

Ermittlung der Güte und Power ausgewählter
GOF-Tests mit Fokussierung auf den Randbereich
einer Verteilung

Alexandra Knebel

Betreuer der LMU: Prof. Dr. Göran Kauermann

Betreuer der BayernLB: Prof. Dr. Matthias Fischer



Institut für Statistik

Ludwig-Maximilians-Universität München

7. Januar 2020

Eidesstattliche Erklärung

Hiermit erkläre ich, Alexandra Knebel, dass ich die Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken und Inhalte sind als solche gekennzeichnet. Die Arbeit habe ich bisher keinem anderen Prüfungsamt in gleicher oder vergleichbarer Form vorgelegt. Sie wurde bisher nicht veröffentlicht.

Alexandra Knebel

Abstract

Die Verwendung von Verteilungen, die als heavy-tailed charakterisiert werden, kann in vielen Bereichen, wie zum Beispiel bei der Berechnung des operationellen Risikos im Banken- oder Versicherungssektor, von großem Nutzen sein. Um die Anpassungsgüte einer solchen Verteilung an tatsächliche Schadensdaten beurteilen zu können, eignen sich sogenannte Goodness-of-Fit-Tests. Nicht jeder Test verfügt jedoch über die gleiche statistische Power. Das korrekte Klassifizieren einer vorliegenden Verteilung kann besonders im Tailbereich nicht von allen GOF-Tests mit identischer Power geleistet werden. Diese Arbeit widmet sich der Ermittlung der Power von vier ausgewählten GOF-Tests bezüglich vier verschiedener Verteilungen, die an reale Schadensdaten aus dem operationellen Risiko angepasst werden. Nach einer theoretischen Einführung folgt eine Simulationsstudie. Die Studie basiert auf zwei Datensätzen mit realen Schadensdaten. Die Testpower wird anhand der vier unterschiedlichen Verteilungen, von light- bis super-heavy-tailed, gemessen. Aus jeder Verteilung werden Zufallsdaten simuliert, wobei die Verteilungsparameter mittels Maximum-Likelihood-Schätzung aus den Schadensdaten ermittelt werden. Schlussendlich werden die Simulationsdaten mit jedem der GOF-Tests auf ihre Anpassung getestet. Dieser Vorgang wird sowohl für die kompletten Daten als auch für eine linkstrunkierte Version der Daten realisiert. Es zeigt sich aufgrund der Testergebnisse, dass die für die jeweiligen Daten geschätzten Verteilungsparameter einen sehr großen Einfluss auf die Power des jeweiligen Tests haben. Nicht zuletzt deshalb fällt es schwer, eine globale Aussage über den Test mit der höchsten zu Power zu treffen. Sicherlich lässt sich jedoch festhalten, dass es stets empfehlenswert ist, wichtige Entscheidungen nicht allein auf Basis eines einzelnen Testverfahrens zu abzuleiten.

Inhaltsverzeichnis

1	Einleitung	1
2	Theoretische Einführung der Verteilungsmodelle	2
2.1	Tailverhalten einer Verteilung	3
2.2	Single-Spliced-Modelle	3
2.2.1	Body-Tail-Threshold	4
2.2.2	Trunkierte Verteilungen	4
2.3	Klassifizierung ausgewählter Verteilungen	5
2.3.1	Kennzahlen einer Verteilung	5
2.3.2	Gompertz-Verteilung	5
2.3.3	Logarithmische Normalverteilung	6
2.3.4	American-Stoppa-Verteilung	7
2.3.5	Generalized-Log-Pareto-Verteilung	10
2.4	Schätzung der Parameter	11
3	Begriffe der Testtheorie zur Überprüfung der Anpassungsgüte einer Verteilung	12
3.1	Güte, Signifikanz und Power eines Tests	13
3.2	p-Wert nach Fisher	15
3.3	Konvergenzkriterien für Goodness-of-Fit-Tests	15
3.3.1	Empirische Verteilungsfunktion	15
3.3.2	Satz von Glivenko-Cantelli	16
3.4	Goodness-of-Fit-Tests	16
3.4.1	Kolmogorov-Smirnov-Anpassungstest	16
3.4.2	Cramér-von-Mises-Test	17
3.4.3	Anderson-Darling-Test	18
3.4.4	Quadratic Class Upper Tail Anderson-Darling Test	18
4	Simulationsstudie mit den vorgestellten Testverfahren hinsichtlich der Verteilungen	19
4.1	Datenbasis der Simulationsstudie	20
4.2	Power der Testverfahren bei Non-Spliced-Modellen	22
4.2.1	Vorbereitung	23
4.2.2	Ergebnisse	27
4.3	Power der Testverfahren bei trunkierten Verteilungen	36
4.3.1	Vorbereitung	36
4.3.2	Ergebnisse	37
4.4	Kritische Diskussion der Ergebnisse der Simulationsstudien	42

5	Fazit	43
A	Anhang	VI
A.1	Verteilungsübersicht	VI
A.2	Tabellen	VI
A.3	Parameterschätzer	VI
A.3.1	Datensatz 1	VI
A.3.2	Datensatz 2	VII
A.4	Thresholds	VIII
A.5	Elektronischer Anhang	VIII

Abbildungsverzeichnis

1	Dichte- und Verteilungsfunktion der Gompertz-Verteilung in Abhängigkeit der Parameter	6
2	Dichte- und Verteilungsfunktion der Log-Normalverteilung in Abhängigkeit der Parameter	7
3	Dichtefunktionen der American-Stoppa-Verteilung. Links mit variablem c , rechts mit variablem θ	9
4	Dichte- und Verteilungsfunktion der American-Stoppa-Verteilung mit festem θ und c , λ variabel	9
5	Dichtefunktion der GLP-Verteilung in Abhängigkeit von σ und λ	10
6	Dichte- und Verteilungsfunktion der GLP-Verteilung mit festem λ und σ , β variabel	11
7	Graphische Darstellung der Kerndichte beider Datensätze im Intervall $[0,100000]$	20
8	Graphische Darstellung der Fünf-Punkte-Zusammenfassung des ersten Datensatzes	21
9	Graphische Darstellung der Fünf-Punkte-Zusammenfassung des zweiten Datensatzes	22
10	Anpassung der vier Verteilungen an den Datensatz 1	25
11	Anpassung der vier Verteilungen an den Datensatz 2	26
12	QQ-Plot Log-Normal- vs. American-Stoppa-Verteilung	30
13	QQ-Plots der American-Stoppa-Verteilung vs. Datensatz 1 bzw. Datensatz 2	32
14	QQ-Plots der Verteilungen vs. Datensatz 1	34
15	QQ-Plot der Verteilungen vs. Datensatz 2	35
16	Anpassung der GLP-Verteilungen an den Tail von Datensatz 1 in Abhängigkeit des Thresholds	41

Tabellenverzeichnis

1	Fehler bei Testentscheidungen	14
2	Übersicht von Variationskoeffizient, Schiefe und Kurtosis beider Datensätze	22
3	Ergebnisse der Maximum-Likelihood-Schätzung aus Datensatz 1	24
4	Ergebnisse der Maximum-Likelihood-Schätzung aus Datensatz 2	24
5	Testpower bezüglich der Gompertz-Verteilung	28
6	Testpower bezüglich der Log-Normalverteilung	29
7	Testpower bezüglich der American-Stoppa-Verteilung	31
8	Testpower bezüglich der GLP-Verteilung	32
9	Testpower bezüglich einer linkstrunkierten Gompertz-Verteilung	38
10	Testpower bezüglich einer linkstrunkierten Log-Normalverteilung	39
11	Testpower bezüglich einer linkstrunkierten American-Stoppa-Verteilung	39
12	Testpower bezüglich einer linkstrunkierten GLP-Verteilung	40
13	Momente der behandelten Verteilungen	VI
14	Mittelwert und Varianz der geschätzten Parameter für die Gompertz- und Log-Normalverteilung (Datensatz 1)	VI
15	Mittelwert und Varianz der geschätzten Parameter für die American-Stoppa-Verteilung (Datensatz 1)	VII
16	Mittelwert und Varianz der geschätzten Parameter für die GLP-Verteilung (Datensatz 1)	VII
17	Mittelwert und Varianz der geschätzten Parameter für die Gompertz- und Log-Normalverteilung (Datensatz 2)	VII
18	Mittelwert und Varianz der geschätzten Parameter für die American-Stoppa-Verteilung (Datensatz 2)	VII
19	Mittelwert und Varianz der geschätzten Parameter für die GLP-Verteilung (Datensatz 2)	VIII
20	Empirischer Mittelwert bzw. empirische Varianz der Thresholds (Datensatz 1)	VIII
21	Empirischer Mittelwert bzw. empirische Varianz der Thresholds (Datensatz 2)	VIII

1 Einleitung

Der Randbereich einer Verteilung, im Verlauf der Arbeit auch als Tail bezeichnet, spielt in vielen analytischen Ansätzen eine bedeutende Rolle. Beispielsweise ist es bei der Modellierung von Schadensdaten einer Versicherung oder bei der Berechnung von operationellen Risiken von Banken unverzichtbar, den Tail möglichst exakt zu modellieren, um hohe finanzielle Verluste zu vermeiden. Die Wahl einer geeigneten Verteilung ist für die Modellierung des Tails essentiell. Es existieren bereits unterschiedliche wissenschaftliche Arbeiten, die sich mit dem Vergleich der Power von statistischen Anpassungstests (Goodness of Fit-Tests; im Folgenden mit GOF-Tests bezeichnet) für Verlustdaten beschäftigen. In dem Paper 'Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples' von Chernobai (2005) wird eine Simulationsstudie für unterschiedliche Verlustdatensätze vorgestellt, in der nach einer geeigneten Verteilung zur Modellierung von Verlustdaten gesucht wird. Die Autoren kommen zu dem Schluss, dass sich Verteilungen mit einem heavy tail¹ besser zur Modellierung eignen als solche, die als light-tailed bezeichnet werden.

Die Weiterführung der Bachelorthesis mit dem Titel 'Comparison of the goodness-of-fit tests for truncated distributions' von Lach (2018) konzentriert sich auf den Vergleich der Power von GOF-Tests für linkstrunkierte Verteilungen. Hier werden ebenfalls die Ergebnisse einer Simulationsstudie vorgestellt. Der Unterschied zur vorliegenden Arbeit besteht darin, dass der Simulationsstudie keine realen Daten zugrunde liegen, sondern die Parameter der Verteilungen vorher frei gewählt und nicht geschätzt wurden.

In der vorliegenden Bachelorarbeit werden vier unterschiedliche Verteilungen hinsichtlich ihrer Anpassungsgüte untersucht. Um die Verteilungsparameter bestmöglich für die Anwendung in der Schadensdatenmodellierung zu kalibrieren, werden diese aus realen Schadensdaten geschätzt. Die Anpassung der Verteilungen wird anschließend unter Verwendung der GOF-Tests untersucht. Es existieren verschiedene Tests, die unterschiedlich gute Ergebnisse erbringen. Die Arbeit konzentriert sich auf die Ermittlung der Güte und Power solcher Testverfahren. Für die Analyse werden vier unterschiedliche Verteilungen und vier unterschiedliche Testverfahren betrachtet. Die Verteilungen sind aufgrund ihres Tailverhaltens, von light-tailed bis super-heavy-tailed, ausgewählt worden. In aufsteigender Reihenfolge ihres Tailverhaltens sind dies im Einzelnen die Gompertz-Verteilung, die Log-Normalverteilung, die American-Stoppa-Verteilung und die Generalized-Log-Pareto-Verteilung (GLP-Verteilung). Als GOF-Tests werden der Kolmogorov-Smirnov-Test, der Cramér-von-Mises-Test, der Anderson-Darling-Test und eine abgewandelte Form des Anderson-Darling-Tests mit überproportionaler Gewichtung des Tails, der sogenannte Qua-

¹Eine Einführung der unterschiedlichen Taildefinitionen folgt in Kapitel 2.

dratic Class Upper Tail Anderson-Darling Test, gegenübergestellt.

Kapitel 2 beschäftigt sich zunächst mit den mathematischen Grundlagen der Modellierung von Schadenshöhenverteilungen, wobei eingangs die spezielle Gestalt dieser Verteilungen dargestellt und anschließend das Konzept von Non-Spliced- und Single-Spliced-Modellen erläutert wird. Die vier ausgewählten Verteilungen, anhand derer auch die Evaluation der Anpassungstests erfolgt, werden näher beschrieben. Im anschließenden Kapitel 3 erfolgt dann eine Einführung aller notwendigen Definitionen für die statistischen Testverfahren der GOF-Tests. Dort werden zentrale Begrifflichkeiten der Testtheorie und die ausgewählten GOF-Tests vorgestellt. Nach der Bildung der theoretischen Grundlagen werden in Kapitel 4 die Ergebnisse einer zweigeteilten Simulationsstudie, in welcher die Testperformance für die vier unterschiedlichen Verteilungsszenarien analysiert wird, diskutiert. Die Arbeit schließt mit einem zusammenfassenden Fazit, das die Ergebnisse der Simulationsstudie kritisch würdigt, und einem Ausblick zu weiterführenden Analyseaspekten ab.

2 Theoretische Einführung der Verteilungsmodelle

Um beispielsweise Schadensdaten als Grundlage für die Bewertung von Risiken, inklusive der Prognose künftiger Schadensentstehung, angemessen modellieren zu können, ist die Annahme einer spezifischen Verteilung der Daten notwendig. So können die wohlbekanntesten theoretischen Eigenschaften dieser Verteilung für Deskription und Inferenz genutzt werden. Betrachtet werden können entweder die komplette Verteilung oder eine trunkierte Version, die auf einen bestimmten Bereich der Verteilung eingeschränkt ist. Da der Fokus dieser Arbeit auf dem Tail einer Verteilung liegt, werden sowohl die komplette Verteilung als auch eine linkstrunkierte Version, die alle Beobachtungen links von einem bestimmten Schwellenwert außer Acht lässt, betrachtet. Diese Unterscheidung definiert sogenannte Non-Spliced-Modelle, wenn keine Trunkierung vorgenommen wird, und Single-Spliced-Modelle, wenn eine Trunkierung vorgenommen wird. Bei Non-Spliced-Modellen wird die Schadenshöhe mittels einer einzigen (parametrischen) Verteilung beschrieben. Im Falle des Single-Spliced-Modells wird dem Body des Modells bis zu einem gewissen Schwellenwert (*Threshold*) eine bestimmte Verteilung unterstellt, der Tail wird mit einer anderen Verteilung modelliert. Anschließend werden beide Teile mit einer Gewichtsfunktion verbunden (vgl. Cruz (2015), S. 94). Es besteht auch die Möglichkeit, mittels mehrerer Thresholds weitere Verteilungen hinzuzuziehen und somit gegebenenfalls noch exakter zu modellieren. Die Bezeichnung der Spliced-Modelle erfolgt somit anhand der Häufigkeit ihrer Unterteilung.

Im nächsten Abschnitt wird zunächst auf die Kennzahlen einer Verteilung und ihr Tailverhalten eingegangen. Anschließend werden Ansätze und Begrifflichkeiten der

Non-Spliced- und Single-Spliced-Modelle und die vier ausgewählten Verteilungen vorgestellt.

2.1 Tailverhalten einer Verteilung

Gerade bei der Modellierung von Schadenshöhen ist das Tailverhalten einer Verteilung entscheidend, um ein Modell optimal anzupassen. Geeignet ist hier ein Tail, der sehr langsam gegen Null konvergiert (vgl. Gatto (2014), S.7).

Die nachfolgenden Definitionen erlauben die Klassifikation einer Verteilung bezüglich des Tailverhaltens (vgl. Foss et al. (2019), S. 7f.). Eine Verteilung F wird als heavy-tailed bezeichnet, wenn gilt:

$$\int_{\mathbb{R}} e^{\lambda x} F(dx) = \infty, \forall \lambda > 0 \quad (1)$$

(ebenda.). Die Verteilung F wird als light-tailed klassifiziert, wenn gilt:

$$\int_{\mathbb{R}} e^{\lambda x} F(dx) < \infty, \exists \lambda > 0 \quad (2)$$

(ebd.).

Zu den light-tailed Verteilungen zählen beispielsweise die *Exponential-* und die *Normalverteilung*² (vgl. Gatto (2014), S.15ff.). Durch eine exponentielle Transformation dieser beiden Verteilungen lassen sich zwei heavy-tailed Verteilungen ableiten, die Pareto- und Log-Normalverteilung (ebenda.). Die Begriffe *semi-heavy-tailed* und *super-heavy-tailed*, die ebenfalls in dieser Arbeit Verwendung finden, werden in der Literatur nicht eindeutig abgegrenzt. Die Dichte einer Verteilung, die als semi-heavy-tailed bezeichnet wird, nimmt im Tail exponentiell ab (vgl. Moscadelli (2004)). Analog fällt der Tail einer super-heavy-tailed Verteilung noch langsamer ab als der einer heavy-tailed Verteilung, erreicht jedoch nicht Null.

2.2 Single-Spliced-Modelle

Bei einem Single-Spliced-Modell werden zunächst zwei Intervalle definiert, welche den Body auf das Intervall $(0, \tau]$ und den Tail der Verteilung auf das Intervall (τ, ∞) beschränken. Angelehnt an Cruz (2015), Definition 5.11 (S. 94), und Peters (2015), S. 93, wird die Definition der Dichtefunktion für Single-Spliced-Verteilungen abgeleitet. Eine Zufallsvariable X nennt man Single-Spliced verteilt, wenn ihre Dichtefunktion durch

$$f_{single-spliced}(x; \theta; \tau; r) = \begin{cases} r f_{Body}(x; \theta_{Body}), & 0 < x \leq \tau \\ (1 - r) f_{Tail}(x; \theta_{Tail}), & \tau < x < \infty \end{cases}$$

²Dichte- und Verteilungsfunktion der beiden Verteilungen sind in Fahrmeir (2016), S.253, definiert.

bestimmt ist. In diesem Fall bezeichnet r das Gewicht. Die Summe der Gewichte in Body und Tail muss Eins ergeben, um sicherzustellen, dass die Dichtefunktion wohldefiniert ist (vgl. Stelzer (2018), S.26). Der Parametervektor ist durch $\theta = (\theta_{Body}, \theta_{Tail})$ gegeben. τ bezeichnet den Threshold.

2.2.1 Body-Tail-Threshold

Bei Single-Spliced-Modellen ist vor allem die Wahl des Thresholds von großer Bedeutung für die Anpassungsgüte einer Tail-Verteilung. Um den Threshold zu bestimmen, gibt es eine Vielzahl an Möglichkeiten, die jedoch unabhängig von den gewählten Verteilungen anwendbar sind. Weitere mögliche Ansätze sind die Bestimmung des Thresholds über einen *Mean-Excess-Plot* oder über Faustregeln (vgl. Stelzer (2018), S.90).

2.2.2 Trunkierte Verteilungen

Unter der Trunkierung einer Verteilung versteht man die Begrenzung eines Wahrscheinlichkeitsraumes an einem Threshold (vgl. Stelzer (2018), S.24). Um insbesondere die seltenen, aber großen Schadenshöhen abzudecken, dient eine linkstrunkierte Verteilung. Hier wird dementsprechend der linke Teil der Verteilung außer Acht gelassen und das Verhalten der Verteilung erst ab einem bestimmten Threshold betrachtet³. Bezeichne nun X eine Zufallsvariable und x_1, \dots, x_n die einzelnen Beobachtungen der realisierten Stichprobe. Der Threshold wird mit τ notiert. Für die Verteilungsfunktion einer linkstrunkierten Verteilung gilt:

$$F_{linkstrunkiert}^*(x) = \frac{F(x) - F(\tau)}{1 - F(\tau)} \mathbb{I}_{\{x \geq \tau\}} \quad (3)$$

und für die Dichtefunktion (vgl. Fahrmeir (2016)):

$$f_{linkstrunkiert}^*(x) = \frac{f(x)}{1 - F(\tau)} \mathbb{I}_{\{x \geq \tau\}} \quad (4)$$

Linkstrunkierte Verteilungen und Tests für linkstrunkierte Verteilungen sind im weiteren Verlauf der Arbeit mit * gekennzeichnet.

³In dieser Arbeit ist nur die linkstrunkierte Verteilung von Interesse, daher wird auf die formelle Einführung der rechtstrunkierten und beidseitig trunkierten Verteilung verzichtet. Eine Definition dieser wird zum Beispiel bei Cruz (2015) auf S. 92f. gegeben.

2.3 Klassifizierung ausgewählter Verteilungen

Da die Anpassung der Tail-Verteilung ein wesentlicher Bestandteil dieser Arbeit ist, wird in diesem Kapitel detaillierter auf die die vier ausgewählten Verteilungen eingegangen. Es werden jeweils Parameter, Dichte- und Verteilungsfunktion definiert sowie maßgebliche Charakteristika der Verteilungen vorgestellt. Eine Übersicht der Kennzahlen jeder Verteilung findet sich im Anhang. Die Verteilungen werden nach ihrem Tailverhalten klassifiziert und in aufsteigender Reihenfolge nach light-tailed, semi-heavy-tailed, heavy-tailed und super-heavy-tailed aufgeführt.

2.3.1 Kennzahlen einer Verteilung

Mit Hilfe der Lage- und Streumaße lässt sich die Gestalt einer Verteilung beschreiben. Zu den Lagemaßen zählen der Median, der Modus und das arithmetische Mittel. Als Streumaße bezeichnet man zum Beispiel die Spannweite, auch Range (Abstand zwischen dem kleinsten und größten Wert), die Varianz und die Standardabweichung. Eine detaillierte Darstellung und Definition wird in Fahrmeir (2016) aufgeführt. Zu den Kennzahlen einer Verteilung zählen unter anderem auch die Schiefe und Wölbung (*Kurtosis*) (vgl. Fahrmeir (2016), S.69). Mit deren Hilfe lässt sich eine Verteilung über die Gestalt ihrer Dichtefunktion klassifizieren. Der Graph einer Dichtefunktion kann beispielsweise links- oder rechtsschief (rechtssteil bzw. linkssteil) sein. Unter der Kurtosis versteht man, wie flach oder spitz die Dichtefunktion um den Hochpunkt des Graphen verläuft.

2.3.2 Gompertz-Verteilung

Die Gompertz-Verteilung beschreibt die Wahrscheinlichkeit des Eintritts von Ereignissen in einer aufeinanderfolgenden Reihe von Zeitintervallen (vgl. Lenart (2011)). Das interessierende Ereignis ist in diesem Fall der Eintritt eines Schadensfalles von sehr großer Höhe. Die Zufallsvariable X ist Gompertz-verteilt mit *Skalierungsparameter* $a > 0$ und *Formparameter* $b > 0$, wenn sie die folgende Dichtefunktion besitzt:

$$f_{Gompertz}(x|a, b) = a \exp(bx) \exp\left(-\frac{a}{b}(\exp(bx) - 1)\right) \quad (5)$$

Die zugehörige Verteilungsfunktion (vgl. Lenart (2016), S. 5) hat die Form

$$F_{Gompertz}(x|a, b) = 1 - \exp\left(-\frac{a}{b}(\exp(bx) - 1)\right) \quad (6)$$

In Abbildung 1 ist das Verhalten von Dichte- und Verteilungsfunktion der Gompertz-Verteilung in Abhängigkeit von den Parametern a und b dargestellt. Beide Parameter wirken sich auf den Modus der Verteilung aus. Die Lage des Modus wird maßgeb-

lich durch den Lageparameter a beeinflusst: Mit steigendem a verschiebt sich der Modus nach links. Der Formparameter hingegen staucht die Verteilungskurve mit sinkendem b .

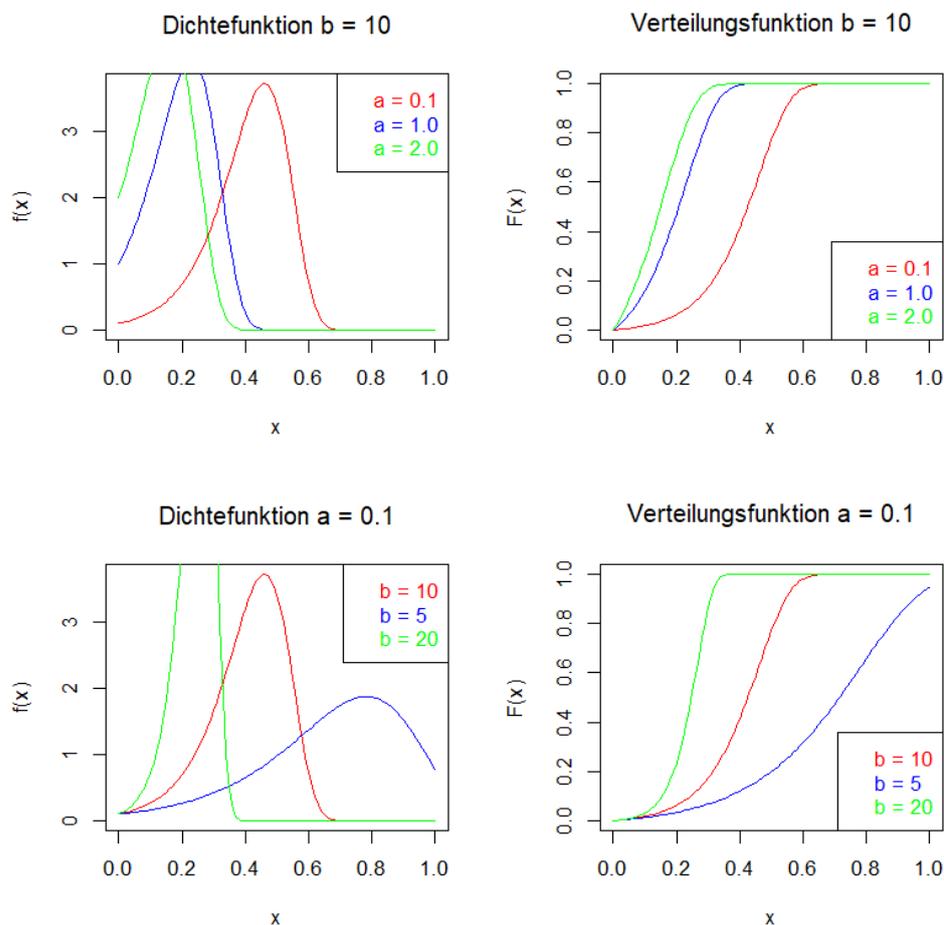


Abbildung 1: Dichte- und Verteilungsfunktion der Gompertz-Verteilung in Abhängigkeit der Parameter

2.3.3 Logarithmische Normalverteilung

Die Logarithmische Normalverteilung (Log-Normalverteilung) ist eine stetige, rechts-schiefe Verteilung von Zufallsvariablen auf \mathbb{R}^+ mit Parametern μ und $\sigma > 0$. Die Log-Normalverteilung wird als semi-heavy-tailed klassifiziert und eignet sich daher ebenfalls prinzipiell zur Modellierung von nicht-negativen Zufallsvariablen mit viel Wahrscheinlichkeitsmasse am rechten Rand (vgl. Fahrmeir (2016), S. 278). Die Log-Normalverteilung erhält man durch die Transformation der normalverteilten Zufallsvariable X zu $Y = \log(X)$. Somit gilt: Die Zufallsvariable X ist Log-normalverteilt (vgl. Cruz (2015), S. 597), wenn ihre Dichte durch

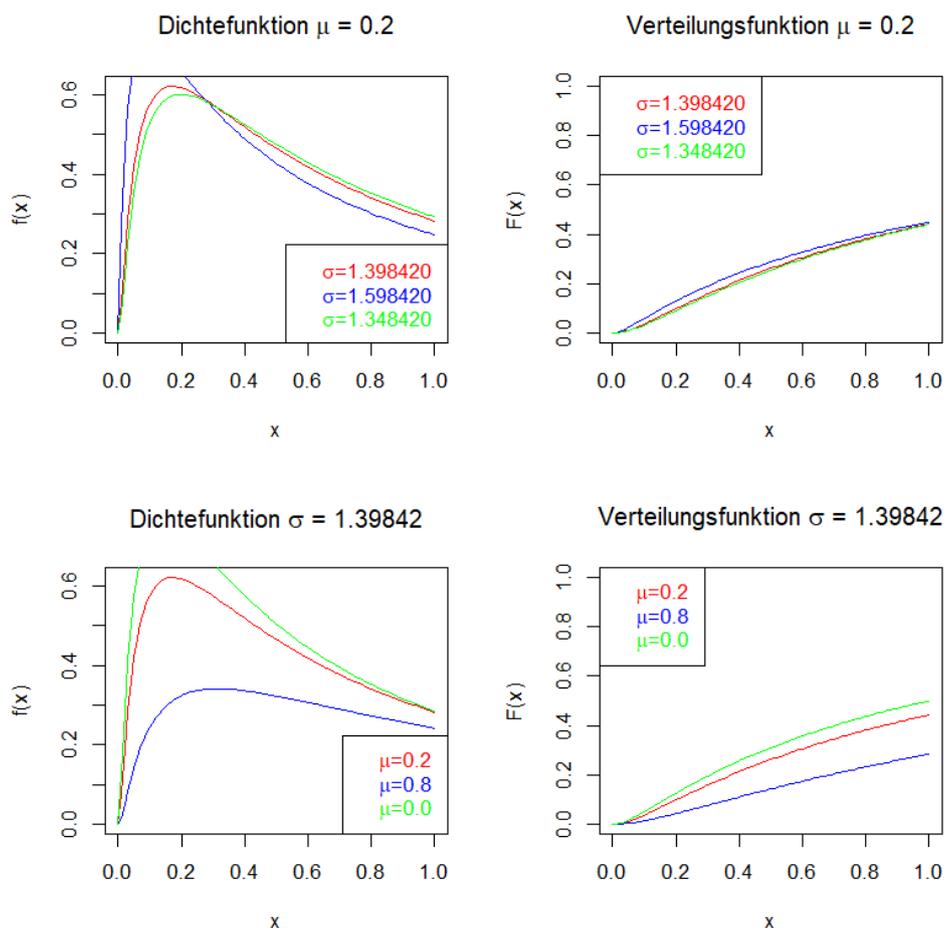


Abbildung 2: Dichte- und Verteilungsfunktion der Log-Normalverteilung in Abhängigkeit der Parameter

$$f_{Log-normal}(x|\mu, \sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right) \quad (7)$$

und ihre Verteilungsfunktion (vgl. Stelzer (2018), S. 50) durch

$$F_{Log-normal}(x|\mu, \sigma) = \Phi\left(\frac{\log(x) - \mu}{\sigma}\right), \quad (8)$$

wobei Φ die Verteilungsfunktion der Normalverteilung bezeichnet, gegeben sind. Die Form der Log-Normalverteilung, abhängig von μ und σ , ist in Abbildung 2 illustriert. Sowohl μ als auch σ beeinflussen den Grad der Steigung des Graphen.

2.3.4 American-Stoppa-Verteilung

Die American-Stoppa-Verteilung ist eine Verallgemeinerung der *Pareto-Verteilung*. Die Pareto-Verteilung zeichnet sich durch einen heavy Tail aus. Die Familie der

Pareto-Verteilungen eignet sich daher insbesondere, um Ereignisse, die sehr selten auftreten, jedoch eine hohe Auswirkung haben, zu modellieren. Das Verhalten der Tailfunktion ist analog zu dem einer Potenzfunktion (vgl. Calderín-Ojeda (2016), S. 818ff.). Eine Zufallsvariable X ist Pareto-verteilt, wenn ihre Dichtefunktion durch

$$f_{Pareto}(x|\theta, c) = \frac{\theta}{x} \left(\frac{c}{x}\right)^\theta \quad (9)$$

und ihre Verteilungsfunktion durch

$$F_{Pareto}(x|\theta, c) = 1 - \left(\frac{c}{x}\right)^\theta \quad (10)$$

mit und $x, c, \theta > 0$ gegeben sind (vgl. Cruz (2015), S.847). Die Pareto-Verteilung dient als Grundlage vieler weiterer Verteilungen, in diesem Fall der American-Stoppa-Verteilung. Durch entsprechende Wahl der Parameter lässt sich die American-Stoppa-Verteilung auf die Pareto-Verteilung zurückführen. Ihre Dichte ist durch

$$f_{American-Stoppa}(x|\theta, \lambda, c) = \frac{\lambda\theta}{c+x} \left(\frac{c}{c+x}\right)^\theta \left(1 - \left(\frac{c}{c+x}\right)^\theta\right)^\lambda \quad (11)$$

mit $c, \theta, \lambda > 0$ gegeben.

Die zugehörige Verteilungsfunktion lautet

$$F_{American-Stoppa}(x|\theta, \lambda, c) = \left(1 - \left(\frac{c}{c+x}\right)^\theta\right)^\lambda \quad (12)$$

mit $x, \theta, \lambda, c > 0$. Für den Fall $\lambda, \theta > 1$ erfüllt die American-Stoppa-Verteilung die Kriterien einer heavy-tailed Verteilung. Die Parameter λ und θ bilden folglich den Formparameter dieser Verteilung. Im Vergleich zur klassischen Pareto-Verteilung ist der Tail der American-Stoppa-Verteilung durch geeignete Wahl dieser Parameter schwerer.

Die Einführung der American-Stoppa-Verteilung diente ursprünglich der Beschreibung einer Einkommensverteilung. Typische Merkmale einer Einkommensverteilung lassen sich auf Merkmale von Schadensdaten übertragen. So sind auch diese entweder *nonmodal*⁴ oder *unimodal*⁵ und positiv asymmetrisch (vgl. Stoppa (1990), S. 34ff.). Die Zufallsvariable X nimmt Werte aus dem Bereich $(0, \infty)$ an, wobei sich ein Großteil der Beobachtungen im Intervall $(0, \tau]$ und ein kleinerer Teil der Beobachtungen im Intervall (τ, ∞) befindet. Der Parameter τ bezeichnet den Threshold zwischen Body und Tail der Verteilung.

⁴Eine nonmodale Verteilungsfunktion besitzt keinen eindeutigen häufigsten Wert.

⁵Eine unimodale Verteilungsfunktion ist gekennzeichnet durch einen Graphen mit einem einzigen, sichtbaren Hochpunkt.

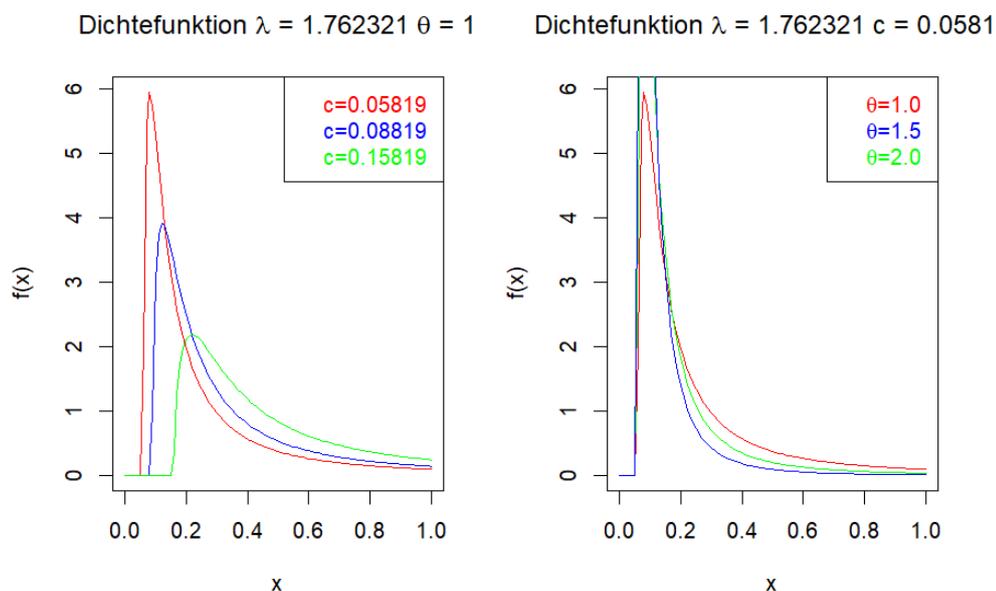


Abbildung 3: Dichtefunktionen der American-Stoppa-Verteilung. Links mit variablem c , rechts mit variablem θ

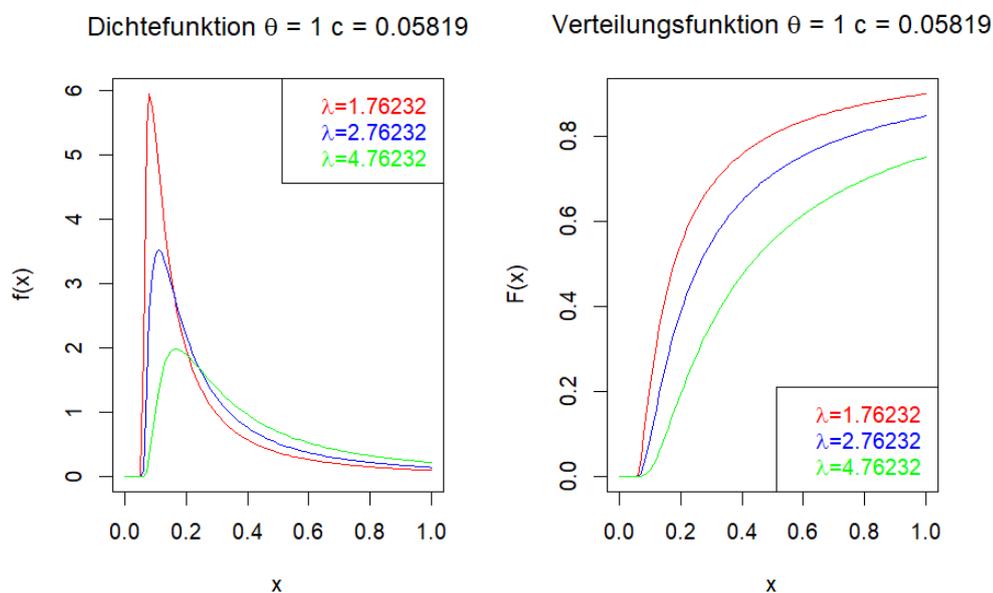


Abbildung 4: Dichte- und Verteilungsfunktion der American-Stoppa-Verteilung mit festem θ und c , λ variabel

In Abbildung 4 sind die Dichte- und Verteilungsfunktion der American-Stoppa-Verteilung illustriert. Hier ist der eben beschriebene Sachverhalt gut zu beobachten - die Veränderung des Parameters λ beeinflusst maßgeblich den Tail der Verteilung.

2.3.5 Generalized-Log-Pareto-Verteilung

Die Generalized-Log-Pareto-Verteilung basiert ebenfalls auf der Pareto-Verteilung⁶. Form- und Skalierungsparameter werden exponentiell zu zwei Formparametern transformiert. Zusätzlich wird ein Skalierungsparameter hinzugefügt, um wieder eine vollständige Log-Pareto-Verteilungsfamilie zu erhalten (vgl. Corman (2008), S. 94ff.). Durch anschließendes Logarithmieren der verallgemeinerten Pareto-Verteilung und Tauschen der Parameter, wie soeben beschrieben, erhält man die Generalized-Log-Pareto-Verteilung, welche als *super-heavy-tailed* zu klassifizieren ist (ebenda.). Dieses Vorgehen kann beliebig oft iteriert werden und führt mit jeder Wiederholung zu einer Generalized-Log-Pareto-Verteilung höherer Ordnung. Mit jeder Iteration wird ein weiterer Formparameter hinzugefügt, wobei der Skalierungsparameter unverändert bleibt. Je höher die Ordnung, umso langsamer konvergiert der Tail der Verteilung gegen Null, was zu *super-super-heavy* Tails führt.

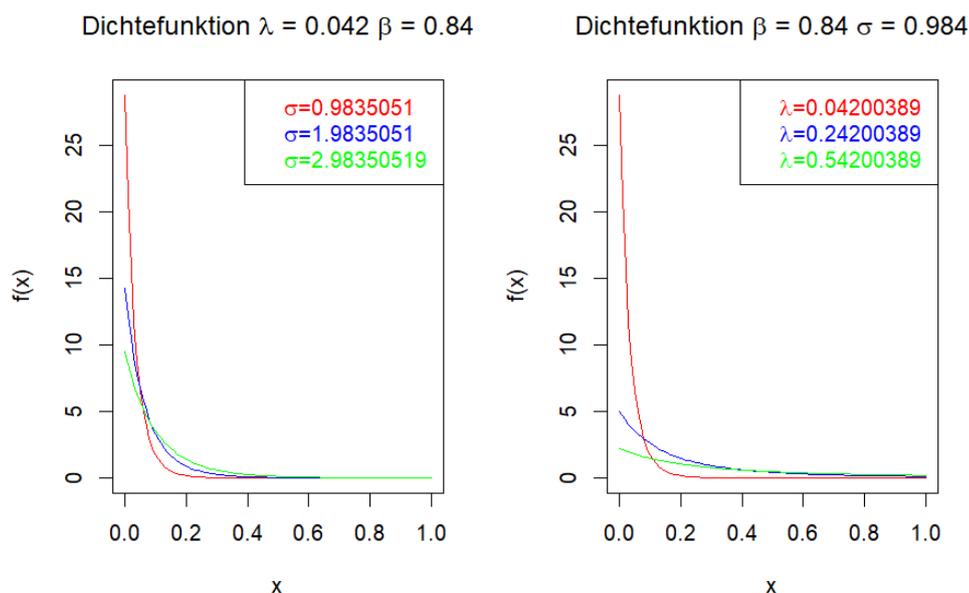


Abbildung 5: Dichtefunktion der GLP-Verteilung in Abhängigkeit von σ und λ

Die folgenden Erläuterungen beschränken sich auf die Generalized-Log-Pareto-Verteilung erster Ordnung. Wendet man die exponentielle Transformation auf die Pareto-Verteilung an, führt dies zu folgender Aussage (vgl. Corman (2008), S.95ff.): Eine Zufallsvariable X folgt der Generalized-Log-Pareto-Verteilung (GLP-Verteilung), wenn sie die Wahrscheinlichkeitsverteilung (ebenda.)

$$F_{GLP}(x|\lambda, \beta, \sigma) = 1 - \left(1 + \frac{1}{\beta} \log\left(1 + \frac{x}{\sigma}\right)\right)^{-\frac{1}{\lambda}} \quad (13)$$

⁶Es gelten die in Kapitel 3.4.3 vorgestellten Verteilungsannahmen der Pareto-Verteilung.

mit $x, \lambda, \beta, \sigma > 0$ und die zugehörige Dichtefunktion⁷

$$f_{GLP}(x|\lambda, \beta, \sigma) = \frac{1}{\lambda\beta\sigma\left(\frac{x}{\sigma} + 1\right)} \left(1 + \frac{1}{\beta} \log\left(\frac{x}{\sigma} + 1\right)\right)^{-\frac{1}{\lambda}-1} \quad (14)$$

besitzt.

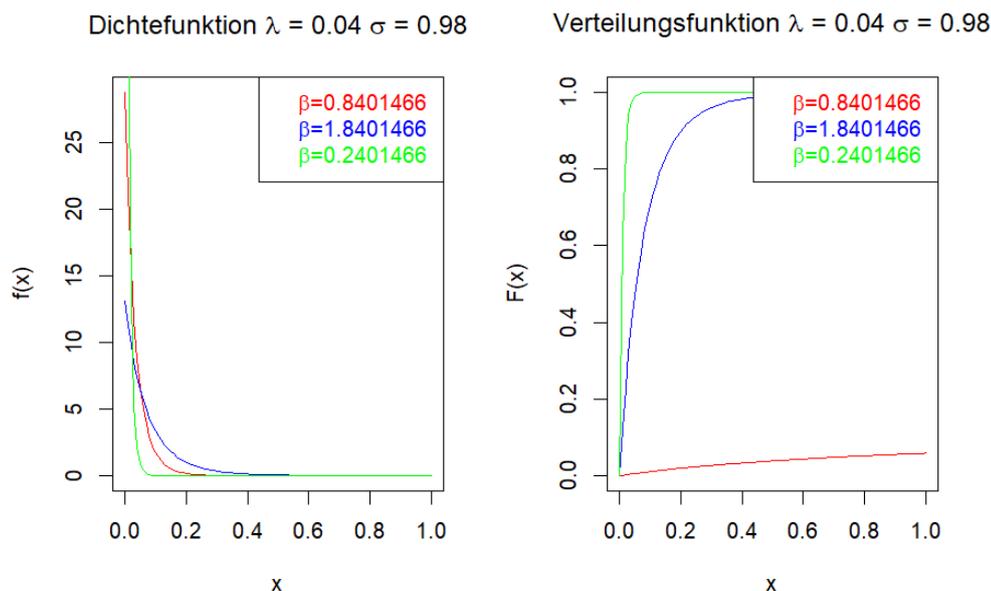


Abbildung 6: Dichte- und Verteilungsfunktion der GLP-Verteilung mit festem λ und σ , β variabel

2.4 Schätzung der Parameter

Die Schätzung der Parameter wird bei allen vier Verteilungen mit der *Maximum-Likelihood-Schätzung* durchgeführt. Die Methode der Maximum-Likelihood-Schätzung wird im Folgenden definiert (vgl. Fahrmeir (2016), S. 349).

Betrachtet wird allgemein die Dichtefunktion der Folge von identisch und unabhängig verteilten Zufallsvariablen X_1, \dots, X_n :

$$f(x_1, \dots, x_n|\theta) = f(x_1|\theta) \cdot \dots \cdot f(x_n|\theta)$$

mit Parametervektor θ . Im ersten Schritt wird die Likelihoodfunktion in Abhängigkeit von θ aufgestellt:

$$L(\theta) = f(x_1, \dots, x_n|\theta)$$

⁷Die zugehörige Dichtefunktion wurde nach Fahrmeir (2016) als Ableitung nach x der Verteilungsfunktion berechnet.

Maximiert man die Likelihoodfunktion, so erhält man zu x_1, \dots, x_n den geschätzten Parameter $\hat{\theta}$, mit dem die Eintrittswahrscheinlichkeit der vorliegenden Beobachtungen der Zufallsvariable maximal wird. Dieser Sachverhalt wird als Maximum-Likelihood-Prinzip bezeichnet. Es gilt somit:

$$L(\hat{\theta}) = \max_{\theta} L(\theta)$$

Das Maximum wird ermittelt, indem die erste Ableitung von L gebildet wird und die Nullstellen der Ableitung berechnet werden. Um diesen Vorgang zu vereinfachen, wird die Likelihoodfunktion üblicherweise zuerst logarithmiert, sodass sich die Log-Likelihoodfunktion ergibt:

$$\ell(\hat{\theta}) = \log(L(\hat{\theta})). \quad (15)$$

3 Begriffe der Testtheorie zur Überprüfung der Anpassungsgüte einer Verteilung

Offensichtlich besteht bei der Modellierung der Schadensdaten stets die Gefahr einer falschen Annahme (vgl. Steinhoff (2008), S.25). Ein solcher Fehler kann zum Beispiel gravierende Auswirkungen auf die Berechnung der Eigenkapitalunterlegung einer Bank oder Versicherung haben (ebenda.). Eine Überschätzung des Risikos kann dazu führen, dass Geldmittel gebunden werden, die in einem anderen Geschäftsfeld benötigt würden. Eine Unterschätzung der Situation dagegen gefährdet den Fortbestand des Unternehmens und führt im schlimmsten Fall zu dessen Insolvenz (ebd.). Um die Anpassungsgüte eines Modells zu testen, können zunächst visuelle Mittel genutzt werden, zum Beispiel ein *QQ-Plot* (ebd.). Auf der x-Achse werden die Quantile der theoretischen Verteilung, auf der y-Achse die Werte der tatsächlich vorliegenden empirischen Verteilung dargestellt. Die Daten und die theoretische Verteilung werden als Punktwolke in einem Koordinatensystem übereinander gelegt. Im theoretischen Fall idealer Anpassung befinden sich alle Beobachtungen auf der Winkelhalbierenden, weil die Quantile der beobachteten Verteilung exakt denen der theoretischen Verteilung entsprechen. Der QQ-Plot wird in der Simulationsstudie in Kapitel 5 erneut aufgegriffen.

Neben visueller Analyse werden GOF-Tests als zentrales Verfahren zur Überprüfung der Anpassungsgüte verwendet. Diese ermöglichen eine beobachterunabhängige Quantifizierung der Anpassungsgüte.

Der Testaufbau beginnt mit der Formulierung der Hypothesen, wobei die Nullhypo-

these H_0 die Gleichheit von empirischer und theoretischer Verteilung fordert. Für diese Arbeit sind die relevanten Hypothesen in Gleichung 21 benannt. Das Signifikanzniveau beziehungsweise der Ablehnbereich, durch welche die Fehlerwahrscheinlichkeit des Test begrenzt wird, werden im nächsten Schritt festgelegt. Anschließend folgt die Berechnung der Teststatistik, deren Wert die Testentscheidung bestimmt. Es handelt sich in diesem Fall um verteilungsfreie Testverfahren, die keine spezifische Verteilung der Teststatistik voraussetzen. Somit wird der Ablehnbereich nicht durch ein theoretisches Quantil, sondern durch die Berechnung des sogenannten p-Werts festgelegt (vgl. Cruz (2015), S. 248). Im Folgenden werden drei häufig genutzte GOF-Tests vorgestellt: der Kolmogorov-Smirnov-Test, der Cramér-von-Mises-Test und der Upper Tail Anderson-Darling-Test. Das Konzept dieser Tests beruht auf dem Vergleich der empirischen Verteilungsfunktion mit einer vermuteten theoretischen Verteilungsfunktion (vgl. Bayer (2012), S.148). Die Methoden lassen sich in *Supremum Class Statistics* und *Quadratic Class Statistics* gruppieren (vgl. Chernobai (2005), S.581ff.). Der Unterschied besteht in der Methodik der Berechnung der Teststatistiken. Bei der erster Klasse wird das Supremum der Divergenz zwischen der theoretischen und empirischen Verteilungsfunktion ermittelt. Zweitere Klasse mittelt und quadriert die Differenz zwischen den beiden Verteilungsfunktionen. Der Kolmogorov-Smirnov-Test ist den Supremum Class Statistiken zuzuordnen, der Cramér-von-Mises-Test und der Upper Tail Anderson-Darling-Test gehören zu den Quadratic Class Statistics.

Bevor die Testmethoden ausführlich diskutiert werden, sollen zunächst die Begriffe Güte, Signifikanz und Power eines Tests definiert werden.

3.1 Güte, Signifikanz und Power eines Tests

Für die Güte eines Anpassungstests sind insbesondere *Signifikanz* und *Power* entscheidend (vgl. Heddrich (2016), S.432). In Tabelle 1 sind die beiden Fehlertypen dargestellt, die bei einer Testentscheidung auftreten können: der Fehler 1. Art und der Fehler 2. Art. Der Fehler 1. Art wird begangen, wenn die Nullhypothese zu Unrecht abgelehnt wird. Unter dem Fehler 2. Art versteht man, dass die Nullhypothese fälschlicherweise beibehalten wird. Es ist leicht einzusehen, dass die beiden Fehlertypen in einem Spannungsverhältnis zueinander stehen, sodass die Begrenzung eines Fehlers den anderen begünstigt (ebd.).

		Realität	
		H_0 ist wahr	H_1 ist wahr
Test- entscheidung	H_0 ist wahr	Testentscheidung korrekt	Fehler 2. Art (β -Fehler)
	H_1 ist wahr	Fehler 1. Art (α -Fehler)	Testentscheidung korrekt

Tabelle 1: Fehler bei Testentscheidungen

Signifikanz und Power eines Tests stehen in direktem Zusammenhang mit dem Fehler 1. Art und dem Fehler 2. Art. Die Signifikanz S entspricht der Wahrscheinlichkeit, dass H_0 korrekterweise beibehalten wird. Formal ist S folgendermaßen definiert (vgl. Cruz (2015), S. 247):

$$S = 1 - \alpha = 1 - \mathbb{P}(x \in R | H_0) \quad (16)$$

α steht hierbei für den Fehler 1. Art, der durch das Signifikanzniveau beschränkt wird. Die Power P hingegen, siehe Gleichung (17), beschreibt die Wahrscheinlichkeit, keinen Fehler 2. Art zu begehen, sprich die Nullhypothese korrekterweise abzulehnen, wenn H_1 der Wahrheit entspricht. Die Power ergibt sich dementsprechend als Komplement des Fehlers 2. Art, der in der Definition als β bezeichnet wird, zu (ebd.):

$$P = 1 - \beta = 1 - \mathbb{P}(x \in A | H_1) \quad (17)$$

Die Power eines Tests hängt von unterschiedlichen Aspekten ab. Dazu zählen vor allem das Signifikanzniveau α und die Eignung von Verteilungsannahme, Testverfahren und Teststatistik für die vorliegenden Daten. Es bestehen in diesem Sinne auch verschiedene Möglichkeiten, die Power eines Tests zu verbessern, indem man beispielsweise ein höheres Signifikanzniveau zulässt. Letzteres impliziert jedoch eine höhere Toleranz gegenüber dem Fehler 1. Art. Um eine neutrale Interpretation der Ergebnisse zu gewährleisten, sollten solche Parameter daher vor der ersten Durchführung des Tests festgelegt und im Nachhinein nicht mehr angepasst werden.

Die typische Herangehensweise besteht darin, mit einem fixierten Signifikanzniveau denjenigen Test unter allen Niveau- α -Tests zu finden, der den geringsten Fehler 2. Art aufweist. Diesem Test wird die höchste statistische Power zugeschrieben (vgl. Heddrich (2016), S. 430ff.).

Die Simulationsstudie, deren Resultate in Kapitel 4 präsentiert werden, beurteilt vier verschiedene Anpassungstests anhand ihrer Power bezüglich der richtigen Verteilungsklassifikation.

3.2 p-Wert nach Fisher

Das zentrale Kriterium für die Entscheidung, ob die Nullhypothese abgelehnt werden kann, ist hier der sogenannte *p-Wert*. Dieser gibt die Wahrscheinlichkeit unter H_0 an, den Wert der Teststatistik oder einen noch mehr in Richtung der Alternativhypothese H_1 weisenden Wert zu beobachten (vgl. Fahrmeir (2016), S. 381ff.). Die Höhe des p-Werts gibt somit das geringste Signifikanzniveau an, zu dem man die Nullhypothese ablehnen kann (vgl. Heddrich (2016), S. 431). Die Berechnung des p-Werts für eine Teststatistik T erfolgt bei allen Anpassungstests identisch:

$$\text{p-Wert} = \mathbb{P}(|T| \geq t | H_0) \quad (18)$$

t bezeichnet dabei den tatsächlich beobachteten Wert der jeweiligen Teststatistik (vgl. Cruz (2015), S.262).

Weiterhin sind der Begriff der empirischen Verteilungsfunktion sowie zwei nützliche Theoreme, nämlich der *Satz von Glivenko-Cantelli* und die *Dvoretzky-Kiefer-Wolfowitz-Ungleichung*, zu beachten.

3.3 Konvergenzkriterien für Goodness-of-Fit-Tests

3.3.1 Empirische Verteilungsfunktion

Die nachfolgend diskutierten Tests auf Anpassungsgüte basieren auf der Divergenz zwischen empirischer Verteilungsfunktion und der angenommenen theoretischen Verteilung. Die empirische Verteilung einer vorliegenden Stichprobe mit Realisationen x_1, \dots, x_n wird wie folgt notiert (vgl. Cruz (2015), S. 248):

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[x_i \leq x] \quad (19)$$

Von dieser Notation wird in den Kapiteln 3.4.1, 3.4.2 und 3.4.3 Gebrauch gemacht. Die empirische Verteilungsfunktion ist Teil der Teststatistik eines jeden Tests, der im Rahmen dieser Arbeit verwendet wird.

3.3.2 Satz von Glivenko-Cantelli

Der Satz von Glivenko-Cantelli bildet eine wichtige Eigenschaft der empirischen Verteilungsfunktion für die nachfolgenden Teststatistiken ab. Mit wachsender Stichprobengröße n konvergiert die empirische Verteilungsfunktion fast sicher⁸ gegen die theoretische Verteilungsfunktion F_0 (vgl. Cruz (2015), S. 249):

$$\|F_n - F_0\|_\infty = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_0(x)| \xrightarrow{f.s.} 0 \quad (20)$$

3.4 Goodness-of-Fit-Tests

Der Testaufbau ist für alle Tests identisch. Man geht von einer Stichprobe mit Beobachtungen (Rohdaten) x_1, \dots, x_n aus. Diese Beobachtungen werden im ersten Schritt aufsteigend sortiert. Nach der Berechnung des arithmetischen Mittels μ und der Standardabweichung σ werden die Beobachtungen standardisiert (vgl. Fahrmeir (2016), S. 49). Dazu wird für jede Beobachtung die Differenz zum arithmetischen Mittel bestimmt und durch die Standardabweichung dividiert. Die empirische Verteilungsfunktion dieser Stichprobe wird mit $\hat{F}(x)$, die Verteilungsfunktion der zu vergleichenden theoretischen Verteilung mit $F_0(x)$ bezeichnet. Auch die Hypothesen, die es zu einem vorgegebenen Signifikanzniveau (vgl. Fahrmeir (2016), S. 373) zu testen gilt, sind bei allen hier untersuchten Tests analog:

$$H_0 : \hat{F}(x) = F_0(x) \quad vs. \quad H_1 : \hat{F}(x) \neq F_0(x). \quad (21)$$

GOF-Tests analysieren die Diskrepanz zwischen den unter Anwendung des Modells zu erwartenden Daten und den tatsächlichen Schadensdaten (vgl. Cruz (2015)).

3.4.1 Kolmogorov-Smirnov-Anpassungstest

Der Kolmogorov-Smirnov-Test ist ein weit verbreiteter Test zum Vergleich zweier Verteilungen. Hier werden einseitiger und beidseitiger Kolmogorov-Smirnov-Test unterschieden. Der einseitige Test bewertet, ob die vorliegende Stichprobe einer bekannten theoretischen Verteilung folgt. Der beidseitige Test prüft zwei vorliegende Stichproben auf Verteilungsgleichheit (vgl. Cruz (2015), S. 261). Für die Simulationsstudie in Kapitel 4 ist der einseitige Kolmogorov-Smirnov-Test (im Folgenden KS-Test) relevant. Er wird in dieser Passage erläutert.

Seien für alle nachfolgend vorgestellten Tests x_1, \dots, x_n beobachtete Schadensdaten, die aus einer Folge von unabhängig und identisch verteilten (iid.) Zufallsschadens-

⁸Für eine genaue Definition verschiedener Konvergenzbegriffe siehe zum Beispiel Henze (2019).

daten X_1, \dots, X_n stammen. Die Hypothesen werden wie in (21) gewählt. Die KS-Teststatistik berechnet den maximalen Wert der Differenz zwischen empirischer Verteilung der vorliegenden Stichprobe und theoretischer Verteilungsfunktion. Im Allgemeinen lautet die Teststatistik:

$$KS = \sqrt{n} \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_0(x) \right| = \max_{1 \leq i \leq n} \left(F_0(x_i) - \frac{i-1}{n}, \frac{i}{n} - F_0(x_i) \right) \quad (22)$$

Die Teststatistik ist also lediglich von der Stichprobengröße n abhängig (ebd.). Für linkstrunkierte Verteilungen ergibt sich die Teststatistik des KS-Tests aus dem Ersetzen von $F_0(x)$ durch die Verteilungsfunktion für linkstrunkierte Verteilungen aus Formel 3 (vgl. Chernobai (2005), S. 581):

$$KS^* = \sqrt{n} \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - \frac{F_0(x) - F_0(\tau)}{1 - F_0(\tau)} \right| \quad (23)$$

Der Parameter τ bezeichnet hier den zuvor berechneten Threshold. Die Nullhypothese wird abgelehnt, falls der p-Wert kleiner als das Signifikanzniveau α ist.

Beim KS-Test ist zu beachten, dass dieser nur exakte kritische Werte liefert, wenn die Modellparameter bekannt sind und nicht geschätzt werden. Zudem ist der einfache KS-Test besser geeignet für eine light-tailed Verteilung. Als Alternative kann hier der Anderson-Darling-Test genutzt werden⁹. Eine Abwandlung des Anderson-Darling-Tests, die sich besonders gut für heavy-tailed Verteilungen eignet, wird in Kapitel 3.4.3 vorgestellt.

3.4.2 Cramér-von-Mises-Test

Der einseitige¹⁰ Cramer-von-Mises Test (CvM-Test) ist eine weitere Methode zum Vergleich der Verteilung einer Stichprobe x_1, \dots, x_n mit einer theoretischen Verteilung F_0 . Die Teststatistik des CvM-Tests ermittelt die quadrierte Differenz zwischen der vermuteten Verteilung und der vorliegenden empirischen Verteilung der Stichprobe (vgl. Bayer (2012), S. 149):

$$w^2 = n \int_{-\infty}^{\infty} \left(\hat{F}_n(x) - F_0(x, \theta) \right) dF(x, \theta) \quad (24)$$

θ steht hierbei für den Parametervektor der jeweiligen Verteilung. Die explizite Berechnung der Teststatistik erfolgt mit Gleichung (25) (ebd.).

$$w^2 = \frac{1}{12n} + \sum_{i=1}^n \left(\frac{2i-1}{2n} - F_0(x_i) \right)^2 \quad (25)$$

⁹Vergleiche Cruz (2015), S. 277 für eine Darstellung des Anderson-Darling-Tests

¹⁰Es gilt die Erklärung von einseitig in Kapitel 3.4.1.

Auch für den CvM-Test lässt sich die Teststatistik für eine trunkierte Verteilung mittels des Thresholds τ berechnen (vgl. Chernobai (2005), S. 583):

$$w^{2*} = n \int_{\tau}^{\infty} \left(\hat{F}_n(x) - F_0(x, \theta) \right) dF(x, \theta) \quad (26)$$

Die untere Integrationsgrenze wird in diesem Fall, im Gegensatz zu Formel 24, durch den Threshold gebildet.

3.4.3 Anderson-Darling-Test

Der Anderson-Darling Test (AD-Test) wird ebenso wie der KS-Test und CvM-Test verwendet, um die Anpassungsgüte einer theoretischen Verteilung an eine empirische Verteilungsfunktion zu bewerten.

Die Teststatistik des AD-Tests ergibt sich aus einer Gewichtsfunktion $\Psi(x)$ und der Teststatistik des CvM-Tests. Für $\Psi(x) \equiv 1$ sind die Teststatistiken identisch.

$$w^2 = n \int_{-\infty}^{\infty} \left(\hat{F}_n(x) - F_0(x, \theta) \right) \Psi(x) dF(x, \theta) \quad (27)$$

Mit der häufig verwendeten Gewichtsfunktion $\Psi = \{F_0(x, \theta)[1 - F_0(x, \theta)]\}^{-1}$ kann die Teststatistik des AD-Tests wie folgt dargestellt werden (vgl. Lenart (2016)):

$$A^2 = n \int_{-\infty}^{\infty} \frac{\left(\hat{F}_n(x) - F_0(x, \theta) \right)}{\left(F_0(x, \theta)[1 - F_0(x, \theta)] \right)} dF(x, \theta) \quad (28)$$

Die Berechnung der Teststatistik erfolgt dann durch

$$T = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) (\log[F_0(x_i, \theta)] + \log[1 - F_0(x_{n-i+1}, \theta)]). \quad (29)$$

3.4.4 Quadratic Class Upper Tail Anderson-Darling Test

Die ursprüngliche Version der Anderson-Darling-Statistik lässt sich mit einer Gewichtsfunktion modifizieren. Die abgewandelte Version ist für Verteilungen mit einem heavy Tail besser geeignet, da die Gewichtung zu einer exakteren Teststatistik im Randbereich und somit zu einer höheren statistischen Power führt (vgl. Chernobai (2005), S. 584f.). Es gibt für diesen Fall wiederum zwei unterschiedliche Möglichkeiten des gewichteten Anderson-Darling-Tests, welche sich für die Verteilungsanpassung im Randbereich der Verteilung eignen (ebd.), einen Supremum Class und einen Quadratic Class Upper Tail Anderson-Darling-Test (AD2up-Test). Vorgestellt wird letztgenannter Test. Die Teststatistik des AD2up-Tests lautet wie folgt (vgl. Cruz (2015), S. 277):

$$AD_{up}^2 = n \int_{-\infty}^{\infty} \frac{(F_0(x) - \hat{F}_n(x))^2}{(1 - \hat{F}_n(x))^2} d\hat{F}_n(x) \quad (30)$$

Die Teststatistik ist mit folgender Gleichung zu berechnen:

$$AD_{up}^2 = \frac{n}{2} - 2 \sum_{i=1}^n F_0(X_{i,n}) - \sum_{i=1}^n \left(2 - \frac{2i-1}{n} \right) \log(1 - F_0(X_{i,n})) \quad (31)$$

Für den Fall, dass linkstrunkierte Verlustdaten vorliegen, verschiebt sich die untere Integrationsgrenze erneut zum Threshold τ (vgl. Chernobai (2005)):

$$AD_{up}^{2*} = n \int_{\tau}^{\infty} \frac{(F_0(x) - \hat{F}_n(x))^2}{(1 - \hat{F}_n(x))^2} d\hat{F}_n(x) \quad (32)$$

Aus den unterschiedlichen Teststatistiken der diskutierten GOF-Tests wird ersichtlich, dass deren Berechnungsansätze jeweils unterschiedliche Aspekte stärker berücksichtigen. Für den konkreten Anwendungsfall ist es daher von großer Bedeutung, denjenigen Test identifizieren zu können, der das Zutreffen der unterstellten Schadenshöhenverteilung am zuverlässigsten überprüft.

Zu diesem Zweck wird im Rahmen dieser Arbeit eine Simulationsstudie mit den unterschiedlichen Verteilungsszenarien aus Kapitel 2.3 durchgeführt. Im folgenden Kapitel sollen die Ergebnisse dieser Analyse präsentiert und kritisch gewürdigt werden.

4 Simulationsstudie mit den vorgestellten Testverfahren hinsichtlich der Verteilungen

Die Simulationsstudie wird mit der Software R (Version 1.1.456 – © 2009-2018 RStudio, Inc.) durchgeführt. Der den nachfolgenden Ergebnissen zugrundeliegende Code befindet sich im elektronischen Anhang.

Für die Log-Normalverteilung sind bereits standardmäßig Funktionen für die Verteilungs-, Dichte- und Quantilsfunktion implementiert. Ebenso kann auf eine Funktion zur Ziehung von Zufallszahlen aus dieser Verteilung zurückgegriffen werden. Für die Gompertz-Verteilung sind solche Funktionen in unterschiedlichen R-Packages verfügbar, jedoch nicht einheitlich definiert. Aus diesem Grund wurden die Funktionen analog zu Kapitel 2.3.2 implementiert. Für die American-Stoppa-Verteilung und die GLP-Verteilung sind keine Implementierungen vorhanden, weshalb diese in Analogie zu den anderen Verteilungen ebenfalls selbst definiert werden.

Die Datengrundlage der Simulationsstudie besteht aus zwei Datensätzen mit Schadensdaten aus dem operationellen Risiko.

4.1 Datenbasis der Simulationsstudie

Vor Beginn der Simulationsstudie werden die Datensätze zunächst deskriptiv analysiert. Es handelt sich bei beiden Datensätzen um quasi-stetige Daten¹¹. In Abbildung 7 ist die empirische Dichte des jeweiligen Datensatzes abgebildet. Da diese für beide Datensätze einen heavy Tail besitzt, wurde zur besseren Übersichtlichkeit die Darstellung der Dichte auf das Intervall $[0,100000]$ beschränkt.

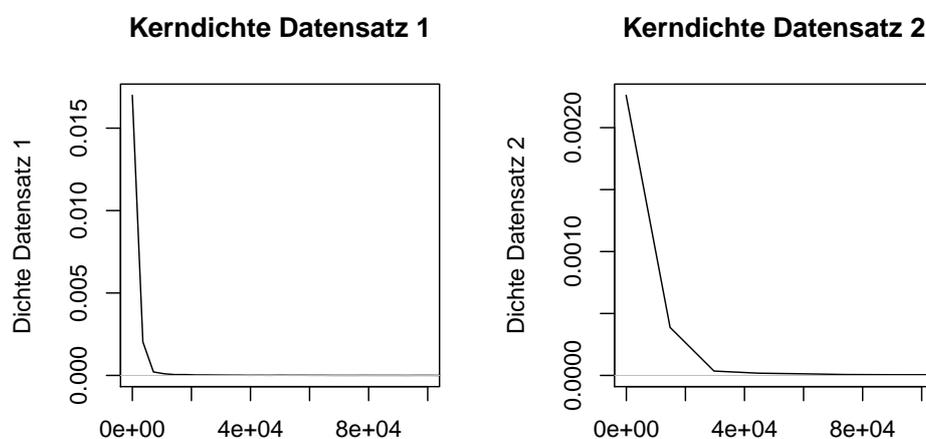


Abbildung 7: Graphische Darstellung der Kerndichte beider Datensätze im Intervall $[0,100000]$

Datensatz 1

Der erste Datensatz enthält mit 14477 Beobachtungen einer Variablen, die Schadensdaten aus dem operationellen Risiko darstellt. Abbildung 8 visualisiert die Fünf-Punkte-Zusammenfassung¹² mittels eines Boxplots. Die Daten oberhalb und unterhalb des Mittelwerts werden in zwei Boxplots separat betrachtet. Das Minimum der Daten liegt bei 5000, der Mittelwert bei 287286 und das Maximum bei 182820000. Die Schiefe der Daten beträgt 30.45. Da $30.45 \gg 1$, ist von einer extrem rechtsschiefen Verteilung auszugehen. Dieses Merkmal wird auch in Abbildung 8 deutlich. Die Kurtosis erfüllt mit $1167.43 \gg 1$ die Kriterien einer sehr spitzen Verteilung. Durch

¹¹Unter quasi-stetigen Daten versteht man Daten, die zwar diskret sind, sich aber in sehr kleinen Intervalle einteilen lassen. Hierzu zählen zum Beispiel Währungsbeträge (vgl. Fahrmeir (2016), S.15).

¹²Die Fünf-Punkte-Zusammenfassung (Minimum, 25%-Quantil, Median, 75%-Quantil, Maximum) wird zum Beispiel bei Fahrmeir (2016), S.62 definiert.

die große Spannweite der Datenpunkte lässt sich die hohe Standardabweichung von 3646554 erklären. Der Variationskoeffizient liegt bei 12.69.

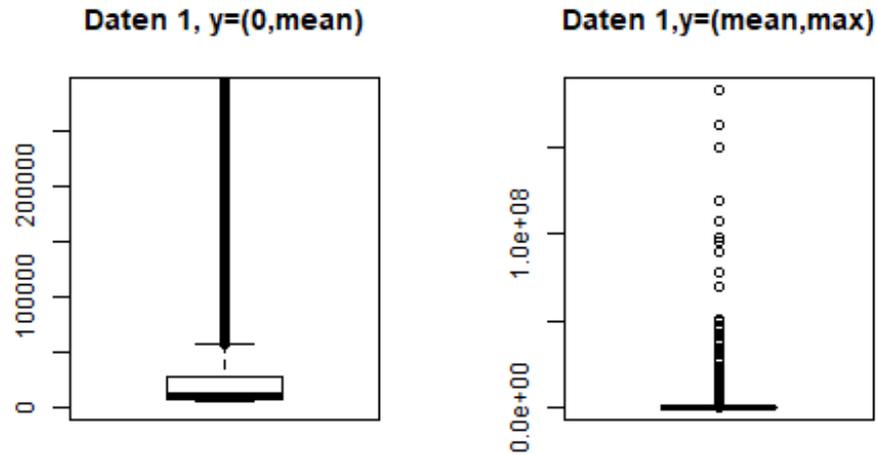


Abbildung 8: Graphische Darstellung der Fünf-Punkte-Zusammenfassung des ersten Datensatzes

Datensatz 2

Der zweite Datensatz der Simulationsstudie enthält ebenfalls eine Variable mit Schadensdaten aus dem operationellen Risiko. Alle Kennzahlen werden analog zu Datensatz 1 berechnet und sind ebenfalls in der Tabelle 2 zusammengefasst. Die Abbildung 9 stellt die Fünf-Punkte-Zusammenfassung des zweiten Datensatzes graphisch dar. Das Minimum der Daten liegt bei 333, der Mittelwert bei 16769 und das Maximum bei 7606221.

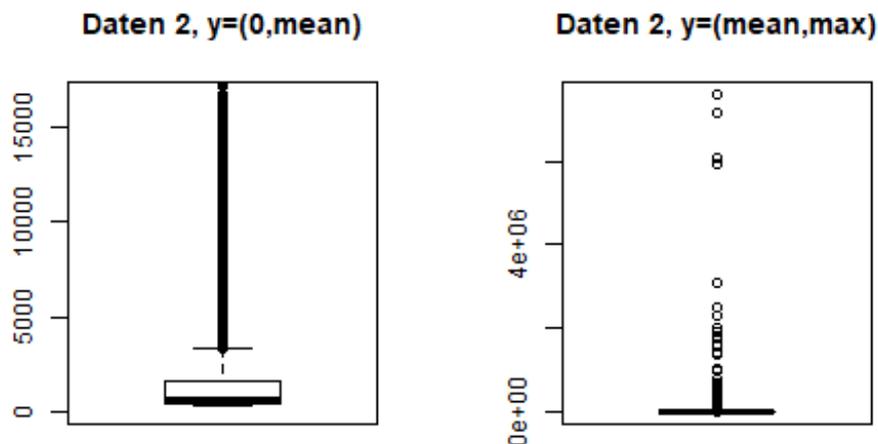


Abbildung 9: Graphische Darstellung der Fünf-Punkte-Zusammenfassung des zweiten Datensatzes

	Datensatz 1	Datensatz 2
Variationskoeffizient	12.69	12.94
Schiefe	30.45	26.17
Kurtosis	1167.43	788.89

Tabelle 2: Übersicht von Variationskoeffizient, Schiefe und Kurtosis beider Datensätze

Die Variationskoeffizienten beider Datensätze unterscheiden sich nur marginal. Auch die Schiefe der Daten liegt bei beiden Datensätzen deutlich über Eins, weshalb von einer extrem rechtsschiefen Verteilung auszugehen ist. Dieses Merkmal wird auch in den Abbildungen 8 und 9 visualisiert. Die Kurtosis erfüllt ebenfalls mit einem Wert $\gg 1$ die Kriterien einer sehr spitzigen Verteilung. Die Kurven der Verteilungen beider Datensätze entsprechen folglich den spezifischen Charakteristika von Daten aus dem operationellen Risiko.

4.2 Power der Testverfahren bei Non-Spliced-Modellen

Ziel dieser Simulationsstudie zu den Non-Spliced-Modellen ist es, vier verschiedene Verteilungsszenarien zu erstellen, in denen die (simulierten) Daten jeweils tatsächlich der zu prüfenden Verteilung folgen, und in diesen Szenarien die Anpassungsgüte der Tests gegen alle vier Verteilungen zu kontrollieren. Idealerweise sollten die Tests die vorliegende Verteilung erkennen und die Nullhypothesen, dass die Daten einer der drei anderen Verteilungen zuzuordnen sind, ablehnen.

4.2.1 Vorbereitung

Um die Verteilungsszenarien möglichst realistisch zu gestalten, werden aus den vorliegenden Rohdaten mittels der Maximum-Likelihood-Methode für alle vier Verteilungen die Maximum-Likelihood-Schätzer zu den jeweiligen Verteilungsparametern berechnet. Die Schätzung der Parameter aus realen Schadensdaten des operationellen Risikos (OpRisk-Daten) erlaubt dann das Ziehen von je 1000 Stichproben mit der Größe $n = 3000$, welche die Verteilung der OpRisk-Daten möglichst realitätsnah abbilden.

Insgesamt lässt sich das mehrstufige Vorgehen also wie folgt zusammenfassen:

1. Berechne für die Parameter der Verteilung i ($i \in \{\text{Gompertz-Verteilung, Log-Normalverteilung, American-Stoppa-Verteilung, GLP-Verteilung}\}$) die Maximum-Likelihood-Schätzer $\hat{\Theta}_i$ aus den Rohdaten
2. Simuliere 1000 Stichproben der Länge 3000, die Verteilung i mit Parametervektor θ_i folgen
3. Berechne für jede dieser Stichproben die Maximum-Likelihood-Schätzer zu den Parametern aller vier Verteilungen
4. Teste für jede Stichprobe und für jede Verteilung, ob die Stichprobe der jeweiligen Verteilung mit den in 3. geschätzten Parametern folgt

Wahl der Startwerte bei der Maximum-Likelihood-Schätzung

Im Rahmen der Simulation stellt sich heraus, dass die Wahl der Startwerte einen entscheidenden Einfluss auf die Maximum-Likelihood-Schätzung in R hat. Ein erster Versuch, dieses Problem für die Gompertz-, Log-Normal- und American-Stoppa-Verteilung zu lösen, besteht in der Ermittlung geeigneter Startwerte mittels der Minimierung des Abstandes zwischen empirischem und theoretischem Mittelwert bzw. Varianz. Dieses Vorgehen ist in Formel 33 dargestellt¹³ und wird in Kapitel 4.4 kritisch diskutiert.

$$\min((EW(X) - \bar{x})^2 + (VAR(x) - s^2)^2) \quad (33)$$

Das Verfahren lässt sich jedoch nicht auf die GLP-Verteilung anwenden, da deren Momente unendlich sind. Außerdem benötigt der Optimierungsalgorithmus seinerseits Startwerte, die wiederum die Schätzung der eigentlichen Startwerte der Pa-

¹³mit \bar{x} als empirischen Mittelwert und s^2 als empirische Varianz

parameter beeinflussen, sodass die Problematik lediglich verlagert wird. Aus diesem Grund wird die Methode als Lösungsansatz verworfen.

Solche Optimierungsprobleme sind generell nicht trivial zu lösen und stellen daher ein Feld lebhafter Forschung dar. Da Parameteroptimierung jedoch außerhalb des Rahmens dieser Arbeit liegt und keine simple Alternativlösung zur Hand ist, werden die Startwerte für die Log-Normal-, American-Stoppa- und GLP-Verteilung einheitlich auf den Wert 1 festgelegt. Die Startwerte für die Gompertz-Verteilung werden ebenfalls fixiert, müssen aber kleiner¹⁴ gewählt werden, damit die Verteilung von ähnlicher Natur wie die der echten Schadensdaten ist.

Die Parametervektoren $\hat{\theta}$ der jeweiligen Verteilung, mit welchen die Stichproben gezogen wurden, ergeben sich somit wie folgt für den Datensatz 1:

$$\begin{aligned}\hat{\Theta}_{Gompertz} &= (\hat{a} = 4.443749e - 04, \hat{b} = -5.931464e - 06) \\ \hat{\Theta}_{Log-Normal} &= (\hat{\mu} = 9.78289, \hat{\sigma} = 1.398420) \\ \hat{\Theta}_{American-Stoppa} &= (\hat{\theta} = 1.239848, \hat{\lambda} = 9322.52, \hat{c} = 6.4587) \\ \hat{\Theta}_{GLP} &= (\lambda = 0.00011, \hat{\beta} = 1794.43, \hat{\sigma} = 1.7692.64)\end{aligned}$$

Tabelle 3: Ergebnisse der Maximum-Likelihood-Schätzung aus Datensatz 1

und für den Datensatz 2:

$$\begin{aligned}\hat{\Theta}_{Gompertz} &= (\hat{a} = 0.0003438942, \hat{b} = -0.00002296459) \\ \hat{\Theta}_{Log-Normal} &= (\hat{\mu} = 6.989496, \hat{\sigma} = 1.353078) \\ \hat{\Theta}_{Stoppa} &= (\hat{\theta} = 1.311310, \hat{\lambda} = 35799.61, \hat{c} = 0.268708) \\ \hat{\Theta}_{GLP} &= (\lambda = 0.00993, \hat{\beta} = 0.000000000000035266, \hat{\sigma} = 1.001659)\end{aligned}$$

Tabelle 4: Ergebnisse der Maximum-Likelihood-Schätzung aus Datensatz 2

Mit den geschätzten Parametern aus Tabelle 3 und 4 werden nun aus jeder Verteilung je 1000 Vektoren aus 3000 Zufallszahlen gezogen. Zur Überprüfung der Parameterschätzungen ist in Abbildung 10 und 11 die Anpassung der Verteilungen mit den geschätzten Parametern aus Tabelle 3 und 4 abgebildet. Die rote Linie stellt die Dichte der jeweiligen theoretischen Verteilung mit den entsprechenden geschätzten Parametern dar, die schwarze Linie repräsentiert die empirische Dichte der Daten. Die Grafiken zur Anpassung machen bereits die Spezifika der Verteilungen deutlich. Die Gompertz-Verteilung scheint im Body der Verteilung eine relativ gute Anpassung zu erzielen, jedoch weniger im Tail. Dies gilt besonders für den zweiten Datensatz. Die Log-Normal- und die American-Stoppa-Verteilung zeigen sowohl im Body

¹⁴ $c(0.00001, 0.00001)$ für Datensatz 1 und $c(0.00001, -0.001)$ für Datensatz 2

als auch im Tail eine adäquate Annäherung an die empirische Verteilung der Daten. Die GLP-Verteilung hingegen gleicht lediglich im Tail der empirischen Verteilung.

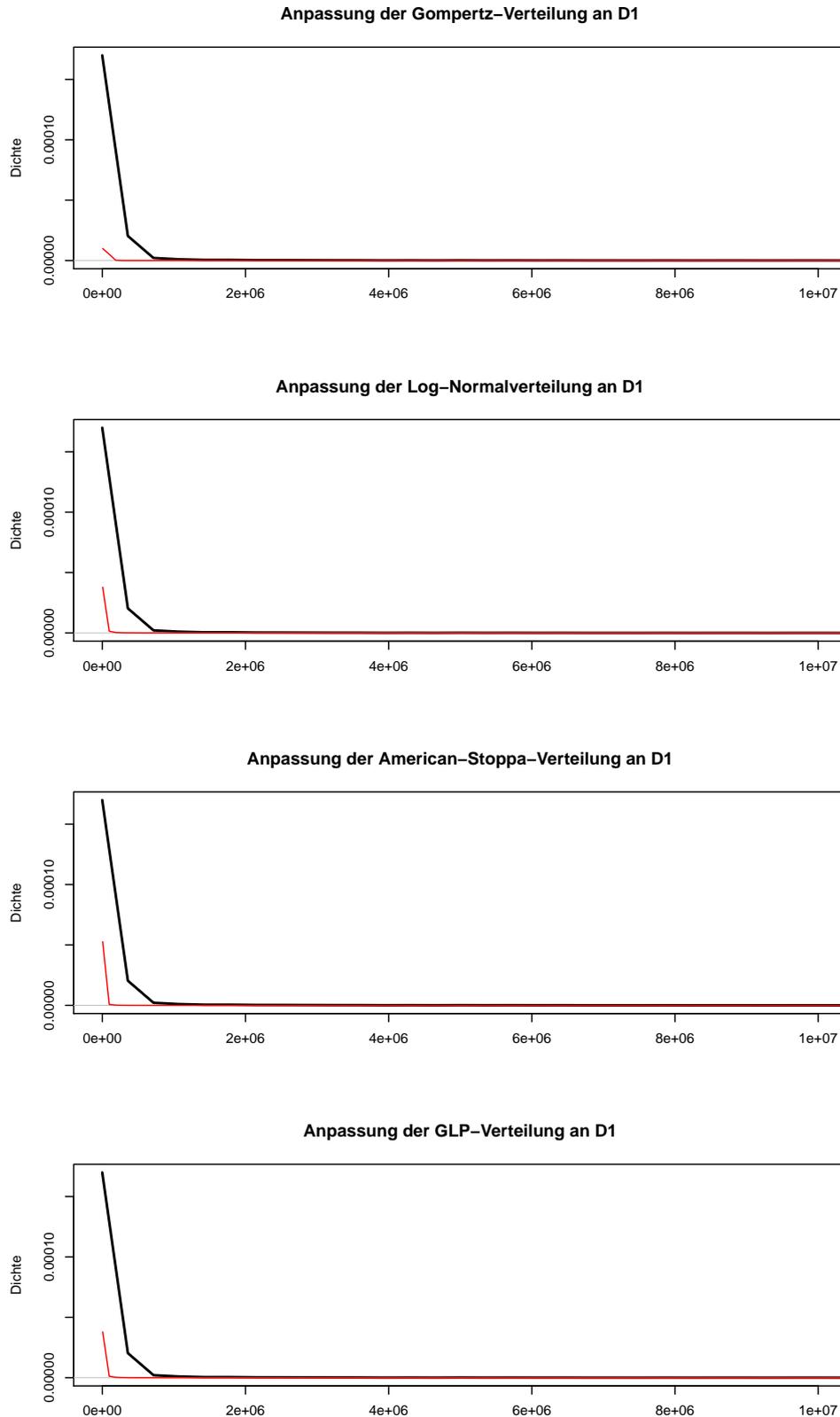


Abbildung 10: Anpassung der vier Verteilungen an den Datensatz 1

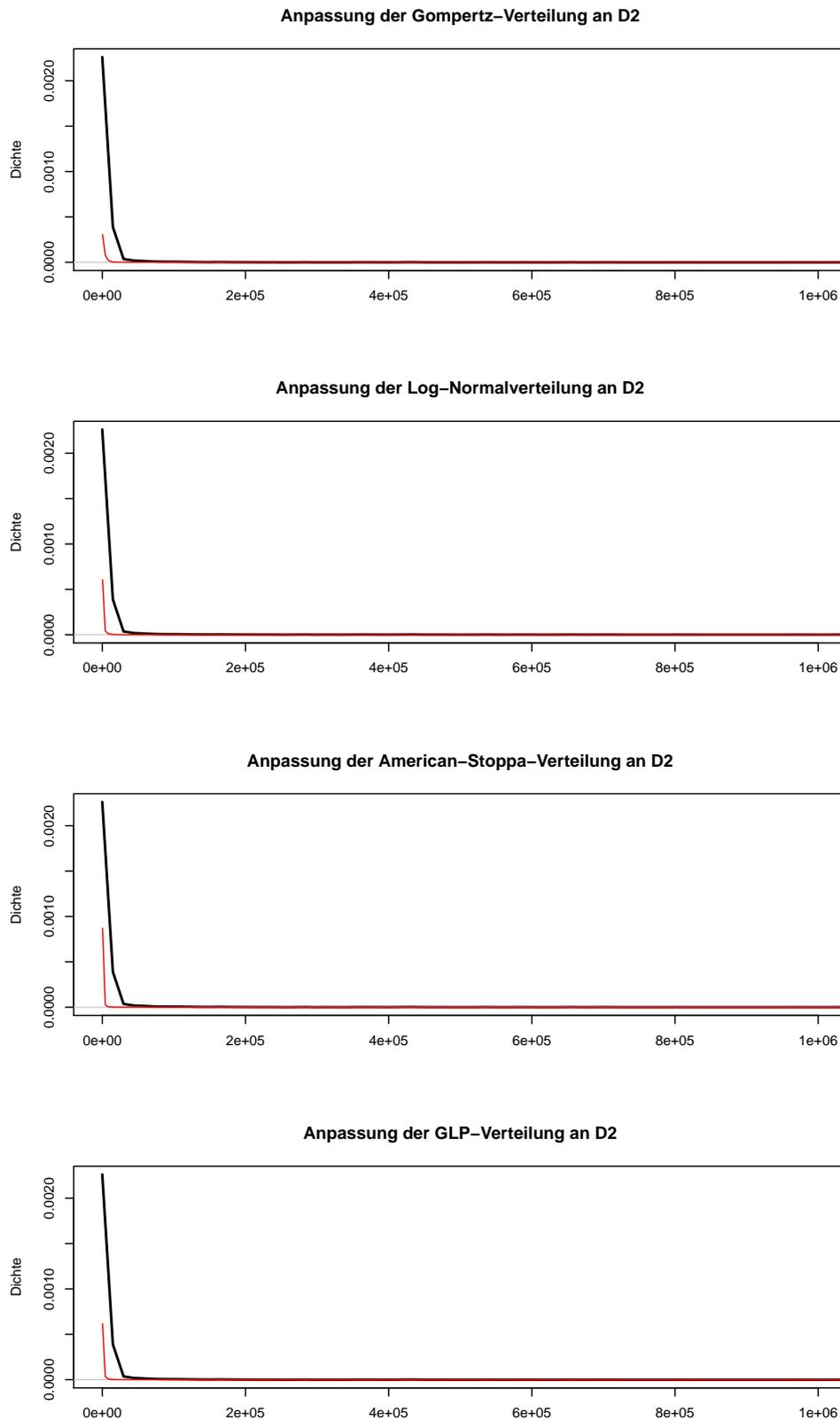


Abbildung 11: Anpassung der vier Verteilungen an den Datensatz 2

Nach der Simulation wird erneut für jede der 1000 Stichproben eine Maximum-

Likelihood-Schätzung durchgeführt. Für jeden Datenvektor, der einer der vier Verteilungen tatsächlich folgt, wurden die Maximum-Likelihood-Schätzer sowohl für diese als auch für die drei anderen Verteilungen berechnet. Dies führt zu den Hypothesen

$$H_0: \hat{F}_n(x) = F_0(x) \quad \text{vs.} \quad H_1: \hat{F}_n(x) \neq F_0(x),$$

wobei $\hat{F}_n(x)$ die empirische Verteilung des Datenvektors beschreibt und $F_0(x)$ für jeweils die Verteilung steht, gegen die getestet wird. Für jedes der vier Verteilungsszenarien werden alle 1000 Stichproben mit dem KS-Test, dem CvM-Test und dem AD-Test gegen jede Verteilung getestet. Als Signifikanzniveau wird einmal $\alpha = 5\%$ und einmal $\alpha = 10\%$ gewählt. Die Nullhypothese wird demnach abgelehnt, falls der p-Wert p kleiner als α ist.

Um die Power eines Tests zu überprüfen, wird die Ablehnquote eines jeden Tests für jedes Hypothesenpaar ermittelt. Idealerweise verfügen die Tests über eine hohe Power, das heißt, eine falsche Nullhypothese wird als solche erkannt.

4.2.2 Ergebnisse

Aufbau der Ergebnistabellen

Die Ergebnisse der Simulationsstudie für Non-Spliced-Modelle sind in vier Tabellen dargestellt. Der Name der Verteilung, aus welcher die betrachteten Stichproben simuliert werden, findet sich in der ersten Zeile der Tabelle. In der ersten Spalte sind die theoretischen Verteilungen, gegen die mit oben genannten Hypothesen getestet wird, aufgelistet:

$$H_0: \hat{F}_n(x) = F_0(x) \quad \text{vs.} \quad H_1: \hat{F}_n(x) \neq F_0(x)$$

In der zweiten Spalte von links befindet sich der jeweilige Datensatz. In der rechten Hälfte einer Tabelle ist das Signifikanzniveau, zu dem getestet wird zu finden. Direkt darunter sind die verwendeten Testverfahren mit den in Kapitel 3 definierten Abkürzungen benannt. Für eine visuelle Prüfung der Anpassung wird außerdem zu jeder Verteilung ein QQ-Plot erstellt, bei dem auf der x-Achse die Quantile des simulierten Datenvektors und auf der y-Achse die Quantile der realen Schadensdaten abgebildet sind.

Mittelwerte und Varianzen aller geschätzten Parameter sind im Anhang unter Punkt A.3.2 zu finden.

Ergebnistabellen und Interpretation

In den Tabellen 5 bis 8 sind die Ergebnisse der Simulationsstudie bezüglich Power der Anpassungstests für nicht-trunkierte Verteilungen abgebildet.

Gompertz-Verteilung							
H_0	Daten	$\alpha = 0.10$			$\alpha = 0.05$		
		KS	CvM	AD	KS	CvM	AD
Gompertz	Daten 1	0.026	0.017	0.016	0.009	0.010	0.005
	Daten 2	0.003	0.001	0.002	0.000	0.000	0.000
Log-Normal	Daten 1	1.000	1.000	1.000	1.000	0.000	1.000
	Daten 2	1.000	1.000	1.000	1.000	1.000	1.000
American- Stoppa	Daten 1	1.000	1.000	1.000	1.000	1.000	1.000
	Daten 2	0.027	0.020	0.036	0.015	0.007	0.025
GLP	Daten 1	1.000	1.000	1.000	1.000	1.000	1.000
	Daten 2	0.009	0.004	0.006	0.002	0.002	0.002

Tabelle 5: Testpower bezüglich der Gompertz-Verteilung

Die Testszenarien bezüglich der Gompertz-Verteilung sind in Tabelle 5 illustriert. Um eine Abhängigkeit der Testergebnisse von den Rohdaten zu vermeiden, wurden zwei verschiedene Datensätze mit Schadensdaten verwendet. Die Daten, welche als Basis für die Tests verwendet wurden, sind je Datensatz 1000 Zufallsvektoren, welche mit den geschätzten Parametern aus Tabelle 3 bzw. Tabelle 4 gezogen wurden. Aus jedem dieser wurden mittels der Maximum-Likelihood-Schätzung die Parameter für die Verteilungen in der rechten Spalte geschätzt. Die Tests wurden zunächst mit dem weniger konservativen Signifikanzniveau von $\alpha = 0.10$ durchgeführt. Alle falschen Nullhypothesen werden unabhängig vom Signifikanzniveau korrekterweise verworfen, sodass eine maximale Power von 100% festgestellt werden kann. Des Weiteren wird die korrekte Nullhypothese bei Datensatz 1 in nahezu alle Fällen beibehalten. Bei Datensatz 2 dagegen ist keiner der Tests in der Lage (unabhängig vom Signifikanzniveau), die Gompertz-Verteilung zu identifizieren. Ein Grund hierfür kann nicht zweifelsfrei festgestellt werden. Eine genauere Untersuchung des Datensatzes, die hier mangels Informationen nicht möglich ist, ist an dieser Stelle sicherlich angebracht. Eine weitere Ursache für diese Auffälligkeit könnte in der Ungenauigkeit der Parameterschätzung liegen. Zudem ist festzuhalten, dass sich die Gompertz-Verteilung grundsätzlich in der Gestalt ihrer Verteilungsfunktion von den übrigen Verteilungen stark unterscheidet. Wie bereits erwähnt, ist die Gompertz-Verteilung als light-tailed Verteilung einzuordnen. Die große Diskrepanz zu den anderen Verteilungsfunktionen kann dazu führen, dass es trivial sein kann, diese Verteilung richtig zu klassifizieren. Hier weisen somit alle drei Test maximale Power auf.

Log-Normalverteilung							
H_0	Daten	$\alpha = 0.10$			$\alpha = 0.05$		
		KS	CvM	AD	KS	CvM	AD
Gompertz	Daten 1	1.000	0.606	1.000	1.000	0.590	1.000
	Daten 2	1.000	1.000	1.000	1.000	1.000	1.000
Log-Normal	Daten 1	0.000	0.000	0.000	0.000	0.000	0.000
	Daten 2	0.000	0.000	0.000	0.000	0.000	0.000
American- Stoppa	Daten 1	0.202	0.266	0.474	0.055	0.068	0.163
	Daten 2	0.147	0.195	0.396	0.035	0.047	0.113
GLP	Daten 1	0.984	0.778	1.000	0.906	0.416	1.000
	Daten 2	0.998	0.963	1.000	0.995	0.790	1.000

Tabelle 6: Testpower bezüglich der Log-Normalverteilung

Tabelle 6 fasst die Ergebnisse der Testsimulation bezüglich der Log-Normalverteilung zusammen. Die wahre Nullhypothese, die besagt, dass die Daten aus einer Log-Normalverteilung stammen, wird von allen drei Tests unabhängig von Datensatz und Signifikanzniveau beibehalten. Jedoch ist auffällig, dass sowohl bei Datensatz 1 als auch bei Datensatz 2 die Nullhypothese $H_0 = \textit{American} - \textit{Stoppa}$ vom KS- und CvM-Test bei einem Signifikanzniveau von $\alpha = 0.05$ in über 90% der Fälle nicht verworfen wird. Der AD Test hingegen verwirft diese etwas häufiger. *berall sind diese Fehlentscheidungen sensibel bezüglich des Signifikanzniveaus: Für $\alpha = 0.10$ wird deutlich häufiger, wenn auch in weit weniger als der Hälfte der Fälle, eine korrekte Entscheidung getroffen. Eine denkbare Erklärung für diese Beobachtung kann sein, dass die Verteilungsfunktionen der Log-Normalverteilung und der American-Stoppa-Verteilung mit den jeweiligen Parametern recht ähnlicher Natur sind. Die Nullhypothese, die besagt, dass die Daten einer GLP-Verteilung folgen, wird am häufigsten vom AD-Test richtigerweise zurückgewiesen. Insgesamt ist die Power des AD-Tests für die Log-normalverteilten Daten für die meisten Hypothesenpaare höher als die der beiden anderen Tests.

QQ-Plot Log-Normalverteilung vs. American-Stoppa-Verteilung

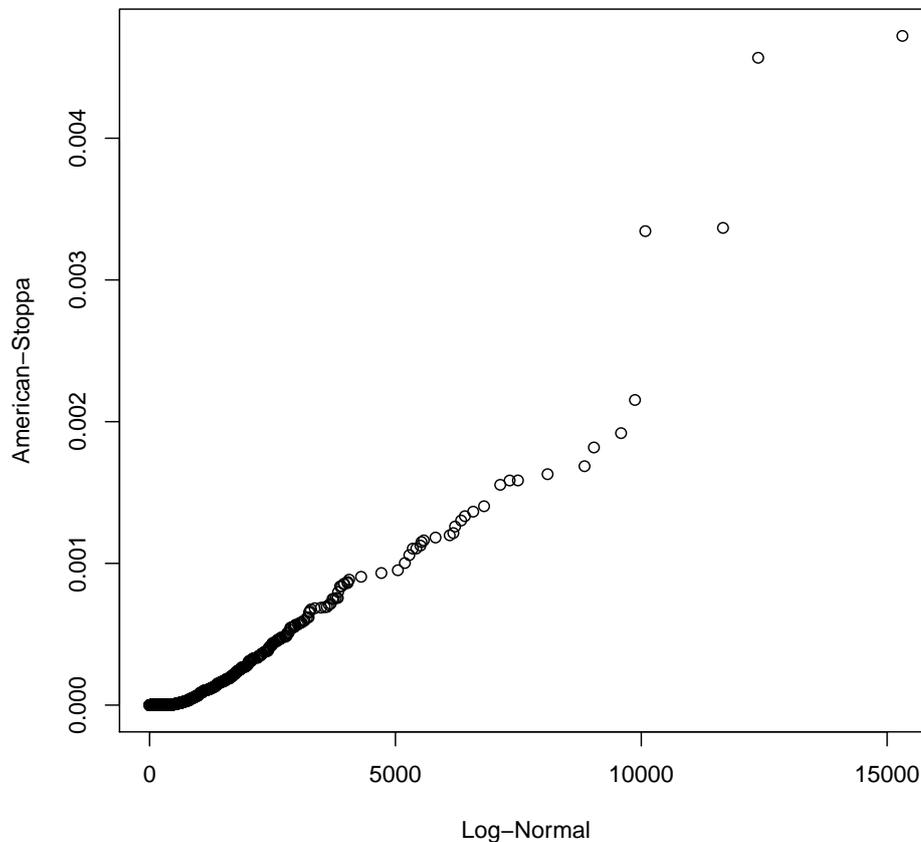


Abbildung 12: QQ-Plot Log-Normal- vs. American-Stoppa-Verteilung

Abbildung 12 zeigt einen QQ-Plot der Log-Normal- und American-Stoppa-Verteilung. Grundlage für diesen Plot ist einer der Zufallsvektoren der Log-Normalverteilung und die Dichte der American-Stoppa-Verteilung mit den geschätzten Parametern aus diesem Log-normalverteilten Zufallsvektor. Die Differenz der Verteilungsfunktionen dieser zwei Verteilungen mit entsprechenden Parametern ist eher als geringfügig einzustufen. Die Quantile der beiden Verteilungen liegen nahe an der winkelhalbierenden Geraden.

American-Stoppa-Verteilung							
H_0	Daten	$\alpha = 0.10$			$\alpha = 0.05$		
		KS	CvM	AD	KS	CvM	AD
Gompertz	Daten 1	1.000	0.982	1.000	1.000	0.766	1.000
	Daten 2	1.000	1.000	1.000	1.000	1.000	1.000
Log-Normal	Daten 1	1.000	1.000	1.000	1.000	1.000	1.000
	Daten 2	1.000	1.000	1.000	1.000	1.000	1.000
American-Stoppa	Daten 1	0.001	0.000	0.000	0.001	0.000	0.000
	Daten 2	0.000	0.000	0.000	0.000	0.000	0.000
GLP	Daten 1	1.000	1.000	1.000	1.000	1.000	1.000
	Daten 2	1.000	1.000	1.000	1.000	1.000	1.000

Tabelle 7: Testpower bezüglich der American-Stoppa-Verteilung

Die Resultate der Testszenerarien in Tabelle 7 stellen die Ergebnisse für die American-Stoppa-Verteilung dar. Alle drei Tests weisen bei einem Signifikanzniveau von $\alpha = 0.05$ meist eine maximale Power hinsichtlich der falschen Nullhypothesen auf. Lediglich beim Test auf die Gompertz-Verteilung mit Datensatz 1 als Grundlage treffen KS- und CvM-Test häufiger eine falsche Testentscheidung. Auch die wahre Nullhypothese $H_0 = \textit{American} - \textit{Stoppa}$ wird von allen Tests gleichermaßen in fast allen Fällen beibehalten. Dass die Fehlerquote beim weniger konservativen Signifikanzniveau von $\alpha = 0.10$ minimal höher liegt, deutet dennoch auf ein schlüssiges Ergebnis bei diesem Hypothesenpaar hin. Eine mögliche Erklärung hierfür könnte eher sein, dass die American-Stopp-Verteilung eine besonders gute Anpassung an die Rohdatensätze erreicht. Diese Vermutung wird durch den QQ-Plot, den Abbildung 13 zeigt, gefestigt. Die Quantile des Datensatzes 1 entsprechen stark denen der American-Stoppa-Verteilung, sodass die Punktfolge nahe der Winkelhalbierenden abgebildet wird.

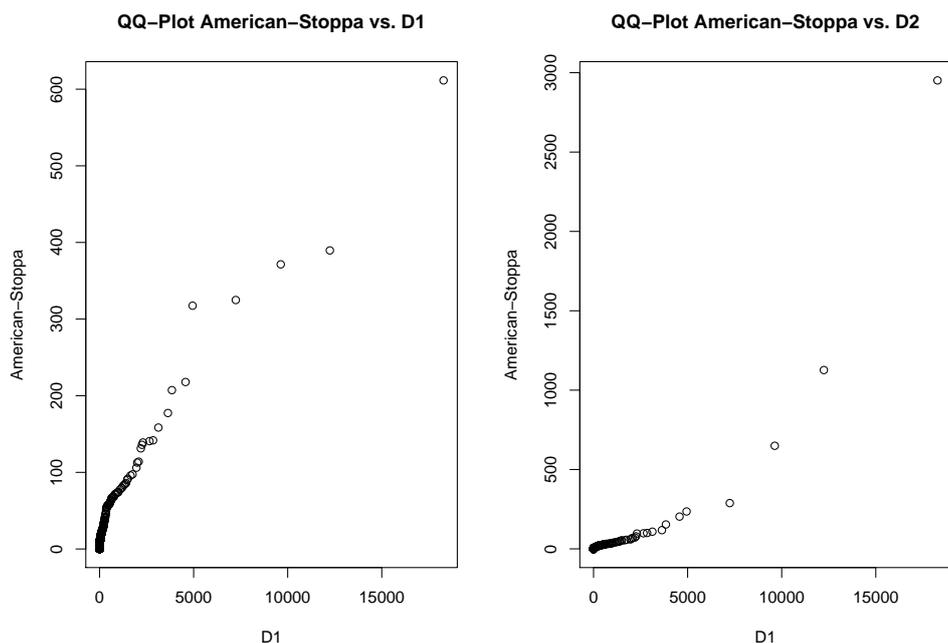


Abbildung 13: QQ-Plots der American-Stoppa-Verteilung vs. Datensatz 1 bzw. Datensatz 2

GLP-Verteilung							
H_0	Daten	$\alpha = 0.10$			$\alpha = 0.05$		
		KS	CvM	AD	KS	CvM	AD
Gompertz	Daten 1	1.000	0.924	1.000	1.000	0.924	1.000
	Daten 2	1.000	0.713	1.000	1.000	0.629	1.000
Log-Normal	Daten 1	0.964	0.986	0.997	0.879	0.949	0.982
	Daten 2	1.000	0.915	1.000	1.000	0.915	1.000
American-Stoppa	Daten 1	0.021	0.021	0.021	0.021	0.021	0.021
	Daten 2	0.207	0.204	0.361	0.077	0.055	0.194
GLP	Daten 1	0.001	0.000	0.001	0.000	0.000	0.000
	Daten 2	0.190	0.270	0.394	0.089	0.122	0.250

Tabelle 8: Testpower bezüglich der GLP-Verteilung

Aus Tabelle 8 lässt sich ein Resümee für die Testpower bezüglich der GLP-Verteilung ziehen. Hier sind die Ergebnisse eher gemischter Natur. Der KS-Test und der AD-Test erreichen bezüglich der Nullhypothese $H_0 = Gompertz$ für beide Signifikanzniveaus maximale Power; der CvM-Test trifft bei für Datensatz eins in den meisten Fällen eine richtige Entscheidung, behält die Hypothese für Datensatz zwei jedoch für etwa zwei Drittel der Stichproben fälschlicherweise bei. Die Log-Normalverteilung wird von allen Tests in den allermeisten Fällen als falsch erkannt. Die Aussage

$H_0 = \textit{American} - \textit{Stoppa}$ wird besonders bei Datensatz 1 fälschlicherweise fast immer beibehalten. Bei Datensatz 2 sind die Ergebnisse etwas besser, wenn ein weniger konservatives Signifikanzniveau gewählt wird. Die GLP-Verteilung als korrekte Verteilung wird bei Datensatz 1 für fast alle Stichproben von allen drei Tests erkannt. Bei Datensatz 2 werden häufiger falsche Testentscheidungen getroffen; hier schneidet der AD-Test am schlechtesten ab.

Insgesamt zeigt es sich schwierig, eine globale Aussage über die Power der drei Tests zu treffen, da diese zwischen den Hypothesenpaaren, Datensätzen und Signifikanzniveaus stark schwankt. Insbesondere scheinen die Erkennung der GLP-Verteilung selbst sowie die Ablehnung der American-Stoppa-Verteilung als theoretische Verteilung Probleme zu bereiten.

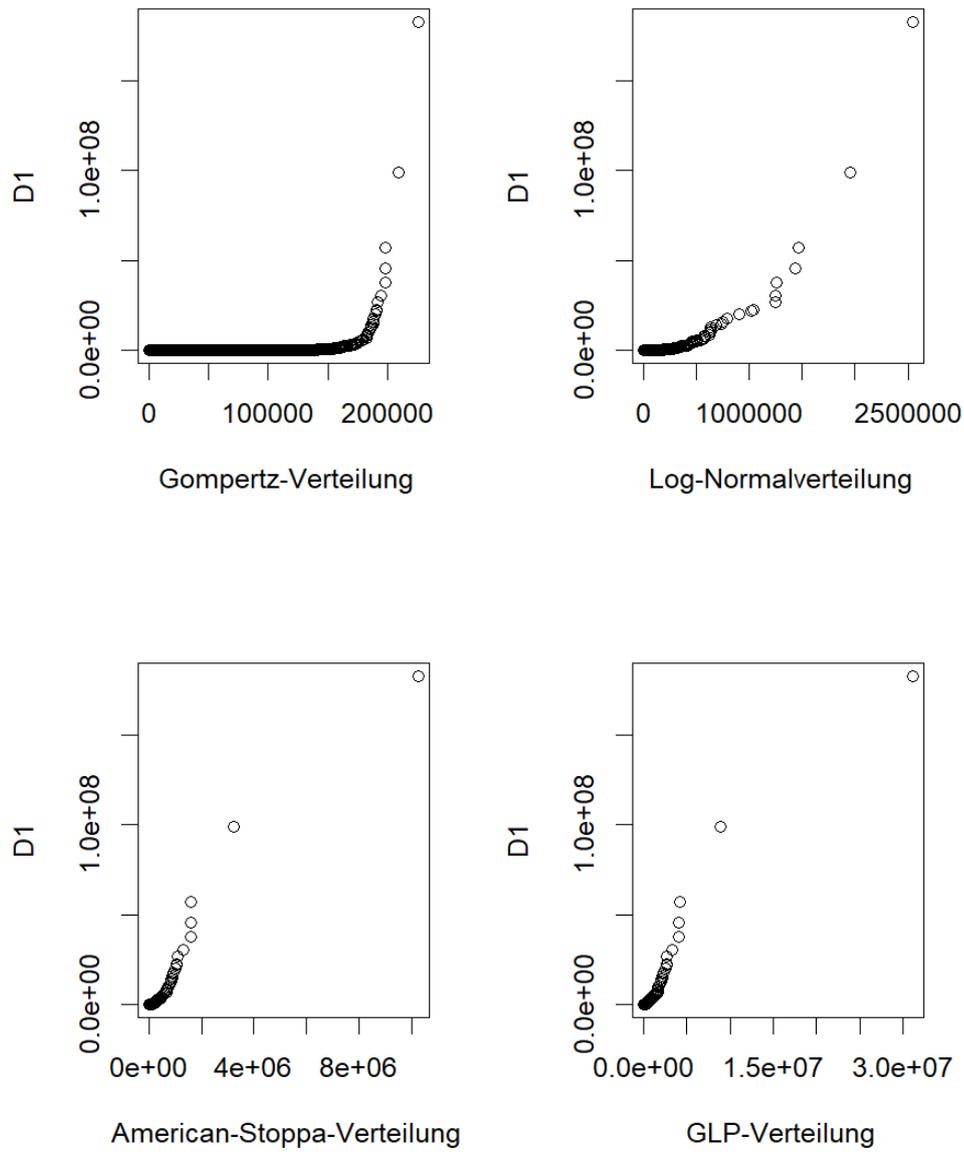


Abbildung 14: QQ-Plots der Verteilungen vs. Datensatz 1

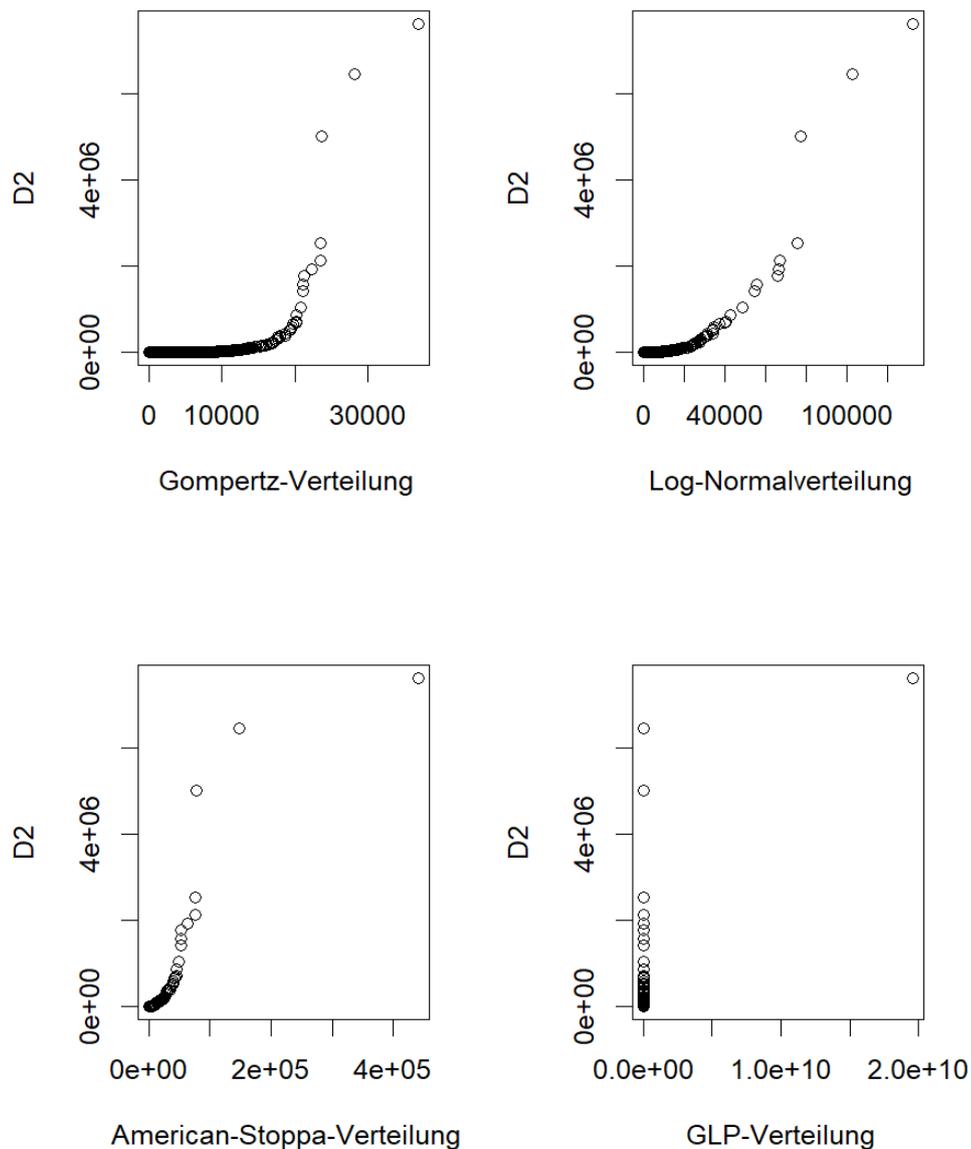


Abbildung 15: QQ-Plot der Verteilungen vs. Datensatz 2

Die Abbildungen 14 und 15 stellen einen visuellen Vergleich der simulierten Daten mit den Datensätzen dar. Auf der x-Achse sind die Quantile des jeweiligen Datenvektors zu sehen, die y-Achse bildet die Quantile der realen Schadensdaten ab. Im idealen Fall entsprechen die angepassten Verteilungen exakt den realen Daten. In diesem Fall würde sich aus der Punktwolke die Winkelhalbierende ergeben. Die Plots in der Abbildung zeigen, dass eine optimale Anpassung an die Daten mit einer einzelnen Verteilung unzureichend ist. Keiner der gezeigten QQ-Plots spricht für eine präzise Übereinstimmung.

Die Log-Normalverteilung und die American-Stoppa-Verteilung weisen zumindest

im linken unten Bereich, in dem ein Großteil der Beobachtungen liegt, eine relativ gute Anpassung auf. Gompertz- und GLP-Verteilung dagegen erreichen keine gute Anpassung an die Daten. Die American-Stoppa-Verteilung bietet, wie bereits diskutiert, im Vergleich zu den drei anderen Verteilungen die beste Anpassung an die Daten.

Generell scheint ein Non-Spliced-Modell die Daten jedoch nicht wirklich abbilden zu können. Im Folgenden wird untersucht, inwiefern die Anpassungstests bei einer linkstrunkierten Verteilung eine angemessene Power erreichen können.

4.3 Power der Testverfahren bei trunkierten Verteilungen

Dieser Abschnitt widmet sich dem Tail einer Verteilung, der mit der linkstrunkierten Version einer Verteilung, wie in Kapitel 2.2.2 erläutert, modelliert wird. Die Güte der Anpassung der vier Verteilungen aus Kapitel 2 wird analog zu Kapitel 4.2 analysiert. Ziel ist es nun also, eine Aussage über die Power der Tests bei trunkierten Verteilungen zu treffen.

4.3.1 Vorbereitung

Als Grundlage für den zweiten Teil der Simulationsstudie dient das R-Package 'truncgof'. Aus den Tests, die in diesem Package enthalten sind, wurden drei Tests für diese Studie ausgewählt: der KS-Test, der CvM-Test und der AD2up-Test. Der Unterschied zwischen AD-Test und AD2up-Test besteht darin, dass der AD2up-Test den Tail stärker gewichtet und sich daher besonders für ein Testszenario mit trunkierten Verteilungen eignet. Die Testfunktionen des Packages 'truncgof' erwarten einen Vektor mit Zufallszahlen größer oder gleich dem Threshold, eine Verteilung inklusive Parameter, gegen die getestet werden soll, und die Übergabe des Thresholds. Der Vektor mit Zufallszahlen wird aus einer kompletten Verteilung gezogen und dann an der Stelle des Thresholds trunkiert.

Somit können die 1000 Vektoren inklusive der geschätzten Parameter aus Teil 1 der Simulationsstudie genauso für Teil 2 der Studie verwendet werden. Für jede einzelne Stichprobe wird nun zunächst ein Threshold berechnet, der Body und Tail trennt. Wie in Kapitel 2.2.1 erwähnt, stehen hierzu mehrere Möglichkeiten zur Verfügung. Für das nachfolgende Testszenario wird eine Ermittlung des Thresholds mittels des empirischen Modalwerts gewählt. Da dieser in den Daten häufig nicht eindeutig ist, werden jene zunächst in eine feste Zahl von Klassen gleicher Breite¹⁵ unterteilt. Als Modalwert wird die obere Schranke derjenigen Klasse festgelegt, welche die meisten Schadensdaten enthält. Die Ermittlung des Thresholds auf diese Weise erweist sich

¹⁵Die Klassenbreite errechnet sich mittels des Quotienten aus der Range und der Wurzel der Stichprobengröße.

für die GLP-Verteilung jedoch als ungeeignet, da in den meisten Fällen zu wenige Daten rechts dieses Wertes liegen. Deshalb wird hier alternativ auf das 85%-Quantil als Threshold zurückgegriffen.

Für die Nullhypothesen

$$H_0: \hat{F}_n^*(x) = F_0^*(x) \quad \text{vs.} \quad H_1: \hat{F}_n^*(x) \neq F_0^*(x)$$

wird erneut jeder Test zum Signifikanzniveau $\alpha = 0.05$ und $\alpha = 0.10$ für jede Verteilungskonstellation mit den geschätzten Parametern aus den Tabellen A.3.1 für den Datensatz 1 und A.3.2 für den Datensatz 2 durchgeführt. Wie bereits in Kapitel 4.2.1 erwähnt, sind zu hohe Differenzen in der Größe der Daten problematisch für die Berechnung einiger Testszenarien. Auch der Wert des Thresholds spielt eine entscheidende Rolle. Ist dieser sehr groß, erreicht der Funktionswert an dieser Stelle für eine andere Verteilung möglicherweise bereits Eins. Dies führt zu einem Abbruch des Tests, da die Teststatistik dann nicht berechnet werden kann. Der empirische Mittelwert und die Varianz der berechneten Thresholds sind im Anhang in den Tabellen 20 und 21 dargestellt. Die Funktion, welche zum Testen und Auswerten der Testergebnisse benutzt wird, gibt zusätzlich zu den Ergebnissen auch die Quote der nicht durchführbaren Tests wieder. Diese wird in Kapitel 4.3.2 kritisch gewürdigt.

4.3.2 Ergebnisse

Aufbau der Ergebnistabellen

Die Ergebnistabellen sind analog zu Kapitel 4.2.1 dargestellt. Es wird eine weitere Spalte 'Quote' hinzugefügt. Die Werte dieser Spalte geben die Anzahl der vollständig durchgeführten Tests, deren Teststatistik berechnet werden kann, in Prozent wieder.

Ergebnistabellen und Interpretation

In den Tabellen 9 bis 12 sind die Testergebnisse der GOF-Tests für linkstrunkierte Verteilungen präsentiert. Alle drei Tests weisen hier eine Erfolgsquote von 100% auf, das heißt, jede Teststatistik kann wie gewünscht berechnet werden. Die Aussage $H_0 = \text{Gompertz}^*$ behalten alle Tests in den meisten Fällen korrekterweise bei. Die zweite Nullhypothese $H_0 = \text{Log} - \text{Normal}^*$ wird unabhängig vom Datensatz bei einem Signifikanzniveau von 10% stets als falsch erkannt; für das konservative Signifikanzniveau liegt die Power bei Datensatz 1 geringfügig niedriger. Der Test gegen die American-Stoppa-Verteilung zeigt dieselbe Auffälligkeit wie im nicht-trunkierten Fall: Bei Datensatz 1 erkennen alle Tests die falsche Nullhypothese, bei Datensatz 2

dagegen wird fast immer eine falsche Testentscheidung getroffen. Auch hier ist keine Ursache ersichtlich. Dasselbe Phänomen lässt sich für die vierte Nullhypothese, $H_0 = GLP^*$, beobachten.

Generell schneiden alle Tests auf linkstrunkierte Verteilungen für Datensatz 1 deutlich besser ab. Die höchste Power, wenn auch mit geringem Abstand, weist hier der CvM-Test auf.

Gompertz-Verteilung*								
H_0	Daten	Quote	$\alpha = 0.10$			$\alpha = 0.05$		
			KS*	CvM*	AD2up*	KS*	CvM*	AD2up*
Gompertz*	Daten 1	100 %	0.032	0.000	0.079	0.011	0.000	0.038
	Daten 2	100 %	0.021	0.018	0.049	0.000	0.000	0.030
Log-Normal*	Daten 1	100 %	1.000	1.000	1.000	0.997	0.998	0.996
	Daten 2	100 %	1.000	1.000	1.000	1.000	1.000	1.000
American- Stoppa*	Daten 1	100 %	1.000	1.000	1.000	1.000	1.000	1.000
	Daten 2	100 %	0.041	0.039	0.130	0.030	0.040	0.010
GLP*	Daten 1	100 %	1.000	1.000	1.000	1.000	1.000	1.000
	Daten 2	100 %	0.023	0.000	0.101	0.000	0.000	0.078

Tabelle 9: Testpower bezüglich einer linkstrunkierten Gompertz-Verteilung

Die Testergebnisse bezüglich der trunkierten Log-Normalverteilung sind in Tabelle 10 abgebildet. Alle Testszenarien werden fehlerfrei durchgeführt. Keine der falschen Nullhypothesen $H_0 = Gompertz^*$, $H_0 = American - Stoppa^*$ und $H_0 = GLP^*$ wird von den drei Tests in angemessener Häufigkeit zurückgewiesen, unerheblich dessen, welches Signifikanzniveau verwendet wird. Auch die wahre Nullhypothese, dass eine Log-Normalverteilung vorliegt, wird häufig fälschlicherweise verworfen. Der AD2up-Test kann hier noch die höchste Treffsicherheit aufweisen und erzielt auch für $H_0 = Gompertz^*$ bessere Ergebnisse.

Insgesamt sind die Ergebnisse jedoch nicht eindeutig. So ist der AD2up-Test fast gar nicht in der Lage, die GLP-Verteilung als falsch einzustufen. Auch die beiden anderen Tests weisen eine gemischte Bilanz auf. Möglicherweise sind die Testergebnisse für Gompertz- und Log-Normalverteilung, aus denen sich kaum globale Aussagen über die Testpower treffen lassen, auch auf den eher light-tailed Charakter beider Verteilungen zurückzuführen. Die daraus resultierende schlechtere Anpassung an den Tail fällt besonders bei der linkstrunkierten Version der Verteilungen ins Gewicht.

Ermittlung der Güte und Power ausgewählter GOF-Tests mit Fokussierung auf
den Randbereich einer Verteilung

Log-Normalverteilung*								
H_0	Daten	Quote	$\alpha = 0.10$			$\alpha = 0.05$		
			KS*	CvM*	AD2up*	KS*	CvM*	AD2up*
Gompertz*	Daten 1	100 %	0.110	0.010	0.200	0.079	0.009	0.142
	Daten 2	100 %	0.000	0.000	0.180	0.000	0.000	0.015
Log-Normal*	Daten 1	100 %	0.492	0.526	0.264	0.379	0.408	0.100
	Daten 2	100 %	0.420	0.473	0.270	0.310	0.290	0.086
American-Stoppa*	Daten 1	100 %	0.231	0.261	0.692	0.162	0.160	0.190
	Daten 2	100 %	0.270	0.269	0.561	0.161	0.150	0.151
GLP*	Daten 1	100 %	0.269	0.240	0.012	0.131	0.149	0.002
	Daten 2	100 %	0.333	0.300	0.001	0.132	0.140	0.000

Tabelle 10: Testpower bezüglich einer linkstrunkierten Log-Normalverteilung

In Tabelle 11 sind die Testergebnisse der linkstrunkierten American-Stoppa-Verteilung dargestellt. Die Quote der korrekt durchgeführten Testdurchläufe ist auch hier mit jeweils über 99% noch sehr hoch. Bei Datensatz 1 wird die Annahme, die trunkierten Daten entstammten einer Gompertz-Verteilung, lediglich vom KS-Test korrekt verworfen. Sowohl der AD2up- als auch der CvM-Test verwerfen dies Nullhypothese nur in wenigen Fällen. Die Nullhypothese $H_0 = \text{Log} - \text{Normal}^*$ wird von allen Tests korrekt zurückgewiesen. Der AD2up-Test zeigt hier die größte Power. Beim KS und CvM Test sinkt die Ablehnquote geringfügig mit einem konservativeren Signifikanzniveau. Auf die Power des AD2up Tests hat dieses keinen Einfluss. Die American-Stoppa-Verteilung wird von allen Tests gleichermaßen korrekt als American-Stoppa-Verteilung erkannt. Die Nullhypothese $H_0 = \text{GLP}^*$ wird von allen drei GOF-Tests verworfen.

American-Stoppa-Verteilung*								
H_0	Daten	Quote	$\alpha = 0.10$			$\alpha = 0.05$		
			KS*	CvM*	AD2up*	KS*	CvM*	AD2up*
Gompertz*	Daten 1	99.99%	0.745	0.033	0.043	0.712	0.021	0.042
	Daten 2	99.99%	0.000	0.323	0.617	0.000	0.242	0.586
Log-Normal*	Daten 1	99.96%	0.989	0.989	1.000	0.916	0.895	1.000
	Daten 2	99.96%	1.000	1.000	1.000	0.947	0.938	1.000
American-Stoppa*	Daten 1	99.99%	0.121	0.081	0.111	0.040	0.050	0.081
	Daten 2	99.99%	0.060	0.061	0.101	0.030	0.021	0.071
GLP*	Daten 1	99.99%	0.929	0.949	1.000	0.869	0.929	1.000
	Daten 2	99.99%	0.939	0.929	0.980	0.913	0.929	0.990

Tabelle 11: Testpower bezüglich einer linkstrunkierten American-Stoppa-Verteilung

Die Testszenarien der GLP-Verteilung weisen in der Simulationsstudie einige Besonderheiten auf. Durch ihren super-heavy Tail und die unendlichen Momente können nicht alle gängigen Verfahren angewendet werden. Auffällig ist zunächst, dass zum Teil nur 23% der Testszenarien ohne Fehlermeldung durchgeführt werden konnten. In Kapitel 4.4 werden mögliche Gründe dafür detailliert diskutiert.

In der Berechnung der Testpower eines jeweiligen Test ist nur die Anzahl der fehlerfreien Testdurchläufe eingeflossen. Die wahre Nullhypothese $H_0 = GLP^*$ wird von allen Tests in einem Großteil der Fälle korrekt beibehalten, wobei der AD2up Test mit einer etwas höheren Power überzeugen kann. Bei Datensatz 1 wird die falsche Nullhypothese $H_0 = American - Stoppa$ von keinem Test in nennenswert vielen Fällen verworfen. Bezüglich der American-Stoppa-Verteilung trifft bei Datensatz 2 lediglich der AD2up Test eine korrekte Entscheidung. Hinsichtlich der Gompertz- und Log-Normalverteilung weist der AD2up Test die höchste Power auf.

GLP-Verteilung*								
H_0	Daten	Quote	$\alpha = 0.10$			$\alpha = 0.05$		
			KS*	CvM*	AD2up*	KS*	CvM*	AD2up*
Gompertz*	Daten 1	92.10%	0.099	0.989	1.000	0.043	0.989	1.000
	Daten 2	08.90%	0.625	0.500	0.750	0.625	0.375	0.750
Log-Normal*	Daten 1	100.0%	0.739	0.792	0.991	0.660	0.729	0.973
	Daten 2	41.00%	0.975	1.000	1.000	0.976	0.976	1.000
American-Stoppa*	Daten 1	100.0%	0.059	0.041	0.010	0.049	0.030	0.060
	Daten 2	94.00%	0.223	0.224	1.000	0.170	0.149	1.000
GLP*	Daten 1	92.10%	0.048	0.039	0.089	0.011	0.012	0.052
	Daten 2	08.90%	0.091	0.051	0.600	0.038	0.049	0.589

Tabelle 12: Testpower bezüglich einer linkstrunkierten GLP-Verteilung

Verhalten der GLP-Verteilung

Aufgrund der hohen Varianz der Schätzer (siehe Tabellen 16 und 19 im Anhang) werden die Spezifika der GLP-Verteilung kritisch betrachtet. Der Tail der GLP-Verteilung konvergiert so langsam gegen Null, dass dieser, abhängig von den Parametern, die Null zum Teil nie erreicht. Diese Eigenschaft ist zwar wesentlich für die Modellierung des oberen Quantils beispielsweise einer Schadenshöhenverteilung, jedoch erweist sie sich für die Modellierung des Bodys der Verteilung als ungeeignet. Je größer der Threshold, umso mehr konvergiert jedoch der Tail der Daten gegen eine GLP-Verteilung mit den geschätzten Parametern.

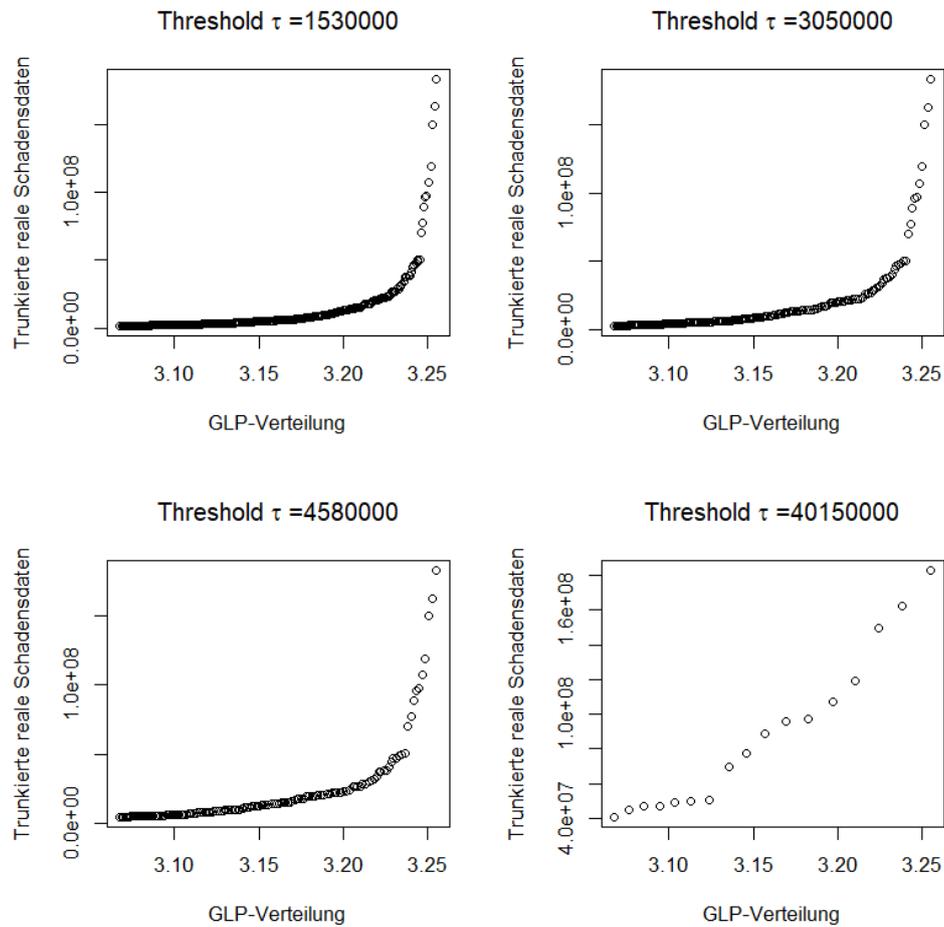


Abbildung 16: Anpassung der GLP-Verteilungen an den Tail von Datensatz 1 in Abhängigkeit des Thresholds

Zur Visualisierung dieses Sachverhalts wurden vier QQ-Plots erstellt, die in Abbildung 16 zu sehen sind. Oben links ist der Threshold aus den Testszenarien gewählt. Mit wachsendem Threshold verbessert sich die Anpassung der Verteilung. Somit ist zu überlegen, ob für eine noch genauere Anpassung entweder eine andere Berechnungsmethode für den Threshold verwendet werden sollte oder sogar zwei Thresholds und somit drei Verteilungen sinnvoll wären.

4.4 Kritische Diskussion der Ergebnisse der Simulationsstudien

Die Ergebnisse der Simulationsstudie zur Power der GOF-Tests bei Non-Spliced-Modellen lassen keine eindeutige Aussage über den besten Test zu. Vielmehr hat sich im Laufe der Studie die Vermutung herausgebildet, dass die größten Einflussfaktoren auf die Testpower durch die Eignung einer Verteilung für die vorliegenden Daten und durch die gewählten Verteilungsparameter zur Ziehung der Zufallsvektoren gegeben sind. Die verwendeten Parameter sind in dieser Studie geschätzt worden, wodurch zusätzliche Ungenauigkeit entsteht. Weiter hat die Wahl der Startwerte, insbesondere bei der Gompertz- und der American-Stoppa-Verteilung, einen entscheidenden Einfluss auf die Maximum-Likelihood-Schätzung der Parameter und damit auch auf die Testergebnisse. Die Minimierungsmethode in Formel 33 ermöglicht zwar durch Approximation der empirischen Momente, zu den Daten passende Startwerte zu finden. Jedoch müssen für diese Optimierung wiederum Startwerte übergeben werden, sodass lediglich ein Problem in ein anderes, nicht weniger schwieriges, Problem überführt wird. Eine mögliche Weiterführung dieser Arbeit müsste dementsprechend klaren Fokus auf die Startwertoptimierung oder weitere Schätzmethoden für die Parameter legen.

Es ist somit festzuhalten, dass keinem der drei Tests für Non-Spliced-Modelle eine global höchste Power zugeschrieben werden kann. Je nach Verteilungskonstellation in der Hypothese können hier unterschiedliche Ergebnisse beobachtet werden. Gleichermaßen scheint die Power stark von der Datenbasis abzuhängen, welche die initiale Parameterschätzung für die Simulation liefert. Es ist also sicherlich ratsam, mehr als einen Test zur Überprüfung der Anpassungsgüte einer Verteilung zu verwenden und die Testergebnisse kritisch im Hinblick auf die Datenbasis zu betrachten. Gleichzeitig sollte aufmerksam mit der Wahl der Startwerte für die Maximum-Likelihood-Schätzung umgegangen werden.

Die Simulationsstudie zur Testpower bei trunkeerten Verteilungen lässt die Annahme zu, dass der AD2up Test bei der Verwendung von heavy- oder super-heavy tailed Verteilungen eine wesentlich höhere Power aufweist als der KS Test und CvM Test. Diese Annahme wurde aufgrund der Ergebnisse in den Tabellen 11 und 12 getroffen. Jedoch ließen sich mit dem Package 'truncgof' nicht alle Testszenarien durchführen. In Tabelle 12 ist Quote der korrekt durchgeführten Tests äußerst gering. Eine Erklärung für diesen Sachverhalt kann an der Teststatistik des AD2up Test aus Gleichung 32 liegen. Die Berechnung erfolgt dann über die Formel

$$AD_{up}^{2*} = -2n \log(1 - z_H) + 2 \sum_{j=1}^n \log(1 - z_j) + \frac{1 - z_H}{n} \sum_{j=1}^n \left(1 + 2(n - j) \frac{1}{1 - z_j} \right) \quad (34)$$

mit $z_H = F_\theta(H)$ und $z_j = F_\theta(x_j)$

$F_\theta(x_j)$ steht hier für die theoretische Verteilungsfunktion. An der Stelle $\log(1 - z_H)$ in Gleichung 34 wird der Logarithmus von 1 minus der theoretischen Verteilungsfunktion mit geschätzten Parametern an der Stelle des Thresholds der empirischen Verteilung berechnet. Ist der Wert der theoretischen Verteilungsfunktion, gegen die getestet wird, an der der Stelle des Thresholds gleich 1, so kann die Teststatistik nicht berechnet werden, da dann $\log(0)$ zu berechnen wäre. Zudem hat die Größen-differenz der Zufallsvektoren bzw. der Dichte der theoretischen Verteilungen mit den geschätzten Parametern einen entscheidenden Einfluss darauf, ob eine Teststatistik berechnet werden kann oder nicht. Ein weitergehender Ansatz könnte auch hinsichtlich des zweiten Teils der Ergebnisse eine Optimierung der Verteilungsparameter sein. Denkbar wäre auch eine Skalierung der Rohdaten vorzunehmen, wobei dann die Frage bliebe, ob die Verteilungen sich für diese Art von Daten eignen.

5 Fazit

Gegenstand der vorliegenden Arbeit war es, eine Evaluierung ausgewählter statistischer Anpassungstests durchzuführen. Das Augenmerk liegt hierbei auf dem Tailverhalten von Verteilungen. Vier Tests, welche die Anpassungsgüte geeigneter Verteilungen für diesen Sachverhalt untersuchen, wurden dazu in einer zweiteiligen Simulationsstudie bezüglich ihrer Power verglichen. Die Zielsetzung war eine analytisch begründete Benennung desjenigen Tests, der die höchste Power erkennen lässt.

Zu Beginn der Arbeit war eine Hinführung zum Thema und die Vermittlung von theoretischen Grundlagen erforderlich. Zur Modellierung wurden Non-Spliced-Modelle und der Tail von Single-Spliced-Modellen betrachtet. Die Beschreibung dieses Tails wurde anhand von linkstrunkierten Verteilungen optimiert.

Vier unterschiedliche Verteilungen wurden detailliert eingeführt, die in der Simulationsstudie für diese Modellierungen verwendet wurden. Der Testaufbau, die Wahl der Hypothesen, die Berechnung der Teststatistik und der Ablehnbereich der vier Tests wurde ausführlich erläutert. Hierbei wurden auch mögliche Fehler und Gütekriterien eines Tests konkretisiert.

Mittels der Simulationsstudie wurde die Evaluierung der ausgewählten Testverfahren durchgeführt. Auf Basis von zwei Datensätzen mit Schadensdaten wurden die vier Verteilungen an realitätsnahe Szenarien aus dem operationellen Risiko angepasst. Im ersten Teil der Simulationsstudie wurde die Power der Tests bezüglich der Anpassung der Verteilungen als Non-Spliced-Modell auf den kompletten Datensätzen der Schadensdaten betrachtet. Im zweiten Teil wurden die Schadensdaten in einer linkstrunkierten Version betrachtet. Darauf aufbauend wurde die Anpassungsgüte mit der Version der Tests für linkstrunkierte Daten beurteilt.

Abschließend wurden die Testergebnisse durch eine Visualisierung der Daten kritisch hinterfragt und differenziert untersucht.

Die Simulationsstudie hat die Kenntnis über die Power der einzelnen Tests für OpRisk-Daten nur unwesentlich verbessert. Es zeigt sich, dass diese stark abhängig von den geschätzten Parametern einer jeweiligen Verteilung ist. Der KS-Test, CvM-Test und AD-Test lieferten für die betrachteten Verteilungen ähnliche Ergebnisse. Alle drei Tests wiesen eine falsche Nullhypothese mit ähnlicher Genauigkeit zurück, ebenso behielten sie eine wahre in etwa gleich häufig bei. Im Einzelnen unterscheidet sich die Power jedoch zwischen den Hypothesenpaaren.

Der AD2up-Test lieferte hingegen bezüglich der Gegenüberstellung der GLP-Verteilung und den anderen drei Verteilungen die besten Ergebnisse. Der Grund hierfür ist, dass dieser Test sich aufgrund einer gesonderten Gewichtung des Tails durch eine Charakterisierung von heavy-tailed Verteilungen auszeichnet. Auf Basis der Studie kann die Annahme getroffen werden, dass der AD2up-Test für einen links-trunkierten Datensatz die höchste Power besitzt, wenn es sich um Schadensdaten aus dem operativen Risiko handelt oder die Verteilungsfunktion der Daten eine ähnliche Gestalt besitzt.

Die gewonnenen Ergebnisse über die Power der einzelnen Tests lassen die Annahme zu, dass ein Test alleine möglicherweise falsche oder ungenaue Ergebnisse liefert. Da ja gerade die Eintrittswahrscheinlichkeit besonders hoher Schadenswerte betrachtet wird, ist es umso empfehlenswerter, unterschiedliche Tests zur Entscheidungsfindung heranzuziehen. Weiter sollte der Fokus auf die exakte Anpassung und geeignete Startwerten für die Maximum-Likelihood-Schätzung gelegt werden. Eine Weiterführung der Evaluierung von verschiedenen Testverfahren könnte ein noch besseres Bild der Einflussfaktoren auf die Testpower ermöglichen. Denkbar hierbei wären etwa eine Startwertoptimierung, alternative Methoden zur Schätzung der Parameter oder die Wahl abweichender Verteilungsmodelle. Auch die Berücksichtigung des vollen Single-Spliced-Modells oder eines Double-Spliced-Modells wären ein Ansatzpunkt.

Insgesamt bleibt festzuhalten, dass zur Beurteilung von GOF-Tests für Verteilungen mit heavy Tails noch weitere Forschung notwendig ist. Dies gilt insbesondere vor dem Hintergrund der Tatsache, dass große Verluste im Banken- und Versicherungsbereich, für welche die hier untersuchte Thematik beispielsweise Anwendung findet, meist Auswirkungen auf die Gesellschaft als Ganzes haben.

A Anhang

A.1 Verteilungsübersicht

	Momente
Gompertz-Verteilung (vgl. Lenart (2016))	$\mathbb{E}[X^k] = \frac{n!}{b^n} e^{\frac{a}{b}} E_1^{n-1}(\frac{a}{b})$
Log-Normalverteilung (vgl. Bayer (2012))	$\mathbb{E}[X^k] = \exp(k\mu + \frac{1}{2}k^2\sigma^2)$
American-Stoppa-Verteilung (vgl. Calderín-Ojeda (2016))	$\mathbb{E}[X^k] = \lambda^k Be(1 - \frac{k}{\theta}, \lambda) - \lambda$ mit $Be(a, b) = \int_0^1 z^{a-1}(1-z)^{b-1} dz$
Paretoverteilung (vgl. Calderín-Ojeda (2016))	$\mathbb{E}[X^k] = \frac{\beta^k}{\prod_{i=1}^k (\alpha-i)}$; $\alpha, \beta > 0$

Tabelle 13: Momente der behandelten Verteilungen

A.2 Tabellen

A.3 Parameterschätzer

A.3.1 Datensatz 1

	Gompertz				Log-Normal			
	$\bar{x}(a)$	$s^2(a)$	$\bar{x}(b)$	$s^2(b)$	$\bar{x}(\mu)$	$s^2(\mu)$	$\bar{x}(\sigma)$	$s^2(\sigma)$
Gompertz	9.9e-06	4.8e-14	1.0e-05	6.4e-14	10.6150	0.0004	1.0809	0.0006
Log-Normal	4.6e-05	3.6e-09	-6.5e-05	5.0e-09	9.7836	6.1e-04	1.3990	0.0003
American-Stoppa	0.0001	2.1e-08	-0.0002	3.1e-08	9.7025	0.0003	1.0348	0.0004
GLP	0.0004	1.4e-07	-3.5e-06	1.2e-12	9.5624	0.0010	1.7310	0.0008

Tabelle 14: Mittelwert und Varianz der geschätzten Parameter für die Gompertz- und Log-Normalverteilung (Datensatz 1)

	American-Stoppa					
	$\bar{x}(\theta)$	$s^2(\theta)$	$\bar{x}(\lambda)$	$s^2(\lambda)$	$\bar{x}(c)$	$s^2(c)$
Gompertz	11755	1.3e+09	1.7298	0.0021	461813400	2.5e+18
Log-Normal	1.4027	0.0089	1.7756	12.8318	1840.20	8.52e+06
American-Stoppa	1.2614	0.00034	139.895	19.08869	209.01	40.1483
GLP	1.1345	0.0101	19.6720	27197.65	17322.23	1.00e+07

Tabelle 15: Mittelwert und Varianz der geschätzten Parameter für die American-Stoppa-Verteilung (Datensatz 1)

	GLP					
	$\bar{x}(\lambda)$	$s^2(\lambda)$	$\bar{x}(\beta)$	$s^2(\beta)$	$\bar{x}(\sigma)$	$s^2(\sigma)$
Gompertz	4.8e-07	3.3e-12	215932	5.7e+09	1972010	4.20e+11
Log-Normal	3.6e-05	2.3e-11	15605.55	2.0e+06	39500.85	7.56e+06
American-Stoppa	2.3e-05	0.0002	17819.99	3.2e+06	45332.23	1.31e+07
GLP	0.0002	5.6e-09	4586.41	1.6e+06	17719.70	1.53e+06

Tabelle 16: Mittelwert und Varianz der geschätzten Parameter für die GLP-Verteilung (Datensatz 1)

A.3.2 Datensatz 2

	Gompertz				Log-Normal			
	$\bar{x}(a)$	$s^2(a)$	$\bar{x}(b)$	$s^2(b)$	$\bar{x}(\mu)$	$s^2(\mu)$	$\bar{x}(\sigma)$	$s^2(\sigma)$
Gompertz	0.0003	7.7e-11	-2.2e-05	3.1e-11	7.4287	0.0006	1.3290	0.0006
Log-Normal	0.0006	2.38e-10	-0.0001	2.3e-11	6.9902	0.0006	1.3536	0.0003
American-Stoppa	0.0008	1.6e-09	-9.3e-05	3.6e-10	6.9073	0.0003	0.9783	0.0003
GLP	0.0016	3.0e-06	-0.0003	1.9e-06	6.2427	2.5569	1.6904	0.0466

Tabelle 17: Mittelwert und Varianz der geschätzten Parameter für die Gompertz- und Log-Normalverteilung (Datensatz 2)

	American-Stoppa					
	$\bar{x}(\theta)$	$s^2(\theta)$	$\bar{x}(\lambda)$	$s^2(\lambda)$	$\bar{x}(c)$	$s^2(c)$
Gompertz	949.87	3.5e+07	0.9573	0.0009	2.9e+06	3.453+14
Log-Normal	1.4807	0.0049	1.4972	0.0048	1153.28	19868.96
American-Stoppa	1.3294	0.0006	603.82	1.7e+05 9.7021	92.0931	
GLP	1.0912	0.0040	1.1015	0.0021	833.37	10892.14

Tabelle 18: Mittelwert und Varianz der geschätzten Parameter für die American-Stoppa-Verteilung (Datensatz 2)

	GLP					
	$\bar{x}(\lambda)$	$s^2(\lambda)$	$\bar{x}(\beta)$	$s^2(\beta)$	$\bar{x}(\sigma)$	$s^2(\sigma)$
Gompertz	0.0003	5.3e-06	6409.26	2.2e+07	39904.32	1.88e+08
Log-Normal	0.0020	4.4e-06	352.84	18914.76	2737.03	36711.13
American-Stoppa	0.0209	0.0031	283.11	38839.32	3845.95	2.80e+06
GLP	0.1425	0.0360	124.59	11250.64	2242.86	3.75e+06

Tabelle 19: Mittelwert und Varianz der geschätzten Parameter für die GLP-Verteilung (Datensatz 2)

A.4 Thresholds

	Gompertz	Log-Normal	American-Stoppa	GLP
Mittelwert	4197.84	52775.90	449786.40	74751.02
Varianz	54986.32	1.2e+09	3.2e+12	11999413

Tabelle 20: Empirischer Mittelwert bzw. empirische Varianz der Thresholds (Datensatz 1)

	Gompertz	Log-Normal	American-Stoppa	GLP
Mittelwert	702.64	2730.45	17269.8	3894.53
Varianz	42909.25	2837418	3.5e+09	29084.26

Tabelle 21: Empirischer Mittelwert bzw. empirische Varianz der Thresholds (Datensatz 2)

A.5 Elektronischer Anhang

Literatur

- Bayer, V. (2012). *Multivariate Modellierung operationeller Risiken in Kreditinstituten*, Springer Fachmedien Wiesbaden GmbH.
- Calderín-Ojeda, E. (2016). Modeling claims data with composite stoppa models, *Scandinavian Actuarial Journal* **9**: 817–836.
- Chernobai, Anna et al. und Svetlozar, R. (2005). *Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples*.
- Corman, Ulf und Reiss, R.-D. (2008). *Generalizing the Pareto to the log-Pareto model and statistical inference*, Springer Science+Business Media.
- Cruz, M. G. (2015). *Fundamental Aspects of Operational Risk and Insurance Analytics*, John Wiley & Sons.
- Fahrmeir, L. (2016). *Statistik - Der Weg zur Datenanalyse*, SpringerSpektrum.
- Foss, S., Korshunov, D. and Zachary, S. (2019). *An Introduction to Heavy-Tailed and Subexponential Distributions*, Springer Science+Business Media.
- Gatto, R. (2014). *Stochastische Modelle der aktuariellen Risikotheorie - Eine mathematische Einführung*, Springer Berlin Heidelberg.
- Heddrich, Jürgen und Sachs, L. (2016). *Angewandte Statistik - Methodensammlung in R*, Springer Berlin Heidelberg.
- Henze, N. (2019). *Stochastik. Eine Einführung mit Grundzügen der Maßtheorie*, Springer Spektrum.
- Lach, Agnieszka und Smaga, L. (2018). *Comparison of the goodness-of-fit tests for truncated distributions*.
- Lenart, A. (2011). *The Gompertz distribution and Maximum Likelihood Estimation of its parameters - a revision*, Max Planck Institute for Demographic Research.
- Lenart, Adam und Missov, T. I. (2016). *Goodness-of-fit tests for the Gompertz distribution*, American Fisheries Society.
- Moscadelli, M. (2004). *The Modelling of Operational Risk: Experience with the Analysis of Daata Collect by the Basel Comittee*, SSRN Electronic Journal.
- Peters, Gareth W und Shevchenko, P. V. (2015). *Adavanced in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*, John Wiley & Sons.

Steinhoff, C. (2008). *Operational Risk*, Cuvillier Verlag.

Stelzer, L. (2018). *Anwendung von Spliced-Modellen zur Quantifizierung
operationeller Risiken*.

Stoppa, G. (1990). *A new Model for Income Size Distributions*, Springer-Verlag
Berlin Heidelberg.