{x - \mu}{\sigma} \right) > 0$$

Hier geldt dat:

- x staat voor de extreme observaties van elk blok
- ξ is de shape parameter
- σ is de parameter voor schaal, $\sigma > 0$.
- μ is de parameter voor locatie, $-\infty < \mu < \infty$

De GEV verdeling kan opgedeeld worden in drie specifieke gevallen met verschillende limieten. De verschillen tussen deze drie gevallen worden hieronder zowel tekstueel als grafisch (in figuur 12) aangeduid:

- GEV type I, Gumbel familie, geldt wanneer $\xi = 0$. In dit geval is er geen limiet aan de linker of rechterkant en kunnen zowel links als rechts extreme waarden optreden. De cumulatieve verdelingsfunctie wordt gegeven als:

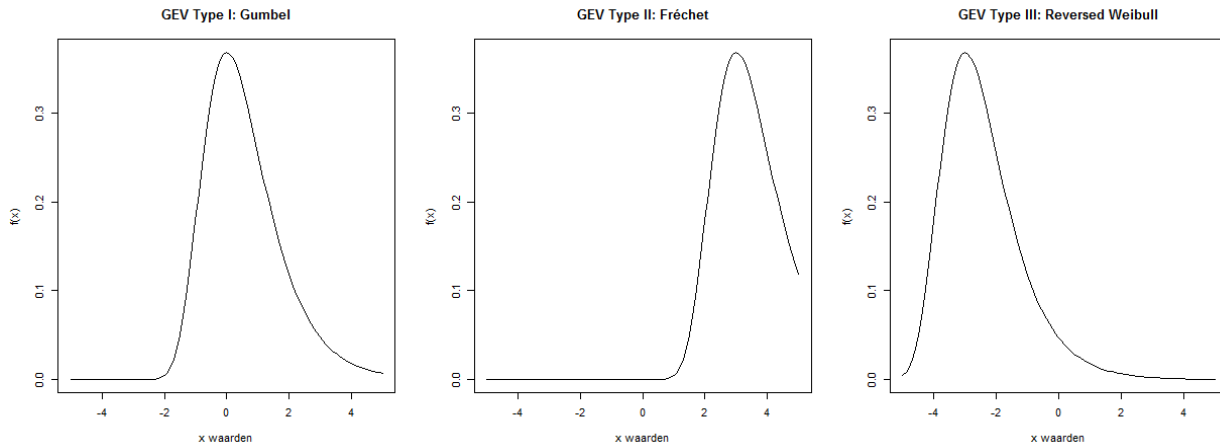
$$F(x; \mu; \sigma; \xi) = \exp \left[e^{-\frac{x-\mu}{\sigma}} \right]$$

- Type II, Fréchet familie, geldt wanneer $\xi > 0$. In dit geval is er linker limiet en liggen de extreme waarden in de rechterstaart. De cumulatieve verdelingsfunctie wordt gegeven als:

$$F(x; \mu; \sigma; \xi) = \begin{cases} \exp \left[- \left(\frac{x - \mu}{\sigma} \right)^{-1/\xi} \right], & x > \mu \\ 0 & x \leq \mu \end{cases}$$

- Type III, de Reversed Weibull familie, geldt wanneer $\xi < 0$. In dit geval is er een rechter limiet en liggen de extreme waarden in de linkerstaart. De cumulatieve verdelingsfunctie wordt nu gegeven als:

$$F(x; \mu; \sigma; \xi) = \begin{cases} \exp \left[- \left(\frac{\mu - x}{\sigma} \right)^{1/\xi} \right], & x < \mu \\ 1 & x \geq \mu \end{cases}$$



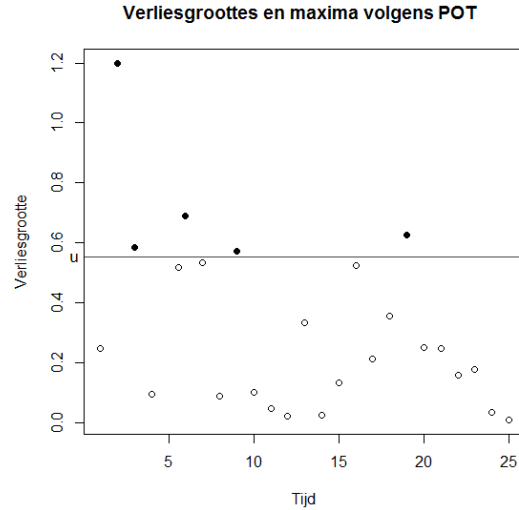
Figuur 12: De drie gevallen van GEV: dichtheidsfuncties van de Gumbel familie, Fréchet familie en Reversed Weibull familie

In de context van operationele verliesincidenten wordt type II, de Fréchet familie gebruikt voor het modelleren. Door de cumulatieve verdelingsfunctie te differentiëren valt af te leiden dat de dichtheidsfunctie voor type II GEV is:

$$f(x; \mu; \sigma; \xi) = \begin{cases} \frac{1}{\sigma \xi} \left(\frac{x - \mu}{\sigma} \right)^{(-1/\xi)-1} \cdot \exp \left[- \left(\frac{x - \mu}{\sigma} \right)^{-1/\xi} \right] & x > \mu \\ 0 & x \leq \mu \end{cases}$$

3.3.2 - Het Peak-Over-Threshold model

In dit EVT model wordt de dataset niet opgedeeld in blokken verdeeld over tijd, maar wordt een drempel u gekozen, waarboven een aantal extreme waarden ligt (zie figuur 13 onderstaand). De verdeling van deze extreme waarden x wordt de conditionele excessen verdelingsfunctie genoemd [7]. Tenzij de data een sterk seizoenscomponent heeft, wordt de POT methode geprefereerd over de BMM methode [8].



Figuur 13: *Extreme waarden volgens de POT methode*

Wanneer de drempel hoog genoeg is, kan de conditionele excessen verdelingsfunctie beschreven worden door de gegeneraliseerde Pareto verdeling (GPD). De cumulatieve verdelingsfunctie en de dichtheidsfunctie van de GPD zijn als volgt:

$$F(x; \mu; \sigma; \xi) = \begin{cases} 1 - \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp\left[-\left(\frac{x - \mu}{\sigma}\right)\right] & \xi = 0 \end{cases}$$

$$f(x; \mu; \sigma; \xi) = \frac{1}{\sigma} \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{(-1/\xi)-1}$$

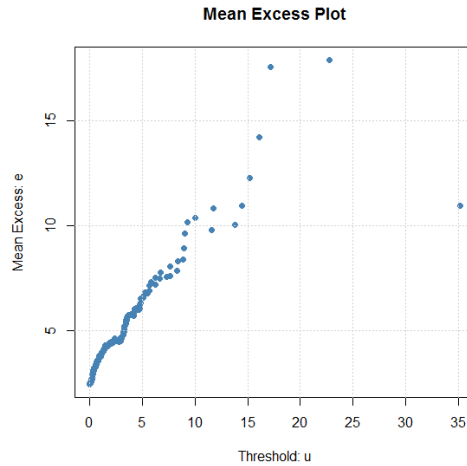
Hierbij geldt ook dat $x \geq \mu$, if $\xi \geq 0$ and $\mu \leq x \leq \mu - \frac{\sigma}{\xi}$. De parameters hebben dezelfde betekenis als de parameters voor de GEV. Ook hier geldt dat $-\infty < \mu < \infty$ en dat $\sigma > 0$. Bij de GPD geldt net als de GEV dat - afhankelijk van de waarde van ξ - drie speciale gevallen te onderscheiden zijn:

- In het geval van $\xi = 0$ vereenvoudigt de GPD naar een exponentiële verdeling:

$$f(x) = \frac{1}{\sigma} \exp\left[-\left(\frac{x - \mu}{\sigma}\right)\right]$$

- In het geval van $\xi > 0$ verkrijgen we zware staarten, ook wel een gereparametriseerde Pareto verdeling. De verdelingsfunctie die verkregen wordt staat bovenaan deze pagina.
- In het geval van $\xi < 0$ vereenvoudigt de GPD naar een uniforme verdeling.

Een belangrijk aspect bij het gebruik van de POT methode is het kiezen van een geschikte drempel: de drempel moet op zo'n manier gekozen worden dat de extreme gebeurtenissen kunnen worden gemodelleerd, maar ook niet zo hoog dat er een enorme variantie in de data aanwezig is, mede door een tekort aan datapunten. De meest gebruikte methode is een visuele methode, namelijk de mean excess plot (zie figuur 14). De mean excess plot wordt verder besproken in hoofdstuk 4.



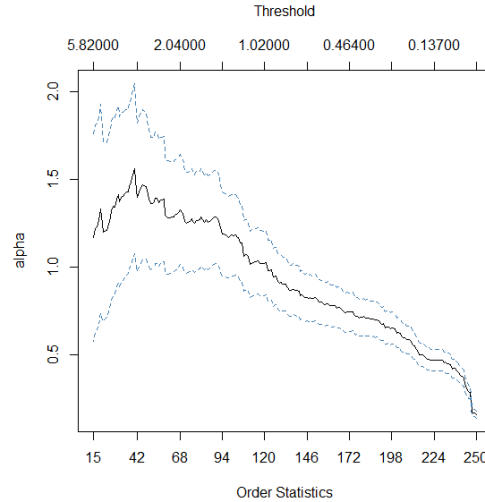
Figuur 14: *mean excess plot voor een GPD verdeling*

3.3.2 – Schatten van parameters

In het geval van EVT kan het schatten van parameters een ingewikkeld proces zijn. In eerste instantie omdat het voorkomt dat de hoeveelheid data in de rechterstaart beperkt is. Mede hierdoor wordt het gebruik van scenario's of externe data aanbevolen door BCBS en CRD [1, 2, 4]. De belangrijkste parameter om te schatten is de shape parameter ξ , omdat deze – zoals in de vorige paragrafen beschreven – de vorm bepaalt. Een belangrijke non-parametrische schatter die gebruikt kan worden is de Hill schatter. Deze schatter voor ξ ziet er als volgt uit:

$$\hat{\xi}^H = \frac{1}{k} \sum_{i=1}^k \ln X_i - \ln X_{(k)}$$

Onderstaand is een voorbeeld gegeven van een Hill plot, waarbij de schattingen voor ξ (alpha in de figuur) worden uitgezet tegen $X_{(k)}$ en de verschillende drempels u .



Figuur 15: Hill Plot voor GPD verdeling

3.4 – De g-and-h verdeling

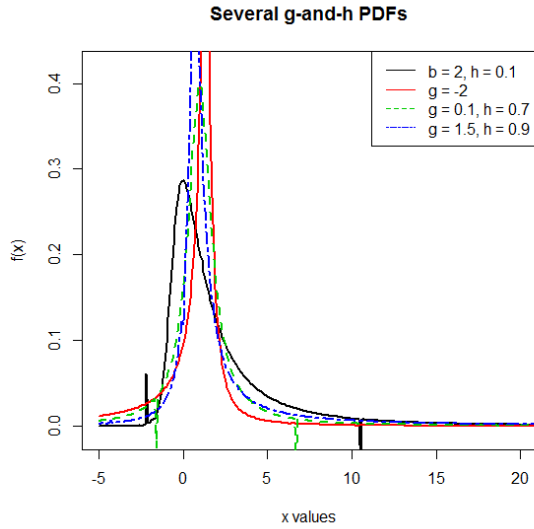
De g-and-h verdeling is een transformatie van een standaard normale verdeling. Deze verdeling heeft het voordeel dat, evenals de alpha-stable verdeling, er vier parameters zijn. Als Z een normaal verdeelde stochast is, dan is X g-and-h verdeeld als X voldoet aan:

$$X = a + b \frac{\exp(gZ - 1)}{g} \exp\left[\frac{hZ^2}{2}\right]$$

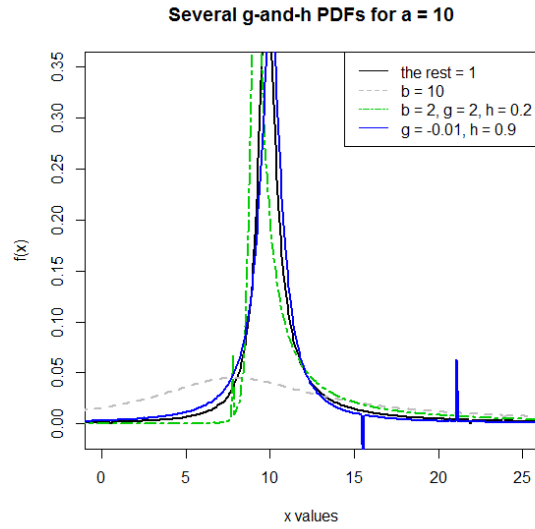
In het geval dat $h = 0$, dan wordt dit een g-verdeling:

$$X = a + b \frac{\exp[gZ - 1]}{g}$$

Naast het feit dat modelleren met vier parameters al meer flexibiliteit geeft dan verreweg de meeste modellen, bestaat hier ook de mogelijkheid om in plaats van g en h als constanten te nemen, g en h te modelleren als polynomen met hogere orden van Z^2 . In [8] wordt bijvoorbeeld gekozen om g constant te nemen en h als polynoom met vier termen, terwijl in [25] wordt gekozen om g en h constant te houden.

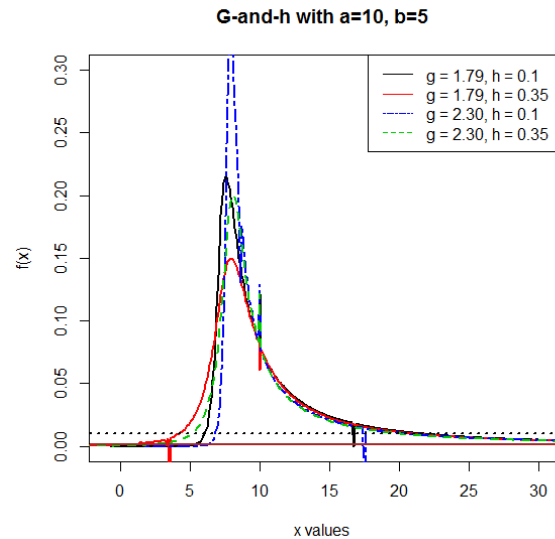


Figuur 16: *g-and-h* verdelingsfuncties met $a = 1$ en $b = 1$



Figuur 17: *g-and-h* verdelingsfuncties met $a = 10$

Bovenstaand, in figuur 16 en figuur 17 zijn een aantal verdelingsfuncties van de *g-and-h* verdeling te zien. In beide figuren zijn voor $g = 1$ een aantal verticale lijnen te zien. Bij inzoomen blijken dit geen verticale lijnen te zijn, maar uitschieters in de zin van pieken. Wat uit de beide figuren in ieder geval blijkt is dat wanneer $g < 0$, de linkerstaart dikker is en voor $g > 0$ is de rechterstaart dikker. Voor het modelleren van operationeel risico zullen dus waarden voor $g > 0$ worden gebruikt. In één specifiek geval bij Dutta & Perry [8] wordt bijvoorbeeld $g \in [1.79; 2.30]$ en $h \in [0.10; 0.35]$ gebruikt. In figuur 18 valt te zien hoe deze verdelingen er ongeveer uitzien bij waarden $a = 10$ en $b = 5$. In het geval dat b hoger wordt, zal het lichaam meer over de x -as verspreid worden dan over de y -as, zoals bij een lage b . De zwarte stippellijn staat voor 0.01, dus in onderstaande gevallen zal vaak 99.9% van de waarden tussen de -1 en 20 liggen. Wanneer gekeken wordt naar de bruine lijn, die staat voor 0.001, dan liggen alles tussen -1 en 80 boven de bruine lijn.



Figuur 18: *G-and-h* verdeling met parameters gebruikt door Dutta & Perry

3.4.1 – Schatten van parameters

Het schatten van de vier parameters voor de g-and-h verdeling ligt iets anders dan het schatten van de vier parameters van bijvoorbeeld de alpha-stable verdeling. De parameter a is in dit geval het makkelijkst te schatten: \hat{a} is namelijk de mediaan $X_{0.5}$ van de dataset. Voor parameter g geldt:

$$g_p = -\left(\frac{1}{Z_p}\right) \cdot \ln\left(\frac{X_{1-p} - X_{0.5}}{X_{0.5} - X_p}\right)$$

Omdat uit deze vergelijking alleen blijkt dat er veel verschillende mogelijke waarden voor g zijn, kan \hat{g} gekozen worden als de mediaan van g_p , wordt gesteld in [26]. Vervolgens geldt dat wanneer bovenstaande vergelijkingen voor de g-and-h verdeling worden gecombineerd, gesteld kan worden dat:

$$\ln \frac{g(X_p - X_{1-p})}{(\exp[gZ_p] - \exp[-gZ_p])} = \ln(b) + h\left(\frac{Z_p^2}{2}\right)$$

Doordat OR data een positieve skewness heeft en een zware rechterstaart, kan de linkerzijde van deze vergelijking het best beschreven worden aan de hand van de upper half spread (UHS):

$$UHS = \frac{g(X_{1-p} - X_{0.5})}{(\exp[-gZ_p] - 1)}$$

Uiteindelijk kunnen dan de parameters \hat{b} en \hat{h} gevonden worden door een OLS regressie van de natuurlijke logaritme van UHS op $\frac{Z_p^2}{2}$. De exponent van de intercept is vervolgens de schatter van b en de regressiecoëfficiënt is de schatter van h. In het geval dat de transformatie van de standaard normale stochast Z naar de g-and-h verdeling niet zorgt voor een goede fit van de data, kunnen de parameters g en h, zoals eerder gesteld, geschreven worden als polynomen van Z^2 . In dit geval kan X geschreven worden als

$$X = a + b \exp\left[g(Z^2)Z - 1\right] \frac{\exp h(Z^2) \frac{Z^2}{2}}{g(Z)}$$

Hier zijn $g(Z^2)$ en $h(Z^2)$ polynomen van Z^2 . Echter, in zowel [8] als [25] wordt aangegeven dat slechts in een klein gedeelte van alle gevallen de fit beter wordt bij het gebruik van deze benadering ten opzichte van het geval dat g en h constant zijn.

4 – Goodness of Fit

Bij het modelleren van OR komt, zoals vaker bij modelleren, model risk kijken: het risico van het kiezen van het verkeerde model. In het geval van ORM is het kiezen van het juiste model zeer belangrijk, omdat het kiezen van het juiste model een grote invloed heeft op de VaR en daarmee ook het uiteindelijke economisch kapitaal. De twee voornaamste methoden voor het inschatten van de GoF zijn:

1. Grafische tests
2. Formele tests

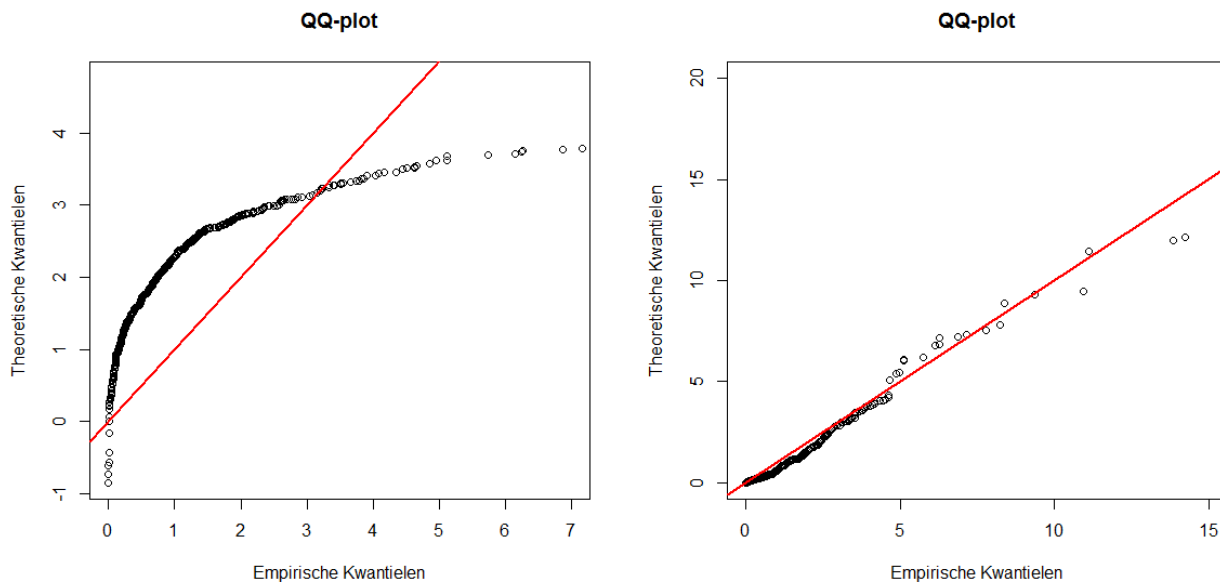
Beide methoden zullen in de hierop volgende paragrafen besproken worden.

4.1 – Grafische tests

Voor het bepalen van GoF worden vaak grafische tests gebruikt. Dit meestal in combinatie met formele tests, maar het gebeurt ook dat formele tests niet bruikbaar zijn en daarom alleen grafische tests gebruikt worden. Twee veel gebruikte grafische tests zijn de QQ-plot (Quantile-Quantile plot) en de mean excess plot.

4.1.1 – QQ-plot

Een QQ-plot is een visuele test om te onderzoeken of een dataset een bepaalde theoretische verdeling volgt. De empirische kwantielen worden geplot tegen de kwantielen van een theoretische verdeling om te testen of de data die specifieke verdeling volgen. Wanneer dit het geval is, volgen de geplotte punten de lijn $y = x$, ofwel een 45° lijn. Ter illustratie zijn hieronder twee QQ-plots weergegeven:



Figuur 19: QQ-plots van operationele verliezen tegen een normale verdeling (links) en een GPD (rechts)

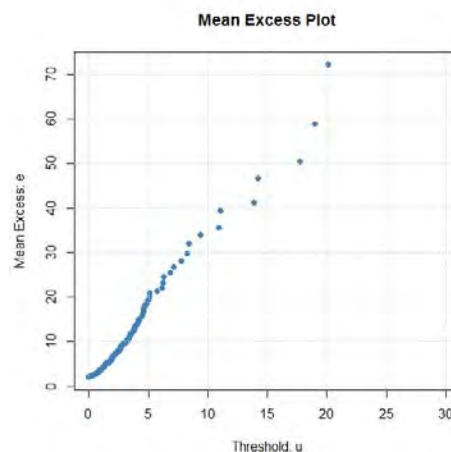
De dataset die gebruikt is voor deze twee plots is een combinatie van uniform, exponentieel en Pareto verdeelde stochasten, zodat de dataset een relatief sterk lichaam heeft en ook een dikke rechterstaart. In de linkerfiguur is deze dataset in een QQ-plot uitgezet tegen een normale verdeling. Hier valt direct te zien dat de dataset een dikke staart heeft, doordat de datapunten eerst boven de rode 45° lijn zitten, en op een gegeven moment terwijl de figuur afnemend stijgend doorloopt onder de rode lijn terecht komt. De rechterfiguur, waar de dataset wordt geplot tegen een GPD verdeling, is beter geschikt om deze dataset te modelleren, omdat de meeste punten rond de rode lijn liggen.

4.1.2 - Mean Excess Plot

Voor de mean excess plot wordt gebruik gemaakt van de mean excess functie, een functie die wordt gedefinieerd als het gemiddelde voor alle verschillen tussen de waarden van de datapunten die groter zijn dan drempel u , voor verschillende waarden van u . De mean excess functie (voor GPD in dit geval) ziet er als volgt uit:

$$e(u) = E [X - u | X > u] = \frac{\sigma}{1 - \xi} + \frac{\xi}{1 - \xi} u$$

Wanneer deze functie tegen verschillende waarden van u wordt geplot, resulteert dit in een mean excess plot. In Figuur 14 (blz. 24) valt een voorbeeld van een mean excess plot te zien, waarbij de populatie getrokken is uit een GPD verdeling met $\xi = 0.7$. Over het algemeen wordt drempel u zo gekozen dat de mean excess plot grofweg lineair is voor $x \geq u$. Een andere methode is om de drempel u zo te nemen, dat observaties x de vijf of tien procent grootste waarden zijn uit de dataset. Het kiezen van u is een belangrijk proces, aangezien de nog onbekende parameters erg gevoelig zijn voor de drempel. Een oplopende lijn suggereert een verdeling met een zware staart.



Figuur 20: Mean Excess Plot van dataset als in figuur 19

In figuur 20 valt een mean excess plot te zien van dezelfde dataset die is gebruikt in figuur 19. Ook hier valt te zien dat de plot erg oplopend is, hetgeen duidt op zware staarten.

4.2 – Formele tests

Formele tests kunnen worden gecombineerd met visuele tests: als er na een grafische test zoals een QQ-plot nog twijfel bestaat of een dataset een bepaalde verdeling volgt, kan een formele test gebruikt worden; op het moment dat uit de visuele test al blijkt dat de dataset een verdeling niet volgt kan de formele test achterwege blijven. Bij formele tests worden een nulhypothese en een alternatieve hypothese gevormd:

- H_0 : De data volgt de gespecificeerde theoretische verdeling
- H_1 : De data volgt de gespecificeerde theoretische verdeling niet

We maken bij formele tests onderscheid tussen twee typen: Chi-kwadraat tests en Empirische verdelingsfunctie (EDF) toetsen.

4.2.1 – Chi-kwadraat tests

Chi-kwadraat tests zijn gebaseerd op het asymptotische resultaat van een chi-kwadraat verdeling van de onderliggende test statistiek. Hier kan onderscheid gemaakt worden tussen discrete en continue stochasten: voor discrete verdelingsfuncties kan de Pearson's Chi-kwadraat toets gebruikt worden en voor continue verdelingsfuncties kan de Likelihood Ratio (LR) toets gebruikt worden.

De Pearson's Chi-kwadraat test wordt dus gebruikt om GoF te testen bij discrete verdelingsfuncties, zoals bijvoorbeeld de Poisson en de binomiale verdeling. De nulhypothese is dat de relatieve frequenties – frequenties van een bepaalde gebeurtenis ten opzichte van andere voorvallen – van de observaties een bepaalde verdeling volgen. In principe wordt de afstand tussen de theoretische frequenties en de daadwerkelijke frequenties. Allereerst moet de dataset opgedeeld worden in K verschillende klassen, in het geval van operationele verliezen zouden dit grootte of frequentie van OR verliezen zijn. Vervolgens wordt de test statistiek uitgerekend aan de hand van:

$$\chi^2 = \sum_{k=1}^K \frac{(n_k - E[n_k])^2}{E[n_k]}$$

Hier zijn n_k en $E[n_k]$ respectievelijk de daadwerkelijke frequentie in een klasse k en de theoretische of verwachte frequentie in een klasse k. Vervolgens wordt de statistiek vergeleken met het 0.95 percentiel (als het betrouwbaarheidsniveau 0.05 is) van de chi-kwadraat verdeling, met als aantal vrijheidsgraden het aantal klassen min 1. De nulhypothese wordt verworpen wanneer de statistiek groter is dan deze waarde. Twee nadelen van de Pearson chi-kwadraat test zijn dat ten eerste de test een dataset moet hebben die groot genoeg is zodat de asymptotische aanpak werkt, en ten tweede de test nogal gevoelig is voor de keuze van het aantal klassen.

De LR test kan gebruikt worden om de GoF te bepalen van continue verdelingen. Bij deze test wordt de likelihood functie waarbij de parameters geschat worden met behulp van MLE vergeleken met de likelihood functie onder de ware parameters van de nulhypothese. Dan wordt bepaald of het waarschijnlijk is of de data tot de populatie hoort. De LR ratio wordt bepaald door onderstaande formule:

$$LR(x) = \frac{\sup_{\theta \in \Theta_0} L(x; \theta)}{\sup_{\theta \in \Theta} L(x; \theta)}$$

Merk op dat het enige verschil tussen de noemer en de teller is dat in de noemer de parameters geschat worden door de MLE en dat dit niet de 'ware' parameters zijn. Aan de hand van $LR(x)$ wordt vervolgens de test statistiek berekend:

$$\chi^2 = -2 \ln LR(x)$$

Onder de nulhypothese is de term $-2 \ln LR \chi^2_d$ verdeeld, met d het aantal vrijheidsgraden, dat gelijkstaat aan de dimensionaliteit van Θ en Θ_0 . De nulhypothese wordt verworpen zodra de waarde van de statistiek groter is dan de 0.95 percentiel waarde van de theoretische chi-kwadraat verdeling. De LR test werkt goed als de ware parameters, dus de parameters onder de nulhypothese, ook daadwerkelijk de parameters zijn van de dataset. Ook voor de LR test is het noodzakelijk dat de dataset groot genoeg is.

4.2.2 - Empirische verdelingsfunctie (EDF) tests

Deze tests gebruiken de EDF en vergelijken deze met de gefitte verdelingsfuncties. De benadering om een EDF te berekenen is vergelijkbaar met de non-parametrische aanpak die is beschreven in hoofdstuk 3. De cumulatieve verdelingsfunctie van de data, dus de EDF kan dan geschreven worden als:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1 \{x_i \leq X\},$$

Met andere woorden, het aantal elementen kleiner dan observatie x , gedeeld door het totaal aantal elementen. Deze functie wordt dus vergeleken met de gefitte verdelingsfuncties. Er wordt dan gekeken naar de verticale verschillen tussen beide. Grofweg kunnen EDF tests verdeeld worden in twee groepen: supremum tests en kwadratische tests. De bekendste supremum test is de Kolmogorov-Smirnov (KS) test. Bij deze test wordt gekeken naar het verschil tussen F_n en F , met andere woorden het verschil tussen de EDF en de gefitte verdelingen. Hierbij noteren we D^+ als het grootste verschil tussen F_n en F en D^- als het grootste verschil tussen F en F_n , vandaar dat de KS test een supremum test is. Vervolgens wordt de KS statistiek berekend door:

$$KS = \sqrt{n} \max \{D^+, D^-\}$$

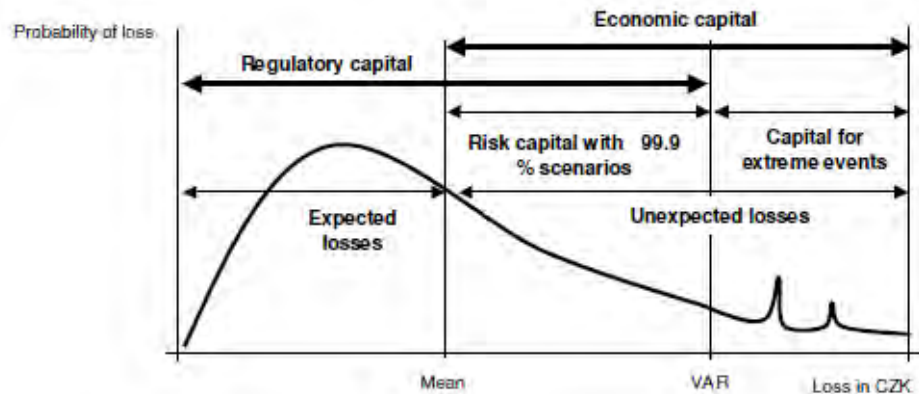
Een ander bekend voorbeeld van een EDF test is de Anderson-Darling (AD) test. De AD test heeft zowel een supremum variant als een kwadratische variant. In de context van OR heeft de AD test een belangrijk voordeel: namelijk dat deze test veel belang hecht aan de staarten van de verdeling. Deze test is dus van belang als er een vermoeden is – in het geval van OR is het geen vermoeden, maar een waarschijnlijkheid – dat de data een zware staart heeft. In [27] wordt aangegeven dat voor OR voornamelijk de rechterstaart van belang is en niet zozeer het fenomeen van zware staarten in het algemeen. Daarom kunnen de supremum statistiek en de kwadratische statistiek uitgedrukt worden als:

$$AD_{up} = \sqrt{n} \sup_x \left| \frac{F_n(x) - F(x)}{1 - F(x)} \right|$$
$$AD_{up}^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{(1 - F(x))^2} dF(x)$$

Deze tests worden de rechterstaart AD tests genoemd. Andere bekende tests zijn de Kuiper test (supremum) en de Cramér-von Mises test (kwadratisch).

5 – Value-at-Risk

Value-at-Risk (VaR) is een maat waarmee risico wordt gemeten, die ook in hoofdstuk 2.4.3 al voorbij kwam. VaR wordt ook wel gedefinieerd als het “zwaarst mogelijke verlies dat kan plaatsvinden gegeven een bepaalde betrouwbaarheid en tijdsinterval” [7]. VaR wordt gebruikt zodat zowel een manager van een bank kan inschatten wat voor verliezen hij/zij kan verwachten in bijvoorbeeld een jaar, terwijl ook toezichthouders de VaR kunnen gebruiken om in te schatten of diezelfde bank genoeg reserves heeft om de grootste verliezen in te kunnen dekken.



Figuur 21: Verdeling vereist kapitaal en economisch kapitaal. Bron: Chalupka & Teplý (2008) [18]

In bovenstaande figuur valt te zien waar de VaR ongeveer geplaatst is. Verreweg het grootste gedeelte van de OR verliesincidenten vindt rond en onder het gemiddelde plaats. 99,9% vindt plaats links van de VaR: dit betekent dat 0,1% van de verliezen hoger kan uitvallen dan deze waarde. Niet alle banken hanteren 99,9%: een hogere waarde als 99,95% of zelfs 99,99% komt voor.

5.1 – Berekenen van Value-at-Risk

Uit bovenstaande tekst komt naar voren in ieder geval het betrouwbaarheidsniveau en het tijdsinterval duidelijk moeten worden voor het berekenen van de VaR. Ook is het nodig dat duidelijk wordt wat de onderliggende valuta is. Daarnaast is duidelijk dat de verdeling voor frequentie en zwaarte van verliesincidenten moeten worden gecombineerd voor het berekenen van de uiteindelijke VaR. Actuariële modellen voor deze berekeningen berusten op een aantal aannames:

1. Gegeven het totaal aantal verliesincidenten, zijn verliesgroottes iid en positieve stochasten.
2. Geconditioneerd op het totaal aantal verliesincidenten n , is de verdeling van de zwaarte van de verliesincidenten onafhankelijk van n .
3. De verdeling van het totaal aan verliesincidenten is onafhankelijk van de verliesgroottes.

Wanneer deze aannames gedaan kunnen worden voor een bepaalde BL-ET combinatie, dan kunnen de verliezen gemodelleerd worden aan de hand van een actuariële model. Nu zijn verliesgroottes X en aantal verliezen N_t in een jaar t , dan volgen de verliezen een stochastisch proces in de vorm:

$$S_t = \sum_{i=1}^{N_t} X_i$$

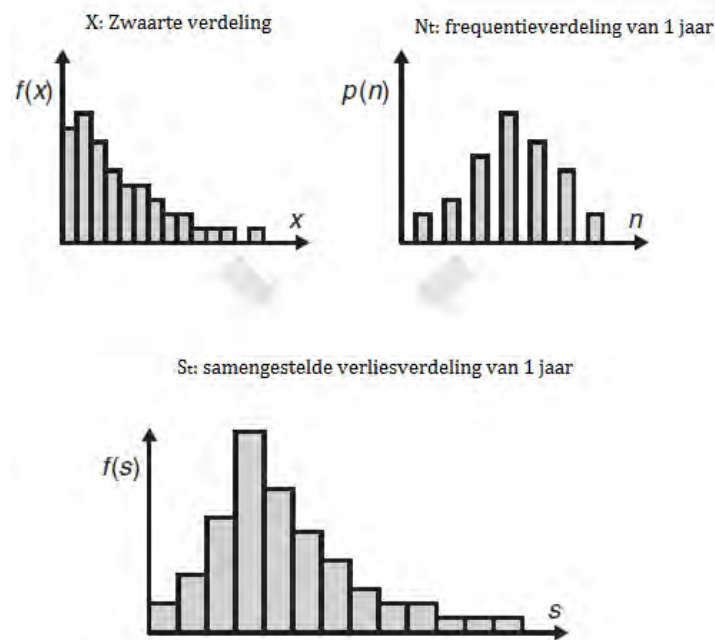
De cumulatieve verdelingsfunctie die bij dit samengesteld proces hoort, kan dan geschreven worden als

$$F_{S_t}(s) = P(S_t \leq s) = \begin{cases} \sum_{n=1}^{\infty} P(N_t = n) \cdot F_X^{n*}(s) & s > 0 \\ P(N_t = 0) & s = 0 \end{cases}$$

Hier staat F_X voor de verdelingsfunctie van X en F_X^{n*} staat voor de n -voudige convolutie van F_X met zichzelf:

$$F_X^{n*}(s) = P\left(\sum_k^n X_k \leq s\right)$$

Chernobai et. al. (2007) beschrijven bovenstaand proces aan de hand van de onderstaande figuur. De continue kansverdeling van de zwaarte van verliesincidenten moet samen worden gebracht met de discrete kansverdeling van de frequentie. Echter, boven valt te zien dat de cumulatieve verdelingsfunctie niet lineair is in zowel X als N , waardoor analytische uitdrukkingen niet bestaan.



Figuur 22: de werking van het actuariële model: samengestelde verliesverdeling (Bron: [7])

Het is mogelijk om de n -voudige convolutie van F_X om te schrijven en te benaderen door de continue verdeling voor de verliezen de discreet te maken:

$$F_X^{n*}(s) = \sum_{x=0}^s F_X^{(n-1)*}(s-x) f_X(x)$$

Echter deze methode is rekenkundig zwaar en vereist een groot aantal stappen om een benadering te geven die realistisch is. Vaak wordt de voorkeur gegeven aan een andere methode, waarbij gebruikt wordt gemaakt van een simulatie. Deze wordt in de volgende paragraaf besproken.

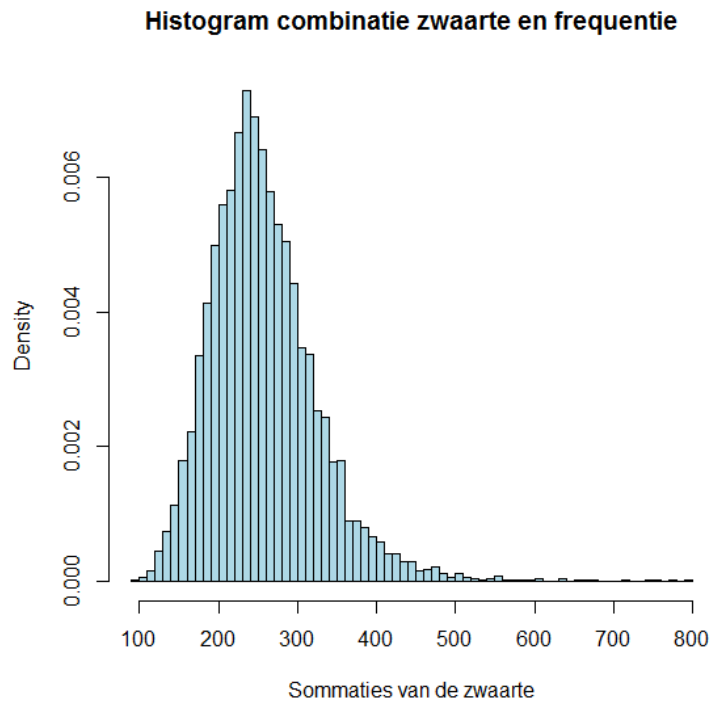
5.2 – Monte Carlo Simulatie

In plaats van een directe berekening, is een benadering als een Monte Carlo simulatie gebruikelijker in de context van OR. Monte Carlo methoden worden vaker gebruikt, vooral op het moment dat een functie niet of moeilijk analytisch uit te drukken is. Er bestaan verschillende vormen van Monte Carlo simulaties. In dit geval draait het om een simulatie waarbij een discrete verdeling en een continue verdeling gecombineerd worden als beschreven in de vorige paragraaf. De stappen die genomen worden kunnen als volgt beschreven worden:

1. Simuleer een groot aantal trekkingen uit de frequentie verdeling met de geschatte parameter, bijvoorbeeld 10.000. Hierdoor ontstaat een vector met trekkingen n_1 tot $n_{10.000}$, die staan voor scenario's van het aantal verliesincidenten in een jaar.
2. Voor elk van deze trekkingen worden n_k trekkingen gedaan uit de zwaarteverdeling, waarbij n_k staat voor de desbetreffende trekking van de frequentieverdeling
3. Voor elk van de scenario's, sommeer de n_k trekkingen uit de zwaarteverdeling. Hierdoor ontstaat een vector met 10.000 elementen, waarbij elk element staat voor een gesimuleerd totaal verlies in een jaar.
4. Sorteert de vector van laag naar hoog. De VaR is nu het $(1-\alpha)$ -percentiel.

Vaak wordt voor het de VaR het 99.5%- of 99.9%-percentiel gekozen. Dit verschilt per bank. Ook kan een nog groter percentiel gekozen worden, zoals bijvoorbeeld door de Rabobank gekozen wordt voor een 99.99% VaR [11]. Zie bijlage 5.2 voor de geschreven functie voor het berekenen van een VaR schatting bij bekende frequentie- en zwaarteverdelingen en parameters.

In een voorbeeld zou kunnen gelden dat bijvoorbeeld $N = 10000$, en wordt de Poisson verdeling als frequentieverdeling gebruikt en wordt de g-and-h als zwaarteverdeling gebruikt. De frequentieparameter wordt 100 genomen en als parameters van de g-and-h worden genomen $a = 1$, $b = 2$, $h = 0.1$ en $g = 1$. Deze specifieke combinatie is ook geplot in hoofdstuk 3.4. In figuur 23 op de volgende bladzijde staat een histogram afgebeeld van de vector die ontstaat uit beide verdelingen. Hier valt te zien dat het overgrote gedeelte rond de 250 ligt. Dit heeft op dit moment geen specifieke waarde, maar dient ter alleen ter illustratie. De grootste waarden liggen rond de 800. In dit specifieke geval heeft de g-and-h verdeling geen bijzonder dikke rechterstaart, maar dient wel goed ter illustratie. Het 99.5%-VaR is 500, 99,9%-VaR is 635 en 99.99%-VaR is 775. Zo valt te zien dat de keuze voor de drempel van VaR veel invloed kan hebben in het uiteindelijke economisch kapitaal. Volgens BIS moet de keuze van deze drempel toegelicht worden door de organisatie [28].



Figuur 23: *histogram van de verdeling van de combinatie van de frequentie en zwaarte verdeling*

Nu duidelijk is hoe van de dataset naar een schatting voor Value-at-Risk toegewerkt kan worden, is het mogelijk meer duidelijkheid te scheppen over de relevantie en invloed van data kwaliteit betreffende OR. Hierover meer in hoofdstuk 6.

6 – Data Kwaliteit

In een consultatief paper van BCBS wordt aangegeven dat “de aard en kwaliteit van OR data verzameld door een AMA bank niet alleen gevolgen heeft voor het proces van kwantificeren van de bank, maar ook voor de OR management beslissingen [28, paragraaf 19].” Alleen al deze stelling geeft aan dat de kwaliteit van de data van groot belang is in het hele OR proces, en enorme gevolgen kan hebben op het economisch kapitaal wat uiteindelijk berekend wordt. Vooral bij het schatten van parameters kunnen onnauwkeurigheden in metingen plaatsvinden met grote gevolgen. Dit wordt aangegeven in hetzelfde paper van BCBS: “Op het moment dat de parameters van een verdeling worden geschat, zou een bank het in beschouwing moeten nemen dat de dataset die gebruikt wordt voor de berekening mogelijk incompleet is. De bank moet kunnen aantonen dat dit geen effect heeft op de nauwkeurigheid en geloofwaardigheid van de geschatte parameters en het uiteindelijke kapitaal [28, paragraaf 35].” Voordat de relevante dimensies kunnen worden aangeduid is het belangrijk dat duidelijk wordt wat het precies rond data verzamelen en data gebruik is, zodat duidelijk is in welke context de data gebruikt wordt. Er moet dus een data kwaliteit strategie worden gevormd.

6.1 – Data Kwaliteit strategie

Voordat in beeld kan komen welke dimensies in het geval van OR verliesdatabases belangrijk zijn, moet er meer duidelijk zijn over de omgeving en de strategie. SAP geeft in een white paper uit 2008 [31] aan welke zes aspecten van data kwaliteit centraal staan bij het bepalen van een strategie. Deze factoren zijn:

1. Context: waar wordt de data voor gebruikt?
2. Opslag: waar bevindt de data zich?
3. Data flow: waar komt data binnen en hoe beweegt data door de organisatie?
4. Werk flow: wat is de interactie tussen werk activiteiten en de gebruikte data?
5. Rentmeesterschap: wie zijn verantwoordelijk voor de data?
6. Doorlopende controle: wat zijn de processen die zorgen voor regelmatige validatie?

De context van operationele verliesdata is duidelijk. In eerste instantie werd het economisch kapitaal berekend door een percentage van het inkomen te gebruiken, waar nu de mogelijkheid bestaat om op basis van historische data het economisch kapitaal te berekenen. Het primaire doel van de data is dus het gebruik voor het berekenen van economisch kapitaal. De data geeft echter ook inzicht in de wellicht zwakke plekken van de organisatie en kan laten zien waar structurele verliezen plaats vinden en verbeteringen nodig zijn.

De andere aspecten zijn in het geval van operationele verliesdatabases ondergeschikt aan de context. Dit betekent echter niet dat ze niet belangrijk zijn; deze aspecten waarborgen uiteraard een goede data kwaliteit. Duidelijke afspraken over wie rechten heeft om data in te voeren en aan te passen, sterker nog wie de verantwoordelijkheid heeft dat de data verwerkt wordt en wie controle uitvoert op de data zijn erg belangrijk. Het is duidelijk dat wanneer dit proces omtrent data verzameling en controle niet nauwkeurig verloopt, er mogelijkheden bestaan dat de data niet geschikt is om een accurate berekening mee uit te voeren. Er zal in deze scriptie niet verder ingegaan worden op het proces rond kwaliteitsbewaking. Voor meer informatie wordt verwezen naar het white paper van SAP Business Objects [31] en het boek “Journey to Data Quality” van Lee, Pipino, Funk en Wang (MIT Press 2007).

Wanneer duidelijk is wat de context van data is, kunnen ook aspecten binnen context vastgesteld worden die helpen bij het beoordelen van data kwaliteit. Dit wordt besproken in de volgende paragraaf.

6.2 – Dimensies van Data Kwaliteit

Binnen de context van data zijn een aantal dimensies bekend die mate van data kwaliteit aangeven. Data kwaliteit is een breed begrip. Elementen die belangrijk zijn in een database systeem van een middelbare school hoeven niet per definitie ook belangrijk te zijn in een operationele verliesdatabase. Het ORX geeft op hun website aan wat voor hun operaties belangrijke dimensies van data kwaliteit zijn, maar dat is specifiek voor hun werkzaamheden, bijvoorbeeld het opstellen van scenario's. De statistiek afdeling van Bank of England geeft een raamwerk voor data kwaliteit [29], waarin de volgens hen belangrijkste dimensies van data kwaliteit moeten worden beoordeeld. Dat zijn de volgende dimensies:

Relevantie

Relevantie is de mate waarin data de huidige en potentiële noden van gebruikers tegemoetkomt. Het geeft aan of alle data die nodig zijn worden geregistreerd en in welke mate de gebruikte concepten (zoals definities en classificaties) in lijn staan met de noden van gebruikers.

Accuraatheid

Accuraatheid is de mate waarin de berekeningen of geschatte waarden afwijken van de exacte of daadwerkelijke waarden.

Tijdigheid en punctualiteit

Tijdigheid van informatie geeft weer wat het tijdsverschil is tussen de beschikbaarheid van de data en het plaatsvinden van de beschreven gebeurtenis.

Beschikbaarheid en duidelijkheid

Beschikbaarheid refereert aan de fysieke condities waarin gebruikers data kunnen bemachtigen: waarheen te gaan, hoe te bestellen, levertijd, duidelijke prijsmethodiek, gemakkelijke marketing voorwaarden (zoals copyright) en de beschikbaarheid van micro of macrodata. Duidelijkheid refereert aan de data informatie omgeving: of de data wordt omringd door geschikte metadata, illustraties als grafieken en kaarten, of informatie betreffende kwaliteit van de data beschikbaar is en de mate waarin aangevulde hulp beschikbaar is.

Vergelijkbaarheid

Vergelijkbaarheid richt zich op het meten van de impact van verschillende in toegepaste statistische concepten, tools en procedures wanneer deze worden vergeleken met geografische gebieden, niet geografische domeinen of over de tijd.

Coherentie

Coherentie van data draait om toereikendheid om betrouwbaar gecombineerd te kunnen worden in verschillende manieren en voor verschillende doeleinden. Het is echter over het algemeen makkelijker een gebrek aan coherentie aan te tonen, dan te laten zijn dat de data coherent is.

Niet al deze dimensies zijn te gebruiken voor de doeleinden in deze scriptie. Het ORX geeft in hun rapport iets andere dimensies aan die zij belangrijk vinden. Alle leden van ORX leveren ieder kwartaal data aan, die volgens het ORX worden beoordeeld wat betreft de volgende dimensies van data kwaliteit:

- De *tijdigheid* van het aanleveren van de data
- Het *format* van de data leveringen

- *Compleetheid* van de data levering
- *Naleving* van de OR Rapportage Standaarden van ORX
- *Reactietijd* ten opzichte van data levering aanvragen door het advieskantoor van ORX.

Het ORX geeft aan in hun Operationele Risk Rapportage Standaarden van 2011 [30] dat aan deze standaarden voldaan moet worden, alsmede aan een accuraatheid eis. Nu is het begrijpelijk dat deze eisen samen met een accuraatheid eis worden gesteld door het ORX, omdat bij het onderzoeken van grote hoeveelheden data het heel tijd kan kosten om de data voor te bereiden als deze van slechte kwaliteit is of als er geen uniforme standaarden en formats gebruikt worden. Echter, voor het onderzoeken van de juiste zwaarteverdeling of het effect van data kwaliteit op economisch kapitaal is een kwaliteitseis als het format van de data levering niet van toepassing. Ook reactietijd of naleving van OR Rapportage Standaarden zijn geen eisen die een directe impact hebben op het economisch kapitaal.

Dimensies die wel directe invloed kunnen hebben op economisch kapitaal voor OR zijn accuraatheid, compleetheid, tijdigheid en relevantie. De eerste twee kunnen gezien worden als relevant voor het totale economische kapitaal en het kapitaal op alle niveaus (BL en ET). Tijdigheid heeft op de lange termijn niet direct invloed op het economisch kapitaal, maar kan wel een verkeerd beeld geven als aangaande consistentie van verliezen. Relevantie heeft vooral te maken met allocatie van verliezen en heeft dus niet per se een grote invloed op het uiteindelijke economische kapitaal. Een laatste dimensie die niet zozeer de kwaliteit van de data weergeeft, maar voor het doeleinde van deze scriptie belangrijk is, betreft de grootte van de dataset. Natuurlijk geldt dat hoe meer data, hoe beter het beeld dat gevormd kan worden over de data en hoe nauwkeuriger de juiste verdeling bepaald kan worden. Dit zijn dus de dimensies die gebruikt zullen worden bij het onderzoeken van de invloed van data kwaliteit op (de berekening van) het economisch kapitaal.

Omdat het niet het doel in deze scriptie is om de data kwaliteit in verliesdatabases te beoordelen, maar alleen om de invloed van deze data kwaliteit te onderzoeken, zal er niet verder stilgestaan bij het beoordelen van de afzonderlijke data kwaliteit dimensies. Voor meer informatie over dit onderwerp zie opnieuw "Journey to Data Quality" van Lee, Pipino, Funk en Wang (MIT Press 2007).

Nu alle nodige informatie is gegeven over regelgeving, modellen voor operationeel risico, technische aspecten van de verdelingen en dimensies van data kwaliteit, kan in hoofdstuk 7 stilgestaan worden bij de genomen stappen in het onderzoek.

7 – Methodologie

Nu de theoretische achtergrond duidelijk is, kan meer duidelijk worden over de methodologie. Omdat er de afgelopen jaren onderzoek heeft plaatsgevonden in ORM betreffende OR data, zou het niet logisch zijn om deze buiten beschouwing te laten. Er kan wat betreft dit hoofdstuk onderscheid gemaakt worden tussen twee delen: het literatuuronderzoek en het praktijkonderzoek.

7.1 – Literatuuronderzoek

Sinds de invoering van Basel II en de introductie van Economisch Kapitaal voor Operationeel Risico is er het nodige onderzoek gedaan naar berekeningsmethodieken. In 2001 kwam al een paper uit van de University of Reading waarin onderzocht wordt of operationele verliezen gemodelleerd kunnen worden op een causale manier aan de hand van Bayesian Belief Networks. Langzamerhand, mede doordat dit is aangegeven door BIS [2], is de focus van het onderzoek richting de Loss Distribution Approach verschoven. Naast onderzoek naar hoe interne data gecombineerd kan worden met externe data, scenario's en opinie van experts (Lambrigger et. al., 2009) er ook onderzoek gedaan naar de meest geschikte verdelingen voor frequentie en zwaarte van verliesincidenten. Zoals eerder gesteld, wat betreft de frequentieverdeling is consensus dat de Poisson verdeling gebruikt kan worden en is er wat betreft de zwaarteverdeling minder consensus. Sterker, op verschillende BL en ET niveaus kunnen verschillende verdelingen het meest geschikt zijn, in de zin dat het kan gebeuren dat voor een BL – ET combinatie Corporate Finance – Interne Fraude de g-and-h verdeling het best past en voor de risico cel Trading & Sales – Execution Delivery & Process Management een Pareto verdeling het best past.

Er worden drie onderzoeken bekeken waarin verschillende zwaarteverdelingen worden geanalyseerd. De onderzoeken van Kabir Dutta en Jason Perry – A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital [8], Milan Rippel en Petr Teplý – Scenario Analysis [10] en Anna Chernobai, Svetlozar Rachev en Frank Fabozzi – A Guide to Basel II Capital Requirements, Models and Analysis [7]. In deze onderzoeken worden een aantal zwaarteverdelingen onderzocht op Goodness of Fit. Deze onderzoeken kunnen worden gebruikt als vergelijking voor het praktijkonderzoek.

7.2 - Praktijkonderzoek

Het praktijkonderzoek zal ook in stappen plaatsvinden. BIS geeft aan dat “het proces van het selecteren van een kansverdeling goed gedocumenteerd en verifieerbaar moet zijn en moet leiden naar een duidelijke en consistente keuze. Om deze reden zou een bank zich aan de volgende punten moeten houden:” [28, par. 196].

- Een onderzoekende data analyse voor elke OR categorie om het statistische profiel van de data beter te begrijpen en de meest geschikte verdeling te kiezen
- Geschikte technieken voor het kiezen van de parameters van de verdelingen
- Geschikte diagnostische technieken voor het evalueren van de kwaliteit van de fit van de verdeling op de data, waarbij voorkeur uitgaat naar degenen met een gevoelige staart.

Er zal dus begonnen worden met een verkennende data analyse (EDA). Het doel van een verkennende data analyse is het beschrijven van de data in een vorm die makkelijk te begrijpen is. Hier wordt een omschrijving gegeven van de data en de bron van de data en wordt gekeken naar de frequenties van risico cellen en kan

gekeken worden voor welke cellen en voor welke BL en ET niveaus genoeg data is om aan een verdeling te fitten. Ook kan men hierdoor een idee krijgen van welke BL – ET combinaties een grote invloed hebben op het uiteindelijke economisch kapitaal. Omdat verschillende risicocellen onderzocht gaan worden, moet ook gecontroleerd worden of de geografische locaties niet een andere onderliggende verdeling hebben. Dit wordt gedaan door QQ-plots te maken en met een formele test, in dit geval een Kolmogorov-Smirnov test en een Anderson Darling test, te controleren of de onderliggende verdelingen hetzelfde zijn. Dit is onderdeel van de EDA.

7.2.1 – Schatten van parameters

Om de parameters te kunnen schatten worden een aantal verschillende vectoren gemaakt, voor alle relevante Business Lines, voor alle relevante Event Types, en voor alle relevante risico cellen. Hoeveel vectoren uiteindelijk onderzocht worden hangt af van de EDA. Vervolgens worden de parameters geschat via de methodes beschreven in hoofdstuk 3. Eerst worden voor alle drie gebieden (BL, ET en risico cel niveau) de parameters van de exponentiële en lognormale verdeling geschat. De exponentiële verdeling zal naar verwachting dienen als voorbeeld met een slechte fit. De lognormale verdeling werd aan het begin van het OR tijdperk veel gebruikt voor het modelleren van economisch kapitaal, vooral omdat de lognormale verdeling goed is in het modelleren van het lichaam van OR data, maar minder geschikt voor de staarten. Om deze reden wordt gesteld [7, 8, 25] dat de lognormale verdeling geschikt wordt voor het modelleren van het lichaam bij EVT. Hier wordt dan enerzijds de GEV gebruikt in het geval van de BMM methode en de GPD in het geval van de POT voor het modelleren van de staarten, ofwel de maxima.

In dit geval wordt bij het schatten van de parameters van EVT in beide gevallen ook de lognormale verdeling gebruikt voor het modelleren van het lichaam. Voor POT zijn van alle te onderzoeken BL, ET en risico cellen mean excess plots gemaakt om de waarde van drempel u te bepalen. Vervolgens worden aan de hand van MLE de parameters voor de GPD boven de drempel en lognormale verdeling onder de drempel bepaald. Voor het schatten van de GPD parameters wordt gebruik gemaakt van de package ‘fExtremes’ in R. Voor het schatten van de parameters van de lognormale verdeling wordt een maximum likelihood methode gebruikt die beschikbaar is in het package ‘MASS’. Voor het schatten van de parameters aan de hand van de BMM methode wordt gebruik gemaakt van een methode die beschikbaar is in het pakket ‘fExtremes’. Vervolgens moeten de datapunten die niet gebruikt worden als maxima onderzocht worden. Deze punten worden onderzocht met dezelfde functie als voor POT, maar het lichaam moet wel eerst geselecteerd worden. De gebruikte en geschreven methoden staan vermeld in bijlage 4.

Voor het schatten van de parameters van de alpha-stable verdeling bestaat geen analytische vorm van MLE. In dit geval wordt een numerieke benadering gedaan via R. De gebruikte methode is vermeld in bijlage 4. Het schatten van deze parameters kan een tijdrovende klus zijn, omdat er geen gesloten vorm is voor de verdelingsfunctie van de alpha-stable verdeling. Er zullen twee gevallen worden onderzocht: één waarbij alleen de restricties worden gehanteerd die bij de verdeling horen en vermeld zijn in hoofdstuk 3 en de parameters geschat worden aan de hand van MLE en één waarbij de parameters geschat worden aan de hand van een kwantielen methode, zoals aangegeven door McCulloch (1986) [33].

Bij het schatten van de parameters van de g-and-h verdeling wordt een van de parameters al bekend bij de verkennende data analyse; de schatter voor a is namelijk de mediaan van de dataset. Vervolgens worden de rest van de parameters geschat als beschreven in paragraaf 3.4.1. Bij het bepalen van de mediaan van g_p wordt een vector voor p gemaakt die van 0.0001 tot 0.9999 loopt met stapgrootte 0.0002, zodat deze niet geëvalueerd wordt bij $p=0.5$. Bij het bekijken van $qnorm(0.5)$ wordt stevast 0 gegeven, dus dit punt moet

ontweken worden. In dit onderzoek is niet naar de mogelijkheid gekeken om g en h polynomen van Z^2 (met Z de standaard normale verdeling) te maken gekeken.

7.2.2 – Bepalen van Goodness of Fit

Nadat voor de drie gebieden de parameters zijn geschat wordt gekeken hoe goed de fit is van elke verdeling op de afzonderlijke ETs, BLs en risico cellen. Dit wordt gedaan zoals beschreven in hoofdstuk 4: zowel aan de hand van grafische methode als aan de hand van formele tests. Voor de grafische methoden worden QQ-plots gebruikt. Hier wordt eerst één plot gemaakt van een van de ETs, BLs of risico cellen met alle verdelingen erin en daarna QQ-plots voor alle categorieën met de verdelingen met de beste fit. Deze selectie wordt aan de hand van de eerste QQ-plot gemaakt, maar voornamelijk aan de hand van de formele tests. Als formele tests worden EDF toetsen gebruikt die zijn beschreven in hoofdstuk 4: de Kolmogorov-Smirnov test voor twee verschillende populaties gebruikt, die ook gebruikt is bij het onderzoeken of de verschillende geografische gebieden samen onderzocht mogen worden, en de Anderson Darling test.

Na het bepalen van goodness of fit wordt voor alle verdeling een VaR schatting gemaakt aan de hand van de methode die is gegeven is bijlage 4.2. De VaR schatting op zichzelf zegt door de aard van de dataset niet bijzonder veel, omdat deze schatting niet vergeleken kan worden met een inkomen en daardoor onder of boven een bepaald percentage ligt. Over deze specifieke aard zal meer duidelijk worden bij hoofdstuk 8.2.1. De VaR schattingen zullen vooral nuttig zijn bij het onderzoeken van data kwaliteit, omdat er dan een duidelijk verschil gemeten kan worden per scenario. Deze scenario's zullen in de volgende paragraaf worden toegelicht.

7.2.3 –Onderzoeken invloed data kwaliteit

In hoofdstuk 6 is duidelijk geworden dat data kwaliteit moeilijk te kwantificeren is en in sommige gevallen zelfs moeilijk te beoordelen is. Het is moeilijk om een cijfer te geven aan de verschillende dimensies, dit geval aan de dimensies accuraatheid en volledigheid. Wat echter wel mogelijk is, betreft een scenario analyse. Via deze weg kan onderzocht worden wat het effect is van een incomplete, onnauwkeurige of niet tijdige/niet punctuele dataset. Onderstaand staat aangegeven hoe dit voor deze dimensies zal plaatsvinden.

Accuraatheid

Om de dimensie accuraatheid te onderzoeken, wordt bekeken wat het gevolg op VaR is bij het afronden van de data. In de dataset worden datapunten tot op tienduizend dollar nauwkeurig weergegeven. Er zal onderzocht worden wat het gevolg is als datapunten op honderdduizend dollar of miljoen dollar nauwkeurig worden gegeven.

Compleetheid

Compleetheid houdt in deze context in: het omvatten van alle gebeurtenissen, dus dat alle operationele verliezen ook in de database zijn opgenomen. In dit geval zullen twee gevallen worden onderzocht: het geval dat de hoogste 5% van de waarnemingen ontbreken en het geval dat een willekeurige 5% van de waarnemingen ontbreken.

Punctualiteit

Bij het registreren van verliesincidenten worden de verliesgroottes vaak genoteerd in de waarde op dat moment. Op het moment van analyseren is de huidige waarde van dat verlies echter relevanter, omdat het

berekenen van economisch kapitaal draait om het afdekken van risico's op dit moment. Dit zal onderzocht worden door niet de huidige waarde van het verlies te onderzoeken, maar de actuele waarde op het moment van verlies.

Voor elke dimensie zullen dezelfde stappen worden uitgevoerd als beschreven in 7.2.1, met in dit geval aangepaste data. De nieuwe VaR schatting wordt gebruikt om een procentuele verandering te kunnen beschouwen ten opzichte van de oorspronkelijke schatting en een conclusie te kunnen trekken over de gevolgen van variërende data kwaliteit.

8 – Resultaten

In dit hoofdstuk worden de resultaten besproken die uit het onderzoek naar voren zijn gekomen. In de eerste paragraaf wordt stilgestaan bij de drie onderzoeken die in het vorige hoofdstuk zijn genoemd, waarna in het praktijkonderzoek na een verkennende data analyse de parameters worden geschat en het kapitaal wordt uitgerekend.

8.1 - Literatuuronderzoek

Eerst wordt het onderzoek van Milan Rippel en Petr Teplý beschouwd. In dit onderzoek worden de g-and-h verdeling en Extreme Value Theory uitgelegd en met elkaar vergeleken. In dit onderzoek wordt de data van een anonieme bank gebruikt die bestaat uit 657 verliesincidenten. In onderstaande tabel wordt de gebruikte data beschreven:

Tabel 5: Data statistieken van Rippel & Teplý in USD

Mean	52882
Median	3945
Standard Deviation	355441
Skewness	17.7
Kurtosis	285

De hoge waarden voor kurtosis, scheefheid en standaard deviatie zijn ook te wijten aan het feit dat er geen onderdrempel wordt gebruikt: 80% van alle datapunten zijn kleiner dan 25340 USD. Na het schatten van de parameters zijn de Kolmogorov-Smirnov test en de Anderson-Darling test gebruikt naast de QQ-plot voor het bepalen van GoF. Het resultaat in dit onderzoek was dat EVT een erg goede fit had met de data, maar dat het economisch kapitaal erg groot was. Waar de BIA en SA ongeveer 15% van het inkomen van de bank innemen, geldt in dit geval dat de g-and-h verdeling 4.43% van het inkomen inneemt en EVT respectievelijk 14.95% en 9.32% voor de Block Maxima en Points Over Threshold Methodes, bij een 99.9%-VaR. In dit geval wordt de g-and-h verdeling geprefereerd boven EVT.

In het onderzoek van Kabir Dutta en Jason Perry worden parametrische verdelingen, EVT en een non-parametrische verdeling gefit op de data. De dataset die gebruikt wordt is de Loss Data Collection Exercise van BIS uit 2004. Waar Rippel en Teplý dus de data van één bank beschouwen, daar beschouwen Dutta en Perry meerdere banken. Wat betreft de parametrische verdelingen wordt gekeken naar de exponentiële, gamma, GPD, log-logistisch, afgeknotte (truncated) lognormale, Weibull, gegeneraliseerde Beta (GB2) en de g-and-h verdeling. Hier zitten dus verdelingen met één, twee of vier parameters in. Dutta en Perry schatten het kapitaal evenals Rippel en Teplý in aan de hand van 99.9%-VaR. Naast een goede statistische fit bekijken Dutta en Perry ook of het kapitaal realistisch is, of de karakteristieken van de gefitte data consistent zijn met de verliesdata, hoe flexibel de methode een wijde range van empirische verliesdata kan modelleren en of de methode makkelijk toe te passen is in de praktijk. Hier komt naar voren dat de g-and-h een goede fit heeft op enterprise niveau, dus over alle datapunten van een organisatie. De GB2, log-logistische, afgeknotte lognormale en GPD hebben een redelijke GoF. De g-and-h en de empirische verdeling geven consistente schattingen over verschillende banken. Dutta en Perry gaan zelfs dusdanig ver dat zij stellen dat de g-and-h verdeling zowel het lichaam als de staart goed kunnen modelleren en hierdoor de g-and-h verdeling een verdeling is die de data goed past zonder af te knotten of te trimmen op een arbitraire of subjectieve manier. In het onderzoek geven

zij aan in hoeveel gevallen elke verdeling een goede fit en een realistische schatting geeft. Ook Dutta en Perry geven dus aan dat de g-and-h verdeling de best beschikbare optie is.

In het laatste onderzoek, van Chernobai et al., wordt de g-and-h verdeling niet besproken. Zij geven verschillende voorbeelden die zijn aangekaart in literatuur tot 2006. Hier wordt wel uitgebreid stilgestaan bij andere parametrische verdelingen, waaronder de alpha-stable, afgeknotte lognormale verdelingen en ook bij EVT. In dit boek worden geen conclusies getrokken over welke verdeling mogelijk te prefereren is, maar worden alleen tips gegeven over hoe het risico kapitaal berekend kan worden en hoe deze verdelingen gebruikt kunnen worden en wat uitkomsten van eerdere onderzoeken waren. In deze onderzoeken tot 2004 komt naar voren dat een lognormale verdeling het best past. Het moet hier wel gezegd worden dat de GoF niet optimaal is en de VaR soms een grote waarde heeft ten opzichte van het inkomen, maar van alle parametrische verdelingen die tot dan getest zijn (exponentieel, Weibull, Pareto en Gumbel) komt lognormaal het sterkst naar voren. Chernobai, Menn, Rachev en Trück onderzochten in 2005 ook de alpha-stable verdeling. Deze verdeling presteerde sterker dan andere parametrische verdelingen in hun onderzoek. EVT wordt in het boek niet vergeleken met parametrische verdelingen. Wel worden de BMM en POT methodes vergeleken en wordt aangeduid dat de POT methode beter presteert dan de BMM methode. Verder wordt ook gerefereerd aan het onderzoek van Dutta en Perry, maar zonder iets over de g-and-h verdeling los te laten. Ook wordt hier vermeld dat de schattingen erg gevoelig zijn voor de keuze van de drempel.

8.2 - Praktijkonderzoek

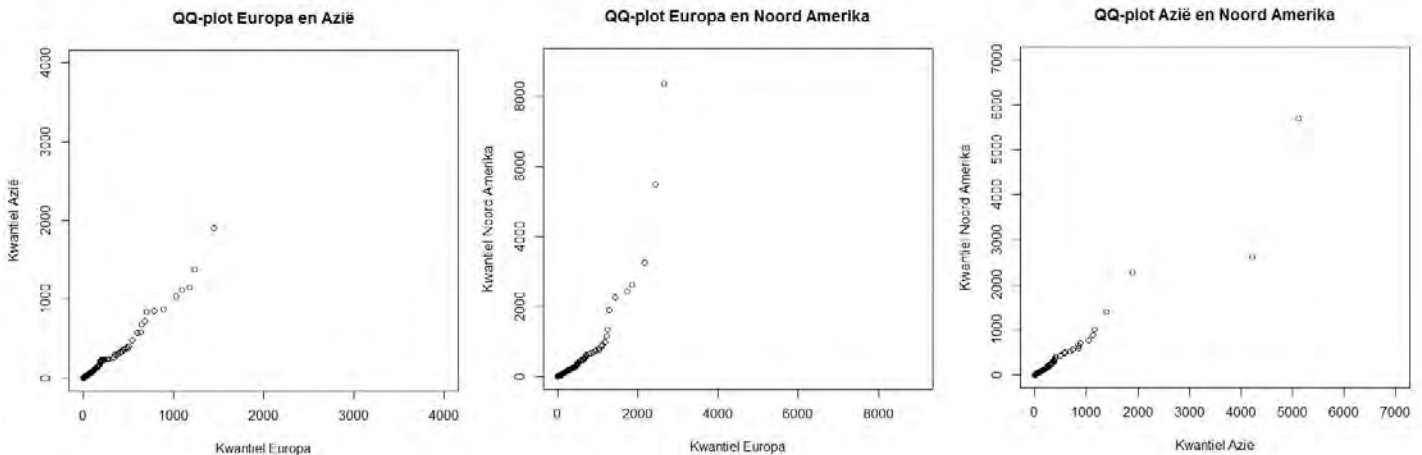
Omdat het niet zinnig is om verdelingen te fitten zonder eerst een duidelijk beeld te hebben van de data en te weten of alle data wel relevant is, wordt eerst een verkennende data analyse uitgevoerd. Ook zal hier gekeken worden naar de algemene aannames die gemaakt moeten worden bij de analyse.

8.2.1 - Verkennende data analyse

De gebruikte dataset is data van [REDACTED]. In totaal zijn in deze database wereldwijd [REDACTED] verliesincidenten opgenomen, die in dollars op tienduizendtallen nauwkeurig gegeven wordt. Echter, hier zitten ook datapunten die [REDACTED] jaar oud zijn. Niet alle datapunten zijn dus interessant en relevant om mee te nemen. In tabel 6 onderstaand valt te zien wat het aantal incidenten is in verschillende periodes. Voor de analyse zullen de jaren 2000 tot 2010 in beschouwing genomen worden. Dit ook omdat hiervoor het hele begrip operationeel risico nog niet gebruikt werd, ook al vonden de incidenten wel plaats. Naast het feit dat in deze database spreiding is over een aantal jaren is er ook spreiding in geografische zin. In tabel 6 valt ook te zien waar de meeste incidenten hebben plaatsgevonden. Er valt te zien dat dit vooral in Europa en Noord Amerika is. Ook in Azië zijn redelijk wat incidenten in de database aanwezig. In de QQ-plots in figuur 24 zijn de kwantielen van de geografische gebieden Azië, Europa en Noord-Amerika tegen elkaar uitgezet. Er valt een diagonale lijn te herkennen in de QQ-plots, ook al geldt voor Europa en Noord-Amerika dat bij hoge waarden afgeweken wordt van deze lijn. QQ-plots van het Midden-Oosten, Oceanië en andere gebieden zijn hier niet weergegeven, maar geven geen ander beeld. In tabel 7 worden de p-waarden resulterend uit een Kolmogorov-Smirnov test gegeven. Alleen de combinatie Europa – Midden Oosten, Midden Oosten – Azië en Noord-Amerika – Oceanië lijken niet dezelfde verdeling te hebben. Echter doordat overwegend uit de KS test komt dat de aanname dat beide onderliggende verdelingen hetzelfde zijn, zullen de gebieden samen onderzocht worden.

Tabel 6: Incidenten per jaar en per regio

Jaar	Aantal Incidenten	Jaar	Aantal Incidenten	Regio	Aantal Incidenten	Regio	Aantal Incidenten 2000-2010
		2004	275	Azië		Azië	293
		2005	303	Caribbean		Caribbean	15
		2006	248	Centraal en Zuid Amerika		Centraal en Zuid Amerika	38
		2007	243	Europa		Europa	577
		2008	247	Midden Oosten		Midden Oosten	69
2000	291	2009	383	Noord Amerika		Noord Amerika	2146
2001	311	2010	285	Oceanië		Oceanië	82
2002	362						
2003	298						



Figuur 24: QQ-plots kwantielen van verschillende geografische gebieden

Tabel 7: p-waarden Kolmogorov-Smirnov test voor onderliggende verdelingen

	Europa	Noord-Amerika	Midden-Oosten	Oceanië	Azië
Europa	1	0,0605	0,6380	0,0481	0,0009
Noord-Amerika	0,0605	1	0,0322	0,3434	0,0241
Midden-Oosten	0,6380	0,0322	1	0,0402	0,6583
Oceanië	0,0481	0,3434	0,0402	1	0,0053
Azië	0,0009	0,0241	0,6583	0,0053	1

Om een verdere indruk te krijgen van de data, zijn twee heat maps gecreëerd: één waarin de frequentie per risico cel wordt weergegeven en één waarin het totale verlies in deze database weergegeven wordt. Hieruit kan opgemaakt worden welke risicocellen of zelfs welke Business Lines of Event Types genoeg data omvatten om onderzocht te worden. Voor de Business Lines en Event Types worden de omschrijvingen aangehouden die

zijn aangegeven door BIS en in bijlage 3 omschreven worden. Onder de Business Line “Others” vallen activiteiten die niet in één van de categorieën past. Ook worden hier niet-bancaire financiële instellingen onder geplaatst, die voor de analyse niet geschikt zijn. Voor het onderzoek zal deze categorie dus niet gebruikt worden, omdat er geen zinnige conclusies getrokken kunnen worden uit een dusdanige verzameling van incidenten. Voor de uiteindelijke analyse zal dus gebruik gemaakt worden van 2532 datapunten.

Tabel 8: *Frequentietabel voor alle risicocellen*

	Agency Services	Asset Management	Commercial Banking	Corporate Finance	Payment & Settlement	Retail Banking	Retail Brokerage	Trading & Sales	Others	Totaal
Business Disruption and System Failure	0	0	0	0	6	8	2	1	5	22
Clients Products and Business Practices	21	206	102	120	42	222	153	177	380	1423
Damage to Physical Assets	0	0	2	0	1	3	0	1	33	40
Employment Practices and Workplace Safety	0	8	9	6	0	19	28	20	32	122
Execution Delivery and Process Management	1	19	19	11	15	41	23	32	95	256
External Fraud	8	44	152	12	14	135	10	8	31	414
Internal Fraud	4	244	146	23	15	234	91	74	112	943
Totaal	34	521	430	172	93	662	307	313	688	3220

Tabel 9: *Totale verlies voor alle risicocellen (in miljoen USD)*

	Agency Services	Asset Management	Commercial Banking	Corporate Finance	Payment & Settlement	Retail Banking	Retail Brokerage	Trading & Sales	Others	Totaal
Business Disruption and System Failure										1032,88
Clients Products and Business Practices										190595,5
Damage to Physical Assets										17613,85
Employment Practices and Workplace Safety										2948,23
Execution Delivery and Process Management										27114,47
External Fraud										40875,54
Internal Fraud										187741,50
Totaal	3862,56	132058,17	62597,73	43201,1	5113,95	88844,85	9308,21	39561,21	83374,23	467922,00

In bovenstaande tabel valt te zien dat de verliezen enorm verschillen per Business Line en ook per Event Type. De verliezen voor Internal Fraud en Clients, Products and Business Practices zijn een stuk groter dan voor de andere Event Types, zoals voor de Business Lines geldt dat Asset Management een stuk groter is dan de rest. Dit komt mede doordat het grootste verlies in de dataset – een verlies van \$67 miljard – een van de verliezen in de risico cel Internal Fraud – Asset Management is. Uit de tabel daarboven blijkt dat er niet genoeg datapunten in alle risico cellen aanwezig zijn om deze te onderzoeken. Om deze reden worden vijf risico cellen geselecteerd om te onderzoeken. De Event Types Business Disruption and System Failure,

Damage to Physical Assets en de Business Lines Agency Services en Payment and Settlement zullen niet afzonderlijk worden onderzocht omdat er niet genoeg datapunten zijn om betrouwbare conclusies te trekken.

Omdat één van de doelen van het onderzoek het vinden van de best passende verdelingen is, volgt logischerwijs dat de karakteristieken van de dataset onderzocht dienen te worden. In tabel 10 en 11 onderstaand worden het gemiddelde, de mediaan, standaard afwijking, kurtosis en scheefheid gegeven voor alle Business Lines en Event Types. Sommige van deze statistieken zullen ook belangrijk zijn bij het schatten van parameters.

Tabel 10: *beschrijvende numerieke statistieken per Business Line*

	Agency Services	Asset Management	Commercial Banking	Corporate Finance	Payment & Settlement	Retail Banking	Retail Brokerage	Trading & Sales
Mean	113,6047	253,4706	145,5761	499,43	54,98871	134,2068	30,3199	126,3936
Median	34,425	14,96	17,795	42,85	14,55	7,23	5,52	9,83
Standard Deviation	226,9979	2980,554	503,0131	3367,6	116,9005	720,2712	94,03816	505,5839
Kurtosis	6,825257	481,7535	98,0241	146,61	27,25484	93,92905	43,9288	125,1351
Skewness	2,700893	21,67204	8,74509	11,872	4,739336	9,244109	6,24531	10,07301

Tabel 11: *beschrijvende numerieke statistieken per Event Type*

	Business Disruption and System Failure	Clients Products and Business Practices	Damage to Physical Assets	Employment Practices and Workplace Safety	Execution Delivery and Process Management	External Fraud	Internal Fraud
Mean	46,94909	133,9392	440,3475	24,16582	105,9159	98,73319	199,0896
Median	6,57	13,31	148,42	5,435	8,43	12,2	12,08
Standard Deviation	88,67079	573,7013	642,5121	50,21711	757,8996	449,113	2247,162
Kurtosis	4,961013	123,7801	3,767549	15,91635	199,869	178,7945	828,8112
Skewness	2,385314	10,19817	2,091192	3,795035	13,71558	12,11389	28,07633

Zoals beschreven in de methodologie zijn vervolgens van de relevante Business Lines, Event Types en risico cellen vectoren gemaakt die onderzocht worden. De relevante categorieën, die eerder al zijn besproken worden voor overzichtelijkheid in tabel 12 weergegeven.

Tabel 12: te onderzoeken vectoren

Event Types	Business Lines	Risico cellen
<ul style="list-style-type: none"> • Clients, Products and Business Practices • Employment Practices and Workplace Safety • Execution Delivery and Process Management • External Fraud • Internal Fraud 	<ul style="list-style-type: none"> • Asset Management • Commercial Banking • Corporate Finance • Retail Banking • Retail Brokerage • Trading & Sales 	<ul style="list-style-type: none"> • Asset Management – Internal Fraud • Commercial Banking – External Fraud • Retail Banking – Clients, Products and Business Practices • Retail Banking – Internal Fraud • Trading & Sales – Clients, Products and Business Practices

8.2.2 – Analyse Event Types

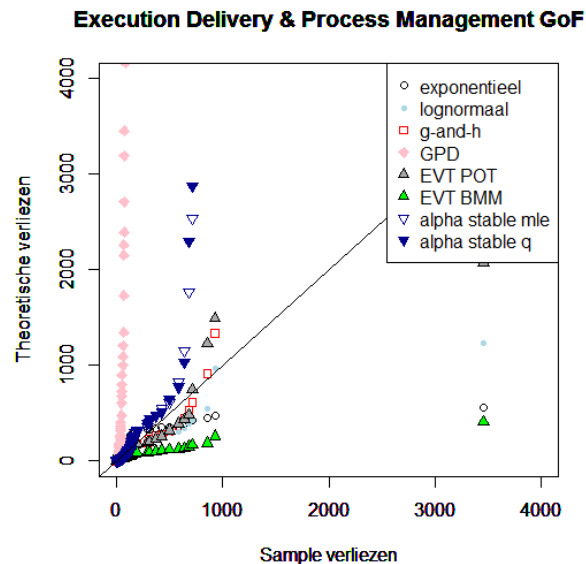
In hoofdstuk 7 is beschreven hoe de parameters zijn geschat. In tabel 13 onderstaand zijn de resultaten te zien. Hier wordt voor EVT bij POT de GPD gebruikt voor de extreme observaties en bij BMM de GEV, zoals beschreven in hoofdstuk 3. Voor de alpha-stable verdeling zijn twee methodes gebruikt voor het schatten van de parameters: een MLE methode en een methode waarbij eerst alpha en beta geschat worden op basis van kwantielen. In tabel 13 valt te zien dat de resultaten voor beide methoden flink verschillen. In onderstaande tabel staan CPBP (Clients Products and Business Processes), EPWS (Employment Practices and Workplace Safety), EDPM (Execution Delivery and Process Management), External Fraud en Internal Fraud beschreven. Zoals beschreven in de vorige paragraaf worden de ETs Business Disruption and System Failure en Damage to Physical Assets niet beschouwd, doordat deze ETs te weinig datapunten bevatten om een zinnige conclusie te trekken.

Tabel 13: parameters van Event Types

Distribution	Distri- bution	Event Types	CPBP	EPWS	EDPM	External Fraud	Internal Fraud
Lognormal	Exp	rate	0,007466	0,041381	0,009442	0,010128	0,0050229
		meanLog	2,8546	2,0409	2,3823	2,7203	2,7935
		sdLog	1,8785	1,4126	1,7334	1,7275	1,8515
EVT POT	EVT POT	threshold	700	50	200	500	750
		shape	0,73202	0,30708	0,63708	0,54009	0,84274
		scale	716,7314	57,05707	840,9791	577,06116	756,69514
		location	1	1	1	1	1
Body	Body	meanLog	2,701634	1,673987	2,213119	2,556778	2,61568
		sdLog	1,703896	1,061619	1,505063	1,529387	1,627454
EVT BMM	EVT BMM	blocks	10	10	10	10	10
		shape	1,2079	0,46408	1,1488	1,3609	1,2079
		scale	204,741	52,95427	128,4574	97,1636	204,741
		location	179,3912	53,71974	114,1881	83,0837	179,3912
Body	Body	meanLog	2,457293	1,782421	2,04175	2,443255	2,457293
		sdLog	1,567687	1,200057	1,40707	1,535132	1,567687

GPD	shape	1,5952	1,1594	1,5901	1,3911	1,5952
	scale	8,4407	4,5827	5,0196	9,1189	8,4407
	location	1	1	1	1	1
g-and-h	A	13,31	5,435	8,43	12,2	12,08
	B	4,61	8,731	10,63	12,5001	12,029
	g	2,4434	2,3276	2,1073	2,1181	2,3633
	h	0,1359	0,0901	0,2901	0,308	0,333
alpha stable (mle)	alpha	0,46423	0,61335	0,515	0,50125	0,48088
	beta	0,9999	0,9999	0,9999	0,9999	0,9999
	gamma	1,9999	1,54146	1,5996	1,9999	1,9999
	delta	0,9999	0,9999	0,9999	0,9999	0,9999
alpha stable (q)	alpha	0,621	0,703	0,699	0,66	0,61
	beta	0,95	0,95	0,95	0,95	0,95
	gamma	10,4303	3,9494	5,9387	8,384	8,6029
	delta	4,1715	2,2188	3,7561	5,26	4,4131

Nu de parameters geschat zijn, moet de Goodness of Fit voor alle verdelingen per ET bepaald worden. Dit vindt plaats aan de hand van zowel grafische als formele tests. In figuur 25 valt een QQ-plot te zien waarbij alle verdelingen geplot zijn tegen de data van de ET Execution Delivery & Process Management. Hier valt te zien dat vooral de alpha-stable verdeling en de GPD een erg slechte fit hebben met de data, omdat deze verdelingen in dit geval te zware staarten hebben. De verdelingen die juist te dunne staarten heen zijn in dit geval EVT BMM en de exponentiële verdeling. Volgens de Anderson Darling test (Tabel 14) en de Kolmogorov-Smirnov test (Tabel 15) is alleen voor EVT POT en de lognormale verdeling niet genoeg bewijs om de nulhypothese dat beide populaties dezelfde onderliggende verdeling hebben te verwerpen. In de QQ-plot valt te zien dat de fit van de g-and-h verdeling niet onderdoet voor EVT POT.



Figuur 25: QQ-plot met de onderzochte verdelingen tegenover de sample verliezen voor EDPM

Tabel 14: *p-waarden uit de Anderson Darling test voor Event Types*

Distribution	CPBP	EPWS	EDPM	External Fraud	Internal Fraud
Exp	0,00	0,00	0,00	0,00	0,00
Lognormal	0,0025	0,298	0,0659	0,26079	0,00251
EVT POT	0,0423	0,14246	0,0791	0,1438	0,00094
EVT BMM	0,00	0,0672	0,00033	0,0122	0,000008
GPD	0,001	0,00043	0,00004	0,00	0,00
alpha stable (mle)	0,00	0,00	0,0001	0,00	0,00
alpha stable (q)	0,00	0,02058	0,01133	0,00024	0,00

Tabel 15: *p-waarden uit de Kolmogorov-Smirnov test voor Event Types*

Distribution	CPBP	EPWS	EDPM	External Fraud	Internal Fraud
Exp	0,00	0,00	0,00	0,00	0,00
Lognormal	0,007648	0,1037	0,2528	0,10370	0,00437
EVT POT	0,01487	0,4701	0,3552	0,4372	0,00513
EVT BMM	0,00	0,0726	0,0649	0,0611	0,002967
GPD	0,00	0,01781	0,00	0,00	0,00
alpha stable (mle)	0,00	0,00	0,00	0,00	0,00
alpha stable (q)	0,00	0,00	0,00	0,00056	0,00

In Tabel 14 en Tabel 15 valt te zien voor welke Event Types er volgens formele test een statistisch goede fit is. In bijlage 5.1 zijn de overige QQ-plots te zien, waar in dit geval steeds EVT, de lognormale verdeling en de g-and-h verdeling worden geplot tegen de data van een specifieke ET. Dit omdat uit de formele tests blijkt dat de exponentiële verdeling, de alpha-stable verdeling en de GPD geen goede fit hebben. Deze verdelingen worden vaak met sterk bewijs verworpen na een formele test. In bijlage 5.1 valt te zien dat de g-and-h verdeling vaak het dichtst bij de lijn $x=y$ ligt. Daarnaast heeft voor EVT vaak POT een betere fit, alleen lijkt bij EPWS BMM een betere fit te hebben, ook al blijkt dit niet uit de formele tests. Voor Internal Fraud lijkt de fit op de QQ-plot voor EVT POT en g-and-h goed, maar bij de formele tests blijkt dit niet. Dit kan echter veroorzaakt zijn door de extreme waarde die aanwezig is bij Internal Fraud. Deze waarde is meer dan 300 maal zo groot als het gemiddelde en 3000 maal zo groot als de mediaan van Internal Fraud.

In Tabel 16 zijn de VaR schattingen voor de ETs geplaatst. Hier valt te zien dat de schattingen voor de exponentiële verdeling en de lognormale verdeling aan de lage kant zijn en de schattingen voor de GPD en alpha-stable verdeling erg hoog zijn. De schattingen voor EVT POT zijn beter voor 5% extreme observaties dan 10% en voor BMM geldt dat jaarlijks beter werkt dan halfjaarlijks.

Tabel 16: Value-at-Risk schattingen per Event Type

Distribution	CPBP	EPWS	EDPM	External Fraud	Internal Fraud
Exp	2,65E+04	7,44E+02	5,55E+03	7,35E+03	2,81E+04
Lognormal	7,88E+04	1,93E+03	1,13E+04	2,02E+04	5,31E+04
EVT POT (5%)	2,47E+05	1,48E+03	1,08E+05	7,34E+04	7,70E+05
EVT POT (10%)	3,77E+05	1,92E+03	2,73E+05	8,42E+04	1,47E+06
EVT BMM (yearly)	5,66E+05	2,94E+03	2,98E+05	1,19E+06	6,73E+05
EVT BMM (half-y)	1,55E+06	3,82E+03	7,03E+05	2,25E+06	1,48E+06
g-and-h	2,13E+05	7,98E+03	1,99E+05	1,27E+06	1,78E+06
GPD	8,09E+08	1,65E+05	6,20E+07	9,26E+06	2,94E+08
alpha stable (mle)	1,80E+11	7,94E+06	3,70E+08	3,36E+09	3,36E+10
alpha stable (q)	8,10E+08	1,00E+08	7143356,0	3,70E+07	4,70E+08

8.2.3 – Analyse Business Lines

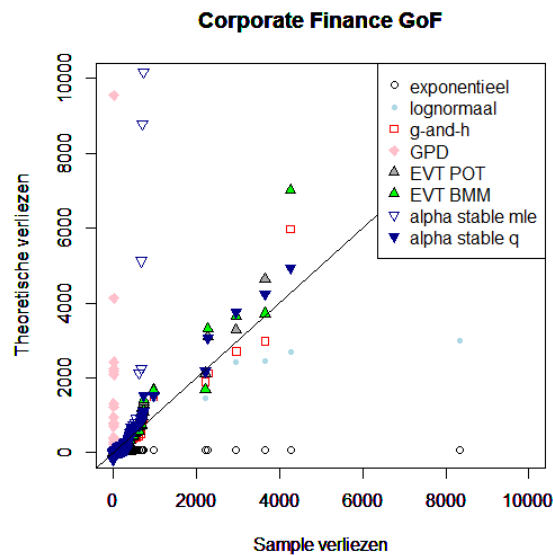
In Tabel 17 zijn de parameters gegeven die horen bij de verdelingen van de Business Lines. De gebruikte methodes zijn gelijk aan de methodes die gebruikt zijn voor de Event Types. De parameters zijn duidelijk anders dan de parameters horend bij ETs.

Tabel 17: parameters van de verdelingen die horen bij Event Types

Distributions	Business Lines	Asset Mgmt	Commercial Banking	Corporate Finance	Retail Banking	Retail Brokerage	Trading & Sales
Exp	rate	0,003945	0,006869	0,002002	0,007451	0,032982	0,007912
Lognormal	meanLog	2,9477	3,1594	3,7095	2,4013	2,0181	2,6795
	sdLog	1,8250	1,8128	2,0347	1,8129	1,4385	1,9395
EVT POT	threshold	400	400	300	450	80	500
	shape	1,0671	0,32217	1,4743	0,7778	0,57999	0,59947
	scale	420	624,1839	294,9764	772,0849	93,63651	397,3104
	location	1	1	1	1	1	1
Body	meanLog	2,696994	2,847318	3,117179	2,234512	1,742971	2,405027
	sdLog	1,539553	1,501785	1,585311	1,564687	1,109954	1,653702
EVT BMM	blocks	10	10	10	10	10	10
	shape	0,81227	0,89098	1,7897	1,199	0,90173	1,0738
	scale	234,1988	226,5082	393,0212	159,233	62,67493	194,8082
	location	231,8738	207,7421	367,2308	119,669	64,30349	147,1438
Body	meanLog	2,611534	2,862488	3,34239	2,059458	1,720024	2,356948
	sdLog	1,546646	1,607313	1,75416	1,49099	1,132231	1,694866
GPD	shape	1,4608	1,4194	1,5903	1,6668	1,1769	1,9108
	scale	11,4701	14,9772	25,4317	4,7909	4,377	5,5415

location	1	1	1	1	1	1	
g-and-h	A	14,96	17,795	42,85	7,23	5,52	9,83
	B	9,9443	10,543	32,76	9,1	4,367	20,590
	g	2,2344	2,4318	2,3124	2,4767	2,0702	2,671
	h	0,441	0,338	0,42	0,368	0,285	0,12
alpha stable (mle)	alpha	0,46188	0,44075	0,3673	0,51256	0,61568	0,46663
	beta	0,9999	0,9	0,9999	0,999	0,9999	0,9999
	gamma	1,9	1,9999	1,9999	1,58447	1,5015	1,76801
	delta	0,999	0,9999	0,9999	0,99	0,9999	0,9999
alpha stable (q)	alpha	0,644	0,625	0,777	0,566	0,69	0,553
	beta	0,95	0,95	0,95	0,95	0,95	0,95
	gamma	10,5136	13,6319	40,359	4,2823	3,2908	7,6725
	delta	6,0128	5,8793	14,025	3,112	2,9141	2,3137

In figuur 26, Tabel 18 en Tabel 19 valt te zien dat in het geval van Corporate Finance er meerdere verdelingen zijn die een goede fit hebben. Ook hier blijkt de g-and-h verdeling een goede fit te hebben. Dit valt te zien doordat de lognormale verdeling en beide EVT verdelingen een statistisch goede fit hebben volgens de formele tests en ook op de QQ-plot er goed uitkomen. Verder valt op dat BMM voor BLs een betere fit heeft dan voor ETs en dat de alpha-stable verdeling die geschat is aan de hand van de kwantielen methode ook een betere fit heeft dan voor ETs. Echter komt de alpha-stable verdeling, op één keer na, niet door de formele tests. Ook bij de BLs hebben de GPD en de alpha-stable verdeling geschat aan de hand van MLE geen goede fit. Deze zullen verder dus ook niet beschouwd worden en ook niet meegenomen in de QQ-plots in bijlage 5.2. Hier valt te zien dat evenals uit de formele tests uit de QQ-plots volgt dat voor Retail Banking geen van de verdelingen een goede fit heeft. In de QQ-plots lijkt de alpha-stable soms een goede fit te hebben, bij de formele tests wordt de alpha-stable verdeling alleen bij de BL Corporate Finance niet verworpen te worden.



Figuur 26: QQ-plot met de onderzochte verdelingen tegenover de sample verliezen voor Corporate Finance

Tabel 18: *p-waarden uit de Anderson Darling test voor Business Lines*

Distributions	Asset Mgmt	Commercial Banking	Corporate Finance	Retail Banking	Retail Brokerage	Trading & Sales
Exp	0,00	0,00	0,00	0,00	0,00	0,00
Lognormal	0,02107	0,08488	0,053453	0,00001	0,06897	0,01066
EVT POT	0,15452	0,06642	0,10808	0,00007	0,05855	0,09138
EVT BMM	0,05597	0,13714	0,35809	0,0003	0,03827	0,00825
GPD	0,00	0,00	0,00	0,00	0,0001	0,00
alpha stable (mle)	0,00	0,00	0,00	0,00	0,00	0,00
alpha stable (q)	0,00	0,00	0,06008	0,00	0,00347	0,0001

Tabel 19: *p-waarden uit de Kolmogorov-Smirnov test voor Business Lines*

Distributions	Asset Mgmt	Commercial Banking	Corporate Finance	Retail Banking	Retail Brokerage	Trading & Sales
Exp	0,00	0,00	0,00	0,00	0,00	0,00
Lognormal	0,1079	0,1847	0,17995	0,000028	0,058460	0,06166
EVT POT	0,4862	0,07414	0,0712	0,000926	0,2209	0,1633
EVT BMM	0,1661	0,5003	0,45	0,00498	0,02319	0,0413
GPD	0,00	0,00	0,00	0,00	0,00	0,000298
alpha stable (mle)	0,00	0,00	0,00	0,00	0,00	0,00
alpha stable (q)	0,00	0,00017	0,07503	0,00	0,0002125	0,00

Vervolgens vallen onderstaand in Tabel 20 de schattingen voor BLs te zien. Evenals voor de ETs valt op dat de schattingen voor de exponentiële verdeling en lognormale verdeling vaak aan de lage kant zijn en voor de GPD en alpha-stable verdeling zijn de schattingen opnieuw vaak hoog. Hier komt wel naar voren dat hoewel de alpha-stable verdeling die geschat is aan de hand van de kwantielen methode hoge VaR schattingen heeft, deze wel beter presteert dan de alpha-stable verdeling die geschat is aan de hand van MLE.

Tabel 20: *Value-at-Risk schattingen per Business Line*

Distribution	Asset Management	Commercial Banking	Corporate Finance	Retail Banking	Retail Brokerage	Trading & Sales
Exp	2,25E+04	1,10E+04	9,28E+02	1,42E+04	1,72E+03	7,69E+03
Lognormal	4,08E+04	3,82E+04	1,24E+05	2,36E+04	2,92E+03	4,18E+04
EVT POT (5%)	1,13E+06	2,24E+04	5,97E+06	3,03E+05	9,70E+03	4,66E+04
EVT POT (10%)	2,91E+06	2,96E+04	1,40E+07	5,89E+05	1,70E+04	8,10E+04
EVT BMM (yearly)	6,79E+05	1,09E+5	4,61E+07	4,41E+05	3,65E+05	2,92E+05
EVT BMM (half-y)	1,69E+06	2,20E+05	1,61E+08	1,32E+06	7,07E+05	6,28E+05
g-and-h	2,43E+06	1,53E+05	2,55E+06	3,25E+06	1,18E+05	8,44E+05
GPD	4,50E+07	8,23E+07	7,64E+07	5,98E+08	1,01E+06	4,21E+08

alpha stable (mle)	2,36E+10	8,97E+10	6,74E+11	5,45E+09	6,46E+06	2,95E+10
alpha stable (q)	1,58E+08	3,49E+08	6,54E+06	1,34E+09	3,54E+06	6,67E+08

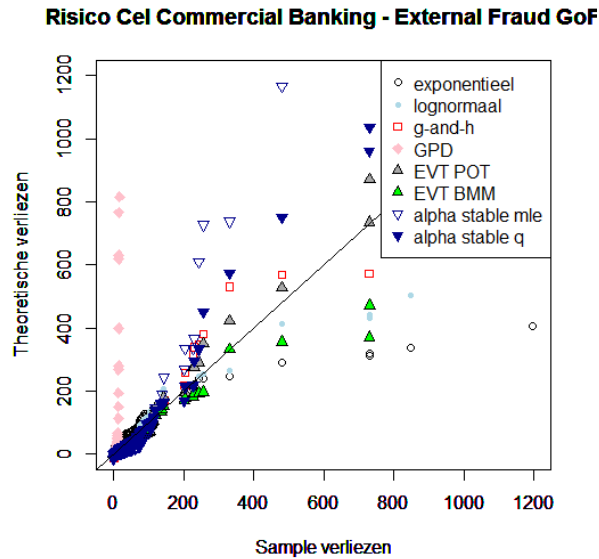
8.2.4 – Analyse Risico Cellen

Ook voor de risico cellen geldt dat het schatten van de parameters op dezelfde manier heeft plaatsgevonden als bij de BLs en de ETs. Wel geldt dat doordat het risico cellen betreft, en dus steeds een selectie uit BLs en ETs betreft, minder data punten beschikbaar. Hierdoor is bijvoorbeeld de drempel voor POT lager dan bij BLs en ETs.

Tabel 21: parameters voor de verdelingen van risico cellen

Distributions	Risico Cellen	Asset Management - Internal Fraud	Commercial Banking - External Fraud	Retail Banking - Clients Products and Business Processes	Retail Banking - Internal Fraud	Trading & Sales - Clients Products and Business Processes
Exp	rate	0,0026725	0,013624	0,0036394	0,010281	0,011312
Lognormal	meanLog	2,9320	2,9793	2,8659	2,3510	2,3970
	sdLog	1,7717	1,5631	2,1050	1,7446	1,8870
EVT POT	threshold	300	60	200	100	40
	shape	1,0543	0,92896	0,6541	0,75677	0,57334
	scale	412,5793	51,72149	880,7244	218,85783	159,94415
	location	1	1	1	1	1
	body	meanLog	2,634743	2,261961	2,292701	1,924515
	sdLog	1,411613	1,047914	1,474269	1,257739	1,063606
EVT BMM	blocks	10	10	10	10	10
	shape	1,1711	0,98441	1,3167	1,0375	1,1481
	scale	269,3533	124,63712	288,47	153,3453	107,7028
	location	216,4982	106,96029	195,5638	132,3661	112,8555
	body	meanLog	2,577255	2,718051	2,507562	1,995192
	sdLog	1,409614	1,372706	1,827044	1,392703	1,640533
GPD	shape	1,3506	0,98717	2,0469	1,5955	1,935
	scale	12,1392	16,88095	6,21	4,7809	3,7402
	location	1	1	1	1	1
g-and-h	A	14,44	16,85	10,27	7,405	7,04
	B	24,693	14,511	18	8,2	14,183
	g	2,1504	2,1033	3,0222	2,2953	2,6981
	h	0,2	0,1699	0,301	0,51	0,3
alpha stable (mle)	alpha	0,47873	0,48141	0,44964	0,52299	0,46657
	beta	0,9999	0,9999	0,9999	0,9999	0,9999
	gamma	1,9999	1,9999	1,98538	1,57489	1,21019
	delta	0,9999	0,9999	0,9999	0,9999	0,9999
Alpha stable (q)	alpha	0,64	0,799	0,51	0,613	0,535
	beta	0,95	0,95	0,95	0,95	0,95
	gamma	8,4109	14,4993	6,0173	5,0711	4,2672
	delta	7,2097	6,4935	3,091	2,8517	2,6813

In figuur 27 is een QQ-plot te zien waar de onderzochte verdelingen zijn geplote tegen de risico cel Commercial Banking – External Fraud. De lognormale verdeling heeft een redelijke fit en uit de formele test blijkt ook dat voor deze verdeling de nulhypothese niet verworpen wordt. Hetzelfde geldt voor beide EVT methodes. Voor de alpha-stable verdeling geldt dat de Anderson Darling test de nulhypothese niet verwerpt en de Kolmogorov-Smirnov test wel. De g-and-h verdeling lijkt hier een goede fit te hebben. De GPD en de exponentiële verdeling hebben evenals de alpha-stable verdeling die is geschat aan de hand van MLE geen goede fit.



Figuur 27: QQ-plot met de onderzochte verdelingen tegenover de sample verliezen voor Corporate Finance

Tabel 22: p-waarden uit de Anderson Darling test voor risico cellen

Distributions	Asset Management - Internal Fraud	Commercial Banking - External Fraud	Retail Banking - Clients Products and Business Processes	Retail Banking - Internal Fraud	Trading & Sales - Clients Products and Business Processes
Exp	0,00	0,00	0,00	0,00	0,00
Lognormal	0,06086	0,50718	0,0427	0,02449	0,03505
EVT POT	0,47157	0,149	0,0951	0,1145	0,05038
EVT BMM	0,28582	0,18307	0,05321	0,08637	0,04205
GPD	0,00	0,00	0,00033	0,00283	0,0011
alpha stable (mle)	0,00	0,00	0,00	0,00	0,00
alpha stable (q)	0,1781	0,0579	0,00986	0,00048	0,0135

Tabel 23: *p-waarden uit de Kolmogorov-Smirnov test voor risico cellen*

Distributions	Asset Management - Internal Fraud	Commercial Banking - External Fraud	Retail Banking - Clients Products and Business Processes	Retail Banking - Internal Fraud	Trading & Sales - Clients Products and Business Processes
Exp	0,00	0,00	0,00	0,00	0,00
Lognormal	0,07417	0,2978	0,04564	0,06444	0,03253
EVT POT	0,2101	0,1099	0,1197	0,359	0,129
EVT BMM	0,06314	0,3688	0,0347	0,3598	0,1298
GPD	0,00	0,00	0,0002	0,0352	0,00
alpha stable (mle)	0,00	0,00	0,0006	0,00	0,00
alpha stable (q)	0,161	0,04184	0,006724	0,00002	0,0005

In bijlage 5.3 vallen de QQ-plots voor de risico cellen te zien. Opnieuw lijkt de alpha-stable (q) verdeling acceptabel te presteren, wat tegengesproken wordt door de formele tests. De g-and-h verdeling heeft alleen bij Retail Banking- Internal Fraud geen goede fit. In bovenstaande tabellen valt ook te zien dat EVT slechts twee keer verworpen wordt bij de formele tests, beiden BMM. De Anderson Darling verwerpt de nulhypothese bij Trading & Sales – CPBP en de Kolmogorov-Smirnov test verwerpt de nulhypothese bij Retail Banking – CPBP. In Tabel 24 zijn de VaR schattingen per risico cel weergegeven.

Tabel 24: *Value-at-Risk schattingen per risico cel*

Distribution	AMIF	CBEF	RBCPBP	RBIF	TSCPBP
Exp	2,25E+04	5,55E+03	1,07E+04	1,03E+04	5,10E+03
Lognormal	3,68E+03	3,96E+03	4,34E+03	3,53E+03	3,74E+03
EVT POT (5%)	551537	3,30E+04	1,29E+05	4,62E+04	1,31E+04
EVT POT (10%)	1,09E+06	7,15E+04	2,09E+05	1,04E+05	2,23E+04
EVT BMM (yearly)	7,92E+05	1,17E+05	1,49E+06	1,88E+05	2,70E+05
EVT BMM (half-y)	1,80E+06	1,98E+05	5,69E+06	3,93E+05	4,52E+05
g-and-h	3,31E+05	1,46E+05	6,05E+06	4,03E+06	2,40E+05
GPD	8,18E+06	6,30E+05	1,40E+09	1,50E+08	7,29E+08
alpha stable (mle)	6,24E+09	5,24E+09	2,48E+09	2,27E+09	2,72E+09
alpha stable (q)	4,19E+07	5,63E+06	6,79E+08	2,34E+08	6,09E+08

8.2.5 – Analyse Data Kwaliteit

Accuraatheid

Zoals omschreven in hoofdstuk 7.2.3, wordt bij accuraatheid onderzocht wat het gevolg is van afronden in de dataset. Niet alleen hanteren banken verschillende drempels voor het meenemen van datapunten in de berekening van VaR, ook hanteren banken verschillende niveaus van accuraatheid bij de berekening. In Tabel 25 valt te zien wat de verhoudingen zijn tussen de originele VaR schatting en de schatting wanneer er is afgerond op honderdduizend dollar of op een miljoen dollar, terwijl er in de oorspronkelijke dataset op tienduizend dollar wordt afgerond. In de tabel wordt dit aangegeven met 1^e5 voor honderdduizend dollar of 1^e6 voor afronden op miljoenen dollars.

Tabel 25: verhoudingen originele Value-at-Risk ten opzichte van de nieuwe schatting voor accuraatheid

Afrond		Exp	Log-normal	EVT POT (5%)	EVT POT (10%)	EVT BMM (yearly)	EVT BMM (half-y)	g-and-h	GPD	alpha stable (q)	
ET	CPBP	1e5	1,00	0,98	1,21	1,47	1,67	2,33	3,81	0,04	0,001
		1e6	1,00	0,96	1,27	1,65	0,41	2,48	3,79	0,04	2,48
	EPWS	1e5	1,02	1,01	0,99	1,01	0,41	10,68	2,14	1,08	4,14
		1e6	1,01	1,06	1,01	1,00	3,28	5,16	1,17	1,03	3,73
	EDPM	1e5	1,00	1,01	0,82	0,57	0,55	0,67	0,94	0,02	0,05
		1e6	1,00	0,96	0,86	0,56	0,45	0,47	0,51	0,02	0,07
	Ext. Fraud	1e5	1,00	0,96	0,86	1,11	1,85	3,62	0,73	0,14	0,02
		1e6	1,01	1,00	1,00	1,22	1,55	0,14	0,73	0,15	0,01
	Int. Fraud	1e5	0,99	0,98	0,70	0,77	0,78	1,03	2,82	3,74	0,03
		1e6	1,00	1,06	0,79	0,73	0,64	0,69	2,24	3,65	0,02
BL	AM	1e5	0,98	0,95	1,07	0,87	12,26	1,55	0,83	13,39	0,68
		1e6	0,98	1,03	1,32	0,90	0,92	0,62	1,03	24,03	0,85
	CB	1e5	1,00	1,13	0,94	0,97	69,84	9,57	0,98	0,002	0,78
		1e6	1,00	1,16	1,04	1,01	1,53	1,58	1,37	0,002	0,770
	CF	1e5	1,01	0,88	0,27	0,31	0,92	1,26	1,22	0,13	1,163
		1e6	1,02	0,97	0,24	0,25	0,10	0,06	1,19	0,12	1,04
	Rba	1e5	1,00	1,05	1,35	1,28	5,29	5,34	0,44	0,002	0,02
		1e6	1,00	1,07	1,23	1,46	1,81	1,27	0,54	0,02	0,06
	Rbr	1e5	1,01	1,03	0,85	0,90	0,35	0,39	0,97	20,90	6,79
		1e6	1,02	1,11	0,91	1,03	0,12	0,10	3,66	2,39	0,69
	TS	1e5	0,99	0,87	0,89	0,96	3,52	6,20	1,20	0,24	0,77
		1e6	0,97	0,89	0,89	1,03	0,88	0,89	1,43	0,01	1,05
Risico Cel	AMIF	1e5	0,82	1,02	1,00	1,17	3,67	4,40	1,55	20,90	0,96
		1e6	0,83	1,00	0,92	1,08	0,99	1,23	1,48	20,78	1,19
	CBEF	1e5	0,48	1,05	0,93	0,81	0,69	0,78	1,46	0,05	0,52
		1e6	0,48	1,02	1,08	0,86	0,63	0,79	1,36	0,05	0,51

RBCP BP	1e5	1,21	1,08	0,93	0,84	1,77	13,48	1,30	0,01	0,84
	1e6	1,21	1,01	0,95	0,88	5,06	3,49	1,34	0,01	0,94
RBIF	1e5	0,46	0,90	1,17	0,88	1,56	0,58	1,15	0,58	0,14
	1e6	0,46	0,98	1,11	0,89	2,00	2,78	1,12	2,78	0,17
TSCPB P	1e5	0,69	1,10	1,23	0,95	1,73	0,34	0,98	0,001	0,58
	1e6	0,67	1,01	1,05	0,97	0,63	0,77	1,02	0,001	0,64
Gemiddelde		0,92	1,01	0,97	0,95	4,00	2,65	1,45	3,63	0,99
Maximum		1,21	1,16	1,35	1,65	69,84	13,48	3,81	24,03	6,79
Minimum		0,46	0,87	0,24	0,25	0,10	0,06	0,44	0,0007	0,0010

Een aantal zaken vallen op in Tabel 25. Ten eerste geldt dat voor de exponentiële verdeling, de lognormale verdeling, EVT-POT en de alpha-stable verdeling er gemiddeld gezien weinig nauwkeurigheid verloren gaat. Voor de lognormale verdeling geldt dat bijna overal de waarden rond de 1 liggen, dus voor veel BLs, ETs en risico cellen blijft de nieuwe waarde dicht bij de oorspronkelijke schatting. Voor EVT-POT geldt dat er naar boven en beneden enige uitschieters zijn, maar over het algemeen er geen grote veranderingen plaatsvinden ten opzichte van de oorspronkelijke schattingen. Dit geldt niet voor EVT-BMM, die voor Commercial Banking een schatting geeft die bijna 70 maal groter is dan de oorspronkelijke VaR. Dit komt in het geval van BMM doordat voor Commercial Banking de grootste metingen in de maxima naar boven worden afgerond en de kleinere metingen naar beneden, waardoor de toch al grote schatter voor scale nog extremer is geworden. Opvallend is ook dat de g-and-h verdeling in dit geval vaak een grotere schatting doet dan de oorspronkelijke VaR. Dit komt mede doordat de h parameter vaak hoger wordt geschat dan oorspronkelijk het geval was.

Compleetheid

Zoals omschreven in hoofdstuk 7.2.3, wordt bij compleetheid onderzocht wat het gevolg is van missende waarden in de dataset. Het is niet mogelijk om voor een willekeurige verliesdatabase aan te geven dat 5% of 10% van de waarden niet geregistreerd zijn, maar het is wel mogelijk om aan te geven wat de gevolgen zijn van 5% missende waarden. In Tabel 26 valt te zien wat de verhoudingen zijn tussen de VaR schattingen van de originele data en data waarbij ofwel 5% willekeurige waarden ontbreken ofwel 5% hoogste waarden.

Tabel 26: verhoudingen originele Value-at-Risk ten opzichte van de nieuwe schatting voor compleetheid

Missen van waarden			Exp	Lognormal	EVT POT (5%)	EVT POT (10%)	EVT BMM (yearly)	EVT BMM (half-y)	g-and-h	GPD	alpha stable (q)
ET	CPBP	5% willekeurig	0,95	0,91	1,17	1,76	3,17	4,55	1,13	0,03	0,001
		5% hoogste	0,36	0,32	0,58	1,27	0,76	0,86	0,81	0,00003	0,00044
EPWS		5% willekeurig	1,02	0,96	1,48	1,55	30,08	61,32	3,59	0,13	0,16
		5% hoogste	0,60	0,43	1,36	1,32	6,00	12,54	1,20	0,01	0,03
EDPM		5% willekeurig	0,99	1,11	1,33	0,88	3,31	3,15	0,67	0,02	0,02
		5% hoogste	0,21	0,28	0,05	0,24	0,55	0,53	0,40	0,02	0,005
Ext. Fraud		5% willekeurig	0,92	0,84	1,66	2,04	0,31	0,52	0,20	0,81	0,01

		5% hoogste	0,34	0,31	0,08	0,01	0,09	0,12	0,12	0,34	0,001	
	Int. Fraud	5% willekeurig	0,98	0,97	1,62	0,76	1,34	2,33	0,64	0,05	0,02	
		5% hoogste	0,21	0,28	0,003	0,002	0,370	0,338	0,54	0,006	0,0014	
BL	AM	5% willekeurig	0,98	0,99	1,34	1,16	1,13	1,60	0,45	0,29	0,001	
		5% hoogste	0,18	0,28	0,00	0,67	0,29	0,39	0,16	0,19	0,0004	
	CB	5% willekeurig	1,00	1,15	1,04	0,99	8,55	10,3	1,48	0,002	0,002	
		5% hoogste	0,41	0,37	0,10	0,10	1,33	1,93	0,97	0,0001	0,0002	
	CF	5% willekeurig	1,01	0,77	0,49	0,60	0,11	0,14	0,79	0,11	0,00001	
		5% hoogste	0,21	0,27	0,43	0,24	0,04	0,03	0,54	0,01	0,000002	
	Rba	5% willekeurig	0,89	0,95	0,49	0,50	1,18	1,91	0,16	0,006	0,118	
		5% hoogste	0,20	0,24	0,004	0,003	0,39	0,27	0,09	0,00002	0,00353	
	Rbr	5% willekeurig	1,01	1,02	1,13	0,86	0,13	0,18	0,82	2,21	0,09	
		5% hoogste	0,45	0,41	0,31	0,02	0,03	0,04	0,29	0,005	0,062	
	TS	5% willekeurig	0,80	0,78	0,24	0,17	0,55	0,86	0,98	0,0004	0,020	
		5% hoogste	0,38	0,26	0,06	0,04	0,18	0,21	0,17	0,0039	0,0002	
	Risico Cel	AMIF	5% willekeurig	0,83	0,84	1,48	1,21	0,22	1,21	0,72	0,76	0,0035
		5% hoogste	0,10	0,28	0,19	1,16	0,06	0,06	0,37	0,01	0,0004	
CBEF		5% willekeurig	0,46	1,09	0,93	0,92	0,93	1,28	0,49	0,35	0,00027	
		5% hoogste	0,23	0,44	0,15	0,08	0,15	0,21	0,28	0,83	0,00004	
RBCP BP		5% willekeurig	1,21	1,14	1,07	0,99	5,15	0,04	1,30	0,53	0,073	
		5% hoogste	0,29	0,20	0,08	0,07	0,47	0,05	0,28	0,00	0,029	
RBIF		5% willekeurig	0,45	0,92	1,18	0,85	2,13	3,94	0,25	0,01	0,022	
		5% hoogste	0,11	0,26	0,14	0,08	0,58	1,06	0,12	0,00	0,001	
TSCPBP		5% willekeurig	0,67	0,92	1,15	1,10	3,56	1,10	1,28	0,00	0,073	
		5% hoogste	0,29	0,31	0,23	0,18	0,91	2,45	0,11	0,00	0,004	
		Gemiddelde		0,59	0,63	0,67	0,68	2,31	3,61	0,67	0,21	0,02
		Maximum		1,21	1,15	1,66	2,04	30,08	61,32	3,59	24,03	0,16
		Minimum		0,10	0,20	0,0012	0,0021	0,03	0,03	0,09	0,00002	0,000002

De verwachting bij het weghalen van metingen is, dat over het algemeen de nieuwe schattingen lager zullen liggen, omdat er schattingen worden gedaan met minder metingen. Dit zou zeker het geval moeten zijn bij het weghalen van de hoogste metingen. In dat opzicht is het dus opvallend dat dit niet geldt voor EVT-BMM. Voor deze verdeling liggen de waarden een aantal keer flink hoger dan de oorspronkelijke schattingen. De gemiddelden geven hier een enigszins vertekend beeld door de waarden die 30 en 61 keer groter zijn dan de

oorspronkelijke waarde, maar ook dit geeft aan dat EVT-BMM in dit geval geen betrouwbare schattingen geeft. De lognormale verdeling, EVT-POT en de g-and-h verdeling voldoen wel aan de verwachting.

Punctualiteit

Voor punctualiteit is onderzocht wat het effect op de VaR schatting is wanneer niet de huidige waarde van verliezen worden gebruikt voor de berekening, maar in plaats daarvan de verlieswaarde (dus de waarde van het verlies op het moment van plaatsvinden). In Tabel 27 zijn de verhoudingen tussen de oorspronkelijke VaR schatting en de schatting aan de hand van “gedateerde” verlieswaarden.

Tabel 27: *verhoudingen originele Value-at-Risk ten opzichte van de nieuwe schatting voor punctualiteit*

Verlieswaarde		Exp	Log-normal	EVT POT (5%)	EVT POT (10%)	EVT BMM (yearly)	EVT BMM (half-y)	g-and-h	GPD	alpha stable (q)
ET	CPBP	0,92	0,94	0,85	0,92	0,44	0,23	1,13	0,028	0,15926
	EPWS	8,71	0,84	1,09	1,20	3,25	2,03	0,29	0,016	0,12
	EDPM	0,92	0,89	0,69	0,34	0,45	0,43	0,31	0,02	0,57
	External Fraud	0,95	0,99	0,98	1,72	0,16	0,19	0,17	0,18	1,89
	Internal Fraud	0,94	0,87	0,45	0,46	0,73	0,67	0,28	0,0032	0,2213
BL	Asset Management	0,94	0,96	0,79	0,72	0,10	0,08	0,12	1,46	0,71
	Commercial Banking	0,92	1,16	0,86	0,84	1,84	1,72	3,49	0,0021	0,62
	Corporate Finance	0,99	0,67	0,02	0,01	0,03	0,67	0,03	0,0011	0,86
	Retail Banking	0,93	1,04	1,34	1,52	1,71	1,31	0,36	0,01	0,34
	Retail Brokerage	0,94	0,96	0,57	0,56	0,96	0,15	0,19	0,01	0,0015
	Trading & Sales	0,92	0,84	0,27	0,26	0,84	1,06	0,84	0,02	0,16
Risico Cel	AMIF	0,79	0,91	0,88	0,91	0,34	0,46	0,37	10,12	0,99
	CBEF	0,44	0,99	0,87	0,81	0,79	1,46	0,41	2,16	0,1675
	RBCPBP	1,14	0,92	0,41	0,73	0,40	0,59	0,18	0,0001	0,15
	RBIF	0,42	0,80	0,85	0,68	3,56	7,15	0,04	0,01	0,17
	TSCPBP	0,61	0,89	0,96	0,82	1,64	3,68	1,45	0,00004	0,62
	Gemiddelde	1,34	0,92	0,74	0,78	1,08	1,37	0,60	0,88	0,48
	Maximum	8,71	1,16	1,34	1,72	3,56	7,15	3,49	10,12	1,89
	Minimum	0,42	0,67	0,02	0,01	0,03	0,08	0,03	0,00004	0,0015

Voor punctualiteit zou evenals voor compleetheid moeten gelden dat de nieuwe schattingen vaak lager liggen dan de oorspronkelijke schattingen, doordat de verlieswaarden op het moment van verlies kleiner zijn dan de huidige waarde van verlies. Opnieuw geldt dit niet voor EVT-BMM, maar wordt dit vooral veroorzaakt door uitschieters bij Employment Practices and Workplace Safety en de risico cel Retail Banking – Internal Fraud. Ook voor punctualiteit geldt dat EVT-POT minder variatie laat zien dan de g-and-h verdeling.

9 – Conclusie en discussie

In het begin van deze scriptie werden de onderzoeksvragen en daarmee doelen voor het onderzoek gesteld. Door de scriptie heen is achtergrondinformatie gegeven, is stilgestaan bij verschillende gebruikte verdelingen, berekeningsmethodes en bruikbare dimensies van data kwaliteit voordat uiteindelijk de analyse is uitgevoerd. In dit hoofdstuk zal worden stilgestaan bij wat de resultaten zijn en wat dat inhoudt. Ook zullen enkele laatste opmerkingen gemaakt worden en aanbevingen worden gedaan.

9.1 – Conclusie

Het doel van dit onderzoek was het onderzoeken van de verschillende zwaarteverdelingen die in literatuur genoemd zijn en het onderzoeken van de gevolgen van data kwaliteit op economisch kapitaal voor OR. Aan het begin van dit document zijn twee onderzoeksvragen gesteld:

- *Wat is de best passende verdeling voor zwaarte van verliesincidenten in verliesdatabases en wat is de invloed hiervan op het economisch kapitaal volgens Basel regelgeving?*
- *Wat is de invloed van variërende data kwaliteit op het economisch kapitaal in verliesdatabases?*

Deze vragen zullen in respectievelijk paragrafen 9.1.1 en 9.1.2 beantwoord worden.

9.1.1 – Zwaarteverdelingen

In het onderzoek is naar voren gekomen dat EVT (5%), g-and-h en de lognormale verdeling volgens formele tests goed presteren. Volgens de Anderson Darling test heeft de lognormale verdeling in acht van de zestien gevallen een goede fit, EVT-POT in dertien gevallen, EVT-BMM in acht en de alpha-stable verdeling in twee van de zestien gevallen. Uit de Kolmogorov-Smirnov test komt naar voren dat de lognormale verdeling in tien gevallen goed fit, EVT-POT dertien keer, EVT-BMM acht keer en de alpha-stable verdeling twee van de zestien keer. De exponentiële verdeling en de GPD hebben in geen van de gevallen een goede fit. Uit de QQ-plots blijkt dat de g-and-h verdeling in veertien van de zestien gevallen een goede fit heeft, mede afgaande op de fit van de andere verdelingen. Uit de QQ-plots en uit de schattingen blijkt dat de exponentiële verdeling en de lognormale verdeling lage VaR schattingen hebben en de GPD en alpha-stable verdeling te hoge schattingen hebben. Ook blijkt dat EVT-BMM niet consistent is met goede schattingen, dat wil zeggen dat de schattingen soms vergelijkbaar zijn met EVT-POT en g-and-h verdeling en soms aan de hoge kant. EVT-POT en de g-and-h verdelingen hebben in de meeste gevallen een goede statistische fit en zijn dus ook geschikt om mee te modelleren.

9.1.2 – Data Kwaliteit

In documentatie van BIS wordt aangegeven dat financiële instellingen verplicht zijn om “de juiste controle te verzorgen rondom de database en diens kwaliteit” [28]. Het lijkt intuïtief om te stellen dat slechte data kwaliteit in de verliesdatabase kan leiden tot onnauwkeurige schattingen voor Value-at-Risk. Om dit verder te onderzoeken en tot onderbouwde uitspraken te komen is onderzocht wat het effect is van variërende data kwaliteit. Dit in de vorm van verschillen in afronden, waar is gekeken wat de verschillen in VaR zijn op het moment dat niet op tienduizend dollar, maar op honderdduizend of miljoen dollar wordt afgerond. Ook is gekeken naar wat er gebeurt wanneer er 5% van de metingen niet aanwezig zijn bij de berekening. Er is gekeken naar het missen van 5% willekeurige metingen en 5% hoogste metingen. Als laatste is gekeken naar

het gebruiken van data die niet up-to-date is. Dit wil zeggen dat in plaats van de actuele waarde van verlies, voor de berekening gebruik is gemaakt van de waarde op het moment van verlies. Wanneer in deze context gesproken wordt over een verdeling die goed presteert, wordt bedoeld dat de nieuwe VaR schatting dichtbij de oorspronkelijke schatting ligt.

Na deze analyse kan ten eerste gesteld worden dat EVT-POT, de g-and-h en de lognormale verdeling minder variatie ervaren bij afwisselende data kwaliteit dan de andere onderzochte verdelingen. Daarnaast komt ook naar voren dat variërende data kwaliteit grote schommelingen teweeg kan brengen in de schattingen van Value-at-Risk. Wanneer gesteld wordt dat EVT-POT en g-and-h minder variatie ervaren, betekent dit alsnog grote verschillen.

In het geval van punctualiteit is deze schommeling logisch en wordt een denkfout gemaakt als inderdaad de waarde op het moment van verlies wordt gebruikt in plaats van de actuele waarde. Dat de nieuwe waarden hier 60 tot 74% lager liggen mag dus logisch worden genoemd. Hierdoor komt wel naar voren dat de noodzaak op goede controle – zowel van de verliesdatabase als van de database die gebruikt wordt voor de berekening – groot is. In geval van accuraatheid presteert EVT-POT (5%) het best. Ook de lognormale verdeling presteert goed, maar zoals eerder gesteld zijn de schattingen van de lognormale verdeling aan de lage kant. De g-and-h verdeling presteert hier minder goed dan EVT-POT, gezien de gemiddelde verhouding van bijna 1.5, dus anderhalf keer de oorspronkelijke schatting. Het BIS geeft aan dat financiële instellingen onderbouwing moeten geven voor accuraatheid in de berekening. Dit onderzoek ondersteunt deze stelling. Zelfs al zijn de verschillen minder groot dan bij punctualiteit, in een aantal gevallen komen stijgingen van 30% of dalingen van 80% voor. Wat betreft compleetheid is de verwachting dat bij het weghalen van 5% willekeurige metingen geen enorme schommelingen zullen plaatsvinden, maar in sommige gevallen wel. Dit komt doordat ook bij willekeurige metingen grote waarden kunnen wegvallen of juist lage waarden weg kunnen vallen of de variatie groter kan worden. De verwachting bij het weghalen van de 5% grootste metingen is dat de nieuwe schattingen een stuk lager zullen liggen. Deze verwachtingen blijken inderdaad te kloppen. Voor EVT-POT en g-and-h blijkt dat bij het weghalen van de hoogste waarden vaak er maar 20% van de oorspronkelijke VaR, soms zelfs nog minder, overblijft. Opvallend is dat EVT-BMM regelmatig in beide gevallen grote schattingen krijgt dan oorspronkelijk.

Uiteindelijk kan geconcludeerd worden dat hoewel zowel EVT-POT (5%) en de g-and-h verdelingen een goede fit hebben, EVT-POT beter presteert op het moment data kwaliteit varieert. Dit verschilt uiteraard per BL, ET en risico cel en per situatie kan een andere verdeling gebruikt worden voor het bepalen van economisch kapitaal. Daarnaast wordt duidelijk dat variërende data kwaliteit geen enorme gevolgen hoeft te hebben op de VaR schatting, maar dit vaak wel het geval is.

9.2 – Verdere opmerkingen en aanbevelingen

Op de conclusie sluit de aanbeveling aan dat controle op data kwaliteit ontzettend belangrijk is, omdat de gevolgen op economisch kapitaal enorm kunnen zijn. Een belangrijke aanbeveling is dan ook dat de controls rondom invoeren en doorvoeren naar de berekeningsdatabase goed moeten zijn. Een volgende aanbeveling is dat bij het kiezen van een verdeling rekening gehouden moet worden met niet-optimale data kwaliteit. Er zou dus niet alleen gekeken moeten worden naar de verdeling met de beste fit en de meest realistische schatting, maar ook naar hoe een verdeling reageert met variërende data kwaliteit.

Ten slotte moet gesteld worden dat het resultaat betreffende verdelingen voor deze dataset niet voor elke dataset geldt. In deze dataset zijn alle waarden groter dan één miljoen dollar, nauwkeurig tot op tienduizendtallen. Wanneer er een lagere drempel gebruikt wordt of een andere nauwkeurigheid kunnen de waarden anders zijn. Daarnaast zijn, mede door de aard van de dataset, relatief veel extreme waarden aanwezig en zullen ook deze een grote invloed hebben op de data.

Referenties

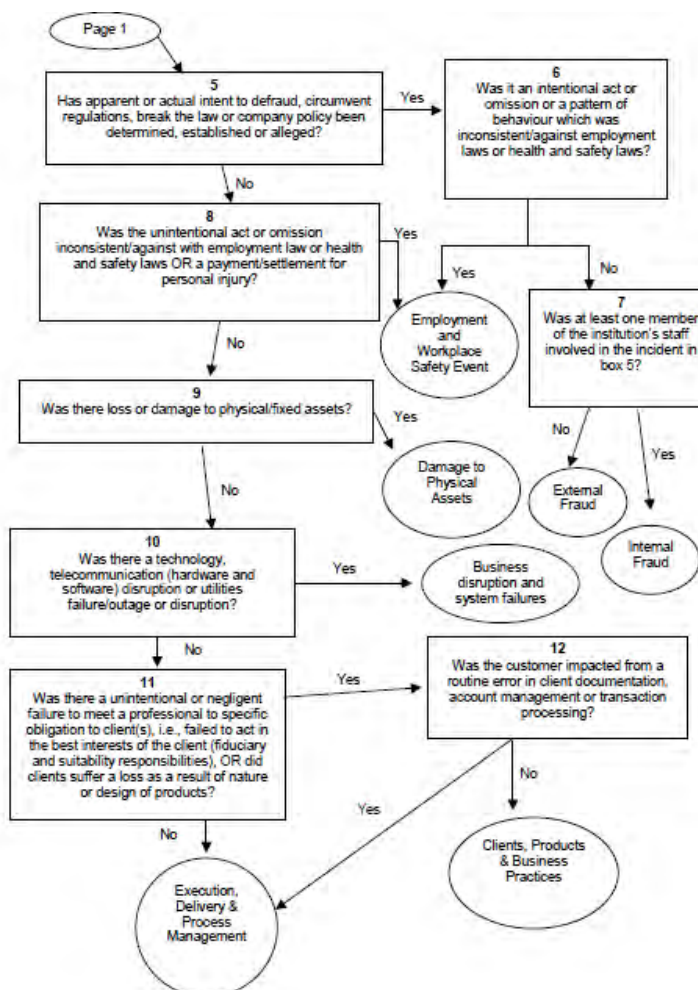
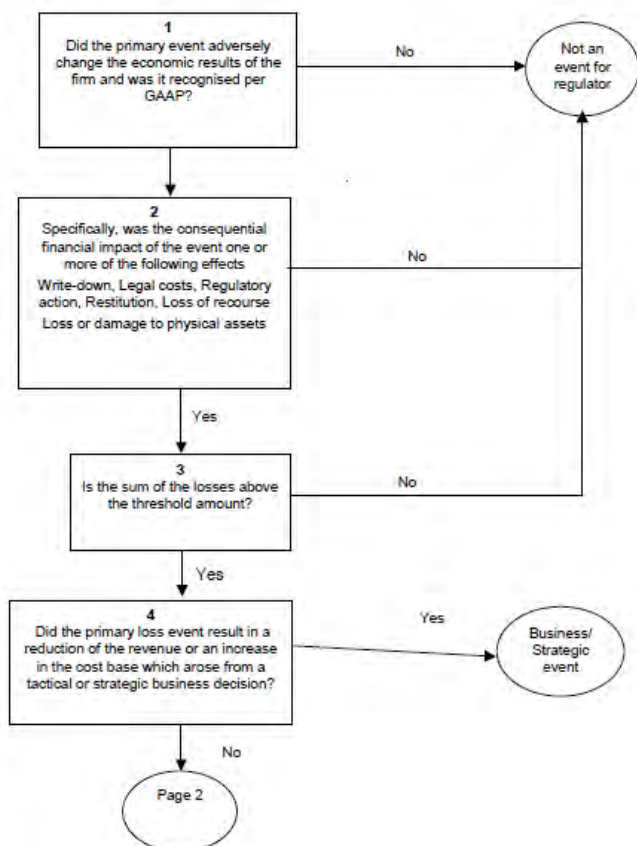
1. Directive 2006/48/EC of the European Parliament and of the Council (p. 232 – 234)
2. BIS (2001): BCBS Consultative Document on Operational Risk
3. BIS (2005): BCBS International Convergence of Capital Measurement and Capital Standards
4. BIS (2009): BCBS Results from the 2008 Loss Data Collection Exercise for Operational Risk
5. K. Böcker & C. Klüppelberg (2010): Multivariate models for Operational Risk
6. A.S. Chernobai, S.T. Rachev (2006): Applying robust models to operational risk modeling
7. A.S. Chernobai, S.T. Rachev & F.J. Fabozzi (2007): Operational Risk: A Guide to Basel II Capital Requirements, Models and Analysis
8. K. Dutta & J. Perry (2007): A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital
9. C. Field & M.G. Genton (2006): The Multivariate g-and-h Distribution
10. M. Rippel & P. Teplý (2011): Operational Risk – Scenario Analysis
11. Jaarverslag Rabobank 2010: Capital Adequacy and Risk Management Report 2010
12. Jaarverslag Van Lanschot Bankiers 2010: Risicobeheer (p. 84-88)
13. Jaarverslag ING 2010
14. Jaarverslag NIBC 2010
15. ISO 31000/ISO 31010 (2009)
16. ISO Guide 73: “Risk Management Vocabulary”
17. A. Resti & A. Sironi (2007): Risk Management and Shareholders’ Value in Banking
18. R. Chalupka & P. Teplý (2008): Operational Risk Management and Implications for Banks’ Economic Capital
19. RMS Special Report (2005): Estimating Losses from the 2004 Southeast Asia Earthquake and Tsunami
20. KPMG whitepaper (2012): Nieuwe perspectieven vragen om actie: een genuanceerde visie op cybercrime
21. BCBS (2002): Operational Risk Data Exercise
22. M. Cruz (2002): Modeling, Measuring and Hedging Operational Risk
23. A. Chappelle, Y. Crama, G. Hübner and J. Peters (2005): Measuring and Managing Operational Risk in the Financial Sector: An Integrated Framework, Technical Report
24. A. Chernobai, C. Menn, S. Rachev, S. Trück (2005): Estimation of Operational Value-at-Risk in the Presence of Minimum Collection Thresholds
25. M. Degen, P. Embrechts & D. Lambrigger (2007): The Quantitative Modeling of Operational Risk: Between g-and-h and EVT
26. D.C. Hoaglin (1985): Summarizing Shape Numerically: The g-and-h distributions.

27. A.S. Chernobai, S.T. Rachev & F.J. Fabozzi (2005): Composites goodness-of-fit tests for left-truncated loss samples
28. BIS (2011): BCBS Operational Risk – Supervisory guidelines for the Advanced Measurement Approaches
29. Bank of England Monetary and Financial Statistics Division (2008) – Data Quality Framework
30. Operational Risk data Exchange Association (2011): Operational Risk Reporting Standards
31. SAP Business Objects (2008): Data Quality Strategy: A Step-By-Step Approach
32. A. Kabašinskas, S. Rachev, L. Sakalauskas, W Sun & I. Belovas (2008): Alpha-Stable Paradigm in Financial Markets
33. J. H. McCulloch (1986) – Simple Consistent Estimators for the Stable Distribution

Bijlagen

Bijlage 1: Classificatie van Event Type

Table 2: Decision trees to determine event categorisation



Bijlage 2: Heat maps van ORX

Onderstaand zijn gegevens afkomstig uit het ORX Operational Risk Report 2010. Het ORX heeft loss data van meer dan zestig verschillende banken. In totaal hebben zij 161.997 loss events van samen € 55 miljard. Onderstaande gegevens zijn een voorbeeld van loss events ingedeeld per Business Line en Event Type. In tabel 3B wordt beschreven welke BL-ET combinaties het meest voorkomen en in tabel 4B wordt beschreven welke combinaties relatief de grootste verliezen tot gevolg hebben. In tabel 5 wordt daarna nog beschreven welke BLs de grootste gemiddelde verliezen hebben per loss event. In deze tabellen staan enkele BLs die niet worden beschreven in het verslag (tabel 4, blz. 14), zoals Clearing, Corporate Items, Private Banking en Multiple Lines. Payment & Settlement wordt niet genoemd.

Table 3b: Distribution of Frequency of Loss by Business Line by Event Type 2009Q1-4

Nos of Events	Internal Fraud	External Fraud	Employment Practices & Workplace Safety	Clients Products & Business Practices	Disasters & Public Safety	Technology & Infrastructure Failures	Execution, Delivery & Process Management	Malicious Damage	Total	% of Total
Corporate Finance	1	25	18	55	14	0	56	0	169	0.62%
Trading & Sales	11	142	105	135	3	114	1,874	0	2,384	8.81%
Retail Banking	692	7,233	2,305	1,672	203	170	3,764	56	16,185	59.83%
Commercial Banking	26	810	86	541	18	32	1,147	0	2,660	9.83%
Clearing	5	186	7	10	0	23	274	0	505	1.87%
Agency Services	10	231	13	60	1	13	1,158	0	1,486	5.49%
Asset Management	7	7	23	124	2	17	506	0	686	2.54%
Retail Brokerage	42	57	96	546	1	12	360	0	1,114	4.12%
Private Banking	28	84	30	423	1	26	668	0	1,260	4.66%
Corporate Items	5	36	229	53	29	36	149	1	538	1.99%
Multiple Lines	4	16	7	16	1	5	17	0	66	0.24%
Total	831	8,827	3,009	3,635	273	448	9,973	57	27,053	
% of Total	3.07%	32.63%	11.12%	13.44%	1.01%	1.66%	36.86%	0.21%		

1% - <3%

3% - 10%

> 10%

Table 4b: Distribution of Gross of Loss by Business Line by Event Type 2009Q1-4

E Mn	Internal Fraud	External Fraud	Employment Practices & Workplace Safety	Clients Products & Business Practices	Disasters & Public Safety	Technology & Infrastructure Failures	Execution, Delivery & Process Management	Malicious Damage	Total	% of Total
Corporate Finance	0	540	4	369	1	0	228	0	1,141	12.53%
Trading & Sales	119	772	39	341	1	17	1,115	0	2,404	26.38%
Retail Banking	99	672	224	442	19	22	624	0	2,105	23.11%
Commercial Banking	46	365	17	142	3	114	303	0	990	10.87%
Clearing	1	20	1	264	0	4	32	0	323	3.54%
Agency Services	14	9	1	22	0	1	220	0	267	2.93%
Asset Management	5	1	4	117	18	2	111	0	256	2.81%
Retail Brokerage	21	7	18	74	0	0	32	0	152	1.67%
Private Banking	28	185	7	516	0	1	63	0	801	8.79%
Corporate Items	0	7	43	37	2	4	107	0	201	2.21%
Multiple Lines	0	343	14	107	0	1	5	0	470	5.16%
Total	334	2,921	371	2,430	44	166	2,839	5	9,110	
% of Total	3.67%	32.06%	4.07%	26.68%	0.48%	1.83%	31.16%	0.06%		

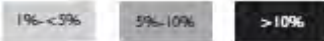


Table 5a: Business Line Risk Ranking 2004-2008

Business Line	Average Loss per Event		Business Line	Gross Loss/100 Gross Income	
	€	Ranking		€	Ranking
Corporate Finance	2,348,769	1	Trading & Sales	3.07	1
Trading & Sales	634,245	2	Corporate Finance	1.99	2
Asset Management	512,597	3	Private Banking	1.94	3
Commercial Banking	461,798	4	Agency Services	1.86	4
Private Banking	333,263	5	Asset Management	1.52	5
Clearing	256,162	6	Commercial Banking	1.25	6
Agency Services	224,643	7	Retail Brokerage	1.09	7
Retail Brokerage	214,363	8	Clearing	1.07	8
Retail Banking	143,892	9	Retail Banking	0.99	9

Table 5b: Business Line Risk Ranking 2009Q1-4

Business Line	Average Loss per Event			Business Line	Gross Loss/€100 Gross Income				
	€	Ranking			€	Ranking			
		Value	Rank			Value	Rank		
Corporate Finance	6,752,175	▲	1	Private Banking	4.12	▲	1	▲	
Trading & Sales	1,008,189	▲	2	Corporate Finance	3.59	▲	2		
Clearing	638,918	▲	3	▲	Trading & Sales	2.34	▼	3	▼
Private Banking	635,612	▲	4	▲	Clearing	2.13	▲	4	▲
Asset Management	373,694	▼	5	▼	Agency Services	1.37	▼	5	▼
Commercial Banking	372,330	▼	6	▼	Asset Management	1.25	▼	6	▼
Agency Services	179,789	▼	7		Retail Brokerage	0.78	▼	7	
Retail Brokerage	136,257	▼	8		Retail Banking	0.70	▼	8	▲
Retail Banking	130,071	▼	9		Commercial Banking	0.70	▼	9	▼

Bijlage 3: Specificatie Business Lines en Event Types

Hieronder zijn de specificaties voor Business Lines en Event Types aangegeven. Hierdoor ontstaat een beeld welke werkzaamheden en gebeurtenissen ingedeeld kunnen worden.

Level 1	Level 2	Activity Groups
Corporate Finance	Corporate Finance	Mergers and Acquisitions, Underwriting, Privatisations, Securitisation, Research, Debt (Government, High Yield), Equity, Syndications, IPO, Secondary Private Placements
	Municipal/Government Finance	
	Merchant Banking	
	Advisory Services	
Trading & Sales	Sales	Fixed Income, equity, foreign exchanges, commodities, credit, funding, own position securities, lending and repos, brokerage, debt, prime brokerage
	Market Making	
	Proprietary Positions	
	Treasury	
Retail Banking	Retail Banking	Retail lending and deposits, banking services, trust and estates
	Private Banking	Private lending and deposits, banking services, trust and estates, investment advice
	Card Services	Merchant/Commercial/Corporate cards, private labels and retail
Commercial Banking	Commercial Banking	Project finance, real estate, export finance, trade finance, factoring, leasing, lends, guarantees, bills of exchange
Payment and Settlement ¹⁵⁴	External Clients	Payments and collections, funds transfer, clearing and settlement
Agency Services	Custody	Escrow, Depository Receipts, Securities lending (Customers) Corporate actions
	Corporate Agency	Issuer and paying agents
	Corporate Trust	
Asset Management	Discretionary Fund Management	Pooled, segregated, retail, institutional, closed, open, private equity
	Non-Discretionary Fund Management	Pooled, segregated, retail, institutional, closed, open
Retail Brokerage	Retail Brokerage	Execution and full service

Event-Type Category (Level 1)	Definition		Activity Examples (Level 3)
Internal fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the law or company policy, excluding diversity/ discrimination events, which involves at least one internal party.	Unauthorised Activity	Transactions not reported (intentional) Trans type unauthorised (w/monetary loss) Mismarking of position (intentional)
		Theft and Fraud	Fraud / credit fraud / worthless deposits Theft / extortion / embezzlement / robbery Misappropriation of assets Malicious destruction of assets Forgery Check kiting Smuggling Account take-over / impersonation / etc. Tax non-compliance / evasion (wilful) Bribes / kickbacks Insider trading (not on firm's account)
External fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent the law, by a third party	Theft and Fraud	Theft/Robbery Forgery Check kiting
		Systems Security	Hacking damage Theft of information (w/monetary loss)
Employment Practices and Workplace Safety	Losses arising from acts inconsistent with employment, health or safety laws or agreements, from payment of personal injury claims, or from diversity / discrimination events	Employee Relations	Compensation, benefit, termination issues Organised labour activity
		Safe Environment	General liability (slip and fall, etc.) Employee health & safety rules events Workers compensation
		Diversity & Discrimination	All discrimination types
Clients, Products & Business Practices	Losses arising from an unintentional or negligent failure to meet a professional obligation to specific clients (including fiduciary and suitability requirements), or from the nature or design of a product.	Suitability, Disclosure & Fiduciary	Fiduciary breaches / guideline violations Suitability / disclosure issues (KYC, etc.) Retail customer disclosure violations Breach of privacy Aggressive sales Account churning Misuse of confidential information Lender Liability

Event-Type Category (Level 1)	Definition	Categories (Level 2)	Activity Examples (Level 3)
		Improper Business or Market Practices	Antitrust Improper trade / market practices Market manipulation Insider trading (on firm's account) Unlicensed activity Money laundering
		Product Flaws	Product defects (unauthorised, etc.) Model errors
		Selection, Sponsorship & Exposure	Failure to investigate client per guidelines Exceeding client exposure limits
		Advisory Activities	Disputes over performance of advisory activities
Damage to Physical Assets	Losses arising from loss or damage to physical assets from natural disaster or other events.	Disasters and other events	Natural disaster losses Human losses from external sources (terrorism, vandalism)
Business disruption and system failures	Losses arising from disruption of business or system failures	Systems	Hardware Software Telecommunications Utility outage / disruptions
Execution, Delivery & Process Management	Losses from failed transaction processing or process management, from relations with trade counterparties and vendors	Transaction Capture, Execution & Maintenance	Miscommunication Data entry, maintenance or loading error Missed deadline or responsibility Model / system misoperation Accounting error / entity attribution error Other task misperformance Delivery failure Collateral management failure Reference Data Maintenance
		Monitoring and Reporting	Failed mandatory reporting obligation Inaccurate external report (loss incurred)
		Customer Intake and Documentation	Client permissions / disclaimers missing Legal documents missing / incomplete
		Customer / Client Account Management	Unapproved access given to accounts Incorrect client records (loss incurred) Negligent loss or damage of client assets
		Trade Counterparties	Non-client counterparty misperformance Misc. non-client counterparty disputes
		Vendors & Suppliers	Outsourcing Vendor disputes

Bijlage 4: Gebruikte en geschreven functies in R

In deze bijlage zijn de belangrijkste functies opgenomen die zijn geschreven en gebruikt voor het schatten van parameters van verschillende verdelingen en voor de Monte Carlo simulatie.

Bijlage 4a: gebruikte functies voor Extreme Value Theory

Onderstaand de gebruikte functies voor het creëren van het lichaam bij het schatten van parameters voor het EVT POT en BMM model. Voor het schatten van de GEV parameters voor BMM is de functie `gevFit` uit het package `fExtremes` gebruikt.

```
pot.bodyfit2<-function(x,u) {

  old.up=which(x>u)
  old.down=which(x<u)
  old.xd=NULL
  for(i in 1:length(old.down)) old.xd[i]=x[old.down[i]]
  print(length(old.xd))
  print(fitdistr(old.xd,"lognormal"))
  res=c(fitdistr(old.xd,"lognormal")$estimate[1], fitdistr(old.xd,"lognormal")$estimate[2])
  res=c(res,gpdFit2(x,u)[1],1,gpdFit2(x,u)[2])
}

bb<-function(x,block=1) {
  data = as.vector(x)
  nblocks = (length(data)%/%block) + 1
  grouping = rep(1:nblocks, rep(block, nblocks))[1:length(data)]
  maxValue = as.vector(tapply(data, grouping, FUN = max))
  maxIndex = as.vector(tapply(as.vector(data), grouping,
    FUN = which.max))
  names(maxValue) = paste(maxIndex)
  maxValue
  rep(0:(nblocks-1)*10+maxIndex)
}

selectbody<-function(data,maxvector){
  result=NULL
  result=as.vector(result)
  for(i in 1:length(data)) {

    help1=0
    for(j in 1:length(maxvector)) {
      if(i != maxvector[j]) {
        help1 = help1 + 1
      }
    }
  }
}
```

```

        if(help1==length(maxvector))
            result=union(result,i)
    }
    result
}

fitbmmbody<-function(x) {
    maxima=NULL
    non.maxima=NULL
    body=NULL
    maxima=bb(x,10)
    non.maxima=selectbody(x,maxima)
    body=x[non.maxima]
    fitdistr(body,'lognormal')
}

gpdFit<-function(x, u = quantile(x, 0.95), type = c("mle", "pwm"), information = c("observed",
    "expected"), title = NULL, description = NULL, ...)
{
    call = match.call()
    type = match.arg(type)
    information = match.arg(information)
    X = x
    xClass = class(x)
    if (xClass == "timeSeries")
        stopifnot(isUnivariate(x))
    x = as.vector(x)
    N = length(x)
    exceedances = x[x > u]
    Names = as.character((1:N)[x > u])
    exceedances = as.vector(exceedances)
    names(exceedances) = Names
    if (type == "mle") {
        fit = .gpdmleFit(x, u, information)
        fit$llh = fit$fit$value
        fit$convergence = fit$fit$convergence
    }
    else if (type == "pwm") {
        fit = .gpdpwmFit(x, u)
        fit$llh = NA
        fit$convergence = NA
    }
    fit$p.less.thresh = fit$prob = 1 - length(x[x > u])/length(x)
    fit$threshold = u
    fit$data = x
    xi = fit$par.ests["xi"]
    beta = fit$par.ests["beta"]
    residuals = log(1 + (xi * (as.vector(exceedances) - u))/beta)/xi
}

```

```

if (is.null(title))
  title = "GPD Parameter Estimation"
if (is.null(description))
  description = description()
parameter = list(u = u, type = type)
if (information == "mle")
  parameter$information = information
new("fGPDFFIT", call = call, method = c("gpd", type), parameter = parameter,
  data = list(x = X, exceedances = exceedances), fit = fit,
  residuals = residuals, title = title, description = description)
parameters=fit$par.ests
parameters
}

```

Bijlage 4b: gebruikte functies voor de g-and-h verdeling

De eerste functies zijn de standaardfuncties die nodig zijn voor berekeningen, zoals de dichtheidsfunctie, cumulatieve kansdichtheid en een functie om trekkingen uit de verdeling te doen.

```

qghd <- function(q,a,b,g,h) {
  Z <- qnorm(q)
  if (g==0) Z else (a+b*(exp(g*Z)-1) * exp((h*Z^2)/2))/g
}

rghd <- function(n,a,b,g,h) {
  qghd(runif(n),a,b,g,h)
}

pghd <- function(p,a,b,g,h,eps=1e-7) {
  manyroots(function(z) { qghd(z,a,b,g,h) - p}, interval=c(eps,1-eps))$root
}

dghd <- function(x,a,b,g,h,log=FALSE,ndep=1e-3,...) {
  if (length(x)>1) return(sapply(x,dghd,a=a,b=b,g=g,h=h,log=log,ndep=ndep,...))
  r <- (pghd(x+ndep,a,b,g,h)-pghd(x,a,b,g,h))/ndep
  if (log) log(r) else r
}

g.p <-function(data,p) {
  #deze functie wordt gebruikt voor het inschatten van de g parameter
  result=NULL
  a1=-(1/qnorm(p))
  a2=quantile(data,(1-p))-median(data)
  a3=median(data)-quantile(data,p)
  if(a2/a3>0) {
    result=a1*log(a2/a3)
  } else NA
  result
}

```

```
UHS <-function(data,p,g) {
  #deze functie wordt gebruikt voor het inschatten van UHS
  s=quantile(data, (1-p))-median(data)
  b=(exp(-g*qnorm(p))-1)
  sol=(g*s/b)
  sol
}
```

Bijlage 4c: gebruikte functies voor de alpha-stable verdeling

Voor het schatten van parameters van de alpha-stable verdeling via MLE kan de volgende functie worden gebruikt:

```
a.s.functie<-function(loss.data,alpha,beta,gamma,delta){
  if(scale>0)
  -sum(stabledist::dstable(loss.data,alpha,beta,gamma,delta,log=TRUE))
  else NA
}

fit2.IF<-mle(a.s.functie,start=list(location=2,scale=2,shape=1),method="L-BFGS-
B",lower=list(scale=0),control=list(ndeps=rep(1e-6,4)))
```

Een alternatief is de kwantielen methode, waarbij startwaarden $\alpha = 1.75$, $\beta = 0$, $\gamma = 1$ en $\delta = 0$:

```
stableFit(dataset, method="q")
```

Bijlage 4d: functies bij Goodness of Fit

Voor de analyse van GoF is gebruik gemaakt van de volgende functies:

- qqplot()
- ks.test()
- adk.test() #package 'adk'

Bijlage 4e: Monte Carlo Simulatie

```
##montecarlo simulatie voor het combineren van een discrete en continue verdeling
mcvar.est<-function(N,freqest=25,method=c("exp","lognormal","alpha-stable","gpd","evt-pot","evt-
bmm","ghd"),theta=list(par1=1,par2=1,par3=1,par4=1,par5=1),n.extr=2) {
  mc.var=seq(0,0.1,length=N)

  ##er wordt via een methode een verdeling geselecteerd
  if(method=="exp") {
    x.freq=rpois(N,freqest)
    for(i in 1:N){
      x.sev=rexp(x.freq[i],theta[1])
      sum=sum(x.sev)
      mc.var[i]=sum
    }
  }
}
```

```

    }
} else if(method=="lognormal") {
  x.freq=rpois(N,freqest)
  for(i in 1:N){
    #geef lmean=theta[1] en lsd=theta[2]
    x.sev=rlnorm(x.freq[i],theta[1],theta[2])
    sum=sum(x.sev)
    mc.var[i]=sum
  }
} else if(method=="alpha-stable") {
  x.freq=rpois(N,freqest)
  alpha=theta[1]
  beta=theta[2]
  gamma=theta[3]
  delta=theta[4]
  for(i in 1:N){
    #geef alpha=theta[1] en beta=theta[2] en gamma=theta[3] en delta=theta[4]
    x.sev=stabledist::rstable(x.freq[i],alpha,beta,gamma,delta)
    sum=sum(x.sev)
    mc.var[i]=sum
  }
} else if(method=="gpd") {
  x.freq=rpois(N,freqest)
  xi=theta[1]
  mu=theta[2]
  sd=theta[3]
  for(i in 1:N){
    x.sev=fExtremes::rgpd(x.freq[i],xi,mu,sd)
    sum=sum(x.sev)
    mc.var[i]=sum
  }
} else if(method=="evt-pot") {
  l.mean=theta[1]
  l.sd=theta[2]
  pot.xi=theta[3]
  pot.mu=theta[4]
  pot.sd=theta[5]
  for(i in 1:N){
    x.freq=rpois(1,12)
    if(x.freq>n.extr){
      x.sev=c(rlnorm(x.freq-
n.extr,l.mean,l.sd),fExtremes::rgpd(n.extr,pot.xi,pot.mu,pot.sd))
      sum=sum(x.sev)
      mc.var[i]=sum
    } else {
      x.sev=c(rlnorm(freqest-
n.extr,l.mean,l.sd),fExtremes::rgpd(n.extr,pot.xi,pot.mu,pot.sd))
      sum=sum(x.sev)
    }
  }
}

```

```

        mc.var[i]=sum
    }
}
} else if(method=="evt-bmm") {
    l.mean=theta[1]
    l.sd=theta[2]
    bmm.xi=theta[3]
    bmm.mu=theta[4]
    bmm.sd=theta[5]
    for(i in 1:N){
        x.freq=rpois(1,12)
        if(x.freq>n.extr){
x.sev=c(rlnorm(x.freq*n.extr,l.mean,l.sd),fExtremes::rgev(n.extr,bmm.xi,bmm.mu,bmm.sd))
            sum=sum(x.sev)
            mc.var[i]=sum
        } else {
            x.sev=c(rlnorm(freqest-
n.extr,l.mean,l.sd),fExtremes::rgev(n.extr,bmm.xi,bmm.mu,bmm.sd))
            sum=sum(x.sev)
            mc.var[i]=sum
        }
    }
} else if(method=="ghd") {
    a=theta[1]
    b=theta[2]
    g=theta[3]
    h=theta[4]
    x.freq=rpois(N,freqest)
    for(i in 1:N){
        x.sev=rghd(x.freq[i],a,b,g,h)
        sum=sum(x.sev)
        mc.var[i]=sum
    }
} else {
    print("give a valid method")
}

hist(mc.var,freq=F,breaks=25,col='lightblue',main="Histogram Value-at-Risk",)
res2=quantile(sort(mc.var),0.999)
print(res2)
}

```

Bijlage 4f: functies voor de analyse voor data kwaliteit

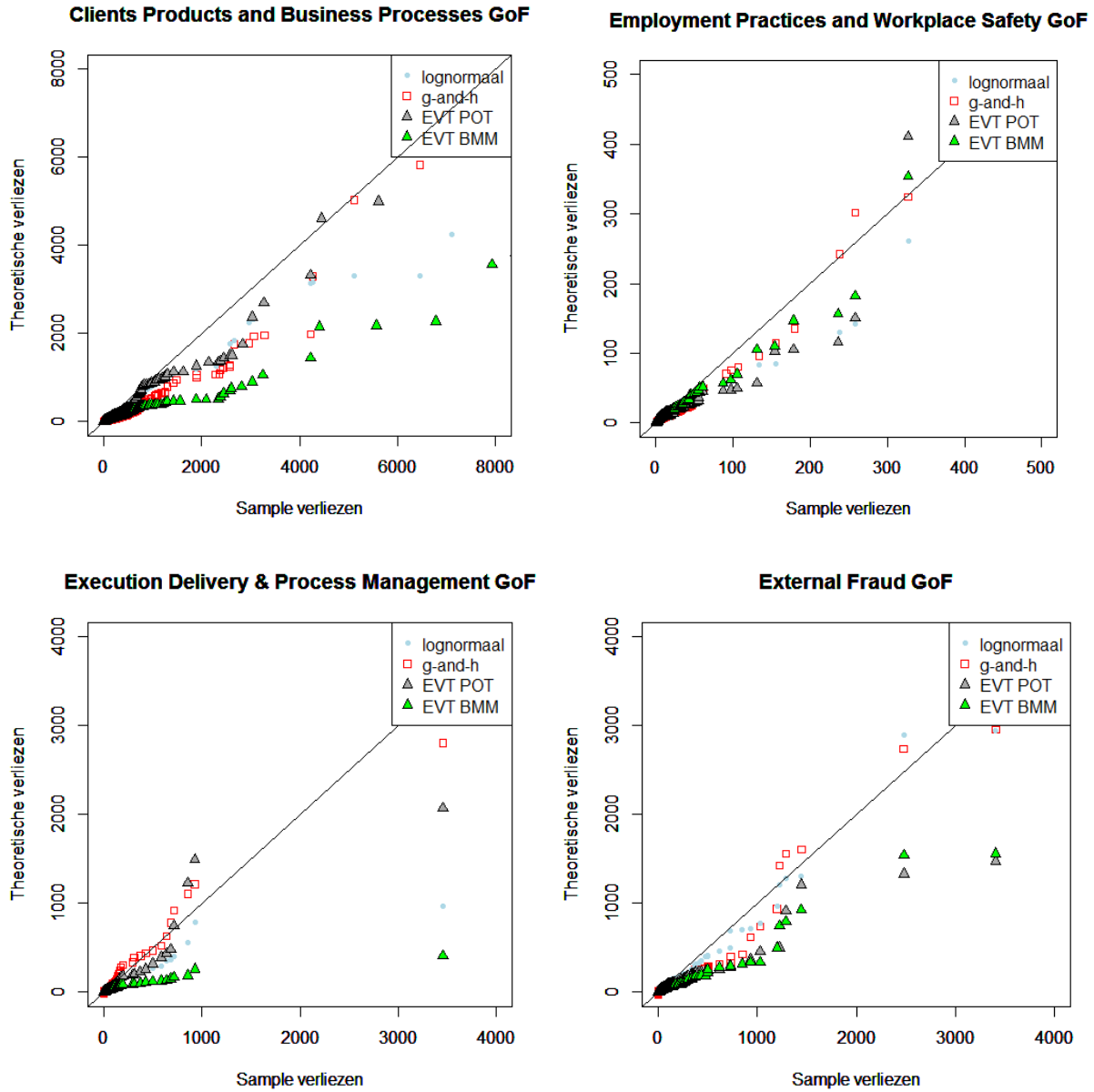
De eerste twee functies zijn geschreven voor het verwijderen van een bepaald percentage van de hoogste of willekeurige waarden in een dataset. Voor punctualiteit is steeds een kolom uit de dataset 'LossAmount' gebruikt in plaats van Loss Amount Current Value. Voor accuraatheid is gebruik gemaakt van de functie round().

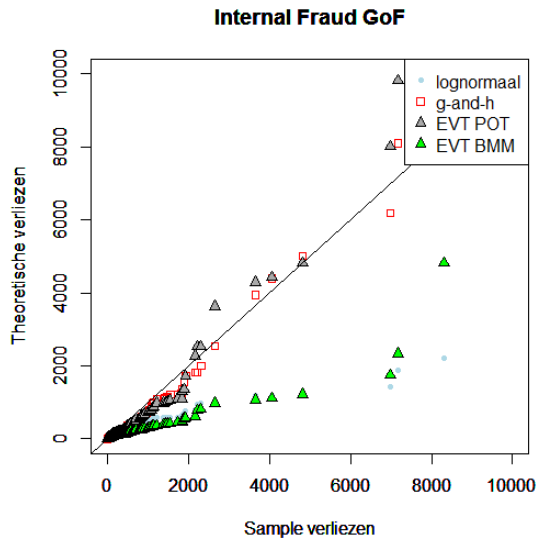
```
remrandomv<-function(x,percentage) {  
  x.a=x  
  n=round(percentage*length(x))  
  for(i in 1:n) {  
    h=round(runif(1,1,length(x.a)))  
    x.a=x.a[-h]  
  }  
  x.a  
}
```

```
remmaxv<-function(x,percentage) {  
  x.a=x  
  n=round(percentage*length(x))  
  for(i in 1:n) {  
    x.a=x.a[which(x.a!=max(x.a))]  
  }  
  x.a  
}
```

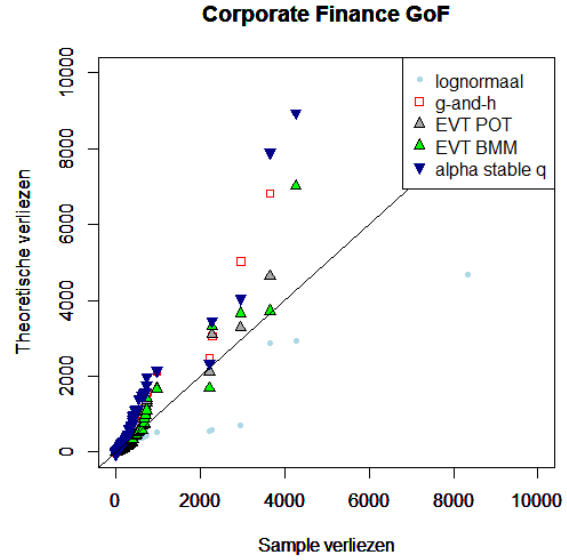
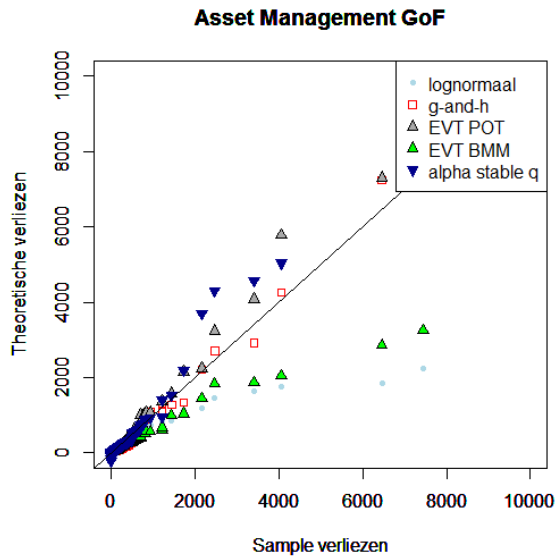

Bijlage 5: QQ-plots voor GoF

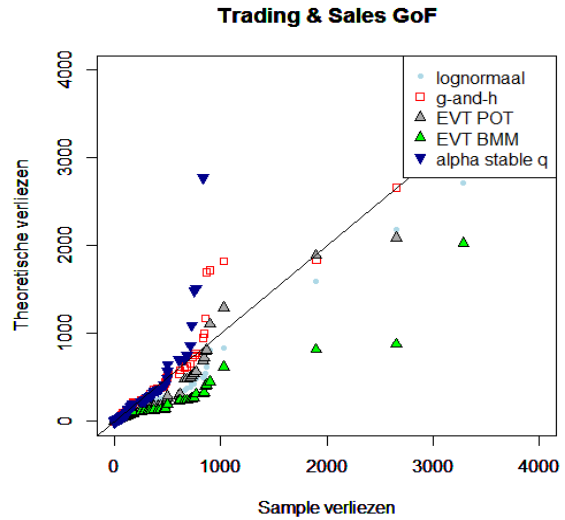
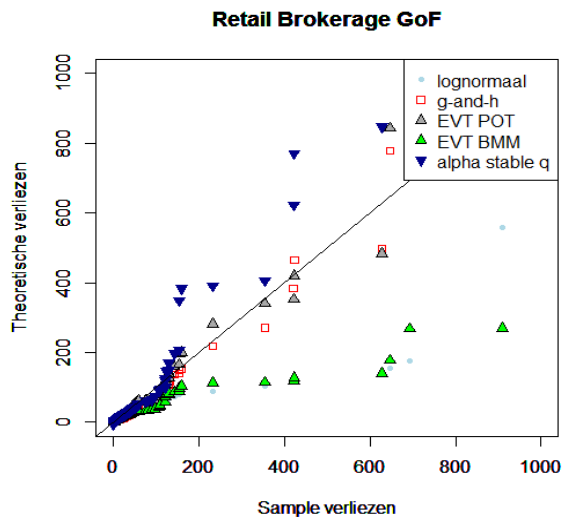
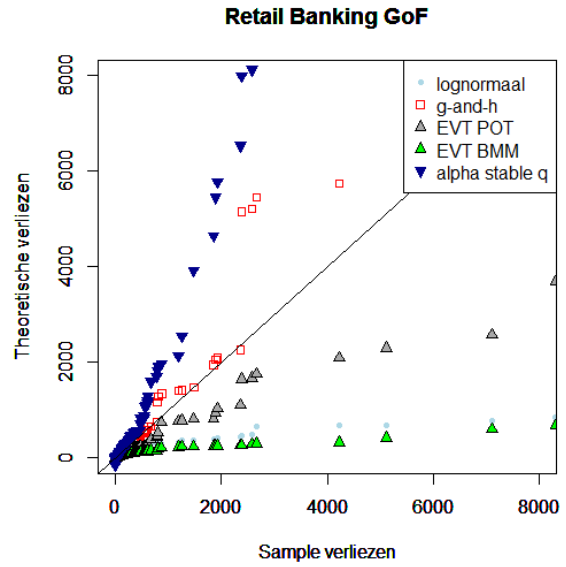
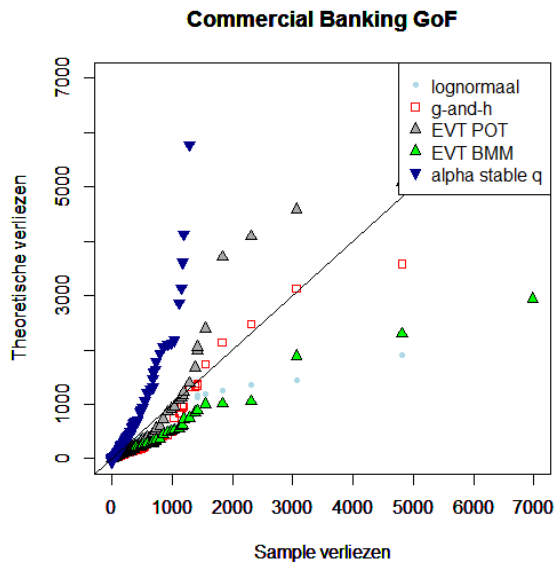
Bijlage 5.1 – QQ-plots voor GoF: Event Types





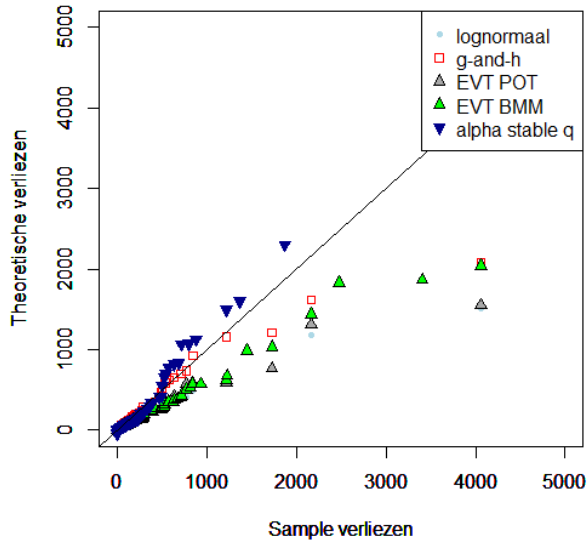
Bijlage 5.2 – QQ-plots voor GoF: Business Lines



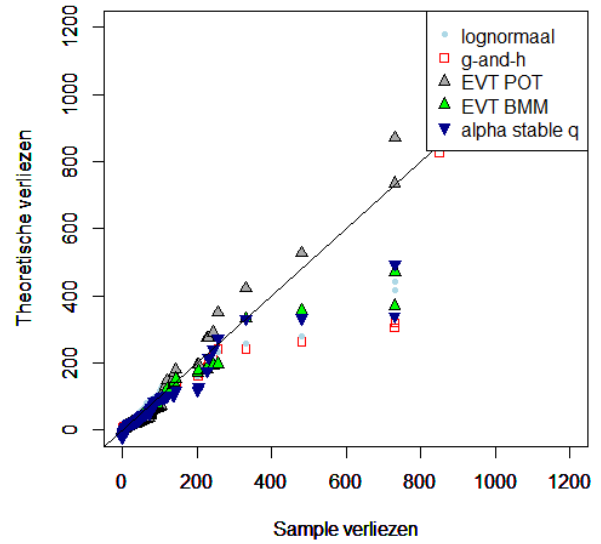


Bijlage 5.3 – QQ-plots voor GoF: Risico Cellen

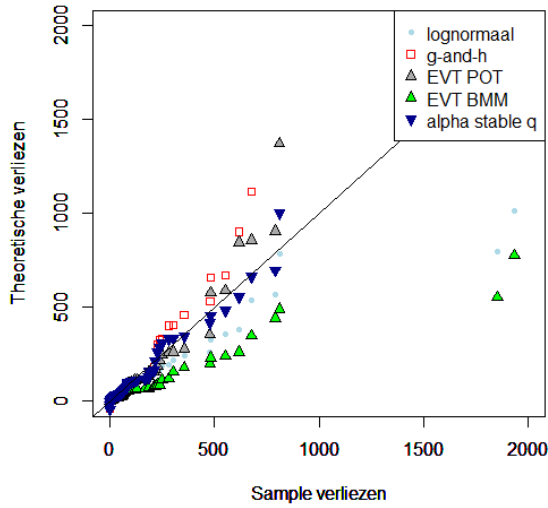
Risico Cel Asset Management - Internal Fraud GoF



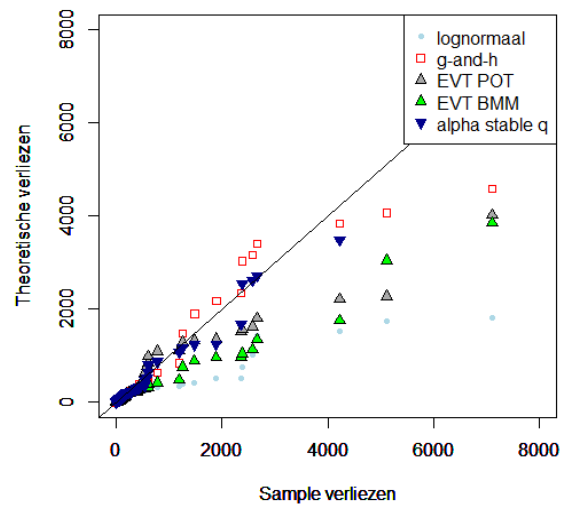
Risico Cel Commercial Banking - External Fraud GoF



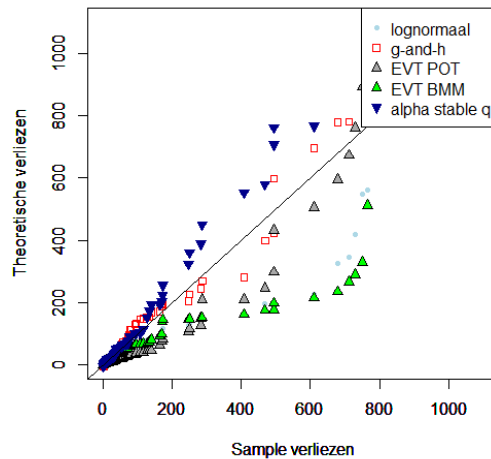
Risico Cel Retail Banking - Internal Fraud GoF



Risico Cel Retail Banking - CPBP GoF



Risico Cel Trading & Sales - CPBP GoF



Bijlage 6 – Gebruikte packages in R

Hieronder staan de packages vermeld die bij de analyse in R zijn gebruikt:

- VGAM
- fExtremes
- fBasics
- stabledist
- MASS
- stats
- graphics
- timeDate
- adk