

Bachelorarbeit Mathematik

Quantifizierung von Operationellen Risiken

Josseline Bauer

Universität Siegen
Fachbereich Mathematik
August 2005

Betreuer: PD Dr. Volkert Paulsen, PD Dr. Edgar Kaufmann

Inhaltsverzeichnis

1	Einleitung	3
2	Übersicht der vorgeschlagenen Ansätze	3
2.1	Basisindikatoransatz	4
2.2	Standardansatz	4
2.3	Fortgeschrittene Messansätze	5
3	Mathematische Grundlagen	8
3.1	Mathematisches Modell	8
3.2	Schätzer	9
3.3	Extremwerttheorie	12
3.4	Copula	20
4	Schätzung	23
4.1	Schadenszahl	23
4.2	Schätzung der Schadenszahlverteilung	24
4.3	Schadenshöhe	25
4.4	Schätzung der 'normalen' Schäden	26
4.5	Schätzung der 'extremen' Schäden	28
5	Verlustverteilungsansatz	40
5.1	Mathematisches Modell	40
5.2	Simulation	41
5.3	Anwendung	44
5.4	Fazit	45
	Anhang	47

1 Einleitung

Der Ausgangspunkt des Risikomanagements ist stets das Erfassen von Risiken. Dazu gibt es eine Vielzahl an Möglichkeiten, zwischen denen man wählen kann. Doch möchte man Risiken nicht nur qualitativ verstehen (zum Beispiel wo diese auftreten oder welche Ursachen vorliegen), sondern auch quantitativ. Im weiteren Verlauf beschäftigen wir uns beispielhaft mit dem bankwirtschaftlichen Risikomanagement und greifen somit das Thema Basel II auf. Gerade die von der Bankenaufsicht vorgeschriebene Eigenkapitalunterlegung fußt auf einer Quantifizierung der Risiken, der eine Bank ausgesetzt ist, d.h. das Ziel ist es, das Risiko anhand eines Kapitalbetrags – der eben dieses absichern soll – zu messen. Insbesondere das Messen von operationellen Risiken rückt mehr und mehr in den Vordergrund und wird explizit von der Bankenaufsicht angesprochen. Der Basler Ausschuss für Bankenaufsicht definiert operationelles Risiko als

“Gefahr von Verlusten, die infolge einer Unzulänglichkeit oder des Versagens von internen Verfahren, Menschen und Systemen oder infolge externer Ereignisse eintreten“.¹

Er schlägt für die Berechnung der Eigenkapitalanforderung für operationelle Risiken drei Ansätze vor: Den Basisindikatoransatz (BIA), den Standardansatz (STA) und die fortgeschrittenen Messansätze (AMA), die alle im Kapitel 2 kurz vorgestellt werden. Der Rest der Arbeit konzentriert sich auf die Ausführung des Loss Distribution Ansatzes als einen der fortgeschrittenen Ansätze.

2 Übersicht der vorgeschlagenen Ansätze

Für die Messung operationeller Risiken nennt der Basler Ausschuss für Bankenaufsicht im wesentlichen drei Vorgehensweisen, die sich bezüglich ihrer Komplexität unterscheiden. Die aufwändigsten Verfahren sind die fortgeschrittenen Messansätze. Gerade viele kleinere und mittlere Banken und Institutionen erfüllen die vom Ausschuss gesetzten Voraussetzungen (noch) nicht und so dürften sich diese Ansätze erst im Laufe der Zeit durchsetzen. Andererseits ist ein größerer Aufwand bei der Berechnung gewöhnlich auch mit einer größeren Genauigkeit verbunden, die wiederum eine angemessenere Kapitalunterlegung in Aussicht stellt.

¹Vgl. Basler Ausschuss für Bankenaufsicht (2004), Seite 127.

2.1 Basisindikatoransatz

Der Basisindikatoransatz als der rudimentärste Ansatz berechnet die Kapitalunterlegung durch Multiplikation einer wirtschaftlichen Kennzahl mit dem sogenannten α -Faktor, der vom Ausschuss auf 15 Prozent festgelegt wurde. Die Kennzahl ist eine Größe, die den Geschäftsumfang widerspiegeln soll. Die Bankenaufsicht nennt als solchen Indikator den mittleren Bruttoertrag der letzten drei Jahre. Durch die Festlegung des α -Faktors von außen wird klar, dass die so berechnete Kapitalunterlegung die individuelle Struktur des Instituts nicht in angemessener Weise berücksichtigen kann. Überdies ist es fragwürdig, ob der Bruttoertrag als Indikator für das operationelle Risiko dienen kann und im proportionalen Verhältnis zu eben diesem steht. Implizit wird unterstellt, dass ein größerer Bruttoertrag automatisch mit einem höheren operationellen Risiko verbunden ist, bzw. ein Institut mit geringerem Bruttoertrag einem kleineren Risiko ausgesetzt ist. Die Formel zur Berechnung der Eigenkapitalanforderung nach dem Basisindikatoransatz lautet:

$$EK_B = \alpha G$$

EK_B Eigenkapitalunterlegung gemäß dem Basisindikatoransatz

α vom Ausschuss auf 15% festgelegter Faktor

G mittlerer Bruttoertrag der letzten drei Jahre (Jahre mit negativem Bruttoertrag sind zu ignorieren)

2.2 Standardansatz

Um den Standardansatz anwenden zu können, müssen im Gegensatz zum BIA sowohl einige qualitative als auch quantitative Voraussetzungen bezüglich des Risikomanagement-Systems erfüllt sein. Gegenüber dem BIA ermöglicht es dieser Ansatz dann aber auch zwischen den Risiken einzelner Geschäftsfelder zu unterscheiden. Innerhalb der einzelnen Geschäftsfelder wird die Eigenkapitalanforderung analog zum BIA berechnet. Die Standardunterteilung des Ausschusses in 8 Geschäftsfelder und die zugehörigen sogenannten β -Faktoren sind Tabelle 1 zu entnehmen. Die Formel lautet hier also:

$$EK_S = \sum_i \beta_i G_i$$

EK_S Eigenkapitalunterlegung gemäß dem Standardansatz

β_i vom Ausschuss festgelegte Faktoren

G_i mittlere Bruttoerträge der Geschäftsbereiche (vergleiche BIA)

Unternehmensfinanzierung/beratung	18%
Handel	18%
Retail-Geschäft	12%
Firmenkundengeschäft	15%
Zahlungsverkehr & Wertpapierabwicklung	18%
Depot- und Treuhandgeschäfte	15%
Vermögensverwaltung	12%
Wertpapierprovisionsgeschäfte	12%

Tabelle 1: Geschäftsfelder mit Beta-Faktoren.
Quelle: BCSB (2004), Seite 130

Neben den bereits beim BIA erwähnten Kritikpunkten fällt auf, dass durch das einfache Aufaddieren eine perfekte Korrelation zwischen den operationellen Risiken der Geschäftsbereiche angenommen wird. Das bedeutet, dass operationelle Risiken zur gleichen Zeit auftreten, was nicht besonders realistisch ist und zu einer eher konservativen Kapitalhinterlegung führt. Zwar ist dies im Sinne der Regulationsbehörde, kann aber eine Fehlallokation des Kapitals innerhalb des Instituts zur Folge haben. Aus diesem Grunde sollte eine Bank daran interessiert sein, einen fortgeschrittenen Messansatz zu verwenden.

2.3 Fortgeschrittene Messansätze

Auf dem Gebiet der fortgeschrittenen Messansätze werden verschiedene Vorgehensweisen diskutiert und (weiter)entwickelt. Aus diesem Grund hat sich der Ausschuss nicht auf ein konkretes Verfahren festgelegt, sondern lässt den Banken freie Hand bei der Wahl des konkreten Ansatzes. Natürlich gibt es auch hier wie beim STA gewisse Anforderungen. Allen fortgeschrittenen Messansätzen gemein ist, dass sie auf den internen Daten der Bank basieren und somit die Grundlage eines individuell auf das Institut zugeschnittenen Modells bilden. Hierbei handelt es sich um ein mathematisches Modell, welches das Risiko bzw. den damit verbundenen Verlust als eine Zufallsvariable ansieht.

Außerdem ist es der Bank erlaubt bei Verwendung eines AMA (advanced measurement approach) die risikomindernde Wirkung von Versicherungen in die Berechnung der Kapitalanforderung mit einzubeziehen. Die Schadensvorfälle werden nicht nur nach ihrer Herkunft gemäß den 8 Geschäftsfeldern aus Tabelle 1 differenziert, sondern zusätzlich 7

verschiedenen Kategorien von operationellen Verlusttypen zugeordnet. Insgesamt können die operationellen Verluste also 56 verschiedenen Zellen zugeordnet werden. Zuerst werden die einzelnen Kombinationen separat betrachtet und später die Ergebnisse aggregiert. Die Klassen der Ereignistypen der operationellen Risiken, wie vom Ausschuss festgelegt, lauten¹ :

- Interner Betrug
- Externer Betrug
- Beschäftigungsverhältnisse und Arbeitsplatzsicherheit
- Klienten-, Produkt- und Geschäftsverhalten
- Schäden an materiellen Aktiva
- Geschäftsunterbrechung und Systemversagen
- Ausführungs-, Liefer- und Prozessmanagement

Zu den fortgeschrittenen Ansätzen gehört unter anderem der IMA (internal measurement approach) und der LDA (loss distribution approach). In der Praxis ist der am häufigsten anzutreffende Ansatz der LDA. Im Vergleich zum IMA bietet er mehr Flexibilität und schätzt den unerwarteten Verlust direkt, d.h. die gesamte Verlustverteilung wird geschätzt, während der IMA von einer analytischen Annäherung des unerwarteten Verlusts innerhalb eines Verlustmodells ausgeht. Der unerwartete Verlust bezeichnet die Differenz zwischen dem 99,9%-Quantil der jährlichen Verlustverteilung und dem erwarteten Verlust, also dem Erwartungswert eben dieser Verteilung. Hierbei muss beachtet werden, dass der erwartete Verlust natürlich auch unterlegt werden muß. Der interne Messansatz kann so verstanden werden, dass das Risiko nach dem Mittelwertprinzip gemessen wird, d.h. der Erwartungswert wird mit einem Faktor multipliziert.

Das Kapitel 7.5 in C. Alexander (2003) beschäftigt sich näher mit dem IMA und enthält auch die folgenden Aussagen. Der Basis-IMA geht davon aus, dass die Schadenshöhe deterministisch ist und die Schadenszahl binomialverteilt. Auf diesen Annahmen aufbauend berechnet er die Eigenkapitalhinterlegung als Anteil des erwarteten Verlusts. Die grundlegende Formel des IMA lautet:

$$EK_{IMA} = \gamma E(L) = \gamma \mu_L,$$

wobei μ_L der Erwartungswert des zufälligen Verlustes L ist.

Die γ -Faktoren sind nicht vom Basler Ausschuss fixiert und müssen selbst kalibriert werden. Eine Methode, wie diese Faktoren geschätzt werden können, wird auf den

¹Quelle: BCBS (2004), Anhang 7.

Seiten 148-156 der oben genannten Quelle beschrieben. Bei der Schätzung wird die Definition des unerwarteten Verlusts ausgenutzt.

$$\gamma \mu_L = EK_{IMA} = q_{99,9\%} - \mu_L \Rightarrow \gamma = \frac{q_{99,9\%} - \mu_L}{\mu_L}$$

EK_{IMA} Eigenkapitalunterlegung gemäß dem Basis-IMA

γ vom Institut zu bestimmender Faktor

μ_L Erwartungswert der Verlustverteilung

$q_{99,9\%}$ das 99,9%-Quantil der Verlustverteilung

Die Verlustverteilung ist nicht bekannt und wird durch eine Annahme über die Schadenszahlverteilung indirekt festgelegt, da wir von einer deterministischen Schadenshöhe ausgehen. Setzt man die Eigenkapitalunterlegung nicht nach dem sogenannten Erwartungswertprinzip an (d.h. die Eigenkapitalunterlegung ist ein Anteil, beziehungsweise Vielfaches des Erwartungswertes), sondern nach dem Standardabweichungsprinzip, ergibt sich mit σ_L als Standardabweichung der Verlustverteilung

$$\zeta \sigma_L = EK_{IMA} = q_{99,9\%} - \mu_L \Rightarrow \zeta = \frac{q_{99,9\%} - \mu_L}{\sigma_L}.$$

Unter verschiedenen Annahmen über die Schadenszahlverteilung ergeben sich die in Tabelle 2 angegebenen Formeln für die Eigenkapitalunterlegung.

Verteilung	Erwartungswert μ_N	Standardabweichung σ_N	EK_{IMA}
Binomial	np	$\sqrt{np(1-p)} \approx \sqrt{np}$	$\zeta X \sqrt{np}$
Poisson	λ	$\sqrt{\lambda}$	$\zeta X \sqrt{\lambda}$
Negativbinomial	$\alpha\beta$	$\beta\sqrt{\alpha}$	$\zeta X \beta \sqrt{\alpha}$
n, p, λ, α und β sind die Parameter der jeweiligen Verteilung; deterministische Schadenshöhe X			

Tabelle 2: Formel für die Eigenkapitalunterlegung nach IMA bei verschiedenen Schadenszahlverteilungsannahmen.

Nachdem man diese Berechnungen für jede der 56 Kombinationen ausgeführt hat, kann man die Eigenkapitalhinterlegungen z.B. aufaddieren, wobei hier wieder eine perfekte Korrelation von impliziert ist. Es ist aber auch eine genauere Modellierung der Beziehungen zwischen den einzelnen Kombinationen möglich.

Im Folgenden wenden wir uns dem LDA zu. Bei diesem Ansatz steht die Verlustverteilung als solche und deren Modellierung im Vordergrund.

3 Mathematische Grundlagen

Zuerst werden wir allgemein die mathematische Modellierung eines Verlustes einführen und die Bedeutungen – übertragen auf den Fall des Verlustverteilungsansatzes – skizzieren. Damit das Modell später an Daten angepasst werden kann, wenden wir uns danach dem Thema Schätzer zu. Die Eigenschaft operationeller Risiken, dass auch extrem hohe Schäden mit einer nicht zu vernachlässigenden Wahrscheinlichkeit vorkommen, findet in einem Paragraphen über Extremwerttheorie seinen Niederschlag. Das Thema Copula soll das Bild vervollständigen.

3.1 Mathematisches Modell

Wie bereits erwähnt, wird das Risiko, bzw. der Verlust mit dem Konzept der Zufallsvariablen verbunden. Für alle weiteren Aussagen wird implizit ein zugehöriger Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ unterstellt. Wir gehen von einem Verlust L aus, der sich aus einer gewissen Anzahl von Einzelschäden X_i zusammensetzt. Diese sowie die Anzahl der Schäden sind zufällig, so dass wir insgesamt einen zufälligen Verlust erhalten. Da diese Überlegung durch einfache Summenbildung auf rein theoretisch beliebig viele Bereiche, in denen Verluste auftreten, übertragen werden kann, reicht es zuerst nur ein Gebiet zu betrachten. Im Kontext des LDA, der die Verlustverteilung der 56 einzelnen Geschäftsfeld-Verlusttyp-Kombinationen modelliert und daraus die Verlustverteilung der Gesamtbank berechnet, konzentrieren wir uns somit zuerst auf eine solche Kombination und behandeln im zweiten Schritt Gesichtspunkte der Aggregation. Um die Verlustverteilung zu modellieren, greift man auf einen versicherungsmathematischen Ansatz zurück. Dazu betrachten wir folgendes wahrscheinlichkeitstheoretische Modell einer zufälligen Summe zur Beschreibung des Verlustes innerhalb eines Geschäftsbereichs:

$$L = \sum_{i=1}^N X_i \tag{1}$$

N zufällige Schadenszahl, unabhängig von den X_i

X_i zufällige Schadenshöhe des i -ten Schadens unter der Bedingung, dass ein Schaden eingetreten ist. Es wird angenommen, dass die X_i unabhängig und identisch verteilt sind.

Unter diesen Annahmen lassen sich folgende grundlegenden Eigenschaften des Summenprozesses herleiten:

- $E(L) = E(N)E(X)$, wobei $X \sim X_i$ dieselbe Verteilung wie X_i besitzt

- $Var(L) = E(N)Var(X) + Var(N)(E(X))^2$.

3.2 Schätzer

Das Thema Schätzer findet man in fast jedem Lehrbuch zur Einführung in die Statistik, wie z.B. in Stahel (2002) oder ausführlicher in Silvey (1975).

Wir gehen von einem sogenannten parametrischen Modell aus und ordnen sowohl der Schadenszahl N als auch der Schadenshöhe X eine parametrische Verteilungsfamilie $\{F_\theta \mid \theta \in \Theta\}$ zu, wobei Θ der Grundraum ist, der alle möglichen Parametervektoren θ der Verteilung umfasst. In den meisten Fällen ist die Verteilung nicht bekannt und soll aufgrund von Daten geschätzt werden. Durch die Festlegung der Familie von Verteilungen ist eine erste Annahme getroffen; durch die Wahl der Parameter soll dann diese Verteilungsfamilie möglichst gut an die beobachteten Werte angepasst werden. Um die Güte eines Schätzers zu beurteilen, stützt man sich auf die folgenden Eigenschaften und Definitionen.

Eigenschaften von Schätzern

Allgemein gilt, dass ein Schätzer $\hat{\theta}$ für θ wiederum selbst eine Zufallsvariable ist, die eine Verteilungsfunktion besitzt. Diese Zufallsvariable kann verschiedene Merkmale aufweisen.

- Eine wichtige Eigenschaft ist die *Erwartungstreue* des Schätzers, d.h. $E(\hat{\theta}) = \theta$. Im Mittel entspricht der Schätzer also dem wahren Parameter θ . Da die Größe $b := E(\hat{\theta}) - \theta$ als Bias bezeichnet wird, heißen erwartungstreue Schätzer auch unbiased.
- Konsequenterweise möchte man überdies zwischen den erwartungstreuen Schätzern eine Beurteilung der Güte, was zu den *MVU-Schätzern* führt (minimum variance unbiased).
- Ähnlich der Erwartungstreue misst der *mittlere quadratische Fehler* (mean standard error) die „Differenz“ zwischen dem Schätzer und dem wahren Parameter. $MSE_{\hat{\theta}} := E((\hat{\theta} - \theta)^2)$.
- Ein weiteres Kriterium ist die *Konsistenz*, d.h. das Verhalten des Schätzers für sehr großen Stichprobenumfang. Der Schätzer heißt konsistent, wenn gilt

$$\hat{\theta}(Y_1, Y_2, \dots, Y_n) \xrightarrow{P} \theta \quad , \text{ d.h.} \\ P(|\hat{\theta}(Y_1, Y_2, \dots, Y_n) - \theta| > \epsilon) \rightarrow 0, \quad \forall \epsilon > 0 \quad \text{und} \quad n \rightarrow \infty.$$

Bemerkung 1. *Ein Schätzer ist eine Funktion der vorliegenden Daten. Diese sind wiederum zufällig (siehe Stahel (2002), Seite 170).*

Um die Parameter in einem gegebenen Modell, d.h. zu einer gewählten Verteilungsfamilie zu bestimmen, gibt es verschiedene Methoden.

Maximum-Likelihood Schätzer

Seien die Daten $\mathbf{y} = (y_1, y_2, \dots, y_m)$ als m unabhängige Realisationen der Zufallsvariablen Y gegeben. Die Wahrscheinlichkeit, dass die Realisationen y_1, y_2, \dots, y_m von einer Zufallsvariablen Y mit Verteilung F_{θ^*} gezogen wurden, spiegelt die Plausibilität des Parameters θ^* als wahrer Parameter der Verteilung wider. Je größer die Wahrscheinlichkeit unter einer Verteilung F_{θ^*} gerade die vorliegenden Realisationen zu erhalten, desto plausibler erscheint die Annahme, dass θ^* der wahre Parameter ist. Wir wollen also θ so bestimmen, dass diese Wahrscheinlichkeit maximal wird. Da fast alle parametrische Verteilungen mehr als einen Parameter besitzen, gehen wir als Verallgemeinerung davon aus, dass es sich bei $\theta \in \mathbb{R}^p$ um einen Vektor handelt, der die Parameter der Verteilung enthält.

Wenn wir annehmen, dass die gemeinsame Verteilung $H_{\theta}(y_1, y_2, \dots, y_m)$ des Zufallsvektors (Y_1, Y_2, \dots, Y_m) eine Dichte $h(\theta; \mathbf{y}) := h_{\theta}(y_1, y_2, \dots, y_m)$ besitzt, kann man die oben angesprochene Wahrscheinlichkeit mit deren Hilfe darstellen. Durch die Notation $h(\theta; \mathbf{y})$ soll ausgedrückt werden, dass wir die Dichte bei gegebenem \mathbf{y} als Funktion in θ betrachten wollen. Man nennt diese Funktion dann auch Likelihood-Funktion. Nutzt man aus, dass die Y_i unabhängig und identisch verteilte Duplikate von Y sind, lautet der erste Ansatz:

$$\max_{\theta \in \Theta} h(\theta; \mathbf{y}) = \max_{\theta \in \Theta} \prod_{i=1}^m f_{\theta}(y_i) \quad (2)$$

Θ Parametergrundraum

f_{θ} Dichte der Verteilung von Y mit Parametervektor θ

Aus praktischen Gründen maximiert man nicht die obige Likelihood-Funktion $h(\theta; \mathbf{y})$, sondern die Log-Likelihood-Funktion $L_{\theta}(y) := \ln h(\theta; \mathbf{y})$. Der Maximum-Likelihood Schätzer $\hat{\theta}$ ergibt sich als Lösung dieses Maximierungsproblems. Es gilt

$$\hat{\theta} := \arg \max_{\theta \in \Theta} h(\theta; \mathbf{y}) \quad \Leftrightarrow \quad \hat{\theta} = \arg \max_{\theta \in \Theta} L_{\theta}(y),$$

da der Logarithmus eine streng monotone Funktion ist. Dieses Maximierungsproblem löst man, indem man den Gradientenvektor der Log-Likelihood-Funktion gleich Null

setzt und aus den daraus resultierenden Gleichungen $\hat{\theta}$ bestimmt. Dieses Gleichungssystem bestehend aus p Gleichungen:

$$\frac{\partial}{\partial \theta_j} L_\theta(y) = \frac{\partial}{\partial \theta_j} \ln h(\theta; \mathbf{y}) = \sum_{i=1}^m \frac{\partial}{\partial \theta_j} \ln f(y_i; \theta) \stackrel{!}{=} 0, \quad j = 1, \dots, p.$$

Die meisten Verteilungen gehören zur Exponentialfamilie, d.h. sie besitzen eine Dichte der Form

$$f_\theta(x) = C(\theta) e^{\theta T(x)} g(x).$$

In diesem Fall erhält man für die Log-Likelihood-Gleichungen

$$\sum_{i=1}^m T(x_i) = m \frac{\partial \ln C(\theta)}{\partial \theta_j}, \quad j = 1, \dots, p.$$

Der Maximum-Likelihood Schätzer (ML Schätzer bzw. MLE für ML estimator) ist der bekannteste der parametrischen Schätzer. Der große Vorteil des Maximum-Likelihood Schätzers ist seine Konsistenz (siehe Silvey (1975), Seite 76).

Momentenmethode

Bei der Momentenmethode verwendet man Kennzahlen der Stichprobe zur Schätzung der entsprechenden theoretischen Kennzahlen, die im wesentlichen Momente der Verteilung sind. Demnach ergibt sich für eine von der Zufallsvariablen Y gezogene unabhängige Stichprobe y_1, y_2, \dots, y_m Folgendes (vergleiche Stahel (2002), Seite 28):

Der Erwartungswert wird geschätzt durch $\bar{y} := \frac{1}{m} \sum_{i=1}^m y_i$, (3a)

die Varianz durch $\hat{s}^2 := \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2$, (3b)

die Schiefe wird geschätzt durch $\hat{s}^3 := \frac{1}{m} \sum_{i=1}^m z_i^2$ (3c)

und die Langschwänzigkeit durch $\hat{s}^4 := \frac{1}{m} \sum_{i=1}^m z_i^3 - 3$ mit $z_i := \frac{y_i - \bar{y}}{\hat{s}}$. (3d)

3.3 Extremwerttheorie

Es sollen einige Aspekte und Ergebnisse der Extremwerttheorie (EVT, extreme value theory) dargelegt werden, die später bei der Modellierung der Schadenshöhenverteilung bestimmend sein werden. Dieser Paragraph basiert hauptsächlich auf Reiss/Thomas (2001), Embrechts/ Klüppelberg/Mikosch (1997) und Menkens (2005). Man beachte, dass die Notationen sich an die beiden letztgenannten Quellen anlehnen.

Wir gehen von Ursprungsdaten x_1, x_2, \dots, x_m als Realisationen unabhängiger und identisch verteilter Zufallsvariablen X_1, X_2, \dots, X_m aus. Die zugehörige Verteilungsfunktion F sei nicht bekannt. Zuerst einige Definitionen und Bezeichnungen, die künftig gebraucht werden:

1. Ordnet man die m Zufallsvariablen X_1, \dots, X_m in absteigender Reihenfolge, erhält man die Ordnungsstatistiken $X_{m:m} \leq \dots \leq X_{2:m} \leq X_{1:m}$, d.h. $X_{1:m}$ ist das Maximum der Zufallsvariablen.
2. Wir bezeichnen allgemein den rechten, bzw. den linken Endpunkt einer Verteilungsfunktion F mit $x_R(F) := \sup\{x \mid F(x) < 1\}$, bzw. $x_L(F) := \inf\{x \mid F(x) > 0\}$.
3. Mit $\bar{F}(x) := 1 - F(x)$ bezeichnen wir die Flankenfunktion von F . Diese gibt die Wahrscheinlichkeit an, dass die Zufallsvariable $X \sim F$ größer als x ist, welche hinsichtlich der Betrachtung von Überschreitungen eine relevante Größe ist.
4. Die verallgemeinerte Inverse von F ist definiert durch

$$F^{\leftarrow}(y) := \inf\{x \mid F(x) \geq y\}.$$

Ist die Verteilungsfunktion stetig und streng monoton, so handelt es sich hierbei um die Umkehrabbildung der Verteilungsfunktion.

5. Für $X \sim F$ definiert man das q -Quantil $x_q := F^{\leftarrow}(q) = \inf\{x \mid F(x) \geq q\}$.
6. Die empirische Verteilungsfunktion basierend auf den m Daten x_1, \dots, x_m ist definiert durch $\hat{F}_m(x) := \frac{1}{m} \sum_{i=1}^m 1_{\{X_i \leq x\}}$.

Überschreitungen

Eine wichtige Frage, mit der sich die Extremwerttheorie auseinandersetzt, ist das Verhalten der Überschreitungen oberhalb einer hohen Schwelle u . Wir nehmen an, dass $u < x_R(F)$.

Es kann sowohl die Menge $\{X_i \mid X_i > u\}$ der Ursprungsdaten, die größer sind als die Schwelle u – die sogenannten Exzedenten –, als auch die Menge der eigentlichen

Überschreitungen $\{Y_i | Y_i := X_i - u, X_i > u\}$ – die sogenannten Exzesse – betrachtet werden. Im Weiteren konzentrieren wir uns hauptsächlich auf die Exzedentent X_i . Es ergeben sich die Verteilungen:

$$F_u(x) := P(X \leq x | X > u) = P(X \leq x, X > u) / P(X > u) = \frac{F(x) - F(u)}{1 - F(u)}$$

Verteilung über der Schwelle u für $x \geq u$

(4)

$$F_{(u)}(y) := P(Y \leq y | X > u) = P(X - u \leq y | X > u)$$

$$= P(X \leq y + u | X > u) = F_u(y + u), \quad y \geq 0$$

Überschreitungsverteilung zur Schwelle u (excess df)

POT und GPD

Mit Hilfe der Verteilung über der Schwelle u lässt sich die sogenannte POT-Stabilität (peaks over threshold) und der POT-Anziehungsbereich definieren (vergleiche Reiss/Thomas (2001), Seite 25):

Definition 1. Eine Verteilungsfunktion F heißt POT-stabil, wenn es zu jedem $u \in \mathbb{R}$ Konstanten a_u, b_u gibt, so dass gilt:

$$F_u(a_u x + b_u) = F(x)$$

Definition 2. Man sagt allgemein F liegt im POT-Anziehungsbereich von G , wenn gilt

$$F_u(a_u x + b_u) \rightarrow G(x), \quad u \rightarrow x_R(F) \quad \text{für geeignete Konstanten } a_u, b_u.$$

Es gibt Verteilungen, die gerade durch die POT-Stabilität ausgezeichnet sind, die sogenannten verallgemeinerten Pareto-Verteilungen (GPD, generalized pareto distribution). Sie ergeben sich als nicht-entartete Grenzverteilung der Überschreitungsdatenverteilung, wenn die Schwelle $u \rightarrow \infty$ strebt.

Das entscheidende Theorem zur Modellierung der Überschreitungsdaten findet man zum Beispiel in Reiss/Thomas (2001), Seite 26.

Satz 1 (Balkeman-de Haan-Pickands). Hat $F_u(a_u x + b_u)$ eine stetige Grenzverteilung für $u \rightarrow x_R(F)$ mit geeigneten Konstanten a_u, b_u , so gilt

$$\sup_{0 \leq x < x_R(F) - u} \left| F_u(x) - G\left(\frac{x - u}{\sigma_u}\right) \right| \rightarrow 0, \quad u \rightarrow x_R(F)$$

mit einer verallgemeinerten Pareto-Verteilung G und geeignetem $\sigma_u > 0$, d.h. die Grenzverteilung der Überschreitungsdatenverteilung ist eine verallgemeinerte Pareto-Verteilung.

Die verallgemeinerten Pareto-Verteilungen lassen sich folgendermaßen definieren durch (α -Parametrisierung):

$$G_0(x) = \begin{cases} 0 & , x < 0 \\ 1 - e^{-x} & , x \geq 0 \end{cases} \quad \text{Exponentialverteilung} \quad (5a)$$

$$G_\alpha(x) = \begin{cases} 0 & , x \leq 1 \\ 1 - x^{-\alpha} & , x \geq 1 \end{cases} \quad \alpha > 0 \quad \text{Pareto-Verteilung} \quad (5b)$$

$$G_\alpha(x) = \begin{cases} 0 & , x < -1 \\ 1 - (-x)^\alpha & , -1 \leq x \leq 0 \\ 1 & , x > 0 \end{cases} \quad \alpha > 0 \quad \text{Betaverteilung.} \quad (5c)$$

Bemerkung 2. Die obige Betaverteilung bezeichnet eine Teilfamilie der üblichen Betaverteilung. Ist im Folgenden die Rede von einer Betaverteilung, so handelt es sich um eine verallgemeinerte Pareto-Verteilung.

Der Parameter α wird als Gestaltungsparameter oder Flankenparameter bezeichnet, der bei der Pareto-Verteilung das Tailverhalten bestimmt. Zur besseren Handhabung der 3 Fälle fasst man diese unter der standardisierten verallgemeinerten Pareto-Verteilung zusammen (ξ -Parametrisierung):

Definition 3.

$$G_\xi(x) := \begin{cases} 1 - e^{-x}, & \text{für } \xi = 0 \\ 1 - (1 + \xi x)^{-\frac{1}{\xi}}, & \text{für } \xi \neq 0 \end{cases} \quad \text{wobei} \quad \begin{cases} 0 < x & \text{für } \xi = 0 \\ 0 < x & \text{für } \xi > 0 \\ 0 < x < \frac{1}{|\xi|} & \text{für } \xi < 0 \end{cases} \quad (6)$$

heißt standardisierte verallgemeinerte Pareto-Verteilung mit Gestaltungsparameter ξ .

Analog zur Erweiterung der Standardnormalverteilung $\mathcal{N}(0, 1)$ lassen sich die standardisierte verallgemeinerte Pareto-Verteilung und die verallgemeinerten Pareto-Verteilungen in der α -Parametrisierung zu einer größeren Klasse ausdehnen, indem man

$\frac{x-\mu}{\sigma}$ statt x in die obigen Definitionen einsetzt. Dann kann man den Definitionsbereich der standardisierten verallgemeinerten Pareto-Verteilung schreiben als

$$D(\xi, \mu, \sigma) = \begin{cases} [\mu, \infty) & \text{falls } \xi \geq 0 \\ [\mu, \mu + \frac{\sigma}{|\xi|}] & \text{falls } \xi < 0 \end{cases}.$$

Wie bereits erwähnt sind verallgemeinerte Pareto-Verteilungen durch ihre POT-Stabilität charakterisiert. Hieraus kann man schließen, dass die Überschreitungen wieder eine verallgemeinerte Pareto-Verteilung besitzen (vergleiche Reiss/Thomas (2001), Seite 26/27).

Lemma 1. *Sei $X \sim G_{\xi, \mu, \sigma}(x)$. Dann folgen die Überschreitungen $Y := X - u$ bezüglich einer hohen Schwelle u wieder einer verallgemeinerten Pareto-Verteilung.*

Beweis. Sei $F(x) := G_{\xi, \mu, \sigma}(x)$. Wegen der POT-Stabilität gilt

$$F_u(a_u x + b_u) = F(x) \Leftrightarrow F_u(y) = F\left(\frac{y - b_u}{a_u}\right) \text{ für geeignete Konstanten } a_u, b_u,$$

und somit

$$\begin{aligned} P(X - u \leq x \mid X > u) &= F_u(x + u) = F\left(\frac{x + u - b_u}{a_u}\right) \\ &= G_{\xi, \mu, \sigma}\left(\frac{x + u - b_u}{a_u}\right) = G_{\xi, a_u \mu + b_u - u, \sigma a_u}(x) \end{aligned}$$

d.h. die Verteilung der Überschreitungen bezüglich einer hohen Schwelle u ist verallgemeinert Pareto-verteilt. □

Die Dichte der verallgemeinerten Pareto-Verteilung lautet (vergleiche Menkens (2005), Seite 30):

$$g_{\xi, \mu, \sigma}(x) = \begin{cases} \frac{1}{\sigma} (1 + \xi \frac{x-\mu}{\sigma})^{-\frac{1}{\xi}-1} & \text{für } \xi \neq 0 \\ \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} & \text{für } \xi = 0 \end{cases} \quad \text{mit } \begin{cases} x \geq \mu & \text{falls } \xi \geq 0 \\ \mu \leq x \leq \mu - \frac{\sigma}{\xi} & \text{falls } \xi < 0. \end{cases}$$

Zur Veranschaulichung sind Dichten der 3 verallgemeinerten Pareto-Verteilungen mit unterschiedlichen Parametern in Abbildung 1 dargestellt.

Zusammenhänge

In der ξ -Parametrisierung ist $G_{\xi, \mu, \sigma}$ eine Pareto-Verteilung für $\xi > 0$ und eine Beta-Verteilung für $\xi < 0$. Es gilt $x_L(G_{\xi, \mu, \sigma}) = \inf\{x \mid G_{\xi}(\frac{x-\mu}{\sigma}) > 0\} = \mu$, denn für $\xi > 0$

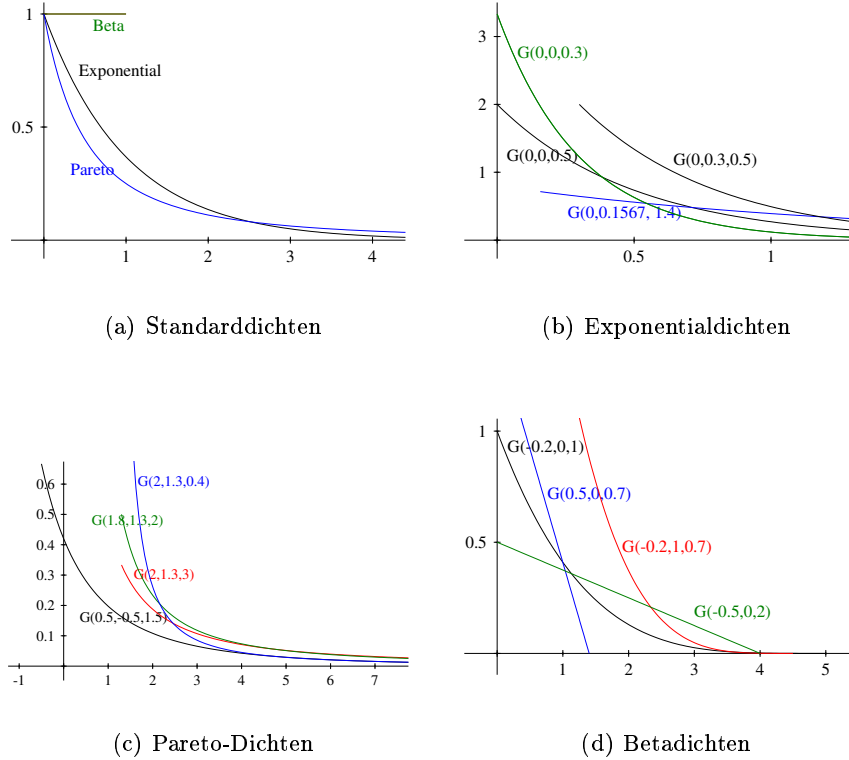


Abbildung 1: Dichten der GPD in der ξ -Parametrisierung

muß $0 \leq \frac{x-\mu}{\sigma} \Leftrightarrow \mu \leq x$, für $\xi < 0$ muß $0 \leq \frac{x-\mu}{\sigma} < \frac{1}{|\xi|} \Leftrightarrow \mu \leq x < \frac{\sigma}{|\xi|} + \mu$ und für $\xi = 0$ muß $0 \leq \frac{x-\mu}{\sigma} \Leftrightarrow \mu \leq x$ gelten, damit die Verteilungsfunktion Werte größer als Null annimmt.

In der α -Parametrisierung gilt für die Pareto-Verteilung mit $\alpha > 0$ $x_L(G_{\alpha,\mu^*,\sigma^*}) = \mu^* + \sigma^*$, da $\frac{x-\mu^*}{\sigma^*} \geq 1 \Leftrightarrow x \geq \mu^* + \sigma^*$ und für die Betaverteilung mit $\alpha < 0$ erhält man wegen $0 \geq \frac{x-\mu^*}{\sigma^*} \geq -1 \Leftrightarrow \mu^* \geq x \geq \mu^* - \sigma^*$, dass der linke Endpunkt gleich $\mu^* - \sigma^*$ und der rechte Endpunkt gleich μ^* ist. Hieran erkennt man eine Besonderheit der Betaverteilung: sie besitzt einen endlichen rechten Endpunkt $x_R < \infty$.

Bemerkung 3. Man beachte, dass $F_u(x)$ als Verteilung über der Schwelle u denselben linken Endpunkt hat wie die Grenzverteilung, deren Lokationsparameter in der ξ -Parametrisierung stets mit dem linken Endpunkt übereinstimmt (siehe oben), d.h. $x_L(G_{\xi,u,\sigma}) = u = x_L(F_u(x)) = \inf \left\{ x \mid \frac{F(x)-F(u)}{1-F(u)} \geq 0 \right\}$.

Die folgenden Beziehungen zwischen den drei verallgemeinerten Pareto-Verteilungen findet man z.B. in Reiss/Thomas (2001) auf Seite 38 .

Lemma 2. Sei $X \sim G_{\alpha,0,\sigma}$, $\alpha > 0$ eine Pareto-verteilte Zufallsvariable.

Dann gilt

- (i) $\ln X \sim G_{0,\ln \sigma, \frac{1}{\alpha}}$, $\alpha > 0$ ist eine exponentialverteilte Zufallsvariable und
- (ii) $-1/X \sim G_{\alpha,0,\frac{1}{\sigma}}$, $\alpha > 0$ eine betaverteilte Zufallsvariable.

Beweis. (i) Einerseits haben wir, dass

$$\begin{aligned} P(X \leq x) &= G_{\alpha,0,\sigma}(x) = 1 - \left(\frac{x}{\sigma}\right)^{-\alpha} \\ \Rightarrow P(\ln X \leq x) &= P(X \leq e^x) = 1 - \left(\frac{e^x}{\sigma}\right)^{-\alpha}. \end{aligned}$$

Andererseits gilt

$$G_{0,\ln \sigma, \frac{1}{\alpha}}(x) = 1 - e^{-\frac{x - \ln \sigma}{1/\alpha}} = 1 - e^{-\alpha(x - \ln \sigma)} = 1 - \left(\frac{e^x}{e^{\ln \sigma}}\right)^{-\alpha} = 1 - \left(\frac{e^x}{\sigma}\right)^{-\alpha}.$$

Hieraus folgt, dass $\ln X \sim G_{0,\ln \sigma, \frac{1}{\alpha}}$.

(ii) Sei $X \sim G_{\alpha,0,\sigma}$ mit $\alpha > 0$. Es gilt

$$P(-1/X \leq y) = P(X \leq -1/y) = 1 - \left(-\frac{1}{\sigma y}\right)^{-\alpha} = 1 - (-\sigma y)^{-(-\alpha)}$$

Also handelt es sich bei $-1/X$ um eine betaverteilte Zufallsvariable mit Gestaltspareter $\alpha > 0$, Lokationspareter 0 und Skalenpareter $1/\sigma$.

□

Die drei verallgemeinerten Pareto-Verteilungen der α -Parametrisierung lassen sich in die standardisierte verallgemeinerte Pareto-Verteilung mit ξ -Parametrisierung überführen:

- für die Pareto-Verteilung durch $G_{\alpha}(1 + \frac{1}{\alpha} x) = G_{\xi}(x)$, wobei $\alpha > 0$, $\xi = \frac{1}{\alpha} > 0$ und
- für die Betaverteilung durch $G_{\alpha}(-1 + (-\frac{1}{\alpha} x)) = G_{\xi}(x)$, wobei $\alpha > 0$, $\xi = -\frac{1}{\alpha} < 0$.

- Für die Exponentialverteilung sind beide Parametrisierungen offensichtlich identisch.

Da einige Resultate in der α -Parametrisierung und andere in der ξ -Parametrisierung präsentiert werden, kann man später auf diese Reparametrisierung zurückgreifen, um sie eventuell leichter nachzuvollziehen.

Flankenmodellierung

Durch Umformen der Definition der Verteilung über der Schwelle u in (4) erhält man für $x \geq u$ die folgende Darstellung der Verteilungsfunktion:

$$F(x) = (1 - F(u)) F_u(x) + F(u), \quad x \geq u. \quad (7)$$

Diese stellt einen sehr nützlichen Zusammenhang her zwischen der Verteilung von X , die modelliert werden soll, und der bedingten Verteilung von X gegeben, dass eine Schwelle überschritten wird. Die bedingte Verteilung können wir durch eine GPD approximieren, wenn die Schwelle u nur hoch genug ist, was gerade der Satz von Balkeman-de Haan-Pickands rechtfertigt.

Bemerkung 4. Die Formel (7) kann man auch durch Bedingen herleiten:

$$\begin{aligned} F(x) &= P(X \leq x | X > u) P(X > u) + P(X \leq x | X \leq u) P(X \leq u) \\ &= F_u(x) (1 - F(u)) + F(u) \quad \text{für } x \geq u, \text{ da dann } P(X \leq x | X \leq u) = 1. \end{aligned}$$

Menkens (2005) beschreibt auf Seite 33 eine andere Sichtweise der Situation. Hier wird davon ausgegangen, dass man die Verteilung der eigentlichen Überschreitungen $Y := X - u$ durch eine GPD approximiert. Dies ergibt sich aus dem Zusammenhang zwischen der Verteilung der Überschreitungsdaten $F_u(x)$ und der Überschreitungsverteilung (vgl.(4)).

$$F_{(u)}(y) = F_u(y + u) = F_u(x) \approx G_{\xi,u,\sigma}(x) = G_{\xi,u,\sigma}(y + u) = G_{\xi,0,\sigma}(y).$$

Hierbei besitzt die GPD, die an die Überschreitungen angepaßt werden kann, den Lokationsparameter und somit den linken Endpunkt 0.

Lemma 3. Gegeben sei eine Verteilungsfunktion F . Handelt es sich bei der Überschreitungsdatenverteilung $F_u(x)$ um eine GPD $G_{\xi,u,\sigma}(x)$ so erhält man als unbedingte Verteilung $F(x) = (1 - F(u)) F_u(x) + F(u)$, $x \geq u$ oberhalb der Schwelle u eine GPD mit demselben Gestaltparameter ξ und Lokations- bzw. Skalenparameter $\tilde{\sigma} = \sigma / (1 - F(u))^{-\xi}$ und $\tilde{\mu} = u - \frac{\sigma}{\xi}$.

Beweis. Es gilt

$$\begin{aligned}
 (1 - F(u)) G_{\xi, u, \sigma}(x) + F(u) &= (1 - F(u)) \left(1 - \left(1 + \xi \frac{x - u}{\sigma} \right)^{-\frac{1}{\xi}} \right) + F(u) \\
 &= 1 - \left((1 - F(u)) \left(1 + \xi \frac{x - u}{\sigma} \right)^{-\frac{1}{\xi}} \right) \\
 &= 1 - \left((1 - F(u))^{-\xi} \left(1 + \xi \frac{x - u}{\sigma} \right) \right)^{-\frac{1}{\xi}} \\
 &= 1 - \left(\bar{F}(u)^{-\xi} \xi \left(\frac{1}{\xi} + \frac{x - u}{\sigma} \right) \right)^{-\frac{1}{\xi}} \\
 &= 1 - \left(\xi \frac{\bar{F}(u)^{-\xi}}{\sigma} \left(\frac{\sigma}{\xi} + x - u \right) \right)^{-\frac{1}{\xi}} \\
 &= 1 - \left(\xi \left(\frac{x - (u - \frac{\sigma}{\xi})}{\sigma / \bar{F}(u)^{-\xi}} \right) \right)^{-\frac{1}{\xi}} \\
 &= G_{\xi, \tilde{\mu}, \tilde{\sigma}}(x), \quad \text{für } x \geq u \\
 &\text{mit } \tilde{\mu} = u - \frac{\sigma}{\xi} \quad \text{und} \quad \tilde{\sigma} = \sigma / \bar{F}(u)^{-\xi}.
 \end{aligned}$$

□

Aus der ξ -Parametrisierung lässt sich durch eine einfache Rechnung das q -Quantil herleiten. Mit einer Wahrscheinlichkeit von q liegen die Werte der Verteilungsfunktion unter dem Wert des q -Quantils, d.h. es wird mit einer Wahrscheinlichkeit von höchstens $1 - q$ überschritten.

$$\begin{aligned}
 G_{\xi}(x) = 1 - (1 + \xi x)^{-\frac{1}{\xi}} = q &\Leftrightarrow G_{\xi}^{-1}(q) = \frac{(1 - q)^{-\xi} - 1}{\xi} = x \quad \text{für } \xi \neq 0, \\
 G_0(x) = 1 - e^{-x} = q &\Leftrightarrow G_0^{-1}(q) = -\ln(1 - q) = x \quad \text{für } \xi = 0.
 \end{aligned}$$

Bemerkung 5. Allgemein lassen sich die Quantile einer Verteilungsfunktion $F_{\mu, \sigma} = F\left(\frac{x - \mu}{\sigma}\right)$ durch die Quantile der standardisierten Verteilungsfunktion ausdrücken und es gilt $F_{\mu, \sigma}^{-1}(y) = \mu + \sigma F^{-1}(y)$. Somit folgt für die Quantile der standardisierten verallgemeinerten Pareto-Verteilung $G_{\xi, \mu, \sigma}$ mit Lokationsparameter μ und Skalenparameter σ :

$$G_{\xi, \mu, \sigma}^{-1}(y) = \mu + \frac{\sigma}{\xi} ((1 - y)^{-\xi} - 1) \quad \text{für } \xi \neq 0 \quad \text{und} \quad G_{0, \mu, \sigma}^{-1}(y) = \mu - \sigma \ln(1 - y) \quad \text{für } \xi = 0.$$

3.4 Copula

Bei der Modellierung von Abhängigkeiten hat der Begriff der Copula in den letzten Jahren zunehmend an Bedeutung gewonnen. Die folgende Definition findet man in Durrleman/Nikeghbali/Roncalli (2000).

Definition 4. Sei $(X_1, \dots, X_d) \in \mathbb{R}^d$ ein d -dimensionaler, reellwertiger Zufallsvektor mit gemeinsamer Verteilung $F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d)$ und Randverteilungen $F_i(x_i) = P(X_i \leq x_i)$, $i = 1, \dots, d$. Eine Verteilungsfunktion $C : [0, 1]^d \rightarrow [0, 1]$ heißt Copula, falls gilt

- $C_n(u_n) := C(1, \dots, 1, u_n, 1, \dots, 1) = u_n$, für $n = 1, \dots, d$ und $u_n \in [0, 1]$
- $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$, für alle Stetigkeitspunkte x_i von F_i .

Eine grundlegende Aussage über Copulas beinhaltet der folgende Satz. Hierzu und zu Lemma 4 vergleiche da Costa Dias (2004).

Satz 2 (Theorem von Sklar). Zu einer d -dimensionale Verteilungsfunktion F mit stetigen Randverteilungen F_i , $i = 1, \dots, d$ existiert eine eindeutig bestimmte Funktion H , so dass gilt

$$F(x_1, \dots, x_d) = H(F_1(x_1), \dots, F_d(x_d)).$$

Die Erweiterung dieser Funktion auf $[0, 1]^d$ ist eine Copula C mit

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

Andererseits ist die zu gegebenen eindimensionalen Verteilungen F_1, \dots, F_d und einer d -dimensionaler Copula C durch

$$F(x_1, \dots, x_d) := C(F_1(x_1), \dots, F_d(x_d))$$

definierte Funktion eine d -dimensionale Verteilungsfunktion, die die Randverteilungen F_1, \dots, F_d besitzt.

Der Grund warum Copulas zur Modellierung besonders geeignet sind, liegt in der Möglichkeit, zu einer mehrdimensionalen Verteilungsfunktion F die Randverteilungen F_i von der Copula C zu trennen und diese beiden Größen separat zu betrachten. Die Copula umfasst hierbei die gesamte Abhängigkeitsstruktur zwischen den einzelnen Verteilungen F_i . Dies erlaubt eine hohe Flexibilität und eine genauere Anpassung des mathematischen Modells an die Realität. Überdies besitzen Copulas das wichtige Merkmal der Invarianz unter monotonen Transformationen (siehe da Costa Dias (2004)).

Lemma 4. *Seien $h_i : \mathbb{R} \rightarrow \mathbb{R}$, $i = 1, \dots, d$ monotone, nicht fallende Funktionen. Dann entspricht die Copula des Zufallsvektors (X_1, \dots, X_d) der Copula, die zu dem Zufallsvektor $(h_1(X_1), \dots, h_d(X_d))$ gehört.*

Von einem praktischen Standpunkt aus betrachtet ist diese Eigenschaft sehr nützlich, da bestimmte Transformationen der Daten, wie z.B. das Logarithmieren, die zugehörige Copula nicht beeinflussen.

Die Formeln zur Schätzung der Parameter der Copula und der Randverteilungen sind da Costa Dias (2004) und Durrleman/Nikeghbali/Roncalli (2000) entnommen. Um diese Parameter zu bestimmen kann man die Maximum-Likelihood Methode anwenden. Unter der Annahme, dass die Copula und die Randverteilungen Dichten besitzen, lässt sich die Dichte der gemeinsamen Verteilung F durch die Dichte der Copula und die Dichten der Randverteilungen darstellen:

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{k=1}^d f_k(x_k),$$

wobei f_k die Dichten der Randverteilungen und

$$c(u_1, \dots, u_d) = \frac{\partial C(u_1, \dots, u_d)}{\partial u_1 \cdots \partial u_d} \quad \text{die Dichte der Copula sind.}$$

Zu einer Stichprobe $\mathbf{x} = (\mathbf{x}^1, \dots, \mathbf{x}^T)$ vom Umfang T mit $\mathbf{x}^t = (x_1^t, \dots, x_d^t)$, $t = 1, \dots, T$ muss die folgende Log-Likelihood Funktion in θ maximiert werden.

$$\begin{aligned} L_\theta(\mathbf{x}) &= \ln \left(\prod_{t=1}^T f(\mathbf{x}^t) \right) \\ &= \sum_{t=1}^T \ln(f(\mathbf{x}^t)) = \sum_{t=1}^T \left(\ln c(F_1(x_1), \dots, F_d(x_d)) + \sum_{k=1}^d \ln(f_k(x_k^t)) \right) \end{aligned} \quad (8)$$

Der Vektor θ enthält sowohl die Parameter der Copula als auch der Randverteilungen, so dass in diesem Fall alle Größen zusammen geschätzt werden. Je nachdem wie viele Randverteilungen vorliegen, ist der resultierende Rechenaufwand sehr groß. Als Alternative kann die sogenannte IFM-Methode angewendet werden, die auf der Trennung der speziell zu den Randverteilungen gehörenden Parameter und den allgemeinen Abhängigkeitsparametern aufbaut. IFM steht für inference functions for margins und diese Methode gliedert sich in zwei Schritte:

- Zuerst werden die Randverteilungsparameter geschätzt und
- dann mit deren Hilfe ein Schätzer für den Vektor der allgemeinen Parameter konstruiert.

Die folgende Form der Log-Likelihood Funktion stellt im Wesentlichen eine andere Schreibweise von (8) dar, die den erklärten Zusammenhang verdeutlichen soll.

$$L_{\theta}(\mathbf{x}) = \sum_{t=1}^T \left(\ln c(F_1(x_1; \theta_1), \dots, F_d(x_d; \theta_d); \alpha) + \sum_{k=1}^d \ln(f_k(x_k^t, \theta_k)) \right),$$

mit den Parametervektoren θ_i , $i = 1, \dots, d$ der Randverteilungen und dem Parametervektor α der Copula. Im ersten Schritt bestimmt man die ML Schätzer

$$\hat{\theta}_k = \arg \max_{\theta \in \Theta} \ln \left(\prod_{t=1}^T f_k(x_k^t; \theta_k) \right) = \arg \max_{\theta \in \Theta} \sum_{t=1}^T \ln(f_k(x_k^t; \theta_k)), \quad k = 1, \dots, d$$

und im zweiten Schritt

$$\hat{\alpha} = \arg \max \sum_{t=1}^T \ln(c(F_1(x_1; \hat{\theta}_1), \dots, F_d(x_d; \hat{\theta}_d); \alpha)).$$

Weitere Ausführungen und auch einen nicht-parametrischen Ansatz findet man in den oben genannten Quellen. Beispiele für Copulas sind:

- Gauß-Copula:

$$C(u_1, \dots, u_d) = \mathcal{N}(\mathbf{0}, R) (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)),$$

wobei $\mathcal{N}(\mathbf{0}, R)$ die d -dimensionale Normalverteilung mit Mittelwertvektor $\mathbf{0}$ und Korrelationsmatrix $R \in \mathbb{R}^{d \times d}$ und Φ die eindimensionale Standardnormalverteilung bezeichnet.

- Gumbel-Copula:

$$C(u_1, \dots, u_d) = \exp \left(- \left(\sum_{i=1}^d (-\ln u_i)^\alpha \right)^{1/\alpha} \right), \quad \alpha \geq 1.$$

- Clayton-Copula:

$$C(u_1, \dots, u_d) = \left(\sum_{i=1}^d (u_i^{-\theta} - 1) + 1 \right)^{-1/\theta}, \quad \theta \in (0, \infty).$$

4 Schätzung

Um eine akkurate Schätzung durchzuführen, bedarf es einer ausreichend großen Datenmenge. Da das Thema operationelle Risiken erst seit einiger Zeit Beachtung findet, ist gerade dieser Punkt einer der Schwachstellen bei der Messung von operationellen Risiken. Mit abnehmendem Umfang der Beobachtungen lässt auch die Aussagekraft eines Schätzers nach, der sich auf diese Beobachtungen stützt.

4.1 Schadenszahl

Die Anzahl der Schäden wird typischerweise als Poisson-verteilte Zufallsvariable angenommen:

$$N : (\Omega, \mathfrak{A}) \rightarrow (\mathbb{N}, \mathbb{P}), \quad \text{mit der Potenzmenge } \mathbb{P} \quad \text{und} \quad P(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

Ein Nachteil der Poisson-Verteilung ist jedoch, dass ihr Erwartungswert mit der Varianz übereinstimmt. Das schränkt nicht nur die Flexibilität des Modells ein, sondern kann je nach Daten auch einfach unzutreffend sein.

Eine Möglichkeit, diesen Mangel zu beheben und gleichzeitig das Konzept sogenannter Risikotreiber einzubauen, ist es, den Intensitätsfaktor λ der Poisson-Verteilung als nicht-deterministisch vorauszusetzen. Ein Risikotreiber im Bereich operationelle Risiken ist zum Beispiel die Anzahl der Mitarbeiter in einer Abteilung: Je weniger Mitarbeiter, desto höher die Gefahr, dass eine Transaktion nicht korrekt ausgeführt wird. Im mathematischen Sinne versteht man unter einem Risikotreiber eine Zufallsvariable, von der der Intensitätsparameter λ der Schadenszahlverteilung abhängt, d.h.

$$\mathbb{P}(N = n | \Lambda = \lambda) = e^{-\lambda} \frac{\lambda^n}{n!}$$

Die unbedingte Schadenszahlverteilung erhält man durch Mischen mit der Verteilung von Λ . Man spricht auch von einem Mischungsmodell.

$$\mathbb{P}(N = n) = \begin{cases} \sum_{\lambda=0}^{\infty} \mathbb{P}(N = n | \Lambda = \lambda) \mathbb{P}(\Lambda = \lambda) & \text{für diskretes } \Lambda \text{ mit Werten aus } \mathbb{N}_0 \\ \int \mathbb{P}(N = n | \Lambda = \lambda) d\mathbb{P}_{\Lambda}(\lambda) & \text{für stetig verteiltes } \Lambda. \end{cases}$$

Der Einfluß des Risikotreibers auf den Schaden wird meist über die Schadenszahlverteilung modelliert, weil sich dann folgender Vorteil ergibt: Wird angenommen, dass der

Risikotreiber Λ gammaverteilt ist mit den Parametern α und β , so erhält man als Resultat, dass die unbedingte Verteilung der Schadenszahl eine Negativbinomialverteilung ist – genauer $N \sim \text{NegBin}\left(\alpha, \frac{1}{\beta+1}\right)$.

$$\begin{aligned} \mathbb{P}(N = n) &= \int_0^\infty e^{-\lambda} \frac{\lambda^n}{n!} \frac{1}{\Gamma(\alpha)} \beta^{-\alpha} \lambda^{\alpha-1} e^{-\frac{1}{\beta}\lambda} d\lambda = \int_0^\infty e^{-\lambda(1+\frac{1}{\beta})} \frac{\lambda^{n+\alpha-1}}{\Gamma(\alpha)n!} \beta^{-\alpha} d\lambda \\ &= \frac{\beta^{-\alpha}}{\Gamma(\alpha)n!} \int_0^\infty e^{-\lambda(1+\frac{1}{\beta})} \lambda^{n+\alpha-1} d\lambda = \frac{\beta^{-\alpha}}{\Gamma(\alpha)n!} \Gamma(\alpha+n) \left(\frac{1}{1+\frac{1}{\beta}}\right)^{\alpha+n} \\ &= \frac{\beta^{-\alpha} \Gamma(\alpha+n)}{\Gamma(\alpha)n!} \left(\frac{1}{1+\frac{1}{\beta}}\right)^\alpha \left(\frac{1}{1+\frac{1}{\beta}}\right)^n = \binom{n+\alpha-1}{n} \left(\frac{1}{\beta+1}\right)^\alpha \left(\frac{1}{1+\frac{1}{\beta}}\right)^n. \end{aligned}$$

4.2 Schätzung der Schadenszahlverteilung

Nachdem man sich für eine Verteilungsfamilie entschieden hat, kann man zu einer Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$ die Parameter der Schadenszahlverteilung mit einer der oben genannten Methoden schätzen.

Unterstellen wir eine Poisson-Verteilung für die zufällige Schadenszahl N , so läßt sich die Dichte auf die in Paragraph 3.2 angesprochene Form bringen, weil die Poisson-Verteilung zur Exponentialfamilie gehört. Es gilt:

$$e^{-\lambda} \frac{\lambda^x}{x!} = \frac{1}{x!} e^{x \ln \lambda - \lambda} = C(\theta) e^{\theta T(x)} g(x) = e^{\theta T(x) - a(\theta)} g(x),$$

wobei

$$a(y) = e^y, \quad T(x) = x, \quad \theta = \ln \lambda, \quad a(\theta) = \lambda \quad \text{und} \quad g(x) = \frac{1}{x!}.$$

Hieraus ergibt sich die Log-Likelihood-Gleichung, bzw. der ML Schätzer

$$\sum_{i=1}^n T(x_i) = n \frac{\partial a}{\partial \theta}(\theta) = n a'(\theta) \quad \Rightarrow \quad \sum_{i=1}^n x_i = n e^\theta = n \lambda \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Bemerkung 6. Der Parametervektor θ ist in diesem Fall eindimensional und man kann eigentlich nicht von partiellen Ableitungen sprechen, da die Poisson-Verteilung nur einen Parameter besitzt. Aus Gründen der Analogie zum Paragraphen über ML Schätzer wurde diese Notation dennoch verwendet.

Für eine negativbinomialverteilte Zufallsvariable $X \sim \text{NegBin}(r, p)$ sind der Erwartungswert und die Varianz gegeben durch

$$E(X) = r(1-p)/p \quad \text{und} \quad \text{Var}(X) = r(1-p)/p^2.$$

In Reiss/Thomas (2001), Seite 108 werden die Parameter der Negativbinomialverteilung durch die Momentenmethode geschätzt. Man ersetzt den Erwartungswert und die Varianz durch \bar{x} und s^2 (vgl. (3a) und (3b)) und löst die sich dann ergebenden Gleichungen in r und p . Als Schätzer erhält man

$$\hat{p} = \bar{x}/s^2, \quad \hat{r} = \bar{x}^2/(s^2 - \bar{x}).$$

4.3 Schadenshöhe

Im Gegensatz zum IMA benötigt man beim LDA eine Schätzung der gesamten Schadenshöhenverteilung, um später mit Hilfe von (1) und der Schadenszahlverteilung die Verteilung des Verlustes L zu berechnen.

Zur Modellierung der Schadenshöhe gehen wir von positiven, unabhängigen und identisch verteilten Zufallsvariablen $X_i : (\Omega, \mathfrak{A}) \rightarrow (\mathbb{R}, \mathbb{B})$ mit der Borelschen σ -Algebra \mathbb{B} aus, d.h. $X_i > 0$, $X_i \sim X \forall i$.

Wenn man die empirischen Daten der Schäden betrachtet, die durch operationelle Risiken verursacht wurden, stellt man fest, dass es Schäden mit

- großer Häufigkeit und geringer Höhe
- mittlerer Häufigkeit und Höhe und
- geringer Häufigkeit und großer Schadenshöhe gibt.

Die letzte Kategorie ist diejenige, die für das Institut die 'gefährlichste' ist, weil schon ein einmaliger Schaden dieser Art zum Ruin führen kann. Die Existenz solcher Schäden spiegelt sich in einer sogenannten schweren Flanke (heavy tail) der Schadenshöhenverteilung wider. Verluste in großer Höhe haben eine nicht zu vernachlässigende Eintrittswahrscheinlichkeit. Die Verteilung von X in ihrer Gesamtheit mit einer einzigen theoretischen Verteilung und hinreichender Genauigkeit zu modellieren ist in den meisten Fällen nicht möglich.

Aus diesem Grund teilt man den Wertebereich der Schadenshöhen in den Bereich der 'normalen' und der 'extremen' Verluste und betrachtet beide getrennt. Als normale Schäden sind solche mit hoher Häufigkeit und geringem Verlust und die mittleren Schäden zu sehen. Nun werden in der Praxis nicht alle operationellen Verluste in eine Datenbank aufgenommen, sondern nur diejenigen ab einer gewissen Höhe l . Darunter liegen keine Schäden vor, folglich ist der Bereich der 'normalen' Schadenshöhen beschränkt. Zu einer gegebenen Schwelle u erhält man somit folgende Partition:

$$\mathcal{N} := \{X_i \mid X_i < u\} \quad \text{und} \quad \mathcal{E} := \{X_i \mid X_i \geq u\}.$$

Man beachte, dass die beiden Bereiche disjunkt sind und aufgrund dieser Annahme eine separate Betrachtung und somit eine Modellierung durch zwei sich nicht überlappende Verteilungsfunktionen möglich ist. Diesen Ansatz, die Schadenshöhenverteilung aus zwei Verteilungen zusammensetzen, findet man explizit in Chapelle et al. (2004) und Di Clemente/Romano (2004). Embrechts/Kaufmann/Samorodnitsky (2002) sprechen dieses Thema nur kurz an, aber erläutern es nicht so ausführlich wie die beiden anderen Quellen.

Für das mathematische Modell der Schadenshöhen ergibt sich somit Folgendes:

1. Annahme: Der zufällige Schaden X eines operationellen Verlustes ist nach unten beschränkt, d.h. $X \geq l$.
2. Annahme: Der Schaden X bedingt danach, dass er geringer als u ist, verhält sich wie eine lognormalverteilte Zufallsvariable Y gegeben, dass Y im Intervall $[l, u]$ liegt, d.h.

$$P(X \in \cdot | X \leq u) = P(Y \in \cdot | Y \in [l, u]) \quad (9)$$

mit einer lognormalverteilten Zufallsvariablen Y .

3. Annahme: Oberhalb der Schwelle u verhält sich die Schadenshöhenverteilung wie eine GPD, d.h.

$$P(X \in \cdot | X > u) = G_{\xi, u, \sigma}(\cdot). \quad (10)$$

Hieraus ergibt sich insgesamt für die Verteilung der Schadenshöhen ($A \in \mathbb{B}$):

$$P(X \in A) = P(X \in A | X > u) P(X > u) + P(X \in A | X \leq u) P(X \leq u). \quad (11)$$

Die Wahl von u ist hierbei ausschlaggebend, da diese die Daten festlegt, auf denen die weitere Schätzung der Parameter basiert. Der Entscheidung über die Höhe dieser Schwelle u wenden wir uns im Abschnitt über die Schätzung für die 'extremen' Schäden zu. Sie sei im folgenden Paragraphen als gegeben vorausgesetzt. Allerdings vermutet man intuitiv, dass u relativ groß sein muss, um eine Abgrenzung zwischen 'normalen' und 'extremen' Schaden darzustellen.

4.4 Schätzung der 'normalen' Schäden

Für die Schätzung der Schadenshöhenverteilung der 'normalen' Schäden lassen sich klassische Schätzmethode verwenden. Allerdings ist darauf zu achten, dass der Bereich der Schadenshöhen von oben durch die Schwelle u begrenzt ist und es in der Realität auch eine untere Schranke gibt, unterhalb derer Daten der eingetretenen Schäden nicht registriert wurden.

Um einen eingeschränkten Wertebereich bei der Schätzung mit einzubeziehen, muss die Dichte der angenommenen Lognormalverteilung so modifiziert werden, dass sie zu der abgeschnittenen Verteilungsfunktion passt. Aus der obigen Formel (9) ergibt sich

$$P(Y \leq y \mid Y \in [l, u]) = \frac{P(l \leq Y \leq y)}{P(Y \in [l, u])} = \frac{F(y) - F(l)}{F(u) - F(l)}$$

$$\Rightarrow f^*(y) = \frac{f(y)}{F(u) - F(l)} \quad \text{mit}$$

Y lognormalverteilte Zufallsvariable

l, u Schranken

F Verteilungsfunktion von Y

f Dichte von F

f^* bedingte Dichte

Die Lognormalverteilung der Zufallsvariablen Y ist gegeben durch $\ln Y \sim \mathcal{N}(\mu, \sigma)$, d.h.

$$F(y) = P(Y \leq y) = P(\ln Y \leq \ln y) = \mathcal{N}(\mu, \sigma)(\ln y) \quad \text{und}$$

$$f(y) = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\ln y} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt.$$

Die Dichte der Lognormalverteilung ist gegeben durch

$$f(y) = \frac{1}{\sqrt{2\pi} \sigma y} e^{-\frac{1}{2}(\frac{\ln y - \mu}{\sigma})^2}.$$

Bemerkung 7. Die Parameter der Lognormalverteilung $\mu := E(\ln Y)$ und $\sigma := \sqrt{\text{Var}(\ln Y)}$ entsprechen nicht dem Erwartungswert, bzw. der Standardabweichung der Zufallsvariablen Y . Diese sind gegeben durch

$$E(Y) = e^\mu e^{\frac{\sigma^2}{2}} \quad \text{und} \quad \sqrt{\text{Var}(Y)} = \sqrt{e^{\sigma^2} (e^{\sigma^2} - 1) e^{2\mu}}.$$

Man beachte, dass sowohl $f(y)$ als auch $F(y)$ von beiden Parametern abhängen.

Sei die untere Schranke l und die obere u bekannt und liegen Beobachtungen $\mathbf{x} = (x_1, \dots, x_n)$ aus dem Intervall $[l, u]$ vor. Die Parameter μ und σ der Lognormalver-

teilung können mittels der Maximum-Likelihood Methode geschätzt werden. Als Log-Likelihood-Funktion beziehungsweise Log-Likelihood-Gleichungen erhält man:

$$\begin{aligned}
 L_\theta(\mathbf{x}) &= \sum_{i=1}^n \ln f^*(x_i) = \sum_{i=1}^n \ln \left(\frac{f(x_i)}{F(u) - F(l)} \right) \\
 \frac{\partial}{\partial \mu} L_\theta(\mathbf{x}) &= \frac{\partial}{\partial \mu} \sum_{i=1}^n \ln \left(\frac{f(x_i)}{F(u) - F(l)} \right) \stackrel{!}{=} 0 \\
 \Rightarrow \frac{\partial}{\partial \sigma} L_\theta(\mathbf{x}) &= \frac{\partial}{\partial \sigma} \sum_{i=1}^n \ln \left(\frac{f(x_i)}{F(u) - F(l)} \right) \stackrel{!}{=} 0.
 \end{aligned}$$

Hier muss man ein numerisches Verfahren zur Lösung der Gleichungen heranziehen.

4.5 Schätzung der 'extremen' Schäden

Allgemein kommt der Modellierung der rechten Flanke ein besonderes Augenmerk zu, weil die meisten Verfahren bei der Messung des Risikos die Schätzung eines hohen Quantils der Verlustverteilung miteinbeziehen. Innerhalb eines Teilssektors ist in erster Linie eine gute Schätzung der gesamten Schadenshöhenverteilung wichtig, weil diese zur Berechnung der Verlustverteilung des Sektors benötigt wird, d.h. ein Quantil der Verteilung der Schadenshöhen wird gar nicht gebildet. Dennoch ist eine gute Modellierung der Schadensverteilung schon auf der untersten Ebene ausschlaggebend, um die extremen Verluste auch auf der Gesamtbankebene nicht zu unterschätzen. Wie im Paragraphen über EVT sind auch hier die Hauptreferenzen Reiss/Thomas (2001) und Embrechts/Klüppelberg/Mikosch (1997).

Anpassung einer GPD an die Überschreitungsverteilung

Im sogenannten POT-Modell wird aufgrund des Balkema-de Haan-Pickands-Theorems die Annahme getroffen $F_u(x) \approx G_{\xi,u,\sigma}(x)$. Die Verteilungsfunktion lässt sich also für $x \geq u$ schätzen, indem man $F(u)$ durch die empirische Verteilungsfunktion an der Stelle u schätzt und $F_u(x)$ durch eine GPD approximiert. So ergibt sich als Schätzer für die Verteilungsfunktion oberhalb der Schranke u mit Hilfe der Formel (7)

$$\hat{F}(x) = \left(1 - \hat{F}_n(u)\right) G_{\xi,u,\sigma}(x) + \hat{F}_n(u).$$

Einen Quantilschätzer für das p -Quantil x_p kann man aus dem Ansatz

$$\hat{F}(x_p) = \left(1 - \hat{F}_n(u)\right) G_{\xi,u,\sigma}(x_p) + \hat{F}_n(u) = p \quad \text{als}$$

$$\hat{x}_p = G_{\xi, u, \sigma}^{-1} \left(\frac{p - \hat{F}_n(u)}{1 - \hat{F}_n(u)} \right) = u + \frac{\sigma}{\xi} \left(\left(1 - \frac{p - \hat{F}_n(u)}{1 - \hat{F}_n(u)} \right)^{-\xi} - 1 \right) \text{ gewinnen.}$$

Dies lässt sich vereinfachen zu

$$\hat{x}_p = u + \frac{\sigma}{\xi} \left(\left(\frac{1-p}{1 - \hat{F}_n(u)} \right)^{-\xi} - 1 \right) = u + \frac{\sigma}{\xi} \left(\left(\frac{n}{k} (1-p) \right)^{-\xi} - 1 \right),$$

wenn man beachtet, dass $1 - \hat{F}_n(u)$ gerade $\frac{k}{n}$ ist. Hierbei gehen wir davon aus, dass es k Überschreitungsdaten innerhalb der n ursprünglichen Daten $x_{n:n} \leq \dots x_{k:n} \leq \dots \leq x_{1:n}$ gibt, die über u liegen. Die Schätzung der Parameter der GPD berücksichtigt nur diese k Daten, weil nur sie als relevant für das Verhalten der rechten Flanke angesehen werden. Die restlichen Daten verhalten sich ja gerade nicht wie eine verallgemeinerte Pareto-Verteilung und können somit auch keine Informationen zur Parameterschätzung von ξ , μ und σ beitragen. Man beachte, dass das Grenzverhalten der Verteilung $F_u(x)$ im Satz von Balkema-de Haan-Pickands für $u \rightarrow \infty$ gilt. Dies bedeutet, dass rein theoretisch die Fehlerrate der Approximation für $k \rightarrow 0$ gegen Null strebt. $k \rightarrow 0$ besagt aber nichts anderes als eine immer kleiner werdende Anzahl von Überschreitungsdaten. Es besteht also ein Konflikt zwischen der Genauigkeit des Modells und der Aussagekraft der geschätzten Parameter, die bei größeren oberen Schwellen abnimmt. Bei der Anpassung der GPD an die Überschreitungsdaten muss deshalb generell ein Gleichgewicht zwischen dem Bias und der Varianz des Parameterschätzers gefunden werden.

Eine wichtige Rolle in der Extremwerttheorie spielt die sogenannte mittlere Exzessfunktion (MEF, Mean Excess Function).

Definition 5.

$$e_F(u) := E(X - u | X > u) = \int x \, dF_{(u)}(x) \quad \text{für } 0 \leq u < x_R(F)$$

heißt mittlere Exzessfunktion der Verteilungsfunktion F , bzw. der Zufallsvariablen X .

Dies ist der Erwartungswert der Überschreitungen $Y = X - u$ bezüglich der Verteilung $F_{(u)}$ als Funktion in u . Betrachtet man die der Größe nach geordneten Daten $x_{n:n} \leq \dots x_{2:n} \leq x_{1:n}$, so kann man die empirische MEF darstellen als

$$e_n(u) := e_{\hat{F}_n}(u) := \frac{1}{k} \sum_{i=1}^k (x_{i:n} - u),$$

wobei $K := \{x_i | x - i > u\} = \{x_{k:n}, \dots, x_{1:n}\}$ die Menge und $k = |K|$ die Anzahl der Beobachtungen oberhalb von u ist.

Offensichtlich stehen die Anzahl der Überschreitungsdaten und die zugehörige Schwelle in einem direkten Verhältnis, denn k hängt von u ab. Jeder Schwelle u lässt sich eine eindeutige Anzahl von Überschreitungsdaten zuordnen, für jede vorgegebene Anzahl lassen sich jedoch theoretisch mehrere Schwellen u_j finden. Als Konvention sei die zugehörige Schwelle durch $u_k := x_{k:n}$ als k -größter Wert der Daten definiert. Den zugehörigen Plot, der die empirische Mean Excess Function (MEF) $e_{\hat{F}_n}(u)$ gegen verschiedene Schwellen u abträgt, nennt man empirischen Mean Excess Plot (MEP). Zu Lemma 5 vergleiche Reiss/Thomas (2001), Seite 53-55 und speziell zum Beweis Han (2003), Seite 19.

Lemma 5. *Für die Mean Excess Function gilt mit $u < x_R(F)$*

(i) $e_{F_{\mu,\sigma}}(u) = \sigma e_F\left(\frac{x-\mu}{\sigma}\right)$

(ii) $e_{G_{\xi,0,\sigma}}(u) = \frac{\sigma-\xi u}{1-\xi}$ ist linear für $0 < \xi < 1$ und konstant für die Exponentialverteilung.

Beweis. (i) Sei $X \sim F$ und $Y := \sigma X + \mu \sim F_{\mu,\sigma}$ mit Lokationsparameter μ und Skalenparameter σ . Dann gilt wegen der Linearität der bedingten Erwartung

$$\begin{aligned} e_{F_{\mu,\sigma}}(u) &= E(Y - u | Y > u) = E\left(\sigma \left(X - \frac{u - \mu}{\sigma}\right) \mid X > \frac{u - \mu}{\sigma}\right) \\ &= \sigma E\left(X - \frac{u - \mu}{\sigma} \mid X > \frac{u - \mu}{\sigma}\right) = \sigma e_F\left(\frac{u - \mu}{\sigma}\right). \end{aligned}$$

(ii) Wegen

$$\begin{aligned} E((X - u) 1_{\{X > u\}}) &= \int_u^\infty (X - u) dF(x) = \lim_{R \rightarrow \infty} \int_u^R (X - u) f(x) dx \\ &= \lim_{R \rightarrow \infty} \left(\left[(x - u) F(x) \right]_u^R - \int_u^R F(x) dx \right) \\ &= \lim_{R \rightarrow \infty} \left((R - u) F(R) - \int_u^R F(x) dx \right) \\ &= \lim_{R \rightarrow \infty} \left(\int_u^R 1 dx - \int_u^R F(x) dx \right) \\ &= \lim_{R \rightarrow \infty} \int_u^R (1 - F(x)) dx = \int_u^\infty (1 - F(x)) dx \end{aligned}$$

lässt sich die mittlere Exzessfunktion grundsätzlich auch in der Form

$$e_F(u) = \frac{E((X - u) 1_{\{X > u\}})}{P(X > u)} = \frac{\int_u^\infty (1 - F(x)) dx}{P(X > u)} \quad \text{schreiben.}$$

Hieraus folgt für eine Exponentialverteilung $F = G_0$

$$\frac{1}{1 - F(u)} \int_u^\infty (1 - F(x)) dx = \frac{1}{e^{-u}} \int_u^\infty e^{-x} dx = \frac{1}{e^{-u}} \left[-e^{-x} \right]_u^\infty = 1$$

$$\Rightarrow e_{G_{0,\mu,\sigma}}(u) = \sigma \quad , \text{ d.h. die MEF ist konstant.}$$

Für $\xi \neq 0$ erhält man

$$e_{G_\xi} = \frac{1}{(1 + \xi u)^{-\frac{1}{\xi}}} \int_u^\infty (1 + \xi x)^{-\frac{1}{\xi}} dx.$$

Das Integral konvergiert nur für $0 < \xi < 1$. Dann gilt

$$\int_u^\infty (1 + \xi x)^{-\frac{1}{\xi}} dx = \frac{1}{1 - \xi} (1 + \xi u)^{-\frac{1}{\xi} + 1}$$

$$\Rightarrow e_{G_{\xi,\mu,\sigma}}(u) = \begin{cases} \frac{1 + \xi u}{1 - \xi}, & 0 < \xi < 1 \\ \infty, & 1 \leq \xi \end{cases}.$$

□

Aus diesem Resultat leitet sich eine Methode zur grafischen Bestimmung der Schwelle u ab. Da wir eine GPD an die Verteilungsfunktion F oberhalb der Schwelle u anpassen wollen, muss sich die oben genannte Eigenschaft der MEF einer GPD auch in den Daten widerspiegeln.

- Plote die empirische MEF, d.h. die Punkte $\{(x_{i:n}, e_n(x_{i:n})) \mid i = 1, \dots, n\}$.
- Bestimme u so, dass für alle $x_{i:n}$ größer als u der MEP als Gerade verläuft.

Die Wahl der Schwelle u erfolgt in den meisten Fällen mit grafischen Hilfsmitteln. Doch gibt es auch rechnerische Methoden. Chapelle et al. (2004), Seite 7 bzw. 12 ermitteln die Schwelle so, dass die beste Anpassung der GPD an die Daten gemäß einer Goodness-of-Fit-Statistik vorliegt. Di Clemente/Romano (2004) wählen u wie folgt $u := \max \{x \mid F(x) < 1 - \frac{N_x}{n}\}$, wobei N_x die Anzahl der Schäden über x ist und n die der gesamten Beobachtungen, sowie F die Verteilung, die die Schadenshöhenverteilung im Zentrum modellieren soll. Dadurch ist die Monotonie der zusammengesetzten Verteilung sichergestellt.

Hat man die Schwelle u geschätzt, so ist dies der Lokationsparameter der GPD, die die Überschreitungsdaten oberhalb der Schwelle modellieren soll. Es müssen also noch der Gestaltungsparameter ξ und der Skalenparameter σ bestimmt werden. Für die Schätzung

des Gestaltparameters stehen mehrere Ansätze zur Verfügung. Der Skalenparameter kann z.B. mit Hilfe eines MLE geschätzt werden.

Maximum-Likelihood Schätzung

An die Überschreitungsdatenverteilung $F_u(x)$ der Daten $\mathbf{x} = (x_1, \dots, x_k)$ soll eine GPD $G_{\xi, u, \sigma}(x)$ angepasst werden. Die Dichte $g(x) = \frac{1}{\xi} (1 + \xi \frac{x-u}{\sigma})^{-\frac{1}{\xi}-1}$, für $x \in D(\xi, u, \sigma)$ dieser GPD führt zur Log-Likelihood-Funktion:

$$\begin{aligned} L_{\theta}(\mathbf{x}) &= \ln \left(\prod_{i=1}^k \frac{1}{\sigma} \left(1 + \xi \frac{x_i - u}{\sigma} \right)^{-\frac{1}{\xi}-1} \right) = \ln \left(\frac{1}{\sigma^k} \prod_{i=1}^k \left(1 + \xi \frac{x_i - u}{\sigma} \right)^{-\frac{1}{\xi}-1} \right) \\ &= -k \ln \sigma - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^k \ln \left(1 + \xi \frac{x_i - u}{\sigma} \right) \Rightarrow \\ \frac{\partial L_{\theta}(\mathbf{x})}{\partial \xi} &= - \left(-\frac{1}{\xi^2} \sum_{i=1}^k \ln \left(1 + \xi \frac{x_i - u}{\sigma} \right) + \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^k \frac{x_i - u}{\sigma (1 + \xi \frac{x_i - u}{\sigma})} \right) \\ &= \frac{1}{\xi^2} \sum_{i=1}^k \ln \left(1 + \xi \frac{x_i - u}{\sigma} \right) - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^k \frac{x_i - u}{\sigma + \xi (x_i - u)} \stackrel{!}{=} 0 \\ \frac{\partial L_{\theta}(\mathbf{x})}{\partial \sigma} &= -\frac{k}{\sigma} + \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^k \frac{\xi (x_i - u)}{\sigma^2 + \sigma \xi (x_i - u)} \\ &= -k \frac{1}{\sigma} + \frac{\xi + 1}{\xi} \frac{\xi}{\sigma} \sum_{i=1}^k \frac{x_i - u}{\sigma + \xi (x_i - u)} \stackrel{!}{=} 0. \end{aligned}$$

Die partiellen Ableitungen der Log-Likelihood-Funktion müssen Null gesetzt werden, als notwendige Bedingung für das Vorliegen eines Maximums. Die resultierende Gleichungen kann man nicht analytisch lösen, sondern muss auf ein numerisches Iterationsverfahren zurückgreifen.

Wahrscheinlichkeits-gewichtete Momentenmethode

Ähnlich der bereits angesprochenen Momentenmethode werden bei der wahrscheinlichkeits-gewichteten Momentenmethode (method of probability-weighted moments) theoretische Momente durch entsprechende Momente der Stichprobe geschätzt. Allerdings kommt hier als zusätzliche Komponente eine Wahrscheinlichkeit hinzu. Da für $\xi \geq 1$ keine Erwartungswerte und endliche Varianzen existieren, ist diese Methode nur begrenzt verwendbar. Die folgenden Aussagen findet man in Embrechts/Klüppelberg/Mikosch (1997), Seite 358.

Definiere für $X \sim G_{\xi,0,\sigma} =: G$

$$w_j(\xi, \sigma) := E(X (\bar{G}(X))^j) = \int x (\bar{G}(x))^j dG(x) \quad \text{und} \quad \hat{w}_j(\xi, \sigma) := \int x (\bar{G}(x))^j d\hat{G}_n(x)$$

als die theoretischen, bzw. empirischen wahrscheinlichkeits-gewichteten Momente von X , wobei $\hat{G}_n(x)$ die empirische Verteilungsfunktion von X ist. Es gilt

$$\begin{aligned} w_j(\xi, \sigma) &= \int x (\bar{G}(x))^j dG(x) = \int_0^1 G^{-1}(y) (1-y)^j dy \\ &= \int_0^1 \frac{\sigma}{\xi} ((1-y)^{-\xi} - 1) (1-y)^j dy \\ &= \frac{\sigma}{\xi} \int_1^0 (z^{-\xi} - 1) z^j (-1) dz = \frac{\sigma}{\xi} \int_0^1 z^{-\xi+j} - z^j dz \\ &= \frac{\sigma}{\xi} \left[\frac{1}{1-\xi+j} z^{1-\xi+j} - \frac{1}{1+j} z^{1+j} \right]_0^1 \\ &= \frac{\sigma}{\xi} \left(\frac{1}{1-\xi+j} - \frac{1}{1+j} \right) = \frac{\sigma}{(1-\xi+j)(1+j)} \end{aligned}$$

für die theoretische Größe und

$$\begin{aligned} \hat{w}_j(\xi, \sigma) &= \int x (\bar{G}(x))^j d \left(\frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}}(\omega) \right) = \frac{1}{n} \sum_{i=1}^n \int x (\bar{G}(x))^j d 1_{\{X_i \leq x\}}(\omega) \\ &= \frac{1}{n} \sum_{i=1}^n X_{i:n} (\bar{G}(X_{i:n}))^j \end{aligned}$$

Für $j = 0, 1$ gilt:

$$\begin{aligned} w_0(\xi, \sigma) &= \frac{\sigma}{1-\xi} \quad \text{und} \quad w_1(\xi, \sigma) = \frac{\sigma}{2(2-\xi)} \\ \Rightarrow \quad w_0(\xi, \sigma) - 2w_1(\xi, \sigma) &= \frac{\sigma}{(1-\xi)(2-\xi)} \end{aligned}$$

und hieraus folgt

$$\begin{aligned} \sigma &= \frac{\sigma^2}{(1-\xi)(2-\xi)} \Big/ \frac{\sigma}{(1-\xi)(2-\xi)} = \frac{2w_0(\xi, \sigma)w_1(\xi, \sigma)}{w_0(\xi, \sigma) - 2w_1(\xi, \sigma)} \\ \text{und} \quad \xi &= 2 - \frac{w_0(\xi, \sigma)}{w_0(\xi, \sigma) - 2w_1(\xi, \sigma)}. \end{aligned}$$

Um einen Schätzer für ξ, σ zu erhalten sind die Gleichungen $w_j(\xi, \sigma) = \hat{w}_j(\xi, \sigma)$ für $j = 0, 1$ zu lösen, d.h. $\hat{\xi}, \hat{\sigma}$ sind so zu bestimmen, dass sie die Lösung dieses Gleichungssystems sind.

Der Pickands Schätzer

Der Pickands Schätzer für $\xi \in \mathbb{R}$ ergibt sich aus folgenden Überlegungen (vgl. Embrechts/Klüppelberg/Mikosch (1997), Seite 327 ff.):

- Sei die verallgemeinerte Extremwertverteilung (EVD, extreme value distribution) gegeben durch

$$H_\xi = \begin{cases} \exp\left(-\left(1 + \xi x\right)^{-\frac{1}{\xi}}\right) & \text{falls } \xi \neq 0 \\ e^{-e^{-x}} & \text{falls } \xi = 0 \end{cases}.$$

Der Definitionsbereich ist für $\xi > 0$ durch $x > -\frac{1}{\xi}$, für $\xi < 0$ durch $x < -\frac{1}{\xi}$ und für $\xi = 0$ durch $x \in \mathbb{R}$ gegeben. Die Extremwertverteilungen ergeben sich als einzige nicht-entartete Grenzverteilungen passend normierter Maxima von unabhängigen und identisch verteilten Zufallsvariablen. Die Beziehung

$$G_\xi(x) = -\ln H_\xi(x)$$

zeigt den engen Zusammenhang zwischen der GPD und der EVD. Analog zum POT-Anziehungsbereich gibt es auch hier einen sogenannten Maximum-Anziehungsbereich (*MDA*, maximum domain of attraction). Die meisten der klassischen Verteilungsfunktionen sind im Maximum-Anziehungsbereich einer EVD. Im Folgenden sei diese Bedingung für die Verteilungsfunktion F der Schadenshöhen erfüllt, d.h. $F \in MDA(H_\xi)$. Sei $U(t) := F^{\leftarrow}\left(1 - \frac{1}{t}\right)$ das $\left(1 - \frac{1}{t}\right)$ -Quantil der Verteilung F .

- Es gilt

$$\lim_{u \uparrow x_R(F)} |F_u(x) - G_{\xi, \sigma}(x)| = 0 \Leftrightarrow F \in MDA(H_\xi) \Leftrightarrow \lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & \text{falls } \xi \neq 0 \\ \frac{\ln x}{\ln y} & \text{falls } \xi = 0 \end{cases}. \quad (12)$$

- Aus (12) erhält man mit $x = 2$, bzw. $x = 1/2$ durch Division

$$\lim_{s \rightarrow \infty} \frac{U(2s) - U(s)}{U(s/2) - U(s)} = \frac{2^\xi - 1}{(1/2)^\xi - 1} = -2^\xi \quad \Leftrightarrow \quad \lim_{s \rightarrow \infty} \frac{U(2s) - U(s)}{U(s) - U(s/2)} = 2^\xi.$$

Für eine positive Funktion mit $\lim_{s \rightarrow \infty} c(s) = 2$ bleibt der Grenzwert erhalten, wenn man die 2 auf der linken Seite durch $c(s)$ ersetzt.

- Für Pareto-verteilte Zufallsvariablen V_1, \dots, V_n bzw. deren entsprechende Ordnungsstatistiken gilt $(X_{k:n})_{k=1, \dots, n} \stackrel{d}{=} (U(V_{k:n}))_{k=1, \dots, n}$, wobei $X_{k:n}$ Ordnungsstatistiken Pareto-verteilter Zufallsvariablen X_1, \dots, X_n sind. Dies kann mit Hilfe der Quantiltransformation gezeigt werden:

$$U(V_k) \stackrel{d}{=} F^{\leftarrow}(F_V(V_k)) \stackrel{d}{=} F^{\leftarrow}(U_k) \stackrel{d}{=} X_k,$$

wobei $F_V(t) = 1 - \frac{1}{t}$ die Verteilungsfunktion der Pareto-Verteilung ist und $U_k \sim U_{[0,1]}$ gleichverteilte Zufallsvariablen auf $[0, 1]$ sind.

- Da es sich bei den $V_{k:n}$ um die empirischen $(1 - \frac{k}{n})$ -Quantile handelt, lässt sich zeigen, dass

$$\frac{k}{n} V_{k:n} \xrightarrow{P} 1, \quad V_{k:n} \xrightarrow{P} \infty, \quad \frac{V_{2k:n}}{V_{k:n}} \xrightarrow{P} \frac{1}{2} \quad \text{für } n \rightarrow \infty.$$

Letzteres ergibt sich durch $2 \frac{V_{2k:n}}{V_{k:n}} = \frac{2k}{n} V_{2k:n} / \frac{k}{n} V_{k:n} \rightarrow 1, n \rightarrow \infty$.

- Insgesamt erhalten wir also

$$\frac{X_{k:n} - X_{2k:n}}{X_{2k:n} - X_{4k:n}} \stackrel{d}{=} \frac{U(V_{k:n}) - U(V_{2k:n})}{U(V_{2k:n}) - U(V_{4k:n})} = \frac{U(\frac{V_{k:n}}{V_{2k:n}} V_{2k:n}) - U(V_{2k:n})}{U(V_{2k:n}) - U(\frac{V_{2k:n}}{V_{4k:n}} V_{2k:n})} \xrightarrow{P} 2^\xi$$

und damit den Pickands Schätzer

$$\hat{\xi}_{k,n}^{(P)} = \frac{1}{\ln 2} \ln \frac{X_{k:n} - X_{2k:n}}{X_{2k:n} - X_{4k:n}}.$$

Wie man bereits an der Notation erkennen kann, hängt dieser Schätzer von k ab, d.h. von der Anzahl der zu berücksichtigenden Daten oberhalb der Schwelle u . Um nun ein geeignetes k zu wählen, greift man ähnlich wie beim MEP auf eine grafische Entscheidungshilfe zurück. Man plottet die Punkte $\{(k, \hat{\xi}_{k,n}^{(P)}) \mid k = 1, \dots, n\}$ und bestimmt k so, dass rechts davon eine Waagerechte zu erkennen ist. Der Pickands Schätzer ist schwach konsistent, d.h. es gilt

$$\hat{\xi}_{k,n}^{(P)} \xrightarrow{P} \xi \quad \text{für } n \rightarrow \infty.$$

Der Hill Schätzer

Dieser Schätzer eignet sich für $\xi > 0$. Im Finanzbereich wird häufig eine Verteilung mit schwerer Flanke unterstellt. Deswegen ist ein positiver Gestaltungsparameter, der einer

Pareto-Verteilung entspricht, durchaus üblich und der Hill Schätzer weit verbreitet. Er ist durch

$$\hat{\xi}_{k,n}^{(H)} := \frac{1}{k} \sum_{j=1}^k \ln \left(\frac{X_{j:n}}{X_{k:n}} \right) = \frac{1}{k} \sum_{j=1}^k \ln(X_{j:n}) - \ln(X_{k:n}) \quad \text{gegeben.}$$

Der Hill Schätzer kann insoweit als natürlicher Schätzer angesehen werden, als dass er sich aus verschiedenen Ansätzen herleiten läßt. Nimmt man zum Beispiel das sogenannte beschränkte Pareto-Modell der α -Parametrisierung mit Lokationsparameter 0 und Skalenparameter u , d.h. als Verteilungsfunktion der Überschreitungsdaten wird die Verteilung $F := G_{\alpha,0,u}$ mit linkem Endpunkt u und Gestaltsparameter $\alpha = \frac{1}{\xi} > 0$ gewählt, so zeigen folgende Überlegungen, dass der MLE dieses Teilmodells gerade der Hill Schätzer ist.

- Wir nehmen an, dass sich die k Beobachtungen der n Ursprungsdaten oberhalb der Schwelle u wie eine Pareto-Verteilung verhalten.
- Besitzen also $X_i \sim X$, $i = 1, \dots, k$ eine Pareto-Verteilung $G_{\alpha,0,u}(x) = F(x) = 1 - \left(\frac{x}{u}\right)^{-\alpha}$, dann gilt mit $C := u^\alpha$ und $Y := \ln X$

$$P(X < x) = \bar{F}(x) = \left(\frac{x}{u}\right)^{-\alpha} = C x^{-\alpha} \quad \Rightarrow \quad P(Y > y) = P(X > e^y) = \left(\frac{e^y}{u}\right)^{-\alpha}.$$

Das heißt, dass es sich bei Y um eine exponentialverteilte Zufallsvariable handelt mit Verteilungsfunktion $F^*(v) = e^{-\alpha v} = e^{-\alpha(y - \ln u)}$. Der MLE des Parameters α der Exponentialverteilung und somit des Gestaltsparameters ist gegeben durch

$$\hat{\alpha} = \left(\frac{1}{k} \sum_{i=1}^k \ln \left(\frac{X_i}{u} \right) \right)^{-1} \quad \Leftrightarrow \quad \hat{\xi} = \frac{1}{k} \sum_{i=1}^k \ln \left(\frac{X_i}{u} \right)$$

Ein andere Ansatz kann über sogenannte regulär variierende Funktionen erfolgen:

- Unter einer *regulär variierenden Funktion mit Index* $\alpha \in \mathbb{R}$ versteht man eine positive, Lebesgue-meßbare Funktion h auf $(0, \infty)$ mit

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\alpha, \quad t > 0.$$

- Eine weitere Analogie zwischen der GPD und der EVD zeigt sich durch eine α -Parametrisierung der EVD (vgl. Reiss/Thomas (2001), Seite 15 ff. und Embrechts/Kluppelber/Mikosch (1997), Seite 121). Mittels dieser Parametrisierung

lässt sich folgende Äquivalenz zeigen (siehe hierzu Embrechts/Klüppelberg/Mikosch (1997), Seite 131):

$$F \in MDA(H_\alpha), \alpha > 0 \quad \Leftrightarrow \quad \lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha}, \quad \alpha, x > 0,$$

d.h. F ist im Maximum-Anziehungsbereich einer sogenannten Fréchetverteilung ($\alpha > 0$), genau dann, wenn die Flankenfunktion $\bar{F}(x)$ regulär variierend mit Index $-\alpha$ ist.

- Hieraus erhalten wir die Darstellung

$$\frac{1}{\bar{F}(t)} \int_t^\infty \ln\left(\frac{x}{t}\right) dF(x) \rightarrow \frac{1}{\alpha}, \quad t \rightarrow \infty.$$

- Ersetzt man die Schwelle u durch die $k + 1$ -te Ordnungsstatistik, so liegen k Beobachtungen oberhalb dieser Schwelle und man erhält als Schätzer für den Gestaltsparameter

$$\hat{\xi} = \frac{1}{\hat{\alpha}} = \frac{1}{k} \int_{X_{k+1:n}}^\infty \ln\left(\frac{x}{X_{k+1:n}}\right) dF_n(x) = \frac{1}{k} \sum_{i=1}^k \ln\left(\frac{X_{i:n}}{X_{k+1:n}}\right).$$

- Man beachte, dass auch hier wieder eine Schätzung im Pareto-Modell, d.h. $\xi = \frac{1}{\alpha} > 0$ vorliegt.

Wie bei der Wahl der Schwelle u und der Schätzung des Gestaltsparameters nach Pickands, nimmt man einen Plot zur Hilfe (der Hill Plot besteht aus den Punkten $\{(k, \hat{\xi}_{k,n}^{(H)}) \mid k = 1, \dots, n\}$ und sollte ab einem gewissen k waagrecht verlaufen).

Der Deckers-Einmahl-de Haan Schätzer

Der Hill Schätzer ist eigentlich nur für die Schätzung von positiven Gestaltsparametern ausgelegt. Als Erweiterung für $\xi \in \mathbb{R}$ kann man den folgenden Schätzer ansehen:

$$\hat{\xi}_{k,n}^{(DEH)} := 1 + H_{k,n}^{(1)} + \frac{1}{2} \left(\frac{(H_{k,n}^{(1)})^2}{H_{k,n}^{(2)}} - 1 \right)^{-1},$$

mit $H_{k,n}^{(1)} := \frac{1}{k} \sum_{i=1}^k \ln\left(\frac{X_{i:n}}{X_{k+1:n}}\right)$ und $H_{k,n}^{(2)} := \frac{1}{k} \sum_{i=1}^k \left(\ln\left(\frac{X_{i:n}}{X_{k+1:n}}\right) \right)^2$.

Man findet diesen Schätzer auch unter der Bezeichnung Momentenschätzer, weil man $H_{k,n}^{(1)}$ und $H_{k,n}^{(2)}$ als empirische Momente auffassen kann.

Mean Excess Function Ansatz

Man kann eine Schätzung von ξ auch direkt aus dem MEP gewinnen, indem man eine lineare Regression auf die MEF oberhalb der Schwelle u anwendet. Aus der Tatsache, dass die MEF einer GPD eine lineare Funktion ist, läßt sich ein Schätzer für den Gestaltsparameter ableiten.

$$e_{G_{\xi,\mu,\sigma}}(u) = \frac{1 + \xi u}{1 - \xi} = \frac{1}{1 - \xi} + \frac{\xi}{1 - \xi} u = \beta_0 + \beta_1 u \Rightarrow \xi = \frac{\beta_1}{1 + \beta_1}.$$

Der Parameter β_1 ist gerade die Steigung der Regressionsgeraden. Dieser Schätzer ist relativ ungenau, da die Punkte des MEP nur bei einer Stichprobe von einer GPD als Gerade zu erkennen sind und in praktischen Fällen eine sehr hohe Volatilität aufweisen können, was die Schätzung im Regressionsmodell ungenau macht – und folglich auch den Schätzer für ξ . Allerdings kann man so einen ersten Startwert für die Iteration des MLE finden.

Eigenschaften der Schätzer

Eine Diskussion über die verschiedenen Schätzer für den Gestaltsparameter findet man in Reiss/Thomas (2001), Seite 140 ff. und in McNeil/Saladin (1997), Seite 8/9. Für Werte des Gestaltsparameters zwischen 0 und 0.4 ist der Schätzer der wahrscheinlichkeits-gewichteten Momentenmethode dem Maximum-Likelihood Schätzer überlegen, weil er eine geringere mittlere quadratische Abweichung besitzt. Dieser Unterschied verwischt allerdings mit zunehmender Stichprobengröße. Ist $\xi \geq 0.5$, so führt die wahrscheinlichkeits-gewichtete Momentenmethode zu einem nicht erwartungstreuen Schätzer, während hier der MLE über den Vorteil der Konsistenz verfügt. Asymptotisch gute Eigenschaften besitzt der MLE für $\xi > -0.5$.

Aussagen zur asymptotischen Verteilung des Maximum-Likelihood Schätzers findet man sowohl in Embrechts/Klüppelberg/Mikosch (1997), Seite 356 als auch bei Reiss/Thomas (2001), Seite 152.

Bezeichnet $X_n \xrightarrow{d} X \Leftrightarrow \lim_{n \rightarrow \infty} P(X_n \leq x) = P(X \leq x)$ die Konvergenz in Verteilung, so gilt mit der bivariaten Standardnormalverteilung $\mathcal{N}(\mu, \Sigma)$:

$$\sqrt{n} \left(\hat{\xi}_n^{MLE} - \xi, \frac{\hat{\sigma}_n^{MLE}}{\sigma} - 1 \right) \xrightarrow{d} \mathcal{N}(0, M^{-1}), \quad n \rightarrow \infty \quad \text{mit } M^{-1} = (1 + \xi) \begin{pmatrix} 1 + \xi & 1 \\ 1 & 2 \end{pmatrix}.$$

Hierbei sind $\hat{\xi}_n^{MLE}$ und $\hat{\sigma}_n^{MLE}$ die MLE zum Stichprobenumfang n . Embrechts/Klüppelberg/Mikosch (1997) untersuchen auch die Eigenschaften des Hill und Pickands Schätzers. Ein wichtiges Resultat ist das asymptotische Verhalten dieser

Schätzer. Für den Pickands Schätzer kann gezeigt werden, dass unter gewissen Bedingungen gilt

$$\sqrt{k} (\xi_{k,n}^{(P)} - \xi) \xrightarrow{d} \mathcal{N}(0, v(\xi)), \quad n \rightarrow \infty,$$

wobei $v(\xi) = \frac{\xi^2 (2^{2\xi+1} + 1)}{(2^{2\xi-1} \ln 2)^2}$.

Das asymptotische Verhalten des Hill Schätzers wird in folgendem Satz verdeutlicht.

Satz 3 (Bias-Varianz-Zusammenhang des Hill Schätzers). *Seien $X_i \sim F$, $i \in \mathbb{N}$ unabhängig und identisch verteilt, $F \in MDA(H_{1/\alpha})$, $\alpha > 0$ und:*

- mit $A(x) := \alpha^{-2} a(U(x))$ und einer messbaren Funktion $a(x)$ ohne Vorzeichenwechsel gelte

$$\lim_{x \rightarrow \infty} \frac{U(tx)/U(x) - t^\xi}{A(x)} = t^\xi \frac{t^\rho - 1}{\rho \xi}.$$

Die Konstante $\rho \leq 0$ wird als Parameter zweiter Ordnung bezeichnet und bestimmt die Konvergenzrate von $\bar{F}(tx)/\bar{F}(x)$ gegen $t^{-\alpha}$.

- es gelte

$$k \rightarrow \infty, \quad \frac{k}{n} \rightarrow 0 \quad \text{und} \quad \sqrt{k} A\left(\frac{n}{k}\right) \rightarrow \lambda \in \mathbb{R}, \quad \text{für } n \rightarrow \infty. \quad (13)$$

Dann gilt für $n \rightarrow \infty$

$$\sqrt{k} (\hat{\alpha}^{(H)} - \alpha) \xrightarrow{d} \mathcal{N}\left(\frac{\alpha^3 \lambda}{\rho - \alpha}, \alpha^2\right).$$

Dieser Satz präzisiert den Zusammenhang zwischen der asymptotischen Varianz, die mit wachsendem k abnimmt, und einem möglichen Bias.

Bemerkung 8. *Die Bedingung (13) ist äquivalent zu folgender Bedingung:*

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(tx)/\bar{F}(x) - t^{-\alpha}}{a(x)} = t^{-\alpha} \frac{t^\rho - 1}{\rho}, \quad t > 0$$

Man beachte, dass es in der Praxis in den seltensten Fällen möglich ist diese Bedingung zu überprüfen.

Allgemein sei darauf hingewiesen, dass die Modellannahmen, auf denen die Schätzer beruhen, erfüllt sein müssen, um aussagekräftige Ergebnisse zu erhalten.

5 Verlustverteilungsansatz

Im Kapitel 4 wurde beschrieben, wie parametrische Modelle an die Schadenszahl- und Schadenshöhenverteilung angepasst werden können. Wie bereits erwähnt wird bei den fortgeschrittenen Messansätzen eine Aufteilung der Schadensfälle der gesamten Bank nach verschiedenen Geschäftsfeldern und Risikotypen (als Unterklassen des operationellen Risikos, siehe Seite 6) vorgenommen, um eine differenziertere Messung zu ermöglichen. Wurden nun die jeweiligen Verteilungen der einzelnen Geschäftsbereich-Risikotyp-Kombination geschätzt, bleibt die Frage, wie man alle diese Informationen zu einer Gesamtverlustverteilung (und späteren Risikokennzahl) komprimieren kann.

5.1 Mathematisches Modell

Nimmt man an, dass \mathcal{K} die Menge der Geschäftsbereich-Risikotyp-Kombinationen ist und $K := |\mathcal{K}|$ die Anzahl der Kombinationen, so läßt sich der Gesamtverlust, der durch die einzelnen Verluste der Kombinationen entsteht, schreiben als

$$L = \sum_{k \in \mathcal{K}} L_k . \quad (14)$$

Der Verlust L_k der k -ten Kombination setzt sich gemäß dem in Kapitel 3 beschriebenen Modell aus mehreren Schäden zusammen, deren Anzahl genau wie die Höhe der Schäden zufällig ist. Innerhalb einer Geschäftsfeld-Risikotyp-Kombination haben wir somit ein kollektives Modell, das den Verlust durch eine Summenverteilung modelliert:

$$L_k = \sum_{i=1}^{N_k} X_{k,i}, \quad k = 1, \dots, K$$

Für jede Kombination wird eine Schadenszahl- und eine Schadenshöhenverteilung benötigt. Verfolgt man einen parametrischen Ansatz, so lassen sich die Parameter dieser Verteilungen durch die in Kapitel 4 beschriebenen Methoden schätzen.

Werden die Kombinationen als unabhängig voneinander angenommen – genauer sind die L_k stochastisch unabhängig –, so kann man die Schadenszahl und Schadenshöhe der k -ten Kombination unabhängig von der Schadenszahl und -höhe der anderen Kombinationen erzeugen. Sind die L_k jedoch nicht stochastisch unabhängig, so muss der Zufallsvektor (L_1, L_2, \dots, L_K) gemäß der angenommenen Abhängigkeitsstruktur als ein Ganzes simuliert werden. Dies ist gerade für eine größere Anzahl K kompliziert. Schon für $K \geq 4$ erweist sich die Schätzung der gemeinsamen Verteilung mit einer bestimmten Abhängigkeitsstruktur als schwer. In diesem Zusammenhang erweist sich das Konzept der Copula als hilfreich.

5.2 Simulation

Eine direkte Berechnung der Verlustverteilung für eine Geschäftsfeld-Ereignistyp-Kombination aus der Schadenszahl- und der Schadenshöhenverteilung ist in den wenigsten Fällen möglich. Wenn man nun alle 56 Kombinationen zusammenfassen will, stößt man auf eine noch größere Komplexität. Zur Berechnung benutzt man deswegen eine sogenannte Monte Carlo Simulation (MCS). Es handelt sich um ein numerisches Verfahren.

Für den Fall der Unabhängigkeit kann man den folgenden Algorithmus für die Simulation der Gesamtverlustverteilung verwenden:

- Erzeuge $N_k^{(1)}, N_k^{(2)}, \dots, N_k^{(n)}$ n Schadenszahlen mit Hilfe der geschätzten Poisson-Verteilung der k -ten Kombination, $k = 1, \dots, K$. Die Zahl n der Simulationen sollte groß sein, zum Beispiel 10.000.
- Die jeweils passende Anzahl an Schadenshöhen

$$X_{k,1}^{(1)}, X_{k,2}^{(1)}, \dots, X_{k,N_k^{(1)}}^{(1)}; X_{k,1}^{(2)}, \dots, X_{k,N_k^{(2)}}^{(2)}; \dots; X_{k,1}^{(n)}, \dots, X_{k,N_k^{(n)}}^{(n)}, \quad k = 1, \dots, K$$

erzeugt man mit Hilfe der Schadenshöhenverteilung der k -ten Kombination.

- Man erhält n simulierte Werte für den Gesamtverlust, indem man

$$L^{(j)} = \sum_{k \in \mathcal{K}} L_k^{(j)}, \quad j = 1, \dots, n$$

berechnet und somit die empirische Verteilungsfunktion des Gesamtverlustes.

Wie bereits erwähnt, scheint eine Unabhängigkeitsannahme zwischen den einzelnen Geschäftsfeld-Ereignistyp-Kombinationen nicht realistisch. Um die Abhängigkeit mit einzubeziehen, ist die Hinzunahme einer Copula sinnvoll. Ein entsprechender Algorithmus, wie ihn Di Clemente/Romano (2004) auf Seite 199/200 beschreiben, sei an dieser Stelle aufgeführt:

Es wird von jeweils S jährlichen Verlustdaten innerhalb einer einzelnen Geschäftsfeld-Ereignistyp-Kombinationen ausgegangen: l_i^j , $i = 1, \dots, m$ $j = 1, \dots, S$ ist der j -te Verlust der i -ten Kombination, wobei m die Anzahl der Geschäftsfeld-Ereignis-Kombinationen ist, die aggregiert werden sollen. Außerdem soll die Abhängigkeitsstruktur durch eine Student's t-Copula modelliert werden. Zuerst wird diese Copula, bzw. deren Parameter, anhand der Daten geschätzt und danach Zufallszahlen mit dieser Abhängigkeitsstruktur erzeugt.

- Transformiere die historischen Daten l_i^j durch $\tilde{l}_i^j = \hat{F}_i(l_i^j) \sim U_{[0,1]}$ zu auf $[0,1]$ gleichverteilten Daten. Man beachte, dass \hat{F}_i die empirisch Verteilungsfunktion der i -ten Kombination ist.
- Transformiere die oben erzeugten Daten \tilde{l}_i^j durch $z_i^j = \Phi^{-1}(\tilde{l}_i^j) \sim \Phi$ zu Standardnormalverteilten Daten.
- Transformiere die oben erzeugten Daten \tilde{l}_i^j durch $c_i^j = t_m^{-1}(\tilde{l}_i^j) \sim t_m$ zu Daten mit Student-t-Verteilung t_m mit m Freiheitsgraden.
- Schätze die Freiheitsgrade der Student's t-Copula durch

$$\hat{\nu} = \arg \max_{\nu} \sum_{j=1}^S \ln c(\tilde{l}_1^j, \dots, \tilde{l}_m^j; \hat{R}, \nu),$$

wobei die zugehörige Korrelationsmatrix \hat{R} wie im nächsten Punkt beschrieben bestimmt wird.

- Die Korrelationsmatrix \hat{R} erhält man als Grenzwert der Folge von Matrizen

$$\hat{R}_0 = \frac{1}{S} \sum_{j=1}^S \mathbf{z}^j \mathbf{z}^{jT} \quad \hat{R}_{k+1} = \frac{\nu + n}{\nu S} \sum_{j=1}^S \frac{\mathbf{c}^j \mathbf{c}^{jT}}{1 + \frac{1}{\nu} \mathbf{c}^{jT} \hat{R}_k^* \mathbf{c}^j},$$

wobei $\mathbf{z}^j = (z_1^j, z_2^j, \dots, z_n^j)^T$ und $\mathbf{c}^j = (c_1^j, c_2^j, \dots, c_n^j)^T$ sowie \hat{R}_k^* die Matrix mit

$$(\hat{R}_k^*)_{u,v} = \frac{(\hat{R}_k)_{u,v}}{\sqrt{(\hat{R}_k)_{u,u}} \sqrt{(\hat{R}_k)_{v,v}}},$$

deren Hauptdiagonale normiert ist. Hierbei ist x^T der transponierte Vektor und es gilt:

$$\hat{R}_k \rightarrow \hat{R}, \quad \text{für } k \rightarrow \infty.$$

Hat man die $\frac{m(m-1)}{2} + 1$ Parameter der Student's t-Copula bestimmt – für die symmetrische Korrelationsmatrix $\frac{m(m-1)}{2}$ und 1 für die Anzahl der Freiheitsgrade –, so lassen sich die benötigten Zufallszahlen wie folgt simulieren:

- Erzeuge m unabhängige standardnormalverteilte Zufallszahlen $\mathbf{z} = (z_1, \dots, z_m)^T$
- Erzeuge eine zu \mathbf{z} unabhängige, Chi-Quadrat-verteilte Zufallsvariable $s \sim \mathcal{X}_{\nu}^2$
- Bestimme die Cholesky Zerlegung der Korrelationsmatrix $R = A A^T$
- Bestimme die Vektoren $\mathbf{y} = A \mathbf{z}$ und $\mathbf{x} = \frac{\sqrt{\nu}}{\sqrt{s}} \mathbf{y}$.

- Der Vektor $\mathbf{u} = (u_1, \dots, u_m)^T$ mit $u_i = t_\nu(x_i)$ für $i = 1, \dots, m$ ist gemäß der Student's t-Copula verteilt.

Die Anpassung einer Copula an die gegebenen Daten ist aufwendig, da bei einer Student's t-Copula z.B. schon bei der Berücksichtigung von 7 der 56 vom Basler Ausschuss vorgeschlagenen Kombinationen 22 Parameter geschätzt werden müssen. Die oben genannte Vorgehensweise um Abhängigkeiten zu modellieren ist also eher theoretisch und es sei eine Methode genannt, die praktikabler erscheint. Hierzu werden die Abhängigkeiten als Abhängigkeiten zwischen den Schadenszahlverteilungen, bzw. den Schadenshöhenverteilungen der einzelnen Kombinationen modelliert. Sei m wieder die Anzahl der Kombinationen über die aggregiert werden soll.

- Erzeuge n Schadenszahlvektoren $(N_1^{(j)}, \dots, N_m^{(j)})$, $j = 1, \dots, n$ der Länge m , deren Randverteilungen den für die jeweilige Kombination geschätzten Schadenszahlverteilungen entsprechen. Die m -dimensionale Verteilung des Vektors wird durch die geschätzten Randverteilungen G_i und eine Copula C_N bestimmt, d.h.

$$C_N(G_1(n_1), \dots, G_m(n_m)) = P(N_1 \leq n_1, \dots, N_m \leq n_m),$$

wobei $N_k \sim G_k$, $k = 1, \dots, m$.

- Bestimme $N^{(j)} := \max_{k=1}^m N_k^{(j)}$, $j = 1, \dots, n$
- Erzeuge für $j = 1, \dots, n$ jeweils $N^{(j)}$ Vektoren der Länge m mit Schadenshöhen, gemäß einer Copula C_X :

$$\begin{aligned} & (X_{1,1}^{(j)}, \dots, X_{m,1}^{(j)}) \\ & \quad \vdots \\ & (X_{1,N^{(j)}}^{(j)}, \dots, X_{m,N^{(j)}}^{(j)}). \end{aligned}$$

Wird die für die k -te Kombination geschätzte Schadenshöhenverteilung mit F_k bezeichnet, bedeutet dies

$$\begin{aligned} C_X(F_1(x_1), \dots, F_m(x_m)) &= P(X_{1,i}^{(j)} \leq x_1, \dots, X_{m,i}^{(j)} \leq x_m), \\ i &= 1, \dots, N^{(j)}, \quad j = 1, \dots, n. \end{aligned}$$

- Für jede Simulation $j = 1, \dots, n$ summiere jeweils die ersten N_k Schadenshöhen, um den Verlust der k -ten Kombination zu erhalten.

$$L_k^{(j)} = \sum_{i=1}^{N_k^{(j)}} X_{k,i}^{(j)}$$

Es kann vorkommen, dass einige der simulierten $X_{k,i}$ nicht in die Berechnung des Verlustes miteinbezogen werden. Es gilt: $X_{k,i} \sim F_k$, $k = 1, \dots, m$ und alle i .

- Ein simulierter Wert des Gesamtverlustes über alle Kombinationen $k = 1, \dots, m$ ergibt sich mit Hilfe der Formel (14). Die n simulierten Gesamtverlustwerte führen auf eine empirische Verteilung des Gesamtverlustes.

5.3 Anwendung

In diesem Paragraphen wird die Vorgehensweise zur Schätzung der Verteilungen anhand simulierter Daten verdeutlicht. Dies bildet die Grundlage für eine spätere Simulation der (Gesamt)verlustverteilung. Die Daten wurden mittels der im Anhang vorliegenden Funktion erzeugt und bearbeitet.

Geht man von einer Poisson-Verteilung für die Modellierung der Schadenszahl aus, so ergibt sich der MLE des Intensitätsparameters als Mittelwert der Beobachtungen. Es wurde von simulierten monatlichen Daten über einen Zeitraum von 3 Jahren ausgegangen, um die Basis der Schätzung der monatlichen Poisson-Verteilung zu erhöhen. Der ermittelte monatliche Intensitätsparameter wurde anschließend auf ein Jahr hochskaliert. Somit ergab sich ein Intensitätsparameter von 1215.333.

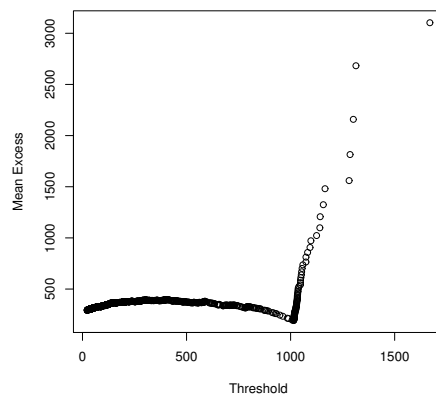


Abbildung 2: Mean Excess Plot

Bei der Schätzung der Schadenshöhenverteilung wurde als erstes ein Mean excess Plot der Daten durchgeführt. Wie man in Abbildung 2 sieht, lässt sich oberhalb der Schwelle $u = 1000$ eine Gerade mit positiver Steigung erkennen. Der MEP ist sehr sensitiv gegenüber Veränderungen der Schwelle u und sollte nur eine erste Einschätzung liefern. Eine genauere Untersuchung zur Wahl der Schwelle ist auf jedem Fall erforderlich (vergleiche Chapelle et al. (2004), Seite 12 ff.).

Einige Ergebnisse der Funktion `fpot`, die die Maximum-Likelihood Schätzer für den Gestalts- und den Skalenparameter bei vorgegebener Schwelle u berechnet, sind in

Tabelle 3 enthalten. Hierbei bezeichnet N_u die Anzahl der Beobachtungen oberhalb der Schwelle u , auf denen die Schätzung der Parameter der verallgemeinerten Pareto-Verteilung basiert. Es sei bemerkt, dass das numerische Verfahren, mit dem die ML Schätzer berechnet werden in den Fällen $u = 1032, 1040$ und 1050 nicht durchführbar ist. Im Folgenden wird die Schwelle $u = 113$ angenommen.

u	N_u	$\hat{\xi}^{MLE}$	$\hat{\sigma}^{MLE}$
1015	75	1.339	12.833
1013	78	1.152	16.329
1031	37	1.75	20.08
1000	79	0.7273	39.5397
970	81	0.4911	80.3368

Tabelle 3: Schätzungen zu verschiedenen Schwellen.

Nun kann man die Parameter der bedingten Lognormalverteilung schätzen, indem man die Maximum-Likelihood Methode auf die bedingte Dichte, wie sie auf Seite 27 angegeben ist, anwendet. Hierzu wurde die Funktion `fitdistr` benutzt. Sie ermöglicht die Maximierung der durch die übergebene Dichte bestimmte Likelihood-Funktion. Allerdings muss man von einigen Ausnahmen abgesehen stets Startwerte der zu schätzenden Parameter übergeben, die für das Iterationsverfahren gebraucht werden. Diese wurden durch die Momentenmethode berechnet als $mls = 5.040625$ und $sdl = 0.9838013$. Es ist zu beachten, dass nur die Daten unterhalb der Schwelle $u = 1013$ zur Gewinnung der Startwerte und Anpassung der bedingten Lognormalverteilung benutzt wurden. Somit ergab sich für die ML Schätzer der Parameter der bedingten Lognormalverteilung $ml = 5.08909442$, $sdl = 1.44464534$.

5.4 Fazit

Mit der Verabschiedung des Basel II Abkommens ist der Bedarf an Verfahren zur Bestimmung der Eigenkapitalhinterlegung gestiegen und mit der expliziten Erwähnung operationeller Risiken insbesondere auf diesem Gebiet.

Die vorliegende Arbeit sollte einen Methodenüberblick zur Quantifizierung von operationellen Risiken geben und ein statistisches Modell aufzeigen, welches auf einem versicherungsmathematischen Grundstein beruhend eine solche Messung ermöglicht. Der Schwerpunkt der angesprochenen Aspekte lag auf der Modellierung der Schadenshöhenverteilung.

Bei der Handhabung operationeller Risiken scheinen zwei Gesichtspunkte von besonderem Interesse: das Auftreten extremer Schäden und die Datenlage. Ersteres kann man durch Techniken der Extremwerttheorie bewältigen. Bei der Analyse der Schadenshöhen wurde das Konzept einer zusammengesetzten Verteilung benutzt. Die Verwendung einer klassischen Lognormalverteilung für das Zentrum der Verteilung und einer verallgemeinerten Pareto-Verteilung für die Flanke gewährleistet eine bessere Modellierung der Schadenshöhenverteilung. In der Literatur findet man auch viele empirische Studien zur Analyse operationeller Risiken mit Hilfe der Extremwerttheorie. Einige davon sind Chapelle et al. (2004), Di Clemente/Romano (2004) und Embrechts/Kaufmann/Samorodnitsky (2002). Das Problem der Datenlage, dass bisher nur wenige Daten zu operationellen Risiken vorliegen, wurde nicht weiter ausgeführt, ebenso wenig wie dessen Bewerkstelligung. Es sei jedoch angemerkt, dass es wichtig ist, eine entsprechende interne Datenbank zur Verfügung zu haben, da es sich gerade bei operationellen Risiken häufig um unternehmensspezifische Risiken handelt.

Anhang

R-Funktionen zur Anpassung der Verteilungsfunktionen

Im Folgenden ist der Quellcode einiger Funktionen gegeben, um die Schätzung der Schadenszahl und -höhe mit Hilfe der statistischen Software R durchzuführen. Viele nützliche Funktionen zu diesem Thema findet man schon in Paketen enthalten.

```
#Schätzung der Schadenszahlverteilung
z<-rpois(36,100)
zpm<-mean(z)
lambda<-12*zpm
rm(list=ls())
#Generiert ein Schadenshöhen-sample einer abgeschnittenen Lognormalvertei-
#lung verbunden mit einer GPD Verteilung oberhalb einer rechten Schranke
#und einem oberen Quantilwert.
#Man beachte, dass die Ergebnisse als Liste zurückgegeben werden.
library(evd) #für die Simulation der GPD
OpRiskSample<-function(anz,alpha=c(0.1,0.9),meanlog=0,sdlog=1,scale=1,
shape=1)
{
left<-qlnorm(alpha[1],meanlog,sdlog)
right<-qlnorm(alpha[2],meanlog,sdlog) #entspricht der Schwelle u
centersample<-rep(NA,anz)
for (i in (1:anz))
{
x<-0
while (x<left | x>right)
x<-rlnorm(1,meanlog,sdlog)
centersample[i]<-x
}
extremsample<-rgpd(anz,loc=right,scale,shape)
randunif<-runif(anz)
sample1<-centersample[randunif<alpha[2]]
sample2<-extremsample[randunif>alpha[2]]
sample<-c(sample1,sample2)
result<-list(sample=sample,anz=anz,alpha=alpha,lognor=c(meanlog,sdlog)
,loc=right,scale=scale,shape=shape)
return(result)
}
mu<-5
```

```
sd<-1.5
alpha<-0.1
alphar<-0.9
sc<-10
sh<-1.5
n<-1000
sim<-OpRiskSample(anz=n,meanlog=mu,sdlog=sd,scale=sc,shape=sh)

#MEP der Daten
meplot(sim$sample)

#Partition der Daten
l<-sim$sample[sim$sample<1013]
h<-sim$sample[sim$sample>=1013]

#Startwertschätzung für den MLE
n<-length(l) # Anzahl der 'normalen' Werte
mls<-mean(log(l)) #Schätzer für meanlog
s<-sum( (log(l)-mls)^2 )
sdls<-sqrt(s/(n-1)) #Schätzer für sdlog

#bedingte Lognormalverteilung
left<-qlnorm(0.1,sim$lognor[1],sim$lognor[2])
right<-qlnorm(0.9,sim$lognor[1],sim$lognor[2])
#entspricht den Werten der Simulation left, right, meanlog, sdlog
#bedingte Dichte
truncdens<-function(x, ml, sdl)
{
dlnorm(x, ml,sdl)/ ( plnorm(right, ml,sdl) - plnorm(left,ml,sdl) )
}
#ML Schätzung
library(MASS) #für die Funktion fitdistr
f<-fitdistr(l, truncdens, start=list(ml=mls, sdl=sdls))
```


Literatur

Die Literaturangaben sind alphabetisch nach den Namen der Autoren geordnet. Bei mehreren Autoren wird nach dem ersten Autor sortiert. In einigen Fällen ist zu beachten, dass das entsprechende Dokument im Internet leicht von dem in Buchform vorliegenden Dokument abweicht. Sämtliche Internetadressen und Links beziehen sich auf August 2005 und können sich zwischenzeitlich geändert haben.

- [1] C. Acerbi, D. Tasche (2005): On the Coherence of Expected Shortfall. *Journal of Banking and Finance*, Vol. 29.
<http://www-m1.mathematik.tu-muenchen.de/m4/pers/tasche/shortfall.pdf>
- [2] C. Alexander (2003): *Operational Risk: Regulation, Analysis and Management*. 1. Aufl. New York: FT Prentice Hall.
- [3] Basler Ausschuss für Bankenaufsicht : Internationale Konvergenz der Eigenkapitalmessung und der Eigenkapitalanforderungen - Überarbeitete Rahmenvereinbarung, Juni 2004.
<http://www.bis.org/publ/bcbs107ger.htm>
- [4] A. Chapelle, Y. Crama, G. Hübner, J.-P. Peters (2004): *Basel II and Operational Risk - Implication for risk measurement and management in the financial sector*. National Bank of Belgium, working paper no 51.
- [5] A. da Costa Dias (2004): *Copula Inference for Finance and Insurance*. Diss., Mathematische Wissenschaften, Eidgenössische Technische Hochschule ETH Zürich, Nr. 15283.
<http://e-collection.ethbib.ethz.ch/show?type=diss&nr=15283>
- [6] A. Di Clemente, C. Romano (2004): *Measuring Portfolio Value-at-Risk by a Copula-EVT based approach*. In: *Operational Risk Modelling and Analysis*. 1. Aufl. Risk Books, Seite 189-206.
<http://www.gloriamundi.org/picsresources/adcr4.pdf>
- [7] H. Drees, L. de Haan, S. Resnick (2000): How to make a Hill plot. *Annals of Statistics*, Vol. 28, No. 1, Seite 254-274.
- [8] V. Durrleman, A. Nikeghbali, T. Roncalli (2000): *Which Copula is the right one?* working paper, Groupe de Recherche Opérationelle, Crédit Lyonnais, France.
<http://gro.creditlyonnais.fr/content/wp/copula-choice.pdf>
- [9] P. Embrechts, C. Klüppelberg, T. Mikosch (1997): *Modelling Extremal Events*. 1. Aufl. New York: Springer.

- [10] P. Embrechts, R. Kaufmann, G. Samorodnitsky (2002): Ruin theory revisited: stochastic models for operational risk. In: Risk Management for Central Bank Foreign Reserves (Eds. C. Bernadell et al.) European Central Bank, Frankfurt a.M., Seite 243-261.
- [11] Z. Han (2003): Actuarial Modelling of Extremal Events using Transformed Generalized Extreme Value Distributions and Generalized Pareto Distributions. <http://www.math.ohio-state.edu/history/phds/abstracts/pdf/Han.Zhongxian.pdf>
- [12] A. J. McNeil, T. Saladin (1997): The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions. Proceedings of 28th International ASTIN Colloquium. <http://www.math.ethz.ch/~mcneil/ftp/cairns.pdf>
- [13] O. Menkens (2005): CF 901 Extreme Value Theory CCFEA/ Department of Economics University of Essex. http://www.essex.ac.uk/ccfea/Courses/CF901/Spring_2005/Week_3_4/CF_901_3.pdf
- [14] R.-D. Reiss, M. Thomas (2001): Statistical Analysis of Extreme Values. 2. Aufl. Berlin: Birkhäuser.
- [15] W. A. Stahel (2002): Statistische Datenanalyse. 4. Aufl. Braunschweig/Wiesbaden: Vieweg Verlag.
- [16] S.D. Silvey (1975): Statistical Inference. London: Chapman and Hall Ltd.
- [17] M.-P. Victoria-Feser (2000): Robust Distribution Estimation with Missing Data. DARP discussion paper No. 57, STICERD, LSE. <http://sticerd.lse.ac.uk/dps/darp/DARP57.pdf>