

# Eine Anwendung der Peak over Threshold Methode im Risikomanagement

Diplomarbeit

vorgelegt von  
Julia Prahm

Betreuer: Privatdozent Dr. Volkert Paulsen  
Institut für Mathematische Statistik  
Fachbereich 10 - Mathematik und Informatik  
Westfälische Wilhelms-Universität Münster



# Inhaltsverzeichnis

<b>Einleitung</b>	<b>9</b>
<b>1 POT-Methode für iid Zufallsvariablen</b>	<b>11</b>
1.1 POT-Methode . . . . .	12
1.2 Verallgemeinerte Paretoverteilung . . . . .	14
1.3 Exkurs: Block-Maxima Methode und GEV-Verteilung . . . . .	16
1.4 Pickands-Balkema-de Haan Theorem . . . . .	17
1.5 Wahl eines geeigneten Thresholds . . . . .	19
1.6 Schätzung der GPD Parameter . . . . .	21
1.7 Value at Risk . . . . .	23
1.7.1 Mathematische Herleitung des VaR . . . . .	23
1.7.2 Einordnung des VaR . . . . .	24
1.8 Expected Shortfall . . . . .	26
<b>2 Anwendung der POT-Methode für iid Zufallsvariablen</b>	<b>29</b>
2.1 Thresholdbestimmung . . . . .	31
2.1.1 mrl-plot . . . . .	31
2.1.2 tcplot . . . . .	33
2.1.3 Thresholdvergleich . . . . .	33
2.2 Schätzen der Parameter . . . . .	35
2.3 Überprüfung der Anpassung . . . . .	36
2.3.1 Histogramm . . . . .	36
2.3.2 Kolmogorov-Smirnov-Test . . . . .	37
2.3.3 Vergleich der ersten Momente, der Varianz und der Schiefe . . . . .	37
2.4 Value at Risk und Expected Shortfall . . . . .	38
2.4.1 Tailplot . . . . .	39
2.4.2 Vergleich GPD und Normalverteilung . . . . .	41
<b>3 POT-Methode für stationäre Zeitreihen</b>	<b>45</b>
3.1 Grundlagen der Zeitreihenanalyse . . . . .	45
3.1.1 Grundbegriffe . . . . .	45
3.1.2 AR- und MA-Prozesse . . . . .	48
3.1.3 ARCH- und GARCH-Prozesse . . . . .	51
3.2 POT-Methode . . . . .	57
3.2.1 Schätzen von $\mu_{t+1}$ und $\sigma_{t+1}$ . . . . .	59
3.2.2 Berechnung des bedingten VaR und des bedingten ES . . . . .	62

<b>4</b>	<b>Anwendung der POT-Methode für stationäre Zeitreihen</b>	<b>65</b>
4.1	Anpassung eines AR(1)-GARCH(1,1)-Prozesses . . . . .	66
4.2	Residuen . . . . .	67
4.2.1	Überprüfung der Unabhängigkeit der Residuen . . . . .	67
4.2.2	Anwendung der POT-Methode auf die Residuen . . . . .	71
4.2.3	Überprüfung der Anpassung . . . . .	74
4.2.4	Value at Risk der Residuen . . . . .	75
4.2.5	Expected Shortfall der Residuen . . . . .	76
4.3	Bedingter Value at Risk . . . . .	76
4.4	Bedingter Expected Shortfall . . . . .	81
4.5	Abschließender Vergleich . . . . .	84
<b>5</b>	<b>Alternative Anwendung</b>	<b>87</b>
5.1	Anpassung eines GARCH(1,1)-Prozesses . . . . .	87
5.2	Residuen . . . . .	88
5.2.1	Überprüfung der Unabhängigkeit der Residuen . . . . .	88
5.2.2	Anwendung der POT-Methode auf die Residuen . . . . .	89
5.2.3	VaR und ES der Residuen . . . . .	92
5.3	Bedingter VaR und ES . . . . .	94
5.4	Abschließender Vergleich . . . . .	99
	<b>Fazit</b>	<b>101</b>
	<b>Anhang</b>	<b>103</b>
	<b>Literaturverzeichnis</b>	<b>105</b>

# Abbildungsverzeichnis

1.1	Leptokurtische Verteilung im Vergleich mit einer Normalverteilung	13
1.2	Verteilungsfunktion und Dichte der verallgemeinerten Paretoverteilung für verschiedene Shape-Parameter . . . . .	15
2.1	Tägliche DAX-Schlusswerte und negative Log-Renditen . . . . .	30
2.2	QQ-Normal Plot der negativen Log-Renditen . . . . .	31
2.3	Mean Residual Life Plot der negativen Log-Renditen . . . . .	32
2.4	Ausschnitt aus dem Mean Residual Life Plot der Daten . . . . .	32
2.5	Threshold Choice Plot für 100 Thresholds ohne Konfidenzintervalle	34
2.6	Threshold Choice Plot für 31 Thresholds mit Konfidenzintervallen	34
2.7	Histogramm und empirische Verteilungsfunktion . . . . .	36
2.8	Tailplot der negativen Log-Renditen . . . . .	40
2.9	Angepasste GPD und Normalverteilung und 99.5% VaR der Verteilungen . . . . .	41
2.10	Histogramm der Daten mit der Dichte einer Konvexkombination aus GPD und Normalverteilung und mit einer angepassten Normalverteilung. . . . .	42
4.1	Negative Log-Renditen der DAX-Abschlusswerte . . . . .	65
4.2	Residuenplot . . . . .	67
4.3	acf-plot der Log-Renditen im Vergleich zum acf-plot der Residuen	69
4.4	acf-plot der quadrierten Log-Renditen im Vergleich zum acf-plot der quadrierten Residuen . . . . .	70
4.5	QQ-Normal Plot der Residuen . . . . .	72
4.6	Mean Residual Life Plot der Residuen . . . . .	73
4.7	Ausschnitt aus dem Mean Residual Life Plot der Residuen . . . . .	73
4.8	Histogramm und empirische Verteilungsfunktion der Residuen . .	74
4.9	Log-Renditen mit den Schätzern für den bedingten VaR . . . . .	77
4.10	Log-Renditen mit dem bedingten VaR unter Annahme GPD verteilter Residuen und unter Annahme normalverteilter Residuen . .	78
4.11	Log-Renditen mit dem bedingten und unbedingten VaR unter Annahme GPD verteilter Residuen bzw. Zufallsvariablen . . . . .	79
4.12	Log-Renditen mit dem bedingten und unbedingten VaR unter Annahme GPD verteilter und normalverteilter Residuen bzw. Zufallsvariablen . . . . .	80
4.13	Log-Renditen mit den Schätzern für den bedingten ES . . . . .	81

---

4.14	Log-Renditen mit dem bedingten ES unter Annahme GPD verteilter Residuen und unter Annahme normalverteilter Residuen . . .	82
4.15	Log-Renditen mit dem bedingten VaR und dem bedingten ES unter Annahme GPD verteilter Residuen . . . . .	83
4.16	Log-Renditen mit dem bedingten und unbedingten VaR und ES .	84
5.1	Residuenplot . . . . .	88
5.2	acf-plot der quadrierten Residuen . . . . .	89
5.3	QQ-Normal-Plot der Residuen . . . . .	90
5.4	Ausschnitt aus dem Mean Residual Life Plot der Residuen. . . . .	91
5.5	Histogramm und empirische Verteilungsfunktion der Residuen . .	92
5.6	Log-Renditen mit den Schätzern für den bedingten GPD VaR der AR(1)-GARCH(1,1)-Anpassung und den bedingten GPD VaR der GARCH(1,1)-Anpassung . . . . .	95
5.7	Log-Renditen mit den Schätzern für den bedingten normal VaR der AR(1)-GARCH(1,1)-Anpassung und den bedingten normal VaR der GARCH(1,1)-Anpassung . . . . .	96
5.8	Log-Renditen mit den Schätzern für den bedingten GPD ES der AR(1)-GARCH(1,1)-Anpassung und den bedingten GPD ES der GARCH(1,1)-Anpassung . . . . .	97
5.9	Log-Renditen mit den Schätzern für den bedingten normal ES der AR(1)-GARCH(1,1)-Anpassung und den bedingten normal ES der GARCH(1,1)-Anpassung . . . . .	98

# Tabellenverzeichnis

2.1	90%- und 95%- Quantile der negativen Log-Renditen . . . . .	35
2.2	Threshold-Vergleich . . . . .	35
2.3	Mittel, Varianz und Schiefe des Datensatzes und der angepassten Verteilung . . . . .	37
2.4	VaR und ES für verschiedene Wahrscheinlichkeiten . . . . .	38
2.5	Vergleich: Value at Risk und Expected Shortfall für die angepasste GPD und die angepasste Normalverteilung . . . . .	39
4.1	Ergebnisse der Pseudo-Maximum-Likelihood Schätzung . . . . .	66
4.2	Ljung-Box-Test Ergebnisse für verschiedene Lag-Längen . . . . .	71
4.3	Value at Risk der Residuen unter Annahme GPD verteilter Residuen (für verschiedene Konfidenzintervalle). . . . .	75
4.4	Value at Risk der Residuen unter Annahme normal verteilter Residuen (für verschiedene Konfidenzintervalle). . . . .	75
4.5	Expected Shortfall der Residuen unter Annahme GPD verteilter Residuen (für verschiedene Konfidenzintervalle). . . . .	76
4.6	Expected Shortfall der Residuen unter Annahme normalverteilter Residuen (für verschiedene Konfidenzintervalle). . . . .	76
4.7	Anzahl der Störungen für die verschiedenen Methoden und Konfidenzintervalle . . . . .	85
5.1	Ergebnisse der Pseudo-Maximum-Likelihood Schätzung . . . . .	87
5.2	Ljung-Box-Test Ergebnisse für verschiedene Lag-Längen . . . . .	90
5.3	Value at Risk der Residuen unter Annahme GPD verteilter Residuen (für verschiedene Konfidenzintervalle). . . . .	92
5.4	Expected Shortfall der Residuen unter Annahme GPD verteilter Residuen (für verschiedene Konfidenzintervalle). . . . .	93
5.5	Value at Risk der Residuen unter Annahme normalverteilter Residuen (für verschiedene Konfidenzintervalle). . . . .	93
5.6	Expected Shortfall der Residuen unter Annahme normalverteilter Residuen (für verschiedene Konfidenzintervalle). . . . .	93





# Einleitung

In der Statistik wird meistens das Hauptaugenmerk auf die Betrachtung von Mittelwerten gelegt. Für viele Bereiche ist das sinnvoll, doch z.B. bei der Investition in eine Aktie, ist es zwar interessant zu wissen wie viel im Durchschnitt an Gewinn oder Verlust zu erwarten ist, aber viel entscheidender ist die Frage nach dem „worst case“. Wie viel kann mit dieser Investition mit welcher Wahrscheinlichkeit im schlimmsten Fall verloren gehen? Auch in vielen anderen Bereichen, wie z.B. in Versicherungen oder bei der Planung von Deichen, ist es entscheidend, sich für den schlimmsten Fall abzusichern. Denn um einen Deich so hoch zu planen, dass im Normalfall keine Überschwemmungen zu befürchten sind, nützt es wenig, die durchschnittliche Wasserhöhe zu kennen. Stattdessen müssen die besonders hohen Wasserstände, oder im Fall der Versicherung die besonders hohen Versicherungsrisiken, abgeschätzt werden. Mit dieser Frage beschäftigt sich die Extremwerttheorie. Dabei werden, wie der Name schon sagt, hauptsächlich die extremen Werte, sogenannte „Ausreißer“ betrachtet. Doch eine Abschätzung dieser Werte ist in vielen Fällen nicht einfach, da Aussagen über einen Bereich der Stichprobe gemacht werden sollen, in denen nur wenige Werte vorliegen. Häufig wird dabei der Fehler gemacht, eine Normalverteilung an die gesamten Daten anzupassen. Diese beschreibt die mittleren Daten zwar meistens recht gut, es ist aber empirisch erwiesen, dass z.B. Renditen von Aktienpreisen deutlich häufiger Ausreißer zeigen, als es von einer Normalverteilung beschrieben wird. An den Rändern, den sogenannten „Tails“, fällt die Normalverteilung zu schnell ab und unterschätzt damit das Risiko großer Verluste. Besser geeignet für die Abschätzung großer Gewinne oder Verluste sind sogenannte „heavy tail“ Verteilungen. Diese besitzen, wie der Name schon vermuten lässt, größere Tails als die Normalverteilung und bieten somit eine bessere Anpassung. Ein gutes Verfahren um eine heavy tail Verteilung z.B. an Renditedaten anzupassen, bietet die Peak over Threshold Methode. Übersetzt heißt Peak Spitze oder Höchstpunkt und ein Threshold ist einfach ein Schwellenwert. Peak over Threshold bedeutet also so etwas wie Höchstpunkt oder Spitze über einem Schwellenwert. Der Name beschreibt im Prinzip auch schon das Vorgehen der Methode. Es wird zunächst ein Schwellenwert im oberen Bereich der Daten gewählt. Dabei ist es entscheidend den Schwellenwert möglichst gut zu wählen, damit es nicht zu viele, aber auch nicht zu wenige Thresholdüberschreitungen gibt. Schließlich sollen „nur“ die extremen Werte betrachtet werden. Ist der Threshold aber so hoch, dass zu wenige Überschreitungen übrig bleiben, ist die Anpassung evtl. zu ungenau. Ist der Threshold gewählt, kann schließlich eine heavy tail Verteilung an die Überschreitungen angepasst werden.

Somit erhält man mit der Peak over Threshold Methode eine Verteilung, die die hohen Werte sehr gut beschreibt. Nur die Verteilung an sich sagt noch nicht sehr viel aus. Daher wird hier eines der, wenn nicht sogar das, inzwischen wichtigste Risikomaß betrachtet, der Value at Risk. Der Value at Risk ist das  $q$ -te Quantil der angepassten Verteilung. Er gibt an, welcher Betrag in einer bestimmten Zeit, mit einer bestimmten Wahrscheinlichkeit, nicht überschritten wird. Darüber hinaus wird der Expected Shortfall betrachtet. Dieser gibt die erwartete Verlustgröße, gegeben dass der Value at Risk überschritten wurde, an.

In dieser Arbeit wird in Kapitel 1 zunächst die Peak over Threshold Methode für den einfachsten Fall, nämlich unabhängig und identisch verteilte Zufallsvariablen, vorgestellt. Darüber hinaus wird in diesem Kapitel der Value at Risk und der Expected Shortfall hergeleitet. Im Anschluss wird die Methode in Kapitel 2 auf die Renditen der täglichen DAX-Abschlusswerte von 1996 bis 2000 angewendet. Zusätzlich wird eine Normalverteilung an die Daten angepasst und damit der Value at Risk und der Expected Shortfall berechnet. Die Ergebnisse werden mit den Ergebnissen aus der Peak over Threshold Methode verglichen. Da die Annahme, unabhängiger Daten in vielen Fällen, vor allem bei der Betrachtung von Finanzdaten, nicht realistisch ist, wird in Kapitel 3 die Peak over Threshold Methode für stationäre Zeitreihen vorgestellt. Dafür wird zunächst eine kleine Einführung in die Grundlagen der Zeitreihenanalyse gegeben. In Kapitel 4 wird die Peak over Threshold Methode für stationäre Zeitreihen wieder auf die Renditen der täglichen DAX-Abschlusswerte von 1996 bis 2000 angewendet. Auch hier wird ein Vergleich mit der Normalverteilung gezogen. Darüber hinaus werden die Ergebnisse mit denen aus Kapitel 2 verglichen. Kapitel 5 bietet schließlich eine alternative Anwendung zu Kapitel 4.

An dieser Stelle möchte ich mich herzlich bei PD Dr. Paulsen für die Bereitstellung dieses interessanten Themas und die gute Betreuung während der Anfertigung dieser Diplomarbeit bedanken. Außerdem gilt mein Dank all denjenigen, die zum Gelingen dieser Arbeit beigetragen haben.

Gemäß § 21 (6) der Diplomprüfungsordnung für den Studiengang Mathematik der Westfälischen Wilhelms-Universität Münster vom 15. Juli 1998 versichere ich, dass ich die vorliegende Diplomarbeit selbständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen und Hilfsmittel benutzt habe.

Münster, den 16. November 2010

Julia Prahm

# 1 POT-Methode für iid Zufallsvariablen

Im ersten Teil sollen unabhängige, identisch verteilte Zufallsvariablen  $X_1, X_2, \dots$  mit einer unbekannten Verteilungsfunktion  $F$  betrachtet werden. Diese Zufallsvariablen können z.B. betriebliche Risiken, Versicherungsrisiken, Kreditverluste oder Renditen von Aktienpreisen darstellen. In dieser Arbeit werden aber vor allem Renditen von Aktienpreisen betrachtet. Vorab sei bemerkt, dass bei einigen der oben genannten Interpretationen, realistisch betrachtet, nicht von unabhängigen Zufallsvariablen ausgegangen werden kann. Dies wird in Kapitel 3 genauer untersucht.

Bei einer Investition in eine Aktie ist der Anleger interessiert daran, große Verluste, die diese Aktie in einer Zeitspanne von z.B. 10 oder 100 Tagen erbringen kann, abzuschätzen. Dazu soll der sogenannte *Value at Risk* (kurz: VaR) betrachtet werden. Der VaR ist das  $q$ -te Quantil der Verteilung  $F$ , für  $0.95 \leq q < 1$

$$VaR_q = F^{-1}(q),$$

wobei  $F^{-1}$  die Inverse von  $F$  ist<sup>1</sup>. Er gibt sozusagen an, welcher Betrag in einer bestimmten Zeit, mit einer bestimmten Wahrscheinlichkeit  $(1 - q)$  nicht überschritten wird. Um den Value at Risk zu berechnen, wird jedoch die zugrunde liegende, unbekannte Verteilungsfunktion  $F$  der Daten benötigt. In vielen Schätzungen wird dazu eine Normalverteilung an die Daten angepasst und damit der Value at Risk berechnet. Das Problem dabei ist, dass die Normalverteilung bei der Art von Daten, die hier betrachtet werden sollen, oft keine sehr gute Anpassung an die extremen Werte der Daten liefert, da diese zu häufig auftreten. Da hier keine Mittelwerte betrachtet werden sollen, sondern große Verluste, ist aber gerade die Anpassung an diese extremen Werte wichtig.

Eine bessere Möglichkeit die sogenannten *tails* (Flanken) der unbekannten, zugrunde liegenden Verteilung zu schätzen, bietet die Peak over Threshold Methode (kurz: POT-Methode). Mit dieser Methode kann eine verallgemeinerte Paretoverteilung an die hohen Werte der Daten angepasst und schließlich der Value at Risk berechnet werden. Dies soll im Folgenden hergeleitet werden. Darüber hinaus soll noch der *Expected Shortfall* (kurz: ES) vorgestellt werden. Er bietet eine Alternative zum Value at Risk. Der Expected Shortfall ist die erwartete Verlustgröße,

---

<sup>1</sup>vgl. [16] (S.4)

gegeben, dass der VaR überschritten wurde<sup>2</sup>

$$ES_q = E[X \mid X > VaR_q].$$

In Kapitel 2 wird für ein konkretes Beispiel mit Hilfe der POT-Methode eine verallgemeinerte Paretoverteilung an einen Datensatz angepasst und der Value at Risk und der Expected Shortfall berechnet.

Da die absoluten Kurswerte einer Aktie keine Auskunft über den Gewinn oder Verlust geben, werden Log-Renditen von Aktien betrachtet<sup>3</sup>. Diese werden zunächst berechnet und dann negiert, da die Methoden, die im Folgenden vorgestellt werden, darauf ausgelegt sind Maxima zu schätzen, in diesem Fall aber die großen Verluste, also Minima, abgeschätzt werden sollen. Dies kann man sich leicht überlegen, denn es gilt<sup>4</sup>:

$$\min_{i \leq n} X_i = -\max_{i \leq n} (-X_i).$$

Im Folgenden sei also

$$X_t = -(\log S_t - \log S_{t-1}),$$

wobei  $S_t$  der Wert einer Aktie am Ende des Tages  $t$ , und  $X_t$  die entsprechende Log-Rendite ist.

## 1.1 POT-Methode

Bei den Anwendungen die hier betrachtet werden, interessiert der Anleger sich vor allem dafür die besonders hohen Werte (die hier durch die gesetzten Voraussetzungen die besonders kleinen Werte sind) abzuschätzen, damit er sich gegen große Verluste absichern kann. Diese hohen Werte liegen in den tails der gesuchten Verteilung. Sie werden oft als *Extremwerte* bezeichnet. Bei der Art von Daten die hier betrachtet werden sollen, treten häufiger Extremwerte auf, als das z.B. von einer Normalverteilung abgebildet wird. Es wird eine Verteilung benötigt, deren tails dicker und breiter sind als die einer Normalverteilung, eine sogenannte *heavy tailed* Verteilung, um diese Extremwerte gut abzuschätzen. Dazu folgende Definition<sup>5</sup>:

### Definition 1

*Eine Verteilung besitzt ein heavy tail, wenn es ein  $j \in \mathbb{N}$  gibt, so dass das  $j$ -te Moment (ausgewertet auf der positiven Halbgerade) unendlich ist.*

---

<sup>2</sup>vgl. [16] (S.4)

<sup>3</sup>Es kann leicht gezeigt werden, dass die Log-Renditen in etwa den „normalen“ Renditen

$X_t = \frac{S_t - S_{t-1}}{S_{t-1}}$  entsprechen.

<sup>4</sup>vgl. [19] (S.9)

<sup>5</sup>vgl. [19] (S.27)

Darüber hinaus treten bei Renditen von Aktien auch die mittleren Werte häufiger auf, als es von einer Normalverteilung abgebildet wird. Eine bessere Anpassung liefert eine sogenannte *thin waist* Verteilung, deren Gipfel höher und schmaler ist als bei einer Normalverteilung<sup>6</sup>. Eine Verteilung mit heavy tails und thin waists wird auch *leptokurtische Verteilung* genannt. Abbildung 1.1 zeigt so eine leptokurtische Verteilung im Vergleich mit einer Normalverteilung.

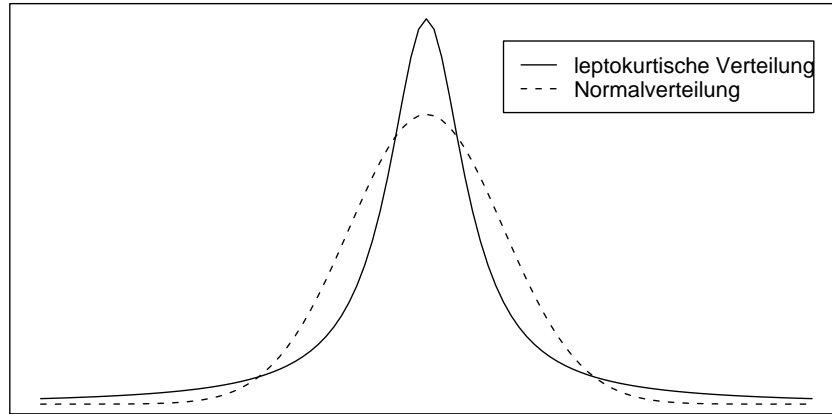


Abbildung 1.1: Leptokurtische Verteilung im Vergleich mit einer Normalverteilung

Da mit der POT-Methode „nur“ der Tail einer Verteilung abgeschätzt wird, wird häufig eine Konvexkombination einer an den Tail angepassten heavy tail Verteilung  $G$  und einer an den Rest der Daten angepassten Normalverteilung  $F$  verwendet<sup>7</sup>

$$(1 - d)F + dG, \quad d \in (0,1) \text{ klein.}$$

Um mit der Peak over Threshold Methode eine geeignete Verteilung an die hohen Werte der Beobachtungen anzupassen, wird zunächst ein Schwellenwert (*Threshold*)  $u$ , der im oberen Teil des Wertebereichs der Daten liegen sollte, vorgegeben. Mit diesem Threshold werden dann für  $x_i > u$  die *Exzesse* über  $u$

$$y_i := x_i - u$$

bestimmt. Die  $x_i$ , die  $x_i > u$  erfüllen, heißen *Exceedances*. Es werden also die großen Werte aus dem Datensatz selektiert.

---

<sup>6</sup>vgl. [5] (S.59)

<sup>7</sup>vgl. [19] (S.32)

Sei nun  $K$  eine Zufallsgröße, die die Anzahl der Exceedances zu einem gegebenen Threshold  $u$  angibt. Da die  $X_i$  iid Zufallsvariablen mit gemeinsamer Verteilungsfunktion  $F$  sind, gilt

$$P(K = k) = \binom{n}{k} p^k (1-p)^{n-k} = B(n, p)(k),$$

wobei  $B(n, p)$  die Binomialverteilung mit Parametern  $n$  und  $p = 1 - F(u)$  ist. Die erwartete Anzahl an Überschreitungen über  $u$  ist also

$$\phi_{n,F}(u) = np = n(1 - F(u))$$

und somit monoton fallend in  $u$ . Wenn  $np$  für  $n \rightarrow \infty$  gegen ein  $\lambda > 0$  konvergiert, kann auch eine  $\text{Poi}(\lambda)$ -Approximation durchgeführt werden. Für einen hohen Threshold  $u$ , also einen Threshold, der nah am rechten Endpunkt liegt, kann die Anzahl  $K$  der Exceedances als poissonverteilte Zufallsgröße aufgefasst werden<sup>8</sup>.

Nun soll die Verteilungsfunktion der Exzesse über  $u$  definiert werden<sup>9</sup>:

**Definition 2**

Sei  $X$  eine Zufallsvariable mit Verteilungsfunktion  $F$ . Die Exzess-Verteilungsfunktion von  $X$  bzw.  $F$  oberhalb des Thresholds  $u$  wird definiert durch

$$F_u(x) = P\{X - u \leq x \mid X > u\} = \frac{F(x + u) - F(u)}{1 - F(u)} \quad (1.1)$$

für  $x \geq 0$  und festes  $u < x_R$ , wobei  $x_R = \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty$  den rechten Endpunkt von  $F$  bezeichnet.

Die Exzess-Verteilung repräsentiert also die Wahrscheinlichkeit, dass ein Verlust den Threshold  $u$  mit höchstens einem Betrag  $x$  überschreitet, gegeben, dass der Threshold überschritten wird.

## 1.2 Verallgemeinerte Paretoverteilung

Das wichtigste Verteilungsmodell für Exzesse über Thresholds ist die verallgemeinerte Paretoverteilung (generalized pareto distribution, kurz: GPD)<sup>10</sup>:

**Definition 3**

Die Verteilungsfunktion der verallgemeinerten Paretoverteilung wird definiert durch

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi \frac{x}{\beta})^{-1/\xi}, & \xi \neq 0, \\ 1 - e^{-x/\beta}, & \xi = 0, \end{cases}$$

wobei  $\beta > 0$ , und  $x \geq 0$  wenn  $\xi \geq 0$  und  $0 \leq x \leq -\frac{\beta}{\xi}$  wenn  $\xi < 0$ .

<sup>8</sup>vgl. [19] (S.6ff)

<sup>9</sup>vgl. [13] (Definition 2.4 S.44)

<sup>10</sup>vgl. [13] (Definition 2.3 S.43)

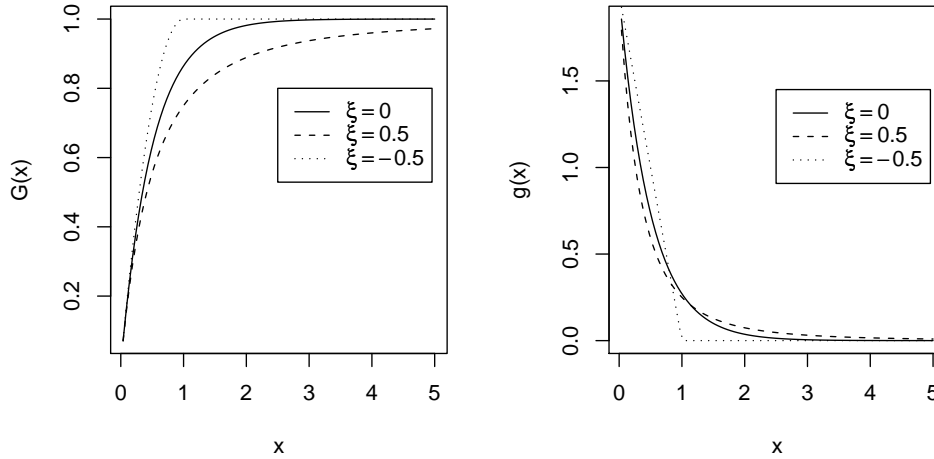


Abbildung 1.2: Verteilungsfunktion und Dichte der verallgemeinerten Paretoverteilung für verschiedene Shape-Parameter

Die Verteilung ist verallgemeinert, in dem Sinne, dass sie verschiedene andere Verteilungen unter einer gemeinsamen parametrischen Form zusammenfasst. Dabei ist  $\xi$  der Form-Parameter (auch: Shape-Parameter) und  $\beta$  ein zusätzlicher Skalierungs-Parameter (auch: Scale-Parameter). Für  $\xi > 0$  ist  $G_{\xi,\beta}$  eine reparametrisierte Version der gewöhnlichen Paretoverteilung,  $\xi = 0$  entspricht einer Exponentialverteilung und  $\xi < 0$  ist bekannt als eine Pareto Typ II Verteilung<sup>11</sup>. Für das Risikomanagement ist vor allem der erste Fall relevant, weil die GPD heavy tailed ist, wenn  $\xi > 0$  ist. Die GPD-Familie kann durch Hinzufügen eines Lage-Parameters (auch: Loc-Parameter)  $\mu$  erweitert werden.  $G_{\xi,\mu,\beta}(x)$  ist dann definiert als  $G_{\xi,\beta}(x - \mu)$ , d.h.

$$G_{\xi,\mu,\beta}(x) = G_{\xi,\beta}(x - \mu) = \begin{cases} 1 - (1 + \xi \frac{x-\mu}{\beta})^{-1/\xi}, & \xi \neq 0, \\ 1 - e^{-\frac{x-\mu}{\beta}}, & \xi = 0. \end{cases}$$

Abbildung 1.2 (links) zeigt die Verteilungsfunktionen der GPD für  $\mu = 0$ ,  $\beta = 0.5$  und  $\xi \in \{-0.5, 0, 0.5\}$ . Abbildung 1.2 (rechts) zeigt dazu die entsprechenden Dichtefunktionen.  $\xi$  bestimmt dabei die Geschwindigkeit, mit der die Dichte für  $x \rightarrow \infty$  gegen 0 konvergiert. In der Abbildung ist zu sehen, dass die Dichte der GPD mit einem Shape-Parameter von 0.5 deutlich größere Tails hat, als die Dichte der GPD mit  $\xi = -0.5$ . Eine Veränderung des Loc-Parameters  $\mu$  würde eine Verschiebung der Verteilungsfunktion bzw. der Dichte parallel zur x-Achse bewirken und  $\beta$  definiert die Stauchung ( $\beta < 1$ ) bzw. die Streckung ( $\beta > 1$ ) der Verteilung.

<sup>11</sup>vgl. [15] (S.3)

## 1.3 Exkurs: Block-Maxima Methode und GEV-Verteilung

Um später Aussagen über das Grenzwertverhalten der Exzesse machen zu können, werden einige Grundlagen der sogenannten Block-Maxima Methode benötigt. Diese soll nun kurz vorgestellt werden. Die Block-Maxima Methode ist die älteste Gruppe von Modellen für Extremwerte. Bei dieser Methode werden die zugrunde liegenden iid Beobachtungen in  $m$  Blöcke der Länge  $n$  eingeteilt und aus jedem Block wird das Maximum ermittelt

$$M_n = \max\{X_1, \dots, X_n\}.$$

Die Blöcke können dabei z.B. Monate oder Jahre darstellen. Am besten geeignet für die Abschätzung solcher Block-Maxima ist die verallgemeinerte Extremwertverteilung<sup>12</sup>:

### Definition 4

*Die Verteilungsfunktion der verallgemeinerten Extremwertverteilung (generalized extreme value distribution, kurz: GEV) wird definiert durch*

$$H_\xi(x) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}), & \xi \neq 0, \\ \exp(-e^{-x}), & \xi = 0, \end{cases}$$

für  $x$  mit  $1 + \xi x > 0$  und  $-\infty < \xi < \infty$ .

Wie bei der GPD ist  $\xi$  wieder der Shape-Parameter. Durch Einbeziehung von entsprechenden Loc- und Scale-Parametern  $\mu$  und  $\psi > 0$  kann die Familie der Verteilungen erweitert werden.  $H_{\xi,\mu,\psi}(x)$  ist dann definiert als  $H_\xi(\frac{x-\mu}{\psi})$ . Der Definitionsbereich besteht in diesem Fall aus allen  $x$  mit  $1 + \xi \frac{x-\mu}{\psi} > 0$ . Man sagt,  $H_{\xi,\mu,\psi}$  ist vom Typ  $H_\xi$ .

Die Extremwert-Dichten stehen in folgendem Zusammenhang mit den verallgemeinerten Pareto Dichten:

$$G(x) = 1 + \log H(x)$$

für  $\log H(x) > -1$ , wobei  $G$  eine verallgemeinerte Paretoverteilung und  $H$  eine entsprechende verallgemeinerte Extremwertverteilung ist<sup>13</sup>.

Ein weiterer Begriff, der später benötigt wird um Aussagen über das Grenzwertverhalten der Exzesse zu machen, ist der des Max-Anziehungsbereiches<sup>14</sup>:

<sup>12</sup>vgl. [13] (Definition 2.1 S.41)

<sup>13</sup>vgl. [19] (S.21)

<sup>14</sup>vgl. [15] (S.4)



**Definition 5**

Seien  $F$  und  $H$  Verteilungsfunktionen.  $F$  liegt im Max-Anziehungsbereich von  $H$  (Schreibweise:  $F \in \mathcal{D}(H)$ ) falls es Folgen von reellen Zahlen  $a_n > 0$  und  $b_n$  gibt, so dass die normalisierte Folge  $(M_n - b_n)/a_n$  in Verteilung gegen  $H$  konvergiert

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} H(x),$$

wobei  $F^n(x) = F^{M_n}(x)$ <sup>15</sup> und  $F^{M_n}(x)$  ist die Verteilungsfunktion von  $M_n$ .

Fisher & Tippet<sup>16</sup> haben gezeigt, dass die Extremwertverteilungen die einzigen nicht-entarteten Grenzverteilungen für annähernd normalisierte Stichproben-Maxima sind. Genauer gesagt kann für eine gegebene Verteilung  $F$  gezeigt werden, dass folgende Aussagen äquivalent sind:

- Die annähernd normalisierten Stichproben-Maxima konvergieren in Verteilung gegen eine nicht entartete Verteilung  $H$ .
- $F$  liegt im Max-Anziehungsbereich einer GEV  $H_\xi$  für irgendein  $\xi$ .

## 1.4 Pickands-Balkema-de Haan Theorem

Nun kann das Pickands-Balkema-de Haan Theorem formuliert werden. Es liefert ein wichtiges Resultat für die Schätzung von Exzessverteilungen oberhalb hoher Thresholds<sup>17</sup>:

**Theorem 6 (Pickands-Balkema-de Haan)**

Für alle  $\xi \in \mathbb{R}$  liegt eine in  $x_R$  stetige Verteilungsfunktion  $F$  in  $\mathcal{D}(H_\xi)$  genau dann, wenn

$$\lim_{u \uparrow x_R} \sup_{0 < x < x_R - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0$$

für eine geeignete positive, messbare Funktion  $\beta$  gilt, wobei

$x_R = \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty$  wieder den rechten Endpunkt von  $F$  bezeichnet.

Diese Aussage wird mit Hilfe des Fisher-Tippet Theorems bewiesen. Ein ausführlicher Beweis ist z.B. in [13] zu finden. Hier soll nur eine Beweisskizze gezeigt werden, die sich weitestgehend an [4] (S.76ff) hält. Darüber hinaus wird gezeigt, dass  $\beta(u) = \beta + \xi(u - \mu)$  gilt, was später noch benötigt wird.

<sup>15</sup>was leicht zu zeigen ist, siehe dazu z.B. [13] (Lemma 1.1 S.2)

<sup>16</sup>vgl. [8]

<sup>17</sup>vgl. [13] (Satz 2.6.2 S.45)

**Beweisskizze**

Sei  $F$  die Verteilungsfunktion von  $X$ . Dann gilt für  $n$  groß genug nach Fisher & Tippett<sup>18</sup>

$$F \in \mathcal{D}(H_\xi) \Leftrightarrow F^n(x) \approx \exp \left[ - \left( 1 + \xi \left( \frac{x - \mu}{\beta} \right) \right)^{-1/\xi} \right]$$

für  $\mu, \beta > 0$  und irgendein  $\xi$ . Wird auf beiden Seiten der Logarithmus berechnet, folgt daraus

$$n \log F(x) \approx - \left[ 1 + \xi \left( \frac{x - \mu}{\beta} \right) \right]^{-1/\xi}. \quad (1.2)$$

Für große Werte von  $x$  impliziert eine Taylor-Entwicklung, dass

$$\log F(x) \approx -(1 - F(x)) \quad (1.3)$$

gilt. Durch Einsetzen von (1.3) in (1.2) und Umstellen der Gleichung ergibt sich

$$F(x) \approx 1 - \frac{1}{n} \left[ 1 + \xi \left( \frac{x - \mu}{\beta} \right) \right]^{-1/\xi}.$$

Also gilt für großes  $u$

$$F(u) \approx 1 - \frac{1}{n} \left[ 1 + \xi \left( \frac{u - \mu}{\beta} \right) \right]^{-1/\xi}$$

und für  $x > 0$

$$F(u + x) \approx 1 - \frac{1}{n} \left[ 1 + \xi \left( \frac{u + x - \mu}{\beta} \right) \right]^{-1/\xi}.$$

Für die Exzess-Verteilungsfunktion ergibt sich damit

$$\begin{aligned} F_u(x) &= P\{X - u \leq x \mid X > u\} \\ &= \frac{F(x + u) - F(u)}{1 - F(u)} \\ &\approx \frac{1 - \frac{1}{n} \left[ 1 + \xi \left( \frac{u + x - \mu}{\beta} \right) \right]^{-1/\xi} - \left( 1 - \frac{1}{n} \left[ 1 + \xi \left( \frac{u - \mu}{\beta} \right) \right]^{-1/\xi} \right)}{1 - \left( 1 - \frac{1}{n} \left[ 1 + \xi \left( \frac{u - \mu}{\beta} \right) \right]^{-1/\xi} \right)} \\ &= \frac{\frac{1}{n} \left[ \left( \frac{\beta + \xi(u - \mu)}{\beta} \right)^{-1/\xi} - \left( \frac{\beta + \xi(u + x - \mu)}{\beta} \right)^{-1/\xi} \right]}{\frac{1}{n} \left( \frac{\beta + \xi(u - \mu)}{\beta} \right)^{-1/\xi}} \\ &= 1 - \left( 1 + \frac{\xi x}{\beta + \xi(u - \mu)} \right)^{-1/\xi} \\ &= G_{\xi, \beta + \xi(u - \mu)}(x). \end{aligned}$$

---

<sup>18</sup>vgl. [8]

Daraus folgt

$$F_u(x) \approx G_{\xi, \beta(u)}(x)$$

mit

$$\beta(u) = \beta + \xi(u - \mu). \quad (1.4)$$

□

Das Pickands-Balkema-de Haan Theorem besagt insbesondere, dass die verallgemeinerte Paretoverteilung für große  $u$  eine gute Approximation der Exzessfunktion  $F_u$  darstellt, wenn  $F$  im Anziehungsbereich einer Extremwertverteilung liegt. Darüber hinaus ist der Shape-Parameter der GPD für die Exzesse derselbe wie der Shape-Parameter der GEV Verteilung für die Maxima. Es kann gezeigt werden, dass im Wesentlichen alle allgemein verwendeten Verteilungen der Statistik für ein  $\xi$  im Max-Anziehungsbereich von  $H_\xi$  liegen. Die Normal-, Exponential-, Gamma- und Lognormalverteilungen z.B. liegen im Max-Anziehungsbereich von  $H_0$ . Diese Verteilungen werden auch *medium tailed* genannt. Für  $\xi < 0$  liegen z.B. die Gleichverteilung und die Betaverteilung im Max-Anziehungsbereich von  $H_\xi$ . Diese Verteilungen heißen *short tailed*. Am interessantesten sind hier jedoch die *heavy tailed* Verteilungen wie z.B. die Pareto-, Burr-, Loggamma-, Cauchy- und die t-Verteilung. Diese liegen für  $\xi > 0$  im Max-Anziehungsbereich von  $H_\xi$ <sup>19</sup>. Zusammenfassend kann festgehalten werden, dass das Pickands-Balkema-de Haan Theorem ein weit anwendbares Resultat ist, dass die GPD als kanonische Verteilung für die Modellierung von Exzessen oberhalb hoher Thresholds qualifiziert.

## 1.5 Wahl eines geeigneten Thresholds

Im Folgenden stellt sich die Frage, wie ein geeigneter Threshold gewählt wird. Bei der Einführung der POT-Methode, wurde bereits erwähnt, dass der gewählte Threshold im oberen Teil des Wertebereichs der Daten liegen sollte, da sonst keine extremen Werte geschätzt werden können. Aus dem Pickands-Balkema-de Haan Theorem resultiert, dass die gesuchte Verteilung der Exzesse durch eine GPD approximiert werden kann, wenn der Threshold groß genug ist. Dies legt nahe, einen relativ hohen Threshold zu wählen, der sehr nah am rechten Endpunkt liegt. Andererseits bleiben bei einem sehr hohen Threshold nur noch sehr wenig Exzesse übrig, mit denen das Modell geschätzt werden kann, was unter Umständen zu einer sehr hohen Varianz führt. Die Wahl eines Thresholds ist somit grundsätzlich ein Kompromiss zwischen der Wahl eines ausreichend hohen Thresholds, so dass das Theorem von Pickands-Balkema-de Haan als tatsächlich exakt betrachtet werden kann, und der Wahl eines ausreichend kleinen Thresholds, so dass genügend Material zum Schätzen der Parameter zur Verfügung steht. Eine optimale, universale Lösung für die Thresholdwahl gibt es also nicht. Es gibt aber zwei

---

<sup>19</sup>vgl. [15] (S.4ff)

graphische Hilfsmittel mit denen geeignete Thresholds gefunden werden können. Eines davon ist der sogenannte Mean Residual Life Plot. Im Weiteren folgt eine kurze Darstellung.

In Definition 2 wurde die Exzess-Verteilungsfunktion von  $X$  bzw.  $F$  oberhalb des Thresholds  $u$

$$F_u(x) = P\{X - u \leq x \mid X > u\} = \frac{F(x + u) - F(u)}{1 - F(u)}$$

für  $x \geq 0$  und festes  $u < x_R$  definiert. Diese Verteilungsfunktion wird häufig auch Residual Life Verteilungsfunktion zum Alter  $u$  genannt. Dabei stellt  $F_u(x)$  die Wahrscheinlichkeit dar, dass die verbleibende Lebenszeit  $\leq x$  ist, wenn das Alter  $u$  erreicht wurde. Im Fall der Lebenszeit-Schätzung ist die im Alter  $u$  zu erwartende verbleibende Lebenszeit von Interesse. Diese wird durch die Mean Residual Life Funktion (kurz: mrl-Funktion) dargestellt<sup>20</sup>:

$$e_f(u) := E(X - u \mid X > u) = \frac{E((X - u)1_{\{X > u\}})}{P(X > u)}, \quad \text{für } u < x_R.$$

Bei der Betrachtung von z.B. Aktienkursen würde die Mean Residual Life Funktion die zu erwartende Kurserhöhung, oder bei Betrachtung negativer Log-Renditen, den zu erwartenden Kursabfall, gegeben, der Kurs hat bereits eine Schwelle  $u$  über- bzw. unterschritten, darstellen. Für Schwellenwerte, die größer sind als der rechte Endpunkt  $x_R$ , ist die mrl-Funktion gleich 0, denn für solche  $u$  ist die Exzess-Verteilungsfunktion gleich 0. Eine für die Thresholdwahl entscheidende Eigenschaft der mrl-Funktion ist, dass die Exzesse ab einem bestimmten Wert für  $u$  annähernd GP(generalized pareto)-verteilt sind und die mrl-Funktion dann linear in  $u$  ist. Dies soll kurz gezeigt werden.

Wenn  $Y$  GP-verteilt ist, mit den Parametern  $\beta$  und  $\xi$ , dann gilt für  $\xi < 1$

$$E(Y) = \frac{\beta}{1 - \xi}. \quad (1.5)$$

Für  $\xi \geq 1$  ist der Erwartungswert unendlich<sup>21</sup>. Sei nun  $u_0$  der kleinste Threshold, für den das GP-Modell gilt, und  $X_1, \dots, X_n$  seien Zufallsgrößen mit der entsprechenden durch  $u$  festgelegten Verteilung. Vorausgesetzt  $\xi < 1$ , gilt dann für jedes  $X_i$

$$E(X_i - u_0 \mid X_i > u_0) = \frac{\beta_{u_0}}{1 - \xi},$$

wobei  $\beta_{u_0}$  den Scale-Parameter, der sich für die Verteilung der Exzesse über den Threshold  $u_0$  ergibt, darstellt. Aber wenn das GP-Modell gültig ist für Exzesse über den Threshold  $u_0$ , sollte es auch gültig sein für alle Thresholds  $u > u_0$ ,

---

<sup>20</sup>vgl. [19] (S.47)

<sup>21</sup>vgl. [4] (S.78ff)

allerdings mit einem anderen durch  $u$  gegebenen Scale-Parameter  $\beta_u$ . Zwischen  $\beta_u$  und  $\beta_{u_0}$  besteht folgender Zusammenhang:

$$\beta_u = \beta_{u_0} + \xi u.$$

Das folgt aus der Beweisskizze des Pickands-Balkema-de Haan Theorem (Theorem 6), wenn anstatt der GPD-Verteilungsfunktion  $G_{\xi,\mu,\beta}$ , die GPD-Verteilungsfunktion  $G_{\xi,\beta}$  eingesetzt wird, d.h. die Verteilungsfunktion, die noch nicht durch einen Loc-Parameter erweitert wurde. Für die mrl-Funktion ergibt sich daraus

$$E(X_i - u \mid X_i > u) = \frac{\beta_u}{1 - \xi} = \frac{\beta_{u_0} + \xi u}{1 - \xi}.$$

Für  $u > u_0$  ist die mrl-Funktion also linear in  $u$ .

Für die Wahl eines geeigneten Thresholds heißt das nun, dass die mrl-Funktion auf ihrem gesamten Wertebereich geplottet werden muss. Da die Verteilung aber für alle Werte unbekannt ist, muss in diesem Fall auf die empirische Variante der mrl-Funktion zurückgegriffen werden<sup>22</sup>:

$$e_n(u) := \frac{\frac{1}{n} \sum_{i=1}^n (x_i - u) 1_{(u,\infty)}(x_i)}{\frac{1}{n} \sum_{i=1}^n 1_{(u,\infty)}(x_i)} \quad \text{für alle } u < x_R.$$

Eine Anwendung des starken Gesetz der großen Zahlen ergibt, dass die empirische mrl-Funktion näherungsweise mit der tatsächlichen übereinstimmt. Somit liefert dieses Verfahren eine gute Hilfestellung für die Thresholdwahl. Allerdings bietet auch diese Vorgehensweise keine Universallösung, denn die Interpretation eines mrl-Plots ist in der Praxis nicht immer eindeutig. Ein konkretes Beispiel eines mrl-Plots folgt in Kapitel 2. Dort wird zudem der Threshold Choice Plot vorgestellt. Dieser bietet eine weitere Möglichkeit um einen geeigneten Threshold zu finden.

## 1.6 Schätzung der GPD Parameter

Nachdem ein geeigneter Threshold gewählt wurde, können die Parameter der GPD-Verteilung geschätzt werden. Durch die Wahl des Thresholds  $u$  ist der Loc-Parameter  $\mu$  bereits festgelegt. Da die Wahrscheinlichkeit, den Threshold zu überschreiten, bei einem Wert, der kleiner ist als  $u$ , gleich 0 ist, ist  $u$  der linke Endpunkt der Exzess-Verteilung. Dieser entspricht bei einer GPD-Verteilung dem Loc-Parameter  $\mu$ . Bleiben also noch der Shape- und der Scale-Parameter zu schätzen. Diese können durch verschiedene Methoden geschätzt werden. Die bekanntesten Methoden sind die Momentenmethode (method of moments, kurz: MoM), die Methode der wahrscheinlichkeitsgewichteten Momente (method of probability weighted moments, kurz: PWM) und die Maximum-Likelihood Methode (kurz: ML).

---

<sup>22</sup>vgl. [19]

Diese drei Methoden wurden 1987 von Hosking & Wallis<sup>23</sup> und 1995 von Rootzén & Tajvidi<sup>24</sup> in Simulations-Studien für GPD-verteilte Daten verglichen. Dabei folgerten Hosking & Wallis aus ihren Simulations-Ergebnissen, dass für eine Empfehlung, welche Schätzmethode besser geeignet ist, in der Praxis einige Spezifikationen des Problems nötig sind. Dabei ist besonders der Wertebereich in dem  $\xi$  wahrscheinlich liegen wird, von Interesse. Nach den Resultaten von Hosking & Wallis zeigen MoM-Schätzer für  $\xi$  nahe 0 und für  $\xi > 0$  die beste Gesamtleistung. Wenn die Wahrscheinlichkeit groß ist, dass  $\xi$  wesentlich kleiner ist als 0, und besonders wenn  $\xi$  kleiner sein könnte als  $-0.2$ , präferieren sie die PWM-Methode wegen ihres geringen systematischen Fehlers. Die ML-Methode erscheint ihnen dagegen nur gerechtfertigt für sehr große Stichproben und  $\xi > 0.2$ . Rootzén & Tajvidi folgerten aus ihrer Studie, dass die PWM-Methode für heavy-tailed Daten mit  $\xi \geq 0.5$  schwer verzerrte Parameter-Schätzer liefert, die dafür aber weniger variabel sind als die ML-Schätzer. Da hier erwartet wird, dass die Daten heavy tailed sind (also  $\xi > 0$  gilt), soll hier nur kurz auf die Maximum-Likelihood Methode eingegangen werden.

Seien  $y_1, \dots, y_k$  die  $k$  Exzesse über den gewählten Threshold  $u$ . Für  $\xi \neq 0$  erhält man aus der GPD-Verteilungsfunktion die Dichtefunktion

$$g(x) = \beta^{-1} \left( 1 + \frac{\xi x}{\beta} \right)^{-(\frac{1}{\xi}+1)}$$

und daraus die Log-Likelihood-Funktion

$$\ell(\beta, \xi) = -k \log \beta - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^k \log \left( 1 + \frac{\xi y_i}{\beta} \right),$$

vorausgesetzt  $\left( 1 + \frac{\xi y_i}{\beta} \right) > 0$  für  $i = 1, \dots, k$ ; sonst gilt  $\ell(\beta, \xi) = -\infty$ . Für  $\xi = 0$  erhält man aus der Dichte

$$g(x) = \beta^{-1} e^{-\frac{x}{\beta}}$$

die Log-Likelihood-Funktion

$$\ell(\beta) = -k \log \beta - \beta^{-1} \sum_{i=1}^k y_i.$$

Die Maximum-Likelihood-Schätzer lassen sich ermitteln, indem  $\hat{\beta}$  und  $\hat{\xi}$  so gewählt werden, dass  $\ell$  ein lokales Maximum annimmt. Für  $\xi > -0.5$  kann gezeigt werden, dass die Regularitätsbedingungen für die Maximum-Likelihood Methode erfüllt sind und dass die ML-Schätzer  $(\hat{\beta}, \hat{\xi})$ , basierend auf einer Stichprobe von  $k$  Exzessen, asymptotisch normalverteilt sind<sup>25</sup>. Konkret haben Hosking & Wallis

---

<sup>23</sup>vgl. [11]

<sup>24</sup>vgl. [20]

<sup>25</sup>vgl. [11]

gezeigt, dass

$$k^{1/2} \begin{pmatrix} \hat{\xi} \\ \hat{\beta} \end{pmatrix} \xrightarrow{d} N \left[ \begin{pmatrix} \xi \\ \beta \end{pmatrix}, \begin{pmatrix} (1+\xi)^2 & \beta(1+\xi) \\ \beta(1+\xi) & 2\beta^2(1+\xi) \end{pmatrix} \right]$$

gilt.

## 1.7 Value at Risk

### 1.7.1 Mathematische Herleitung des VaR

Nachdem gezeigt wurde, wie mit den vorgestellten Verfahren eine verallgemeinerte Paretoverteilung an die Daten angepasst werden kann, kann nun der Value at Risk hergeleitet werden. Wie schon in der Einleitung erwähnt, ist der Value at Risk das  $q$ -te Quantil einer Verteilung  $F$  für  $0.95 \leq q < 1$

$$VaR_q = F^{-1}(q),$$

wobei  $F^{-1}$  die Inverse von  $F$  ist.

Sei nun  $F$  die unbekannte, zugrunde liegende Verteilungsfunktion mit rechtem Endpunkt  $x_R$ . Es wird angenommen, dass  $F$  im Max-Anziehungsbereich einer GEV-Verteilung  $H_\xi$  liegt. Diese Annahme ist gerechtfertigt, da wie schon in Abschnitt 1.4 erwähnt, im Wesentlichen alle in der Statistik verwendeten Verteilungen im Anziehungsbereich einer GEV-Verteilung liegen. Mit Bezug auf das Pickands-Balkema-de Haan Theorem (Theorem 6, Abschnitt 1.4) wird angenommen, dass

$$F_u(y) = G_{\xi,\beta}(y), \quad 0 \leq y < x_R - u \quad (1.6)$$

für einen hohen Threshold  $u$  und irgendein  $\xi \in \mathbb{R}$  und  $\beta > 0$  gilt. Durch Setzen von  $x = u + y$  und Kombinieren dieser Formel mit der Formel der Exzess-Verteilungsfunktion

$$F_u(y) = \frac{F(y+u) - F(u)}{1 - F(u)}$$

gelangt man zu folgender Darstellung

$$F(x) = (1 - F(u))G_{\xi,\beta}(x - u) + F(u). \quad (1.7)$$

Ziel ist es nun, diese Formel zu verwenden, um einen Tail-Schätzer zu konstruieren. Das einzige zusätzliche Element, das dazu benötigt wird, ist ein Schätzer für  $F(u)$ . Ein naheliegender Schätzer für  $F(u)$  ist der empirische Schätzer

$$\frac{n - k}{n}, \quad (1.8)$$

wobei  $k$  die Anzahl der Thresholdüberschreitungen bezeichnet. Es wird die Methode der historischen Simulation (kurz: HS) genutzt. Jetzt könnte sich die Frage stellen, warum die HS Methode nicht direkt genutzt wird um den gesamten Tail von  $F(x)$  zu schätzen. Dies wäre allerdings keine gute Alternative, da die HS Methode bei wenigen Daten, wie z.B. im Tail der Verteilung, relativ unzuverlässig ist. Durch Setzen eines Thresholds  $u$  wird aber gewährleistet, dass genügend Beobachtungen, die  $u$  überschreiten, zur Verfügung stehen um eine angemessene Schätzung von  $F(u)$  durchzuführen. Durch Einsetzen des HS Schätzer (1.8) für  $F(u)$  und der Maximum-Likelihood-Schätzer für die Parameter der GPD in (1.7) ergibt sich

$$\hat{F}(x) = 1 - \frac{k}{n} \left( 1 + \hat{\xi} \frac{x - u}{\hat{\beta}} \right)^{-1/\hat{\xi}} \quad (1.9)$$

für  $x > u$ . Durch Invertieren der Tail-Schätzungs-Formel (1.9) gelangt man schließlich für eine gegebene Wahrscheinlichkeit  $q > F(u)$  zu folgendem VaR-Schätzer

$$\widehat{VaR}_q = u + \frac{\hat{\beta}}{\hat{\xi}} \left( \left( \frac{n}{k} (1 - q) \right)^{-\hat{\xi}} - 1 \right). \quad (1.10)$$

Aus mathematischer Sicht ist der VaR damit hergeleitet<sup>26</sup>.

### 1.7.2 Einordnung des VaR

Bis jetzt wurde gezeigt, wie mit Hilfe der POT-Methode eine GPD-Verteilung an iid Zufallsvariablen, die z.B. Log-Renditen von Aktien darstellen, angepasst und der VaR berechnet werden kann. Im Weiteren stellt sich die Frage, zu welcher Aussage der VaR konkret führt. Eine weniger mathematische Definition des Value at Risk bietet z.B. Meyer<sup>27</sup>:

„Unter Value at Risk wird die

- in Geldeinheiten (der Referenzwährung) ausgedrückte,
- mit einer vorgegebenen Wahrscheinlichkeit (dem Konfidenzniveau)
- in einer bestimmten Zeitperiode (dem Zeithorizont oder der Haltdauer)
- nicht überschrittene absolute Wertänderung (maximaler unrealisierter Verlust)
- einer Position verstanden.“

---

<sup>26</sup>vgl. [16] (S.5ff)

<sup>27</sup>[18] (S.12)



Darüber hinaus stellt Meyer klar, dass der Value at Risk „... als absoluter Wert keine direkte Prognose über die Höhe auftretender Verluste darstellt, sondern lediglich die Schwelle angibt, die auftretende Verluste mit der vorgegebenen Wahrscheinlichkeit unter normalen Marktbedingungen nicht überschreiten werden ...“. Durch seine Struktur ermöglicht es der VaR aber Wahrscheinlichkeitsaussagen über den Eintritt zukünftiger Verluste machen zu können. Viele andere Risikomaße geben nur die Höhe des Verlustes an, der bei einer bestimmten Marktentwicklung eintreten wird. Die Eintrittswahrscheinlichkeit dieser Marktentwicklung kann durch sie aber nicht bestimmt werden<sup>28</sup>. Ein weiterer Vorteil des VaR ist, dass er das Risiko in Form einer in Geldeinheiten ausgedrückten, skalaren Größe darstellt<sup>29</sup>, was ihn leicht kommunizierbar<sup>30</sup> und vergleichbar macht. Darüber hinaus wird ihm eine hohe Verständlichkeit auch bei Nicht-Finanzmathematikern nachgesagt<sup>31</sup>.

Gerade diese Vorteile sind es, die den Value at Risk, als Risikomaß, vor allem im finanzwirtschaftlichen Sektor so populär gemacht haben. Besonders die Morgan Guaranty Trust Company, kurz J.P. Morgan, hat 1994 mit ihrem RiskMetrics<sup>TM</sup> System zur Verbreitung des Value at Risk beigetragen. Seinen Durchbruch erlangte der Value at Risk in der zweiten Hälfte der 90er Jahre, als er im Aufsichtsrecht der Banken durch Basel II als verbindliches Risikomaß vorgeschrieben wurde.

Durch die Verbreitung und die weitläufige Nutzung des Value at Risk kam auch Kritik auf. Mit der Zeit sind zahlreiche Verfahren für die VaR-Bestimmung entwickelt worden. Dabei wurde oft kritisiert, dass viele dieser Verfahren eine Normalverteilung der Daten annehmen. Wie anfangs erwähnt, ist diese Annahme aber in den meisten Fällen nicht gerechtfertigt, da die Daten im Gegensatz zur Normalverteilung heavy tails aufweisen. Dieser Kritikpunkt trifft auf den hier vorgestellten Ansatz der GPD-Anpassung nicht zu, da dieser die Tails der Verteilung besser modelliert. Der Unterschied zwischen dem VaR einer GPD-Anpassung und dem VaR unter Annahme einer Normalverteilung wird im Anwendungsteil (Kapitel 2) später verdeutlicht. Kritik, die allgemein auf den Value at Risk zutrifft, erhoben Artzener u.a. (1997,1999)<sup>32</sup>. Sie kritisierten den Value at Risk als Risikomaß indem sie zeigten, dass der VaR kein kohärentes Risikomaß ist, da er nicht subadditiv ist. Sie wiesen nach, dass das Risiko einer aggregierten Position höher liegt, als die Summe der Einzelrisiken. Diese Beobachtung stellte den VaR als Maß für die interne Risikolimitierung in Frage, da es damit möglich ist, dass alle Geschäftsbereiche ihr Risikolimit einhalten, das Gesamtunternehmen seine Risikobeschränkung aber weit verfehlen kann<sup>33</sup>. Als Alternative schlugen Artzener u.a. (1997,1999) den Expected Shortfall vor, welcher im Gegensatz zum Value at Risk die Eigenschaften

---

<sup>28</sup>vgl. [12] (S.39)

<sup>29</sup>vgl. [5] (S.88)

<sup>30</sup>vgl. [5] (S.89)

<sup>31</sup>vgl. [10] (S.23)

<sup>32</sup>vgl. [1] und [2]

<sup>33</sup>vgl. [10] (S.25ff)

eines kohärenten Risikomaßes besitzt. Dieser wird im nächsten Abschnitt hergeleitet.

## 1.8 Expected Shortfall

Der Expected Shortfall ist die erwartete Verlustgröße, gegeben dass der VaR überschritten wurde

$$ES_q = E[X \mid X > VaR_q].$$

Er bezieht sich auf den  $VaR_q$  durch

$$ES_q = VaR_q + E[X - VaR_q \mid X > VaR_q],$$

wobei der zweite Term der Erwartungswert der Exzess-Verteilung  $F_{VaR_q}(x)$  über den Threshold  $VaR_q$  ist.

Aus (1.4) folgt, dass

$$F_{VaR_q}(x) = G_{\xi, \beta + \xi(VaR_q - u)}(x) \quad (1.11)$$

gilt, das heißt, dass die GPD-Anpassung für  $F_{VaR_q}$  den Shape-Parameter  $\xi$  und den Scale-Parameter  $\beta + \xi(VaR_q - u)$  besitzt. Daraus folgt mit (1.5), dass die GPD-Verteilung in (1.11) folgenden Erwartungswert hat

$$E[X - VaR_q \mid X > VaR_q] = \frac{\beta + \xi(VaR_q - u)}{1 - \xi},$$

vorausgesetzt  $\xi < 1$ . Es sei bemerkt, dass

$$\frac{ES_q}{VaR_q} = \frac{1}{1 - \xi} + \frac{\beta - \xi u}{(1 - \xi)VaR_q} \quad (1.12)$$

gilt. In dem Fall, wo die zugrunde liegende Verteilung einen unendlichen rechten Endpunkt hat, lohnt es sich dieses Verhältnis genauer zu betrachten. Denn in diesem Fall ist das Verhältnis größtenteils festgelegt durch den Faktor  $\frac{1}{1 - \xi}$ . Der zweite Term auf der rechten Seite von (1.12) wird nebensächlich klein, wenn die Wahrscheinlichkeit  $q$  näher gegen 1 geht. Diese Beobachtung unterstreicht die Wichtigkeit des Shape-Parameters  $\xi$  in der Tail-Schätzung. Sie legt fest, wie weit sich die beiden Risikomaße in den extremen Regionen der Verlust-Verteilung unterscheiden<sup>34</sup>. Den Expected Shortfall erhält man nun durch Austauschen von Daten-bezogenen Schätzern für alles, was in (1.12) unbekannt ist

$$\widehat{ES}_q = \frac{\widehat{VaR}_q}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1 - \hat{\xi}}.$$

---

<sup>34</sup>vgl. [16] (S.7ff)

---

Der ES im Gegensatz zum VaR subadditiv. Darüber hinaus ist er für die Risikomessung von Bedeutung, da der Value at Risk „nur“ eine Schwelle angibt, die mit einer vorgegebenen Wahrscheinlichkeit nicht über- bzw. unterschritten wird. In vielen Anwendungen ist es aber von Interesse zu wissen, wie groß der erwartete Verlust, in dem Fall in dem die Schwelle überschritten wurde, ist. Da der Expected Shortfall wie der Value at Risk relativ leicht berechnet werden kann, bietet er in vielen Fällen eine gute Alternative zur Value at Risk Berechnung.



## 2 Anwendung der POT-Methode für iid Zufallsvariablen

In diesem Kapitel wird das in Kapitel 1 beschriebene Verfahren in einem Anwendungsbeispiel genauer beschrieben. Dazu wird die POT-Methode im Folgenden auf die täglichen DAX Abschlusswerte vom 01.01.1996 bis zum 31.12.2000 angewendet. Für die Aufarbeitung der Daten wurde das Programm R verwendet.

Zuerst werden die Daten in R eingelesen. In diesem Fall wurden die historischen DAX Abschlusswerte der Internetseite Yahoo Deutschland Finanzen<sup>1</sup> entnommen. Da bei Aktienindizes vor allem die Abweichungen, die sich von einem Tag zum nächsten ergeben, von Interesse sind, sollen hier die Log-Renditen der Daten betrachtet werden. Wie schon in Kapitel 1 erwähnt, müssen diese noch negiert werden, da hier große Verluste (also Minima) abgeschätzt werden sollen, mit den verwendeten Methoden aber Maxima geschätzt werden. Es sei also:

$$X_t = -(logS_t - logS_{t-1}),$$

wobei  $S_t$  der Abschlusswert des DAX am Tag  $t$  und  $X_t$  die entsprechende negative Log-Rendite ist. Abbildung 2.1(oben) zeigt die DAX Abschlusswerte in dem beobachteten Zeitraum. Zu erkennen ist der leichte Kursverlust im Oktober 1997 aufgrund der Asienkrise und der DAX-Absturz im Oktober 1998 aufgrund der Finanzkrisen in Russland, Asien und Lateinamerika. Im Jahr 2000 sind die ersten Kursrückgänge aufgrund des Platzens der sogenannten Dotcom-Blase zu erkennen. Abbildung 2.1(unten) zeigt die negativen Log-Renditen.

Nun soll eine Verteilung an die Daten angepasst werden. Zunächst wird geprüft, ob eine Normalverteilung eine gute Anpassung an die Daten bieten würde. Abbildung 2.2 zeigt dazu den QQ-Plot der negativen Log-Renditen gegen die Normalverteilung. Der QQ-Plot setzt sich dabei aus den Punkten

$$\left\{ \left( \hat{F}^{-1} \left( \frac{i}{n+1} \right), x_{(i)} \right) : i = 1, \dots, n \right\}$$

zusammen. Wobei  $\hat{F}$  in diesem Fall die Standardnormalverteilung und  $x_{(1)}, \dots, x_{(n)}$  den geordneten Datensatz der negativen Log-Renditen darstellt. Wenn die Verteilung die Daten gut beschreibt, sollte dieser Plot keine deutlichen Abweichungen

---

<sup>1</sup><http://de.finance.yahoo.com>

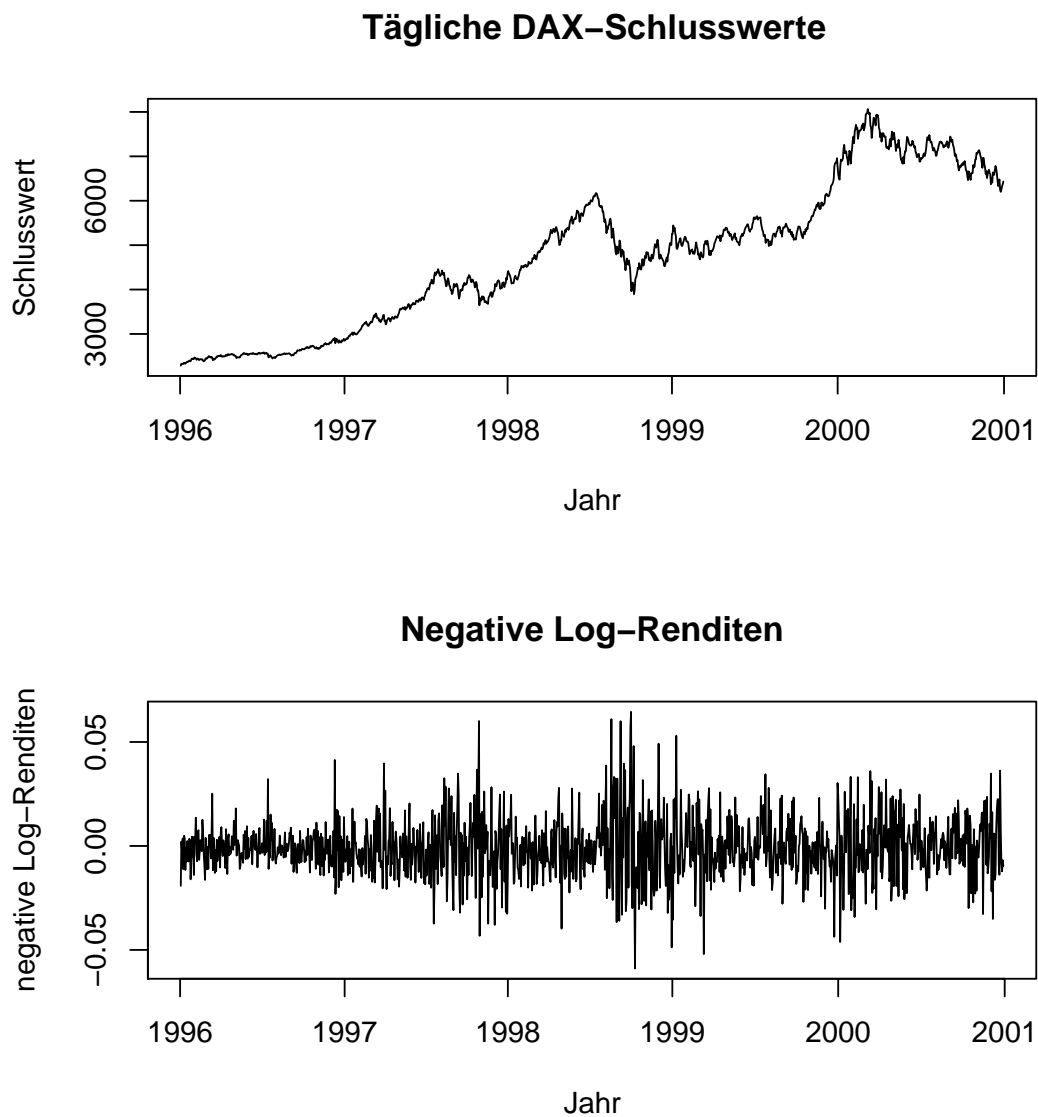


Abbildung 2.1: Tägliche DAX-Schlusswerte und negative Log-Renditen vom 01.01.1996 bis zum 31.12.2000.

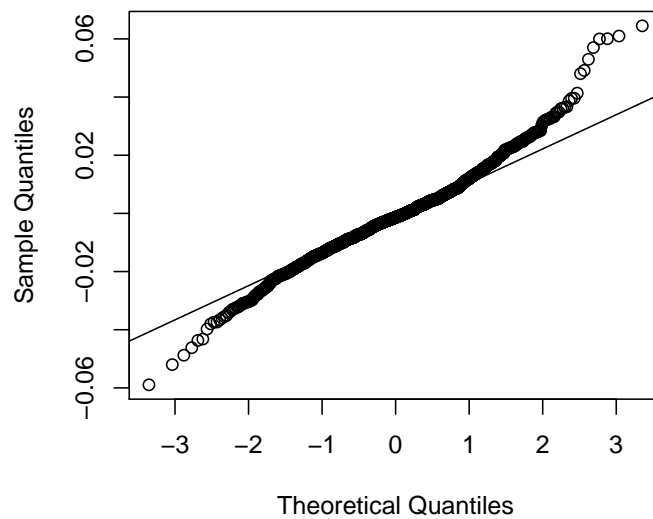


Abbildung 2.2: QQ-Normal Plot der negativen Log-Renditen

von einer Geraden haben. Hier ist aber zu sehen, dass die Punkte an den Rändern deutlich von der Linie abweichen. Das zeigt, dass eine Normalverteilung gerade in den Tails keine gute Abschätzung der Daten liefert, und dass die gesuchte Verteilung deutlich „heavier tailed“ ist als eine Normalverteilung.

## 2.1 Thresholdbestimmung

### 2.1.1 mrl-plot

Um eine verallgemeinerte Paretoverteilung anzupassen, muss zunächst ein Threshold bestimmt werden. Dazu lässt man sich in R einen Mean Residual Life Plot<sup>2</sup> der negativen Log-Renditen ausgeben, siehe Abbildung 2.3. Da hier noch nicht sehr deutlich zu erkennen ist, ab wann der Plot annähernd linear verläuft, sollte der interessante Teil noch einmal vergrößert ausgegeben werden, siehe Abbildung 2.4. Hier würde zunächst wahrscheinlich einen Threshold von 0.0395 (blaue Linie) gewählt werden, da der Plot ab diesem Wert annähernd linear verläuft. Weil dieser Threshold aber, wie sich später herausstellen wird, deutlich zu wenig Exzesse liefert, versucht man einen niedrigeren Threshold zu finden. Dafür bietet sich z.B. ein Threshold von 0.0218 (rote Linie) an. Wie schon im Theorieteil erwähnt, ist es nicht immer einfach einen Threshold zu wählen, da ein möglichst hoher Threshold

<sup>2</sup>siehe Abschnitt 1.5

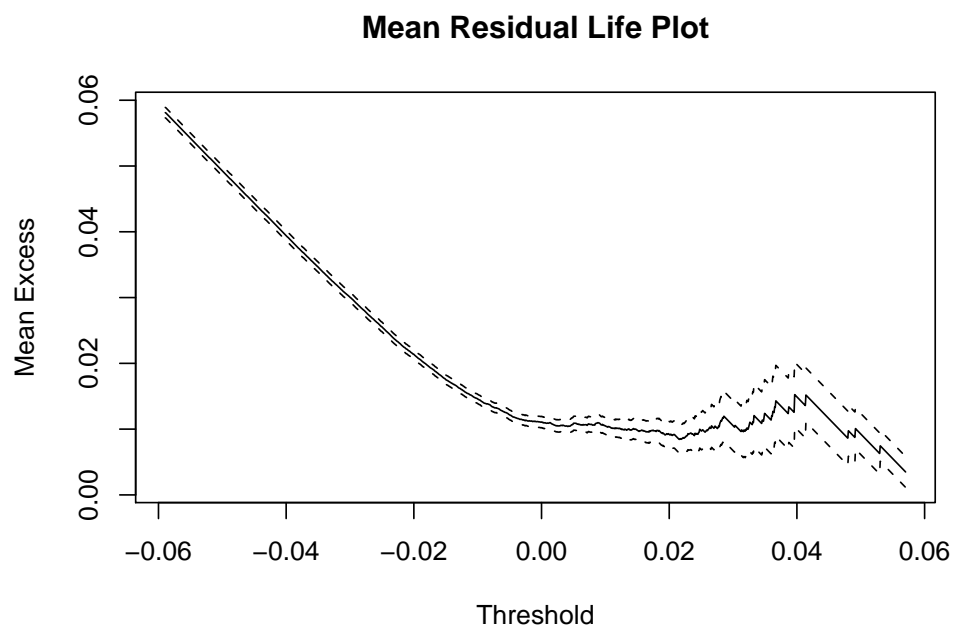


Abbildung 2.3: Mean Residual Life Plot der negativen Log-Renditen

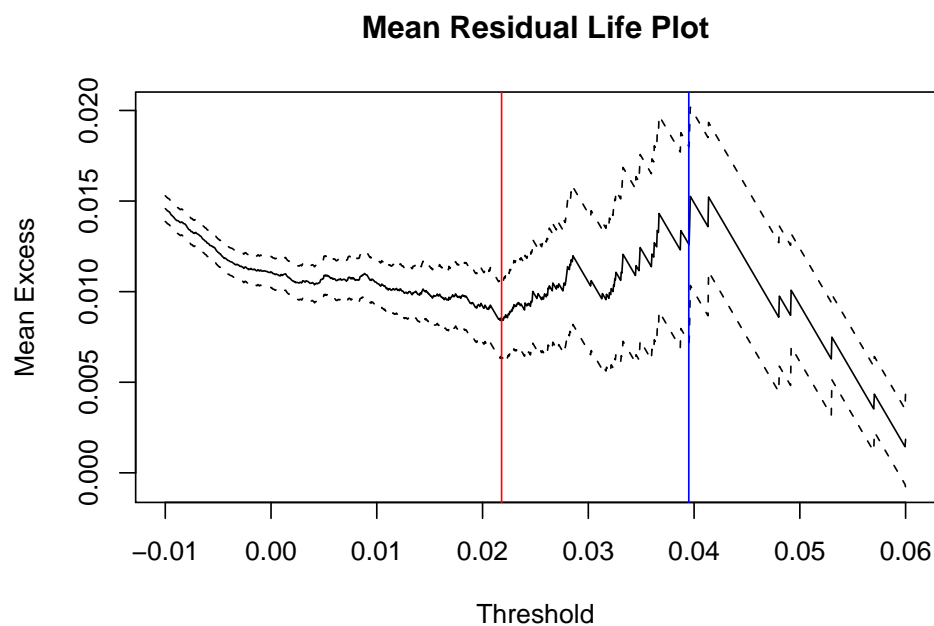


Abbildung 2.4: Ausschnitt aus dem Mean Residual Life Plot der Daten



gewählt werden sollte, damit das Pickands-Balkema-de Haan Theorem als exakt betrachtet werden kann, aber der Threshold auch noch genügend Überschreitungen bieten muss, da sonst keine Verteilung an die Exzesse angepasst werden kann. Deshalb sollen hier zunächst mehrere mögliche Thresholds gesammelt und später verglichen werden, um dann einen geeigneten Threshold zu wählen.

### 2.1.2 tcplot

Eine weitere Möglichkeit in R einen geeigneten Threshold zu finden, bietet der sogenannte Threshold Choice Plot (kurz tcplot). Die Funktion tcplot führt in R für eine angegebene Anzahl an Thresholds aus einem angegebenen Intervall Maximum-Likelihood-Schätzungen für POT-Modelle durch. Für jeden zu schätzenden Parameter wird dann eine Graphik ausgegeben, in der die geschätzten Parameterwerte samt Konfidenzintervall zu einem vorgegebenen Niveau über den Threshold abgetragen werden. Dann gibt es einen minimalen Threshold  $u_0$  ab dem die Approximation der Daten durch eine GPD greift. Für alle größeren Schwellenwerte sollten die Schätzwerte für den Shape-Parameter annähernd konstant bleiben. Für den Scale-Parameter  $\beta_u$  mit  $u > u_0$  gilt nach (1.4) wieder

$$\beta_u = \beta_{u_0} + \xi(u - u_0).$$

Wenn man zur Betrachtung des reparametrisierten Schätzers

$$\beta^* = \beta_u - \xi u$$

übergeht, ist auch dieser Parameter konstant für  $u > u_0$ <sup>3</sup>. Für die Thresholdwahl bedeutet dies, dass aus den tcplots der Threshold ausgewählt wird, ab dem beide Parameter annähernd konstant bleiben.

Für das Beispiel wurden zuerst Maximum-Likelihood-Schätzungen für 100 Thresholds im Intervall von  $[0, 0.04]$ , ohne Konfidenzintervalle, ausgegeben (Abbildung 2.5). Hier ist zu erkennen, dass die geschätzten Scale- und Shape-Parameter ab einem Wert von 0.0335 annähernd konstant sind. Abbildung 2.6 zeigt Maximum-Likelihood-Schätzungen für 31 Thresholds im Intervall  $[0, 0.03]$  mit Konfidenzintervallen. In diesen Abbildungen kann ein annähernd konstanter Verlauf der geschätzten Parameter ab einem Threshold von 0.02 beobachtet werden.

### 2.1.3 Thresholdvergleich

Schließlich wurden mit den zwei Methoden insgesamt vier Thresholds bestimmt. Es soll im Folgenden ein Threshold für die weiteren Berechnungen ausgesucht

---

<sup>3</sup>vgl. [4] (S.83)

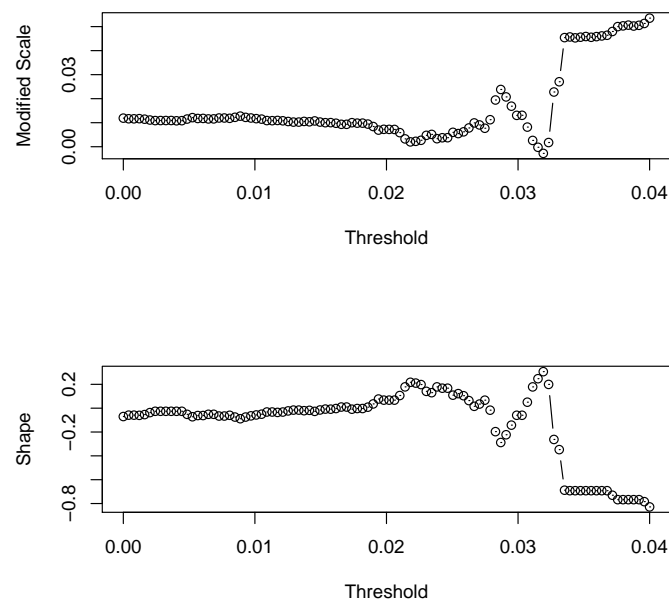


Abbildung 2.5: Threshold Choice Plot für 100 Thresholds im Intervall von  $[0, 0.04]$  ohne Konfidenzintervalle

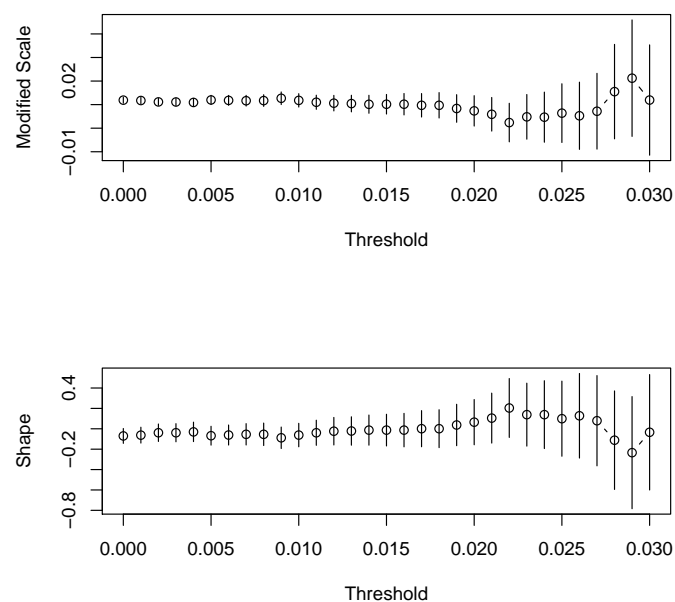


Abbildung 2.6: Threshold Choice Plot für 31 Thresholds im Intervall von  $[0, 0.03]$  mit Konfidenzintervallen

werden. Entscheidend ist hierbei die Anzahl der Werte, die über dem Threshold liegen, damit genügend Werte für die Schätzung einer Tail-Verteilung zur Verfügung stehen. Einen guten Anhaltspunkt bieten hierbei die Quantile der Daten. Ein guter Threshold sollte ungefähr zwischen dem 90%- und dem 95%-Quantil der Daten liegen. Die 90%- und 95%-Quantile der negativen Log-Renditen sind in Tabelle 2.1 abgebildet.

90%	95%
0.01698462	0.02369838

Tabelle 2.1: 90%- und 95%- Quantile der negativen Log-Renditen

	Threshold	Anzahl der Exzesse	Liegt der Threshold innerhalb der Quantile?
mrl-plot	0.0395	11	Nein, weit darüber.
	0.0218	85	Ja.
tcplot	0.0335	19	Nein, weit darüber.
	0.02	96	Ja.

Tabelle 2.2: Threshold-Vergleich

Tabelle 2.2 zeigt einen Vergleich der ausgewählten Thresholds, mit der Anzahl der Exzesse. Zwei der gewählten Thresholds (0.0218 und 0.02) passen genau in das von den Quantilen aufgespannte Intervall. Das sind auch die einzigen Thresholds, die eine angemessene Anzahl an Exzessen bieten. Die Thresholds 0.0395 und 0.0335 bietet mit nur 11 bzw. 19 Überschreitungen keine Alternative. Darüber hinaus liegen sie auch weit über dem von den Quantilen aufgespannten Intervall. Für die weiteren Berechnungen soll nun ein Threshold von 0.0218 benutzt werden.

## 2.2 Schätzen der Parameter

Nachdem ein Threshold gewählt wurde können mit R die restlichen Parameter der GPD geschätzt werden. R führt dazu eine Maximum-Likelihood-Schätzung durch und gibt folgende Werte aus:

Threshold:	0.0218	
Anzahl der Exzesse:	85	
Parameterschätzer:	$\xi$	0.227585836
	$\beta$	0.006636448
Standardabweichung der Schätzer:	$\xi$	0.149105745
	$\beta$	0.001144116
Varianz-Kovarianzmatrix:	0.0222325232	-1.179388e-04
	-0.0001179388	1.309001e-06

Wie schon im Theorieteil erwähnt, ist der gewählte Threshold automatisch der Loc-Parameter. Der geschätzte Shape-Parameter  $\xi$  beträgt 0.227585836 und der Scale-Parameter wurde auf 0.006636448 geschätzt. Darüber hinaus gibt R die Standardabweichungen der Schätzer an. Diese ist im Vergleich zu den Werten relativ hoch. Eine Abweichung von 0.149105745 bei einem Shape-Parameter von 0.227585836 würde aber nicht bewirken, dass  $\xi$  negativ oder 0 wird, d.h. die heavy tailed Form der angepassten Verteilung würde erhalten bleiben. Mit den drei geschätzten Parametern wurde schließlich eine GPD Verteilung an den Tail der Daten angepasst.

## 2.3 Überprüfung der Anpassung

### 2.3.1 Histogramm

Mit Hilfe eines Histogramms soll die Güte der Anpassung überprüft werden. Hierfür wird auf Grundlage der Datenbasis ein Histogramm berechnet. R teilt dafür den Datensatz in verschiedene Klassen auf. Um ein aussagekräftiges Histogramm zu erhalten ist es sinnvoll, die Klassen gleich groß zu wählen. R gibt dann eine Graphik aus, bei der auf der x-Achse die Klassen abgetragen werden und über jeder Klasse wird ein Rechteck gezeichnet, dessen Höhe die absolute Häufigkeit der Daten in dieser Klasse beschreibt. Siehe Abbildung 2.7 (links). Die blaue Kurve

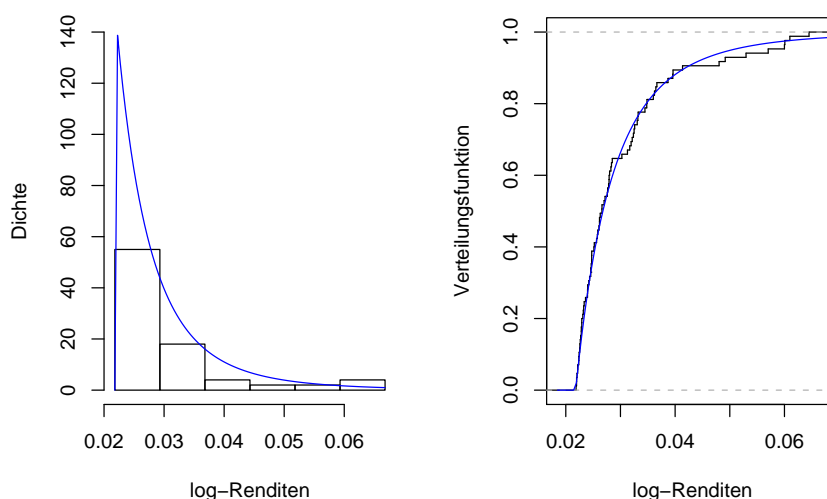


Abbildung 2.7: Histogramm der negativen Log-Renditen mit der Dichte der geschätzten Verteilung (links) und empirische Verteilungsfunktion der Daten mit der geschätzten Verteilungsfunktion (rechts).

in der Abbildung zeigt zum Vergleich die Dichte der geschätzten GPD-Verteilung. Hier ist zu sehen, dass die Verteilung schon recht gut an die Daten angepasst ist. Die sehr hohen Werte werden zwar nach wie vor leicht unterschätzt, aber insgesamt passt sich die Kurve ziemlich gut dem Datensatz an. Abbildung 2.7 (rechts) zeigt die empirische Verteilungsfunktion

$$\hat{F}_{n,x}(t) := \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, t]}(x_i),$$

wobei  $x = (x_1, \dots, x_n)$  der Datensatz der negativen Log-Renditen ist, zusammen mit der geschätzten Verteilungsfunktion. Auch in dieser Abbildung ist zu erkennen, dass die beiden Kurven gut aufeinander liegen, d.h. die Anpassung der Verteilungsfunktion ist gut gelungen.

### 2.3.2 Kolmogorov-Smirnov-Test

Eine weitere Möglichkeit, die Güte der Schätzer zu überprüfen, bietet der Kolmogorov-Smirnov-Test, der auf dem Vergleich der empirischen Verteilungsfunktion mit der geschätzten Verteilungsfunktion basiert. Dieser gibt im Fall der vorgenommenen Anpassung einen p-Wert von 0.9629 aus, d.h. zu praktisch allen üblichen Niveaus kann die Nullhypothese der Verteilungsgleichheit nicht abgelehnt werden. Das ist ein weiteres Indiz dafür, dass die Anpassung gut gelungen ist.

### 2.3.3 Vergleich der ersten Momente, der Varianz und der Schiefe

Schließlich soll die Anpassung rechnerisch mithilfe der ersten Momente, der Varianz und der Schiefe nachgeprüft werden. Dazu wird mit R jeweils das erste Moment, die Varianz und die Schiefe des Datensatzes und der angepassten Verteilung berechnet. Die Ergebnisse sind in Tabelle 2.3 abgebildet. Das arithmetische Mittel

	Datensatz	Verteilung
Mittel	0.03028352	0.0303918259
Varianz	0.0000994638	0.0001354914
Schiefe	1.854026	5.712416

Tabelle 2.3: Mittel, Varianz und Schiefe des Datensatzes und der angepassten Verteilung

des Datensatzes und das erste Moment der Verteilung weichen erst bei der vierten Nachkommastelle voneinander ab. Ebenso ähnlich sind sich die Varianz der Daten und die Varianz der Verteilung, welche ebenfalls erst ab der vierten Nachkommastelle voneinander abweichen. Lediglich die Schiefe der Verteilung weicht deutlich

von der Schiefe des Datensatzes ab. Beide Werte sind allerdings größer als Null, d.h. die angepasste Verteilung und die Daten sind wie erwartet rechtsschief.

Insgesamt hat das Histogramm, die empirische Verteilungsfunktion, der Kolmogorov-Smirnov-Test und die Berechnungen der ersten Momente, der Varianz und der Schiefe gezeigt, dass die GPD Verteilung mit den geschätzten Parametern eine gute Anpassung an die Daten liefert.

## 2.4 Value at Risk und Expected Shortfall

Mithilfe der geschätzten Parameter kann jetzt der Value at Risk

$$\widehat{VaR}_q = u + \frac{\hat{\beta}}{\hat{\xi}} \left( \left( \frac{n}{k} (1 - q) \right)^{-\hat{\xi}} - 1 \right)$$

und der Expected Shortfall

$$\widehat{ES}_q = \frac{\widehat{VaR}_q}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1 - \hat{\xi}}$$

berechnet werden. Tabelle 2.4 zeigt die Ergebnisse. Es ist zu sehen, dass der Value

	Value at Risk	Expected Shortfall
95%	0.02387964	0.03308421
99%	0.03769910	0.05097547
99.5%	0.04539856	0.06094352
99.9%	0.06873728	0.09115881
99.99%	0.12115548	0.15902162

Tabelle 2.4: VaR und ES für verschiedene Wahrscheinlichkeiten

at Risk und der Expected Shortfall stark vom Konfidenzintervall abhängen. Der Value at Risk für ein Konfidenzintervall von 99% und eine Haltedauer von z.B. 10 Tagen, kann so interpretiert werden, dass innerhalb der nächsten 10 Tage mit einer Wahrscheinlichkeit von einem Prozent eine negative Log-Rendite von mehr als 0.03769910 zu erwarten ist. Der erwartete Verlust, gegeben, dass der VaR überschritten wurde entspricht in diesem Fall 0.05097547. Im Vergleich dazu gibt der VaR für ein Konfidenzintervall von 99.99% an, dass innerhalb der nächsten 10 Tage mit einer Wahrscheinlichkeit von 0.01% eine negative Log-Rendite von mehr als 0.12115548 zu erwarten ist. Das entspräche einem Kursverlust um mehr als 12%. Dieser Wert ist deutlich höher als der  $VaR_{0.99}$ , er wird aber auch mit einer deutlich geringeren Wahrscheinlichkeit erwartet. Die größte negative Log-Rendite des Datensatzes beträgt im Vergleich dazu 0.06449672, d.h. der 99.99% VaR und auch der 99.9% VaR wurde in dem betrachteten Zeitraum nicht überschritten.

Da die größte negative Log-Rendite aber größer ist als 0.04539856, folgt dass der  $VaR_{0.995}$  in dem betrachteten Zeitraum mindestens einmal überschritten wurde. Bei Betrachtung der zehn größten Werte des Datensatzes, ist festzustellen dass der 99.5% VaR zwischen 1996 und 2000 insgesamt achtmal überschritten wurde.

Zum Vergleich kann noch der Value at Risk und der Expected Shortfall für die verschiedenen Wahrscheinlichkeiten unter der Annahme einer Normalverteilung der Daten berechnet werden. Dazu wird mit Hilfe des Erwartungswertes  $\mu_n = -0.0008242146$  und der Standardabweichung  $\sigma_n = 0.01436555$  der negativen Log-Renditen eine Normalverteilung an die Daten angepasst. Der Vergleich ist in Tabelle 2.5 abgebildet. Es ist deutlich zu erkennen, dass der Value at Risk und der Expected Shortfall bei Annahme einer verallgemeinerten Paretoverteilung größer ist als bei Annahme einer Normalverteilung. Der Unterschied wird mit wachsendem Konfidenzintervall größer. Dies zeigt, dass der Value at Risk unter der Annahme einer Normalverteilung das Risiko großer Verluste unterschätzt. Wenn z.B. der 99% Value at Risk und einen Zeitraum von 100 Handelstagen betrachtet werden, gibt der  $VaR_{0.99}$  unter Annahme einer Normalverteilung an, dass an einem von 100 Handelstagen ein relativer Verlust (negative Log-Rendite) von mehr als 0.03259505 zu erwarten ist. Der  $VaR_{0.99}$  unter Annahme einer GPD, gibt im Vergleich dazu an, dass an einem von 100 Handelstagen ein relativer Verlust von mehr als 0.03769910 zu erwarten ist. D.h. bei gleicher Wahrscheinlichkeit, wird mit dem GPD-VaR ein höherer Verlust geschätzt.

	Value at Risk		Expected Shortfall	
	GPD	Normalverteilung	GPD	Normalverteilung
95%	0.02387964	0.02280501	0.03308421	0.02880779
99%	0.03769910	0.03259505	0.05097547	0.03746305
99.5%	0.04539856	0.03617899	0.06094352	0.04072021
99.9%	0.06873728	0.04356867	0.09115881	0.04754588
99.99%	0.12115548	0.05260150	0.15902162	0.05604152

Tabelle 2.5: Vergleich: Value at Risk und Expected Shortfall für die angepasste GPD und die angepasste Normalverteilung

### 2.4.1 Tailplot

Abbildung 2.8 zeigt den sogenannten Tail-Plot zur ML-Schätzung. Die Punkte in der Abbildung sind die 85 Thresholdüberschreitungen und die Kurve durch die Punkte ist der Tailschätzer

$$\hat{F}(x) = \frac{k}{n} \left( 1 + \hat{\xi} \frac{x - u}{\hat{\beta}} \right)^{-1/\hat{\xi}},$$

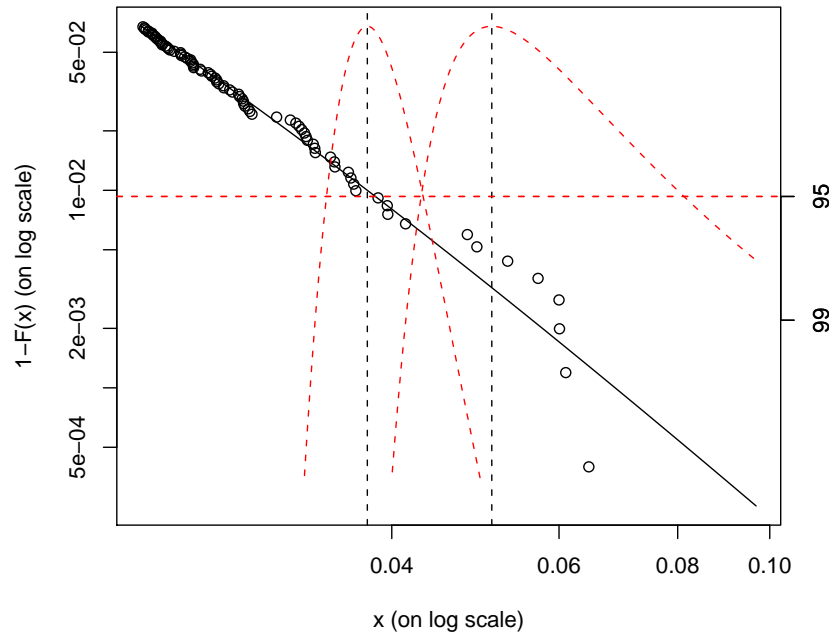


Abbildung 2.8: Tailplot der negativen Log-Renditen

wobei  $k$  wieder die Anzahl der Thresholdüberschreitungen bezeichnet. Auch hier ist zu sehen, dass sich der Schätzer gut an die Daten anschmiegt. Die erste gestrichelte, vertikale Linie markiert das 99%-Quantil der verallgemeinerten Paretoverteilung. Die linke gestrichelte Kurve ist die Profile Likelihood Kurve für den VaR als Parameter des GPD-Modells. Sie wird konstruiert durch Nutzung des Likelihood Ratio Tests. Für weitere Informationen zur Profile Likelihood Funktion und zum Likelihood Ratio Test siehe z.B. [4]. Die horizontale, gestrichelte Linie dient der Anschauung eines 95% Konfidenzintervalls für die Schätzer. Die Schnittpunkte der Profile Likelihood Kurve mit dieser Linie sind die Grenzen eines 95%-Intervalls für den Schätzer. Im Fall der vorliegenden Daten liegt der 99%-Quantilschätzer bei 0.03769910 und die Schnittpunkte der Profile Likelihood Kurve mit der horizontalen Linie liegen bei 0.03418402 und 0.04307676, d.h. der 99%-VaR liegt mit einer Wahrscheinlichkeit von 95% im Intervall  $[0.03418402, 0.04307676]$ . Die zweite vertikale, gestrichelte Linie markiert den Punktschätzer des 99% Expected Shortfalls. Dieser liegt bei 0.05097547. Die Schnittpunkte der Profile Likelihood Kurve des ES mit der horizontalen Linie liegen bei 0.04318442 und 0.08107574, d.h. der 99% Expected Shortfall liegt mit einer Wahrscheinlichkeit von 95% im Intervall  $[0.04318442, 0.08107574]$ .



### 2.4.2 Vergleich GPD und Normalverteilung

Im Folgenden soll die angepasste verallgemeinerte Paretoverteilung noch einmal mit einer angepassten Normalverteilung verglichen werden.

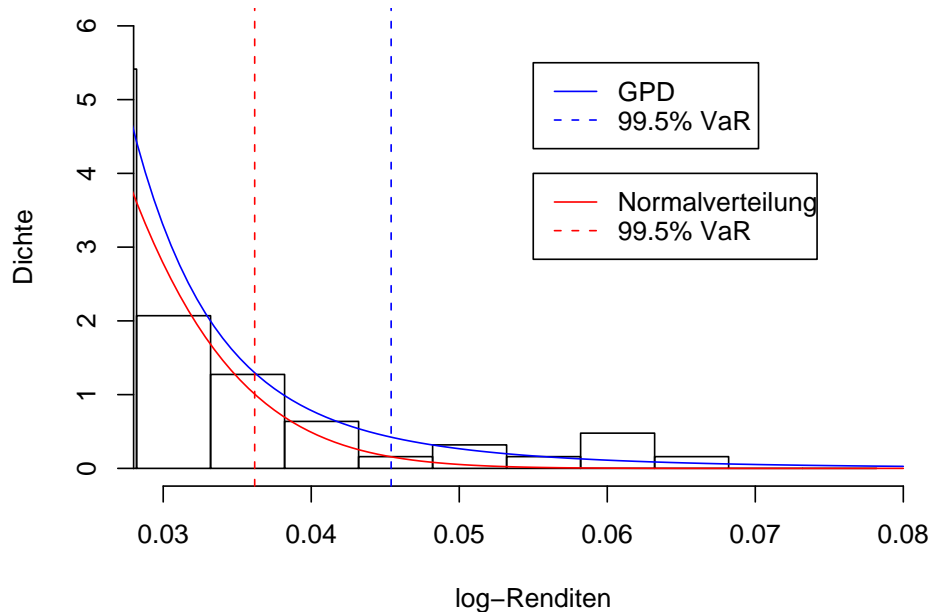


Abbildung 2.9: Angepasste GPD und Normalverteilung und 99.5% Value at Risk der Verteilungen.

Abbildung 2.9 zeigt ein Histogramm der Thresholdüberschreitungen mit den Dichten der angepassten verallgemeinerten Paretoverteilung  $G_{\hat{\xi}, \hat{\mu}, \hat{\beta}}$  (blaue Kurve) und der in Abschnitt 2.4 angepassten Normalverteilung  $\mathcal{N}_{\mu_n, \sigma_n}$  (rote Kurve). Die gestrichelten vertikalen Linien markieren jeweils den 99.5% Value at Risk der Verteilungen. In der Abbildung ist wieder zu erkennen, dass die Normalverteilung die hohen Werte deutlich unterschätzt und somit ist auch der VaR unter Annahme einer Normalverteilung zu niedrig. Die GPD beschreibt die hohen Werte (oder in diesem Beispiel die niedrigen Werte, denn es werden negative Log-Renditen betrachtet) der Daten deutlich besser und da die Anpassung besser ist, ist das Ergebnis ein „verlässlicherer“ Value at Risk. Zusammenfassend kann festgehalten werden, dass eine verallgemeinerte Paretoverteilung die großen Verluste des DAX deutlich besser modelliert als eine Normalverteilung. Der „interessante“ Teil der Daten ist durch die hier vorgestellte Methode gut modelliert worden.

Der Vollständigkeit halber soll nun noch eine Verteilung an den kompletten Daten-

satz angepasst werden. Hierfür wird, wie schon in Abschnitt 1.1 erwähnt, am besten eine Konvexkombination aus einer GPD und einer Normalverteilung verwendet. Dafür muss eine Normalverteilung an den linken und mittleren Teil der Daten, also alle Daten die kleiner sind als 0.0218, angepasst werden. Der Erwartungswert dieser Daten beträgt  $\mu_m = -0.003082248$  und die Varianz  $\sigma_m = 0.01178198$ .

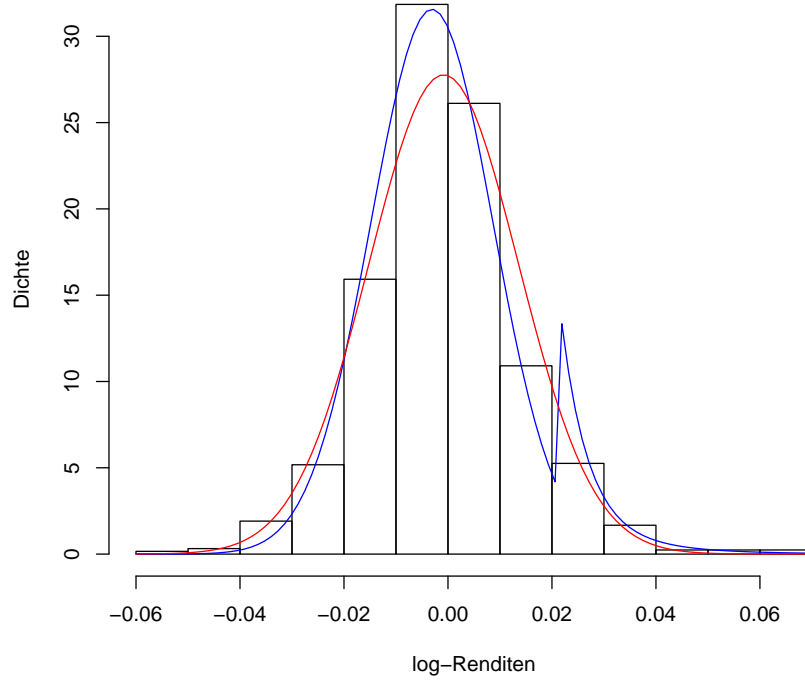


Abbildung 2.10: Histogramm der Daten mit der Dichte einer Konvexkombination aus GPD und Normalverteilung (blau) und mit einer angepassten Normalverteilung (rot).

Abbildung 2.10 zeigt einen Ausschnitt aus dem Histogramm der gesamten Daten mit der Dichte der Konvexkombination

$$(1 - d)\mathcal{N}_{\mu_m, \sigma_m} + dG_{\hat{\xi}, \hat{\mu}, \hat{\beta}}$$

(blaue Kurve), wobei  $\mathcal{N}_{\mu_m, \sigma_m}$  die angepasste Normalverteilung für den linken und mittleren Teil der Daten und  $G_{\hat{\xi}, \hat{\mu}, \hat{\beta}}$  die angepasste GPD mit den geschätzten Parametern ist.  $d$  ist dabei die relative Anzahl der Thresholdüberschreitungen. Sie ergibt sich aus

$$d = \frac{85}{1256} = 0.06767516.$$

Die rote Kurve in Abbildung 2.10 zeigt zum Vergleich die Dichte der, an den gesamten Datensatz angepassten Normalverteilung  $\mathcal{N}_{\mu_n, \sigma_n}$ . In der Abbildungen ist zu sehen, dass die Konvexkombination die Daten insgesamt besser beschreibt. Vor allem die sehr großen Werte werden bei dieser Verteilung noch gut abgeschätzt, während sie durch die Normalverteilung, wie schon in Abbildung 2.9 zu erkennen war, unterschätzt werden. Aber auch die mittleren Werte der Daten werden durch die Konvexkombination besser abgebildet. Die Normalverteilung unterschätzt die Daten dort leicht. Lediglich im linken Tail, also für die sehr kleinen Werte, liefert die Normalverteilung eine bessere Anpassung als die Konvexkombination, welche die Daten dort leicht unterschätzt. Um hier eine bessere Anpassung der Konvexkombination zu erhalten, könnte das komplette POT-Verfahren noch einmal auf die kleinsten Werte der Daten (die eigentlich die größten sind, da negative Log-Renditen betrachtet werden) angewendet werden. Da der linke Tail in diesem Beispiel die größten Gewinne darstellt, welche für gewöhnlich nicht abgesichert werden müssen, wird hier nicht weiter darauf eingegangen. Insgesamt zeigt das Histogramm, dass eine leptokurtische Verteilung mit thin waists und heavy tails eine bessere Anpassung an die Daten liefert, als eine Normalverteilung.

Damit wurde nun an den Datensatz der negativen Log-Renditen der DAX Abschlusswerte vom 01.01.1996 bis zum 31.12.00 eine verallgemeinerte GPD-Verteilung mit den Parametern

- $\xi = 0.227585836$
- $\mu = 0.0218$
- $\beta = 0.006636448$

angepasst. Die Überprüfung der Anpassung mit Hilfe eines Histogramms, des Kolmogorov-Smirnov-Tests und dem Vergleich des Erwartungswertes, der Varianz und der Schiefe der Daten und der geschätzten Verteilung haben ergeben, dass die Anpassung gut gelungen ist. Hier müssen allerdings Abstriche gemacht werden, da in diesem ersten Teil unabhängige und identisch verteilte Zufallsvariablen also in diesem Beispiel unabhängige, identisch verteilte negative Log-Renditen vorausgesetzt wurden. Es ist aber bekannt, dass Log-Renditen von Aktienkursen nicht unbedingt als unabhängig angesehen werden können.

Im Vergleich zur Normalverteilung haben die graphischen Mittel gezeigt, dass eine verallgemeinerte Paretoverteilung eine deutlich bessere Anpassung an die DAX Log-Renditen liefert. Der mit den Parametern der GPD geschätzte Value at Risk ist im Vergleich zum Value at Risk unter Annahme einer Normalverteilung deutlich konservativer.



## 3 POT-Methode für stationäre Zeitreihen

Wie im ersten Teil dieser Arbeit gesehen, bietet die POT-Methode eine sehr gute Möglichkeit um Verluste von Renditezeitreihen abzuschätzen. Obwohl die Ergebnisse für den iid Fall schon recht gut waren, gibt es noch Verbesserungsmöglichkeiten. Ein häufiger Kritikpunkt ist die Annahme, dass die Daten unabhängig voneinander sind. Denn bei der genauen Betrachtung von Finanzzeitreihen ist festzustellen, dass große Veränderungen dazu tendieren von großen Veränderungen gefolgt zu werden und kleine Schwankungen werden häufig von kleinen Schwankungen gefolgt. Es sind also Phasen hoher Volatilität und Phasen niedriger Volatilität zu erkennen, was im Widerspruch zur iid Annahme steht. Eine bessere Modellierung der Zeitreihe erhält man durch die Verwendung von Modellen, in denen sich die Varianz ständig verändert. Diese Modelle werden auch heteroskedastisch genannt. Sie stehen im Gegensatz zu den homoskedastischen Modellen, bei denen die Varianz konstant ist. Die wichtigsten Modelle für Renditezeitreihen sind die so genannten AutoRegressive-Conditional-Heteroskedasticity (kurz ARCH) und die generalized ARCH (kurz GARCH) Prozesse. Diese sollen im Folgenden kurz mit einigen Eigenschaften vorgestellt werden. Zunächst folgt dazu eine kurze Einführung in die Grundlagen der Zeitreihenanalyse, in der vor allem einige für die Modellierung von ARCH- und GARCH-Prozessen wichtige Begriffe, wie Stationarität und weißes Rauschen, definiert werden. Danach wird eine Methode vorgestellt, mit der der Value at Risk und der Expected Shortfall einer stationären Zeitreihe, mit Hilfe eines GARCH-Prozesses, berechnet werden kann.

### 3.1 Grundlagen der Zeitreihenanalyse

#### 3.1.1 Grundbegriffe

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum. Ein *stochastischer Prozess*  $(X_t)_{t \in I}$  ist eine Familie von Zufallsvariablen definiert auf diesem Wahrscheinlichkeitsraum. *Zeitreihen* sind spezielle stochastische Prozesse. Sie laufen in diskreter Zeit ab.  $I$  ist in diesem Fall eine höchstens abzählbare Menge. Für festes  $\omega \in \Omega$  heißt

$(Y(\omega)_t)_{t \in I}$  *Pfad* oder *Realisation* des stochastischen Prozesses  $Y$ . Im Folgenden sei  $I = \mathbb{Z}$  und  $(X_t)$  stehe für den stochastischen Prozess  $(X_t)_{t \in \mathbb{Z}}$ .

Die *Erwartungswertfunktion* sei definiert durch

$$\mu(t) = E(X_t), \quad t \in \mathbb{Z}$$

und

$$\sigma_t^2 = \sigma^2(t) = \text{Var}(X_t), \quad t \in \mathbb{Z}$$

definiere, falls sie existiert, die *Varianzfunktion* von  $(X_t)_{t \in \mathbb{Z}}$ . Darüber hinaus sei

$$\gamma(t, s) = E((X_t - \mu(t))(X_s - \mu(s))), \quad t, s \in \mathbb{Z}$$

die *Autokovarianzfunktion* von  $(X_t)_{t \in \mathbb{Z}}$ .

Die Erwartungswertfunktion gibt als Funktion der Zeit  $t$  an, welcher Wert des stochastischen Prozesses im Zeitpunkt  $t$  zu erwarten ist. Dabei kann  $E(X_t)$  als durchschnittliche Rendite im Zeitpunkt  $t$  über verschiedene mögliche historische Verläufe aufgefasst werden.

## Stationäre Zeitreihen

Eine wichtige Eigenschaft für das Schätzen von stochastischen Prozessen ist die Stationarität. Dabei wird zwischen strikter und schwacher Stationarität unterschieden<sup>1</sup>:

### Definition 7

Eine Zeitreihe  $(X_t)_{t \in \mathbb{Z}}$  ist strikt stationär, falls

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+k}, \dots, X_{t_n+k})$$

für alle  $t_1, \dots, t_n, k \in \mathbb{Z}$  und alle  $n \in \mathbb{N}$  gilt.

### Definition 8

Eine Zeitreihe  $(X_t)_{t \in \mathbb{Z}}$  ist schwach stationär, wenn die ersten beiden Momente existieren und den Bedingungen

$$\begin{aligned} \mu(t) &= \mu, & t &\in \mathbb{Z}, \\ \gamma(t, s) &= \gamma(t+k, s+k), & t, s, k &\in \mathbb{Z} \end{aligned}$$

genügen.

Stationarität bedeutet also, dass sich die „Struktur“ der Zeitreihe im Zeitablauf nicht verändert. Systematische Änderungen des Erwartungswertes, der Varianz

---

<sup>1</sup>vgl. [17] (S.126)

oder der Kovarianz zwischen äquidistanten Beobachtungen sind widersprüchlich zur Stationarität, d.h. falls sie definiert sind gilt  $E(X_t) = \mu(t) = \mu$  und  $Var(X_t) = \sigma_t^2 = \sigma^2$  für alle  $t \in T$  und alle Kovarianzen sind invariant in Bezug auf die Verschiebung um  $s$  Zeiteinheiten.

Anschaulich wird Stationarität häufig so erklärt, dass, wenn man ein Fenster geeigneter Breite über die Graphik einer stationären Zeitreihe gleiten lässt, sich das Bild nicht qualitativ verändert<sup>2</sup>. Diese Definition ist allerdings sehr vage, da nicht definiert ist, was ein Fenster geeigneter Breite ist, und wann eine qualitative Veränderung vorliegt.

Es kann leicht nachgeprüft werden, dass eine strikt stationäre Zeitreihe mit endlicher Varianz schwach stationär ist. Da die Varianz einer Zeitreihe aber nicht immer endlich ist, gibt es auch strikt stationäre Prozesse, die nicht schwach stationär sind<sup>3</sup>.

Aus Definition 8 folgt, dass die Kovarianz von  $X_t$  und  $X_s$  nur von der Zeitdifferenz  $|s - t|$  abhängt, denn für alle  $s, t$  gilt

$$\gamma(t - s, 0) = \gamma(t, s) = \gamma(s, t) = \gamma(s - t, 0).$$

Die Zeitdifferenz  $|s - t|$  wird auch *Lag* genannt.

Für einen schwach stationären Prozess kann die Autokovarianzfunktion als Funktion in einer Variablen geschrieben werden:

$$\gamma(h) := \gamma(h, 0), \quad \forall h \in \mathbb{Z},$$

dabei sei  $\gamma(0) = Var(X_t), \forall t \in \mathbb{Z}$ .

Mit dieser Definition kann die *Autokorrelationsfunktion*  $\rho(h)$  eines schwach stationären Prozesses  $(X_t)_{t \in \mathbb{Z}}$  definiert werden:

$$\rho(h) = \rho(X_h, X_0) = \frac{\gamma(h)}{\gamma(0)}, \quad \forall h \in \mathbb{Z}.$$

$\rho(h)$  heißt dann *Autokorrelation zum Lag  $h$* .

## Weißes Rauschen

Eine weitere Definition, die für die Einführung von ARCH und GARCH Prozessen benötigt wird, ist die des weißen Rauschens<sup>4</sup>:

---

<sup>2</sup>vgl. [21] (S115)

<sup>3</sup>vgl. [21] (S.115)

<sup>4</sup>vgl. [17] (Definition 4.4 und 4.5 S.127)

**Definition 9**

$(X_t)_{t \in \mathbb{Z}}$  heißt *weißes Rauschen* (kurz: *WN*, für *white noise*), wenn  $(X_t)_{t \in \mathbb{Z}}$  schwach stationär ist mit Autokorrelationsfunktion

$$\rho(h) = \begin{cases} 1, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

**Definition 10**

$(X_t)_{t \in \mathbb{Z}}$  heißt *striktes weißes Rauschen* (kurz: *SWN*, für *strict white noise*), wenn  $(X_t)_{t \in \mathbb{Z}}$  eine Reihe von unabhängig identisch verteilten Zufallsvariablen mit endlicher Varianz ist.

Striktes weißes Rauschen ist damit eine deutlich stärkere Bedingung, da hier Unabhängigkeit und nicht nur Unkorreliertheit der Zufallsvariablen gefordert wird. Bei Finanzzeitreihen modelliert weißes Rauschen den Teil, der als zufällig angesehen wird. Im Folgenden steht  $(X_t) \sim WN(0, \sigma^2)$  dafür, dass der stochastische Prozess  $(X_t)$  weißes Rauschen mit Erwartungswert Null und Varianz  $\sigma^2$  ist, und  $(X_t) \sim SWN(0, \sigma^2)$  bedeutet entsprechend, dass  $(X_t)$  striktes weißes Rauschen mit Erwartungswert Null und Varianz  $\sigma^2$  ist.

**3.1.2 AR- und MA-Prozesse**

Als nächstes sollen die sogenannten *autoregressiven* und *Moving-Average* Prozesse eingeführt werden. Diese sind relativ einfache Prozesse, die wegen ihrer Linearität für die Beschreibung von Renditen in den meisten Fällen nicht so gut geeignet sind, aber trotzdem einen wichtigen Teil der Kapitalmarktforschung abdecken.

**MA-Prozesse****Definition 11**

Ein stochastischer Prozess  $(X_t)_{t \in \mathbb{Z}}$  heißt *Moving-Average-Prozess der Ordnung  $q$*  (kurz: *MA( $q$ )-Prozess*), wenn er sich in der Form

$$X_t = \sum_{j=0}^q \theta_j \epsilon_{t-j}, \quad t \in \mathbb{Z}$$

mit  $\theta_q \neq 0$  und  $\theta_0 = 1$  darstellen lässt.  $(\epsilon_t)_{t \in \mathbb{Z}}$  ist dabei weißes Rauschen<sup>5</sup>.

Es kann leicht gezeigt werden, dass MA( $q$ )-Prozesse immer schwach stationär sind<sup>6</sup>. Da MA-Prozesse für das spätere Vorgehen nicht weiter benötigt werden, seien sie hier nur der Vollständigkeit halber kurz erwähnt.

<sup>5</sup>vgl. [21] (Definition 4.17 S.129)

<sup>6</sup>siehe z.B. [21] (S.130ff)



## AR-Prozesse

Eine weitere Art von Prozessen sind die autoregressiven Prozesse<sup>7</sup>:

### Definition 12

Ein stochastischer Prozess  $(X_t)_{t \in \mathbb{Z}}$  heißt autoregressiver Prozess der Ordnung  $p$  (kurz:  $AR(p)$ -Prozess), wenn es weißes Rauschen  $(\epsilon_t)_{t \in \mathbb{Z}}$  und reelle Zahlen  $\phi_0, \dots, \phi_p$  mit  $\phi_0 = 1$  und  $\phi_p \neq 0$  gibt, so dass

$$\sum_{j=0}^p \phi_j X_{t-j} = \epsilon_t, \quad t \in \mathbb{Z}$$

gilt.

Bei einer  $AR(p)$ -Zeitreihe hängt  $X_t$  also linear von den  $p$  vorangehenden Werten  $X_{t-1}, X_{t-2}, \dots, X_{t-p}$  und einer zufälligen Störung  $\epsilon_t$  ab.

$AR(p)$ -Prozesse sind etwas komplizierter als  $MA(q)$ -Prozesse, da sie nur unter bestimmten Bedingungen stationär sind. Es kann gezeigt werden, dass ein  $AR(p)$ -Prozess stationär ist, wenn alle Wurzeln der charakteristischen Gleichung außerhalb des Einheitskreises liegen, d.h. betragsmäßig größer als Eins sind<sup>8</sup>. Der Einfachheit halber, und da für die späteren Anwendungen nur ein  $AR(1)$ -Prozess benötigt wird, werden hier nur einige Eigenschaften von  $AR(1)$ -Prozessen aufgeführt.

## Stationarität des $AR(1)$ -Prozesses

Der  $AR(1)$ -Prozess hat die Darstellung

$$X_t = \phi X_{t-1} + \epsilon_t, \quad \forall t. \quad (3.1)$$

Es kann leicht gezeigt werden, dass für  $|\phi| \geq 1$  kein schwach stationärer Prozess  $X_t$  existiert, der die Differenzengleichung (3.1) erfüllt. Für  $|\phi| < 1$  existiert aber ein eindeutiger, schwach stationärer Prozess, der (3.1) erfüllt. In diesem Fall gilt

$$\begin{aligned} X_t &= \phi X_{t-1} + \epsilon_t \\ &= \phi(\phi X_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi(\phi(\phi X_{t-3} + \epsilon_{t-2}) + \epsilon_{t-1}) + \epsilon_t \\ &\vdots \\ &= \phi^{k+1} X_{t-k-1} + \sum_{i=0}^k \phi^i \epsilon_{t-i}. \end{aligned}$$

<sup>7</sup>vgl. [21] (Definition 4.19 S.131)

<sup>8</sup>siehe z.B. [21]

Dann kann gezeigt werden, dass die Bedingung  $|\phi| < 1$  sicherstellt, dass der erste Term für  $k \rightarrow \infty$  verschwindet und der zweite Term konvergiert. Der Prozess

$$X_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$$

ist die einzige Lösung der Gleichung (3.1) und es ist leicht zu zeigen, dass diese schwach stationär ist<sup>9</sup>.

## ARMA-Prozesse

Eine Kombination aus AR- und MA-Prozessen sind die ARMA-Prozesse<sup>10</sup>:

### Definition 13

Sei  $(\epsilon_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma_\epsilon^2)$ . Der stochastische Prozess  $(X_t)_{t \in \mathbb{Z}}$  heißt autoregressiver Moving-Average-Prozess der Ordnung  $p$  und  $q$  (kurz: ARMA( $p, q$ )-Prozess) mit Erwartungswert Null, wenn er schwach stationär ist und folgende Darstellung besitzt

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \quad \forall t \in \mathbb{Z}.$$

$(X_t)$  heißt ARMA-Prozess mit Erwartungswert  $\mu$ , wenn der zentrierte Prozess  $(X_t - \mu)_{t \in \mathbb{Z}}$  ein ARMA-Prozess mit Erwartungswert Null ist.

Ob der ARMA-Prozess strikt stationär ist oder nicht, hängt nach dieser Definition nur von den  $\epsilon_t$  ab. Diese werden häufig auch als *Innovationen* bezeichnet.

## Invertierbarkeit

Unter bestimmten Bedingungen können die hier vorgestellten stochastischen Prozesse ineinander überführt werden. Diese Eigenschaft wird häufig *Invertierbarkeit* genannt. Es kann gezeigt werden, dass ein AR(1)-Prozess mit  $|\phi| < 1$  als ein MA( $\infty$ )-Prozess dargestellt werden kann und allgemein, dass ein AR( $p$ )-Prozess genau dann in einen MA( $\infty$ )-Prozess überführt werden kann, wenn  $(X_t)$  stationär ist. Umgekehrt kann ein MA( $q$ )-Prozess als AR( $\infty$ )-Prozess dargestellt werden, falls alle Nullstellen des charakteristischen Polynoms betragsmäßig größer sind als Eins. ARMA( $p, q$ )-Prozesse lassen sich ebenfalls in AR( $\infty$ )- oder MA( $\infty$ )-Prozesse überführen<sup>11</sup>.

<sup>9</sup>vgl. [17] (S.129ff)

<sup>10</sup>vgl. [17] (Definition 4.7 S.128)

<sup>11</sup>vgl. [21] (S.135ff)

### 3.1.3 ARCH- und GARCH-Prozesse

Schließlich werden die AutoRegressive-Conditional-Heteroskedasticity-Prozesse (kurz: ARCH-Prozesse) vorgestellt. Diese wurden 1982 von Engle<sup>12</sup> eingeführt und bilden eines der wichtigsten Modelle um tägliche Renditezeitreihen zu schätzen. Darüber hinaus bieten sie eine Möglichkeit, die bedingten Varianzen  $Var(X_t | X_{t-1})$  von  $X_{t-1}$  abhängig zu machen.

#### ARCH-Prozesse

##### Definition 14

Sei  $(Z_t)_{t \in \mathbb{Z}} \sim SWN(0, 1)$ . Ein Prozess  $(X_t)_{t \in \mathbb{Z}}$  heißt  $ARCH(p)$ -Prozess, wenn er strikt stationär ist und für alle  $t \in \mathbb{Z}$  und einen Prozess  $(\sigma_t)_{t \in \mathbb{Z}}$  mit strikt positiven Werten die Gleichungen

$$X_t = \sigma_t Z_t, \quad (3.2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 \quad (3.3)$$

erfüllt, wobei  $\alpha_0 > 0$  und  $\alpha_i \geq 0$  für  $i = 1, \dots, p$  gilt<sup>13</sup>.

Um ARCH-Prozesse genauer zu betrachten, empfiehlt es sich, zunächst die Eigenschaften eines  $ARCH(1)$ -Prozesses zu beobachten. Die Resultate können später leicht auf den allgemeinen Fall erweitert werden. Der  $ARCH(1)$ -Prozess ist von der Form

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2, \end{aligned}$$

wobei  $\alpha_0 > 0$  und  $\alpha_1 \geq 0$ .

#### Stationarität des $ARCH(1)$ -Prozesses

##### Strikte Stationarität

Für die Analyse von Renditezeitreihen spielt die Stationarität eine wichtige Rolle, daher ist es sinnvoll Bedingungen für die Stationarität eines  $ARCH(1)$ -Prozesses zu ermitteln. McNeil, Frey und Embrechts<sup>14</sup> verwenden dafür sogenannte stochastische Rekurrenzgleichungen. Sie betrachteten den quadrierten  $ARCH(1)$ -Prozess  $X_t^2 = \sigma_t^2 Z_t^2$  mit  $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2$ . Dieser Prozess ist von der Form

$$X_t^2 = \alpha_0 Z_t^2 + \alpha_1 Z_t^2 X_{t-1}^2. \quad (3.4)$$

<sup>12</sup>[7]

<sup>13</sup>vgl. [17] (Definition 4.16 S.139)

<sup>14</sup>vgl. [17] (S.141ff)

Diese Gleichung ist eine *stochastische Rekurrenzgleichung*. Eine detaillierte Analyse stochastischer Rekurrenzgleichungen findet der interessierte Leser z.B. in Embrechts u.a.<sup>15</sup>. Da die stochastischen Rekurrenzgleichungen hier nur als Hilfsmittel dienen um Bedingungen für die Stationarität von ARCH-Prozessen zu erhalten, werden hier nur die nötigsten Eigenschaften erwähnt. Gleichung (3.4) ist ein spezielles Beispiel einer Klasse von Rekurrenzgleichungen der Form

$$Y_t = A_t Y_{t-1} + B_t, \quad (3.5)$$

wobei  $(A_t)_{t \in \mathbb{Z}}$  und  $(B_t)_{t \in \mathbb{Z}}$  Folgen von unabhängig und identisch verteilten Zufallsvariablen sind. Embrechts u.a.<sup>16</sup> haben gezeigt, dass

$$E(\ln^+ | B_t |) < \infty \quad \text{und} \quad E(\ln | A_t |) < 0, \quad (3.6)$$

wobei  $\ln^+ x = \max(0, \ln x)$ , hinreichende Bedingungen für eine Lösung der Gleichung sind. Die einzige Lösung ist

$$Y_t = B_t + \sum_{i=1}^{\infty} B_{t-i} \prod_{j=0}^{i-1} A_{t-j}, \quad (3.7)$$

wobei die Summe fast sicher, absolut konvergiert. Zu diesem Ergebnis gelangt man durch k-fache Iteration der Gleichung (3.5)

$$\begin{aligned} Y_t &= A_t(A_{t-1}Y_{t-2} + B_{t-1}) + B_t \\ &= B_t + \sum_{i=1}^k B_{t-i} \prod_{j=0}^{i-1} A_{t-j} + Y_{t-k-1} \prod_{i=0}^k A_{t-i}. \end{aligned}$$

Die Bedingungen (3.6) versichern, dass der mittlere Term auf der rechten Seite absolut konvergiert und der Endterm verschwindet, da das starke Gesetz der großen Zahlen zusammen mit der Bedingung  $E(\ln | A_t |) < 0$  ergibt, dass

$$\frac{1}{k+1} \sum_{i=0}^k \ln | A_{t-i} | \xrightarrow{f.s.} E(\ln | A_t |) < 0.$$

Daraus folgt

$$\prod_{i=0}^k | A_{t-i} | = \exp \left( \sum_{i=0}^k \ln | A_{t-i} | \right) \xrightarrow{f.s.} 0.$$

Die Lösung (3.7) ist ein strikt stationärer Prozess und die Bedingung  $E(\ln | A_t |) < 0$  entwickelt sich damit zu einem entscheidenden Faktor für die strikte Stationarität von ARCH-Prozessen.

Der quadrierte ARCH(1)-Prozess (3.4) ist eine stochastische Rekurrenzgleichung der Form (3.5) mit  $A_t = \alpha_1 Z_t^2$  und  $B_t = \alpha_0 Z_t^2$ . Die Bedingungen (3.6) verwandeln

---

<sup>15</sup>[6] (S.455ff)

<sup>16</sup>[6]

sich damit zu  $E(\ln^+ |\alpha_0 Z_t^2|) < \infty$ , was durch die Definition des ARCH(1)-Prozesses bereits erfüllt ist, und  $E(\ln(\alpha_1 Z_t^2)) < 0$ . Es kann also gezeigt werden, dass  $E(\ln(\alpha_1 Z_t^2)) < 0$  eine notwendige und hinreichende Bedingung für die strikte Stationarität eines ARCH(1)-Prozesses ist. Die strikte Stationarität des ARCH(1)-Prozesses hängt damit von der Verteilung der Innovationen  $Z_t$  ab. Die Lösung der Gleichung (3.4) hat die Form

$$X_t^2 = \alpha_0 \sum_{i=0}^{\infty} \alpha_1^i \prod_{j=0}^i Z_{t-j}^2.$$

### Schwache Stationarität

Damit wäre die Frage, wann ein ARCH(1)-Prozess strikt stationär ist, geklärt. Bleibt noch die Frage nach der schwachen Stationarität. McNeil, Frey und Embrechts<sup>17</sup> haben gezeigt, dass ein ARCH(1)-Prozess genau dann ein schwach stationäres weißes Rauschen ist, wenn  $\alpha_1 < 1$  gilt. In diesem Fall ist die unbedingte Varianz des Prozesses durch  $\sigma_t^2 = \frac{\alpha_0}{1-\alpha_1}$  gegeben. Dies wird im Folgenden kurz, zusammen mit einigen weiteren Eigenschaften des ARCH(1)-Prozesses, gezeigt.

### Weitere Eigenschaften des ARCH(1)-Prozesses

Für einen ARCH(1)-Prozess ist leicht zu zeigen, dass für  $t \in \mathbb{Z}$  gilt

$$\begin{aligned} E(X_t) &= 0, \\ E(X_t | \mathcal{F}_{t-1}) &= 0. \end{aligned}$$

Für die unbedingte Varianz gilt

$$\sigma_t^2 = \text{Var}(X_t) = E(X_t^2) = E(E(X_t^2 | \mathcal{F}_{t-1})) = E(\alpha_0 + \alpha_1 X_{t-1}^2) = \alpha_0 + \alpha_1 E(X_{t-1}^2).$$

Da  $X_t$  ein stationärer Prozess mit  $E(X_t) = 0$ ,  $\sigma_t^2 = \sigma_{t-1}^2 = E(X_{t-1}^2)$  ist, folgt  $\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_t^2$  und daraus folgt

$$\sigma_t^2 = \frac{\alpha_0}{1 - \alpha_1}.$$

Damit die Varianz von  $X_t$  positiv ist, muss  $0 \leq \alpha_1 < 1$  gelten<sup>18</sup>.

Häufig wird angenommen, dass die  $Z_t$  standardnormalverteilt oder standard-Student-t verteilt sind. Trotz dieser Annahme ist die Kurtosis eines ARCH(1)-Prozesses größer als die einer Normalverteilung. Das hat zur Folge, dass ein ARCH-Prozess eher „Ausreißer“ zeigt als ein gaußsches weißes Rauschen. Dies soll nun

<sup>17</sup>vgl. [17] (Proposition 4.18 S.142)

<sup>18</sup>vgl. [22] (S.104ff)

kurz gezeigt werden. Sei  $X_t$  ein stationärer ARCH(1)-Prozess mit standardnormalverteilten Innovationen  $Z_t$  und  $E(X_t^4) = m_4 < \infty$ . Dann gilt

$$\begin{aligned} m_4 &= E(X_t^4) = E(\sigma_t^4 Z_t^4) = E(\sigma_t^4) E(Z_t^4) = E(Z_t^4) E((\alpha_0 + \alpha_1 X_{t-1}^2)^2) \\ &= 3(\alpha_0^2 + 2\alpha_0\alpha_1 E(X_{t-1}^2) + \alpha_1^2 E(X_{t-1}^4)). \end{aligned}$$

Da  $E(X_{t-1}^2) = \frac{\alpha_0}{1-\alpha_1}$ , wie oben gesehen, und  $E(X_{t-1}^4) = m_4$ , folgt

$$m_4 = 3\left(\alpha_0^2 + 2\alpha_0\alpha_1 \frac{\alpha_0}{1-\alpha_1} + \alpha_1^2 m_4\right) = 3\alpha_0^2 \left(1 + 2\frac{\alpha_1}{1-\alpha_1}\right) + 3\alpha_1^2 m_4.$$

Auflösen der Gleichung ergibt schließlich

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}.$$

Damit das vierte Moment positiv ist, muss  $1 - 3\alpha_1^2 > 0$  gelten, d.h.  $0 \leq \alpha_1^2 < \frac{1}{3}$ . Mit diesem Ergebnis kann nun die Kurtosis berechnet werden:

$$\frac{E(X_t^4)}{(\sigma_t^2)^2} = 3 \frac{\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} * \frac{(1-\alpha_1)^2}{\alpha_0^2} = 3 \frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3.$$

Also ist die Tailverteilung von  $X_t$  heavier tailed als die einer Normalverteilung, obwohl angenommen wurde, dass die  $Z_t$  standardnormalverteilt sind. Diese Eigenschaft ist sehr nützlich für die Modellierung von Renditezeitreihen, da diese bekanntlich auch heavier tailed sind als eine Normalverteilung.

Eine weitere nützliche Eigenschaft von ARCH Prozessen ist, dass sie sogenannte Volatilitätsschwankungen modellieren, denn angenommen  $X_t$  sei ein schwach stationäres weißes Rauschen<sup>19</sup>, dann gilt  $E(X_t^2) < \infty$  und

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = E(\sigma_t^2 Z_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2 \text{Var}(Z_t) = \sigma_t^2.$$

Daraus folgt, dass die bedingte Varianz  $\sigma_t$ , auch *Volatilität* genannt, eine sich kontinuierlich verändernde Funktion der vorherigen quadrierten Werte des Prozesses ist. Wenn einer oder mehrere der  $|X_{t-1}|, \dots, |X_{t-p}|$  besonders groß ist, dann wird  $X_t$  gewissermaßen aus einer Verteilung mit großer Varianz gezogen und kann daher selbst groß sein. Auf diese Weise erzeugt das Modell sogenannte *Volatilitätscluster*<sup>20</sup>, d.h. Phasen hoher Volatilität und Phasen niedriger Volatilität. Dieses Phänomen ist bei „echten“ Renditezeitreihen häufig zu beobachten.

Der Name ARCH bezieht sich auf die Struktur des Prozesses. Das Modell ist autoregressiv, da  $X_t$  von den vorherigen  $X_{t-i}$  abhängt, und bedingt heteroskedastisch, da sich die bedingte Varianz kontinuierlich verändert<sup>21</sup>.

Die vorgestellten Eigenschaften gelten auch für allgemeine ARCH-Modelle, aber die Berechnungen werden komplizierter für größere  $p$ .

<sup>19</sup>was, wie zuvor gezeigt, für einen ARCH(1)-Prozess unter der Bedingung  $\alpha_1 < 1$  gilt

<sup>20</sup>vgl. [17] (S.139ff)

<sup>21</sup>vgl. [17] (S.139)

## GARCH-Prozesse

In einem ARCH-Modell kann die Anzahl der Parameter, die zu schätzen sind, leicht sehr groß werden. Daher wurde eine Erweiterung mit weniger Parametern aber dennoch guter Datenanpassung entwickelt. Bollerslev<sup>22</sup> führte 1986 die sogenannten GARCH-Prozesse (generalized ARCH-Prozesse) ein.

### Definition 15

Sei  $(Z_t)_{t \in \mathbb{Z}} \sim SWN(0, 1)$ . Der Prozess  $(X_t)_{t \in \mathbb{Z}}$  heißt  $GARCH(p, q)$ -Prozess, wenn er strikt stationär ist und für alle  $t \in \mathbb{Z}$  und einen strikt positiven Prozess  $(\sigma_t)_{t \in \mathbb{Z}}$  die Gleichungen

$$X_t = \sigma_t Z_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

erfüllt, wobei  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$  für  $i = 1, \dots, p$  und  $\beta_j \geq 0$  für  $j = 1, \dots, q$ .

Der GARCH-Prozess ist ein verallgemeinerter ARCH-Prozess in dem Sinne, dass die quadrierte Volatilität  $\sigma_t^2$  genauso von den vorherigen quadrierten Volatilitäten abhängen kann, wie von den vorherigen quadrierten Werten des Prozesses. Dabei ist ein  $GARCH(p, 0)$ -Prozess gerade ein  $ARCH(p)$ -Prozess.  $GARCH(1, 1)$ -Prozesse eignen sich besonders gut für die Modellierung von Aktienrenditezeitreihen, weil in diesen Modellen hohe Volatilitäten dazu tendieren beständig zu sein, da  $|X_t|$  groß ist, wenn  $|X_{t-1}|$  oder  $\sigma_{t-1}$  groß ist<sup>23</sup>.

Wie bei den ARCH-Prozessen ist es auch bei GARCH-Prozessen sinnvoll sich zunächst auf den einfachen Fall zu konzentrieren. Die Ergebnisse können dann leicht auf Prozesse höherer Ordnung erweitert werden. Darüber hinaus werden in den meisten Anwendungen ohnehin GARCH-Prozesse niedriger Ordnung verwendet.

## Stationarität des $GARCH(1, 1)$ -Prozesses

### Strikte Stationarität

Für einen  $GARCH(1, 1)$ -Prozess gilt nach Definition 15

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 Z_{t-1}^2 + \beta \sigma_{t-1}^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta) \sigma_{t-1}^2.$$

Dies ist wieder eine stochastische Rekurrenzgleichung der Form  $Y_t = A_t Y_{t-1} + B_t$ , allerdings in diesem Fall für  $Y_t = \sigma_t^2$  anstatt für  $X_t^2$ . Dabei sei

$$\underbrace{\sigma_t^2}_{=: Y_t} = \underbrace{(\alpha_1 Z_{t-1}^2 + \beta)}_{=: A_t} \underbrace{\sigma_{t-1}^2}_{=: Y_{t-1}} + \underbrace{\alpha_0}_{=: B_t}. \quad (3.8)$$

<sup>22</sup>[3]

<sup>23</sup>vgl. [17] (S.145)

Die Bedingung  $E(\ln |A_t|) < 0$  für die strikte Stationarität der stochastischen Rekurrenzgleichung wird durch Einsetzen von  $A_t$  zu  $E(\ln(\alpha_1 Z_{t-1}^2 + \beta)) < 0$ . Die einzige Lösung von (3.8) sieht in diesem Fall so aus

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{i=1}^{\infty} \prod_{j=1}^i (\alpha_1 Z_{t-j}^2 + \beta).$$

Wenn  $(\sigma_t^2)_{t \in \mathbb{Z}}$  ein strikt stationärer Prozess ist, ist  $(X_t)_{t \in \mathbb{Z}}$  ebenfalls strikt stationär, da  $X_t = \sigma_t Z_t$  gilt und  $(Z_t)_{t \in \mathbb{Z}}$  striktes weißes Rauschen ist. Die Lösung der GARCH(1,1) Definitionsgleichung ist damit

$$X_t = Z_t \sqrt{\alpha_0 \left( 1 + \sum_{i=1}^{\infty} \prod_{j=1}^i (\alpha_1 Z_{t-j}^2 + \beta) \right)}.$$

### Schwache Stationarität

Ähnlich wie für den ARCH(1)-Prozess kann gezeigt werden, dass der GARCH(1,1)-Prozess schwach stationäres weißes Rauschen ist, genau dann wenn  $\alpha_1 + \beta < 1$  gilt<sup>24</sup>.

### Weitere Eigenschaften des GARCH(1,1)-Prozesses

Wie beim ARCH(1)-Prozess ist der unbedingte Erwartungswert des GARCH(1,1)-Prozesses Null. Für die unbedingte Varianz gilt

$$\sigma_t^2 = \text{Var}(X_t) = E(X_t^2) = \frac{\alpha_0}{1 - \alpha_1 - \beta}.$$

Auch bei GARCH-Prozessen wird häufig angenommen, dass die Innovationen  $Z_t$  standardnormal- oder standard-Student-t-verteilt sind. Die Kurtosis eines GARCH-Prozesses ist aber wie beim ARCH-Prozess größer als die einer Normalverteilung. Dies zeigt man ähnlich wie beim ARCH-Prozess. Da die Berechnungen nicht schwierig, aber relativ lang sind, werden hier nur die Ergebnisse aufgezeigt. Für eine ausführliche Rechnung siehe z.B. [21] (S.175ff). Für einen schwach stationären GARCH(1,1)-Prozess  $X_t$  mit  $\sigma_t^2 < \infty$ ,  $Z_t \sim N(0, 1)$  und  $E(X_t^4) < \infty$  gilt

$$E(X_t^4) = E(Z_t^4)E(\sigma_t^4) = 3E((\alpha_0 + \alpha_1 X_{t-1}^2 + \beta \sigma_{t-1}^2)^2).$$

Durch einige Umformungen ergibt sich

$$E(X_t^4) = 3 \frac{\alpha_0^2(1 + \alpha_1 + \beta)}{(1 - 3\alpha_1^2 - 2\alpha_1\beta - \beta^2)(1 - \alpha_1 - \beta)}. \quad (3.9)$$

<sup>24</sup>vgl. [17] (Proposition 4.21 S.146)



Der Zähler in (3.9) ist positiv. Durch die Voraussetzung, dass der Prozess schwach stationär ist, ist  $1 - \alpha_1 - \beta$  ebenfalls positiv. Für die Existenz des vierten Momentes muss also gelten, dass

$$3\alpha_1^2 + \beta^2 + 2\alpha_1\beta < 1.$$

Die Kurtosis ergibt sich dann aus

$$\begin{aligned} \frac{E(X_t^4)}{(E(X_t^2))^2} &= 3 \frac{\alpha_0^2(1 + \alpha_1 + \beta)}{(1 - 3\alpha_1^2 - 2\alpha_1\beta - \beta^2)(1 - \alpha_1 - \beta)} \frac{(1 - \alpha_1 - \beta)^2}{\alpha_0^2} \\ &= \dots \\ &= 3 + \frac{6\alpha_1^2}{(1 - 3\alpha_1^2 - \beta^2 - 2\alpha_1\beta)} > 3. \end{aligned}$$

Damit wurden im Wesentlichen die für die Kapitalmarktforschung wichtigen, stochastischen Prozesse mit einigen Eigenschaften vorgestellt. Weitere Eigenschaften der vorgestellten Prozessen sind z.B. in [21] oder [17] zu finden.

## 3.2 POT-Methode

Nun kann die POT-Methode für stationäre Zeitreihen vorgestellt werden. Im Folgenden wird also nicht mehr angenommen, dass die Zeitreihe iid ist. Stattdessen werden stationäre Zeitreihen betrachtet. McNeil und Frey<sup>25</sup> stellten 2000 eine Methode vor, bei der die VaR Schätzungen für stationäre Zeitreihen mit Hilfe eines GARCH(1,1)- und eines AR(1)-Prozesses auf den iid Fall zurückgeführt werden können. Dieser Ansatz soll hier gezeigt werden und später in einem Anwendungsteil mit anderen Ansätzen verglichen werden.

Sei  $(X_t)_{t \in \mathbb{Z}}$  eine strikt stationäre Zeitreihe, die die Log-Renditen eines Aktienkurses darstellt. Es wird angenommen, dass  $X_t$  von der Form

$$X_t = \mu_t + \sigma_t Z_t$$

ist, wobei die Innovationen  $Z_t$  striktes weißes Rauschen mit Erwartungswert Null und Varianz Eins sind. Die Randverteilungsfunktion der  $Z_t$  wird mit  $F_Z(z)$  bezeichnet.  $\mathcal{G}_{t-1}$  sei die bis zur Zeit  $t - 1$  verfügbare Information und es wird angenommen, dass  $\mu_t$  und  $\sigma_t$  messbar bezüglich  $\mathcal{G}_{t-1}$  sind. Darüber hinaus sei  $F_X(x)$  die Randverteilung von  $X_t$  und  $F_{X_{t+1}+\dots+X_{t+k}|\mathcal{G}_t}(x)$  sei die  $k$ -Tages-Rendite, gegeben dem Wissen der Renditen bis zu und einschließlich dem Tag  $t$ .

Wie im iid Fall sollen wieder der Value at Risk und der Expected Shortfall berechnet werden. Weil die Log-Renditen in diesem Kapitel nicht mehr als iid angesehen werden, können die einzelnen Beobachtungen voneinander abhängen. Daher wird der bedingte Value at Risk und der bedingte Expected Shortfall betrachtet.

---

<sup>25</sup>[14]

**Definition 16**

a) Der (in Kapitel 1 vorgestellte) unbedingte Value at Risk der Randverteilung  $F_X(x)$  mit  $0 < q < 1$  wird im Folgenden bezeichnet mit

$$x_q = \inf \{x \in \mathbb{R} : F_X(x) \geq q\}.$$

b) Darüber hinaus wird der bedingte Value at Risk der voraussagenden Verteilung  $F_{X_{t+1}+\dots+X_{t+k}|\mathcal{G}_t}(x)$  definiert durch

$$x_q^t(k) = \inf \{x \in \mathbb{R} : F_{X_{t+1}+\dots+X_{t+k}|\mathcal{G}_t}(x) \geq q\}.$$

**Definition 17**

a) Der (in Kapitel 1 vorgestellte) unbedingte Expected Shortfall wird im Folgenden bezeichnet mit

$$S_q = E[X \mid X > x_q].$$

b) Der bedingte Expected Shortfall wird definiert durch

$$S_q^t(k) = E \left[ \sum_{j=1}^k X_{t+j} \mid \sum_{j=1}^k X_{t+j} > x_q^t(k), \mathcal{G}_t \right].$$

Im Folgenden sollen 1-Schritt Vorhersagen betrachtet werden, d.h. es gelte  $k = 1$ . Der bedingte Value at Risk hat dann die Gestalt

$$x_q^t = \inf \{x \in \mathbb{R} : F_{X_{t+1}|\mathcal{G}_t}(x) \geq q\}$$

und der bedingte Expected Shortfall wird zu

$$S_q^t = E[X_{t+1} \mid X_{t+1} > x_q^t, \mathcal{G}_t].$$

Für die Verteilung  $F_{X_{t+1}|\mathcal{G}_t}(x)$  kann leicht gezeigt werden, dass

$$F_{X_{t+1}|\mathcal{G}_t}(x) = F_Z \left( \frac{x - \mu_{t+1}}{\sigma_{t+1}} \right) \quad (3.10)$$

gilt. Mit (3.10) lassen sich  $x_q^t$  und  $S_q^t$  vereinfachen zu

$$x_q^t = \mu_{t+1} + \sigma_{t+1} z_q, \quad (3.11)$$

$$S_q^t = \mu_{t+1} + \sigma_{t+1} E[Z \mid Z > z_q], \quad (3.12)$$

wobei

$$z_q = \inf \{z \in \mathbb{R} : F_Z(z) \geq q\}$$

der Value at Risk der Randverteilung von  $Z_t$  ist.

Um den Value at Risk und den Expected Shortfall für stationäre Zeitreihen zu schätzen, müssen also die Parameter  $\mu_{t+1}$ ,  $\sigma_{t+1}$ ,  $z_q$  und  $E[Z \mid Z > z_q]$  geschätzt werden. In vielen Ansätzen wird dabei unterstellt, dass die Innovationen  $Z_t$  standardnormalverteilt sind, so dass  $z_q = \Phi^{-1}(q)$  ist, wobei  $\Phi(z)$  die Standardnormalverteilung bezeichnet. McNeil und Frey<sup>26</sup> haben allerdings gezeigt, dass dieser Ansatz die bedingten Quantile unterschätzt, da die Verteilung der Innovationen in den meisten Fällen heavier tailed sind als die Normalverteilung. Sie schlugen stattdessen einen alternativen Ansatz vor, den sie bedingte Extremwerttheorie nannten. Diese Methode sieht vor, mit Hilfe der Pseudo-Maximum-Likelihood Methode einen AR(1)-Prozess mit GARCH(1,1)-Fehlern an die Renditezeitreihe anzupassen. Dabei sollte das Modell keine Annahmen über die Verteilung  $F_Z(z)$  machen. Nach der Anpassung werden mit Hilfe der Schätzer des AR(1)-GARCH(1,1)-Prozesses die Innovationen  $z_t$  geschätzt. Diese werden auch *Residuen* genannt. Wenn die Anpassung gelungen ist sollten die Residuen iid sein. Die 1-Schritt Vorhersagen  $\mu_{t+1}$  und  $\sigma_{t+1}$  können dann rekursiv berechnet werden. In einem zweiten Schritt kann die POT-Methode für iid Zeitreihen auf die Residuen angewendet werden um eine GPD-Verteilung anzupassen und den VaR und den ES der Residuen zu berechnen. Den Value at Risk und den Expected Shortfall der stationären Zeitreihe erhält man dann durch Zusammensetzen der geschätzten Parameter. Dieses Verfahren wird im Folgenden ausführlich beschrieben.

### 3.2.1 Schätzen von $\mu_{t+1}$ und $\sigma_{t+1}$

Die Log-Renditen  $(x_{t-n+1}, \dots, x_{t-1}, x_t)$  werden nun als Realisation eines AR(1)-GARCH(1,1)-Prozess aufgefasst. Der Einfachheit halber wird ein konstanter Datenspeicher  $n$  festgelegt, so dass die vorliegenden Daten am Tag  $t$  aus den letzten  $n$  negativen Log-Renditen  $(x_{t-n+1}, \dots, x_{t-1}, x_t)$  bestehen. Darüber hinaus wird die Mittelwert-bereinigte Folge

$$\epsilon_t = X_t - \mu_t \quad (3.13)$$

betrachtet. Es wird also ein AR(1)-Prozess der Form

$$X_t = \mu_t + \epsilon_t \quad (3.14)$$

betrachtet, wobei

$$\epsilon_t = \sigma_t Z_t \quad (3.15)$$

ein GARCH(1,1)-Prozess mit

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (3.16)$$

für  $\alpha_0 > 0$ ,  $\alpha_1 > 0$  und  $\beta > 0$  ist und für  $\mu_t$  gilt

$$\mu_t = \phi X_{t-1}. \quad (3.17)$$

---

<sup>26</sup>[14]

## Stationarität

Zunächst müssen Bedingungen für die Stationarität der Anpassung aufgestellt werden. Da in diesem Kapitel angenommen wird, dass  $X_t$  eine strikt stationäre Zeitreihe ist, muss auch die GARCH-Anpassung strikt stationär sein, d.h. sie muss der Bedingung

$$E(\log(\alpha_1 Z_{t-1}^2 + \beta)) < 0$$

genügen. Weil für die Parameterschätzung durch die Verwendung der Pseudo-Maximum-Likelihood-Methode indirekt angenommen wird, dass die  $Z_t$  standardnormalverteilt sind, reicht es, in diesem Fall die Bedingung für schwache Stationarität eines GARCH-Prozesses zu zeigen, denn da die Logarithmusfunktion konvex ist und  $Z_t$  integrierbar ist, kann die Jensensche Ungleichung für Erwartungswerte angewendet werden. Es gilt

$$0 > E(\log(\alpha_1 Z_{t-1}^2 + \beta)) \geq \log(E(\alpha_1 Z_{t-1}^2 + \beta)) = \log(\alpha_1 E Z_{t-1}^2 + \beta) = \log(\alpha_1 + \beta),$$

wobei die zweite Ungleichheit aus der Jensen-Ungleichung folgt und die letzte Gleichheit aus der Annahme, dass  $Z_t$  standardnormalverteilt ist. Da die Logarithmusfunktion  $\log(x)$  nur kleiner als Null ist, wenn  $x$  kleiner als Eins ist, ist

$$\alpha_1 + \beta < 1$$

eine hinreichende Bedingung für die strikte Stationarität der Anpassung.

## Pseudo-Maximum-Likelihood Methode

Da die Verteilung der Innovationen unbekannt ist, wird für die Schätzung der Parameter des AR(1)-GARCH(1,1)-Prozesses die Pseudo-Maximum-Likelihood-Methode (kurz: PML-Methode) verwendet. Das bedeutet, für die Maximum-Likelihood-Schätzung wird angenommen, dass die  $Z_t$  standardnormalverteilt sind, obwohl diese Annahme für die betrachteten Daten unrealistisch ist. Es kann aber gezeigt werden, dass die PML-Methode trotz dieser Annahme einen konsistenten, asymptotisch normalen Schätzer liefert<sup>27</sup>.

Es wird angenommen, dass die Stationaritätsbedingung  $\alpha_1 + \beta < 1$  erfüllt ist und dass  $E\epsilon_t^4 < \infty$  existiert. Da die  $Z_t$ , nach Annahme, iid und standardnormalverteilt sind, ist die Verteilung von  $X_t$  bedingt auf  $\mathcal{F}_{t-1}$  eine Normalverteilung mit Erwartungswert  $\mu_t = \phi X_{t-1}$  und Varianz  $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$ , d.h. es gilt

$$f(X_t | \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma_t^2}\right),$$

---

<sup>27</sup>vgl. [9]

wobei  $\epsilon_t = X_t - \phi X_{t-1}$  ist.

Für den GARCH(1,1)-Prozess gilt dann

$$L(\phi, \alpha_0, \alpha_1, \beta \mid x) = f_{X_0}(x_0) \prod_{t=1}^n f(x_t \mid \mathcal{F}_{t-1}).$$

Dabei wird der erste Wert als deterministischer Startwert angesehen. Wie bei der Maximum-Likelihood-Schätzung üblich, wird die Log-Likelihood-Funktion betrachtet. Dabei kann  $f_{X_0}(x_0)$  als feste Konstante bei der Optimierung vernachlässigt werden. Es ergibt sich

$$\log L(\phi, \alpha_0, \alpha_1, \beta \mid x) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log \sigma_t^2 - \frac{1}{2} \sum_{t=1}^T \frac{\epsilon_t^2}{\sigma_t^2}, \quad (3.18)$$

wobei  $\epsilon_t = x_t - \phi x_{t-1}$  gilt. Den Maximum-Likelihood-Schätzer  $\hat{\theta} = (\hat{\phi}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta})^T$  erhält man schließlich durch die Maximierung von (3.18) über den zulässigen Parameterraum.

Mit dem PML-Schätzer  $\hat{\theta}$  können aus (3.16) und (3.17) rekursiv Schätzer für den bedingten Mittelwert

$$(\hat{\mu}_{t-n+1}, \dots, \hat{\mu}_t)$$

und die bedingte Varianz

$$(\hat{\sigma}_{t-n+1}, \dots, \hat{\sigma}_t)$$

berechnet werden. Weil  $\sigma_t$  von  $\sigma_{t-1}$  abhängt, muss allerdings zunächst ein Startwert für  $\sigma_t$  gewählt werden. Häufig wird dafür die Stichprobenvarianz, oder Null gewählt.

### Residuen $z_t$

Als nächstes werden die Residuen  $z_t$  berechnet. Sie werden benötigt um die GARCH-Modellierung zu überprüfen. Wenn das angepasste Modell haltbar ist, sollten die  $z_t$  iid sein. Darüber hinaus werden sie später benötigt, um Schätzer für  $z_q$  und  $E[Z \mid Z > z_q]$  zu erhalten. Da

$$X_t = \mu_t + \sigma_t Z_t$$

gilt, folgt für die Residuen

$$(z_{t-n+1}, \dots, z_t) = \left( \frac{x_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \dots, \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} \right).$$

Nachdem die Residuen berechnet wurden, kann z.B mit einem Autokorrelationsfunktions-Plot oder einem Ljung-Box-Test<sup>28</sup> kontrolliert werden, ob die Residuen iid sind. Darüber hinaus muss die Stationaritätsbedingung  $\alpha_1 + \beta < 1$  überprüft werden. Ist das Ergebnis zufriedenstellend, können die 1-Schritt Vorhersagen  $\hat{\mu}_{t+1}$  und  $\hat{\sigma}_{t+1}$  berechnet werden. Ansonsten muss die AR(1)-GARCH(1,1)-Anpassung wiederholt werden bis man akzeptable Schätzer erhält.

$\hat{\mu}_{t+1}$  und  $\hat{\sigma}_{t+1}$  ergeben sich schließlich aus

$$\begin{aligned}\hat{\mu}_{t+1} &= \hat{\phi}x_t, \\ \hat{\sigma}_{t+1}^2 &= \hat{\alpha}_0 + \hat{\alpha}_1\hat{\epsilon}_t^2 + \hat{\beta}\hat{\sigma}_t^2,\end{aligned}$$

wobei  $\hat{\epsilon}_t = x_t - \hat{\mu}_t$ .

Damit wären  $\mu_{t+1}$  und  $\sigma_{t+1}$  geschätzt. Es fehlen noch Schätzer für  $z_q$  und  $E[Z | Z > z_q]$  um den bedingten VaR und den bedingten ES zu ermitteln. Da die  $Z_t$  aber iid sind, kann die POT-Methode für iid Renditefolgen auf sie angewendet und eine GPD-Verteilung angepasst werden.

### 3.2.2 Berechnung des bedingten VaR und des bedingten ES

Nachdem die Parameter des GARCH-Prozesses geschätzt und damit Schätzer für  $\mu_{t+1}$  und  $\sigma_{t+1}$  ermittelt wurden, wird in einem zweiten Schritt die POT-Methode für iid Renditefolgen, auf die Residuen  $Z_t$  angewendet um Schätzer für den Value at Risk und den Expected Shortfall der Residuen zu erhalten. Dabei wird, wie in Kapitel 1 beschrieben, mit Hilfe einer Maximum-Likelihood-Schätzung eine GPD-Verteilung an die hohen Werte der Residuen angepasst und mit dieser Verteilung der Value at Risk und der Expected Shortfall bestimmt.

Wenn schließlich Schätzer für den VaR und den ES der Residuen bestimmt worden sind, können die geschätzten Parameter zusammengeführt werden um die Schätzer für den bedingten VaR und den bedingten ES der stationären Renditezeitreihe zu erhalten

$$\begin{aligned}x_q^t &= \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}z_q, \\ S_q^t &= \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}E[Z | Z > z_q].\end{aligned}$$

Damit wurde gezeigt, wie mit Hilfe der POT-Methode Schätzer für den bedingten VaR und den bedingten ES einer stationären Zeitreihe berechnet werden können. Im nächsten Kapitel wird diese Methode auf die täglichen Abschlusswerte des DAX von 1996 bis 2001 angewendet. Die Ergebnisse werden später mit den Ergebnissen der unbedingten Extremwerttheorie aus Kapitel 2 verglichen. Darüber hinaus wird der Ansatz standardnormalverteilter Innovationen bei stationären Zeitreihen,

---

<sup>28</sup>siehe Anwendung, Kapitel 4

---

der in diesem Kapitel kurz angesprochen wurde, ebenfalls auf die Abschlusswerte angewendet und mit den anderen Ergebnissen verglichen.





## 4 Anwendung der POT-Methode für stationäre Zeitreihen

In diesem Anwendungsteil wird die POT-Methode für stationäre Zeitreihen auf den Datensatz aus Kapitel 2, also die täglichen DAX-Abschlusswerte von 1996 bis 2001, angewendet. Wie schon am Anfang von Kapitel 3 erwähnt wurde, ist die Annahme, dass tägliche DAX-Abschlusswerte unabhängig voneinander sind, nicht sehr realistisch. Betrachtet man die Werte genauer, ist festzustellen, dass hohe Werte eher auftreten, wenn der Kurs vorher auch schon relativ hoch war und ebenso treten niedrige Werte häufiger auf, wenn der Kurs insgesamt sehr niedrig ist. Diese sogenannten Volatilitätscluster sind bei näherer Betrachtung der negativen Log-Renditen der DAX-Abschlusswerte auch in Abbildung 4.1 zu erkennen. In der Zeit von 1996 bis Mitte 1997 ist z.B. eine deutlich geringere Volatilität als in der Zeit um 1999 zu erkennen. Realistischer ist es also in diesem Fall von stationären Zeitreihen auszugehen.

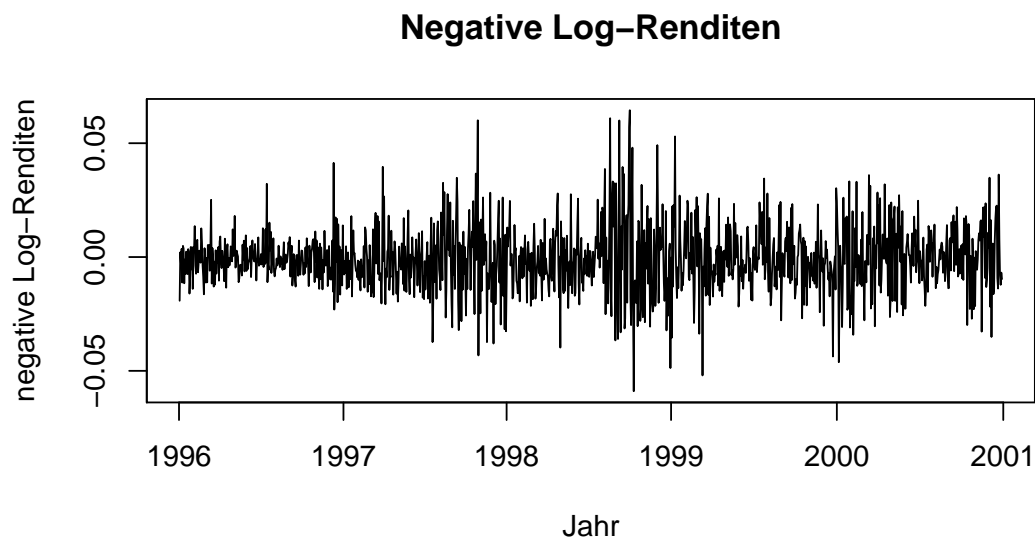


Abbildung 4.1: Negative Log-Renditen der DAX-Abschlusswerte vom 01.01.1996 bis zum 31.12.2000.

## 4.1 Anpassung eines AR(1)-GARCH(1,1)-Prozesses

Zunächst wird, wie im Theorieteil 3.2 beschrieben, ein AR(1)-GARCH(1,1)-Prozess an die Log-Renditen angepasst. R führt dazu eine Pseudo-Maximum-Likelihood Schätzung durch und gibt folgende Parameter aus:

	Schätzer	Standard-abweichung	t-Wert	Signifikanzcode
$\phi$	0.01494	0.03021	0.495	0.6208
$\alpha_0$	0.000002398	0.0000009827	2.440	0.0147 *
$\alpha_1$	0.09199	0.01689	5.448	0.000000051 ***
$\beta$	0.9	0.01743	51.634	$< 2 * 10^{-16}$ ***

Tabelle 4.1: Ergebnisse der Pseudo-Maximum-Likelihood Schätzung

Die Standardabweichung des geschätzten  $\phi$ -Parameters ist im Vergleich zum Schätzer relativ hoch. Auch der Signifikanzcode ist sehr hoch. Das bedeutet, dass dieser Parameter nicht sehr signifikant in das Modell eingeht. Um die Signifikanz zu verbessern, wird häufig geraten Prozesse niedrigerer Ordnung an die Daten anzupassen. Da aber ohnehin nur ein AR-Prozess 1.Ordnung gewählt wurde und für dieses Beispiel AR(1)-GARCH(1,1)-Prozesse betrachtet werden sollen, kann auch mit diesem Parameter weiter gerechnet werden.<sup>1</sup> Darüber hinaus sind  $\alpha_1$  und  $\beta$  sehr gut geschätzt. Sie haben eine verhältnismäßig kleine Standardabweichung und erhalten beim Signifikanzcode die beste Bewertung (drei Sterne).

Mithilfe der Schätzer für  $\phi$ ,  $\alpha_0$ ,  $\alpha_1$  und  $\beta$ , die man durch die Pseudo-Maximum-Likelihood-Schätzung erhalten hat, können rekursiv die Schätzer für  $(\sigma_{t-n+1}, \dots, \sigma_t)$  und  $(\mu_{t-n+1}, \dots, \mu_t)$  durch

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \\ \mu_t &= \phi X_{t-1}\end{aligned}$$

berechnet werden. Als Startwert für  $\sigma_t$  wurde in diesem Fall Null gewählt. Man erhält also den Vektor  $(\hat{\mu}_{t-n+1}, \dots, \hat{\mu}_t)$  als Schätzer für den bedingten Mittelwert und  $(\hat{\sigma}_{t-n+1}, \dots, \hat{\sigma}_t)$  als Schätzer für die bedingte Varianz. Wobei beide Vektoren in diesem Beispiel eine Länge von 1256 Werten haben.

<sup>1</sup>In Kapitel 5 wird später eine alternative Anwendung gezeigt, bei der auf den AR(1)-Prozess verzichtet wird.

## 4.2 Residuen

Als nächstes werden die Residuen berechnet. Diese erhält man, wie in Kapitel 3.2.1 beschrieben, durch

$$(z_{t-n+1}, \dots, z_t) = \left( \frac{x_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \dots, \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} \right).$$

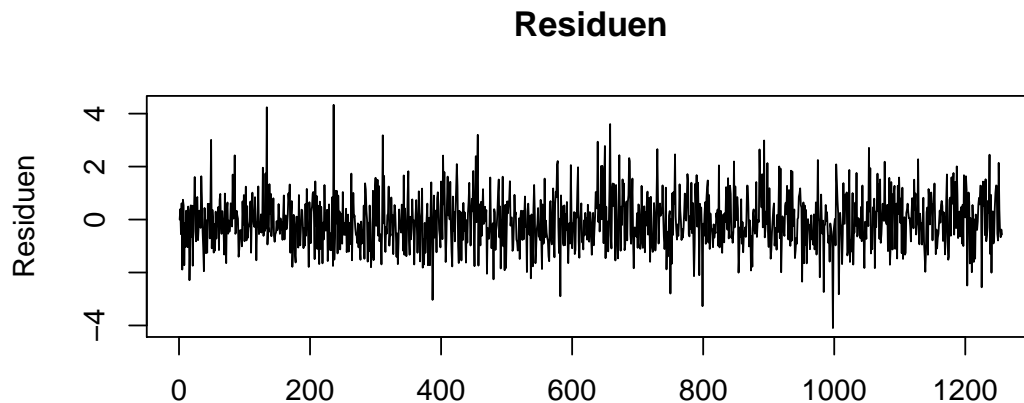


Abbildung 4.2: Residuenplot

Abbildung 4.2 zeigt einen Plot der geschätzten Residuen. Im Vergleich zum Plot der negativen Log-Renditen (Abbildung 4.1) ist die Struktur der Residuen „einheitlicher“. Die Schwankungen über die Zeit sind, bis auf einige „Ausreißer“ ungefähr gleich, wohingegen die Log-Renditen Abschnitte mit kleinen Schwankungen und Abschnitte mit höheren Schwankungen zeigen.

### 4.2.1 Überprüfung der Unabhängigkeit der Residuen

Wichtig ist, dass die Residuen unabhängig voneinander sind. Die Überprüfung der Unabhängigkeit ist nämlich eine Überprüfung der Qualität der Anpassung. Wenn die Residuen als annähernd unabhängig angesehen werden können, ist die Anpassung gut gelungen.

#### acf-plot

Um die Unabhängigkeit zu überprüfen, wird ein sogenannter *autocorrelation-function-plot* (kurz: acf-plot) der Residuen erstellt. Ein acf-plot plottet die empirische

Autokorrelationsfunktion

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)},$$

wobei  $\hat{\gamma}(0) > 0$  und

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x}_n)(x_t - \bar{x}_n)$$

die empirische Autokovarianzfunktion ist, mit

$$\bar{x}_n = \frac{1}{n} \sum_{t=1}^n x_t.$$

Die Autokorrelation sollte relativ nah bei Null liegen. Daher gibt R zusätzlich ein Konfidenzintervall  $(-1.96/\sqrt{n}, 1.96/\sqrt{n})$  um Null aus (blaue gestrichelte Linien). Wenn die Autokorrelationen, ab einem möglichst kleinen Wert, in dem vorgegebenen Intervall liegen, kann von unkorrelierten Daten ausgegangen werden. Zusätzlich zum Datensatz kann dabei noch eine maximale Zeitverschiebung  $h$  angegeben werden. Standardmäßig gibt R eine Lag-Länge von 30 aus. Eine weitere häufig verwendete Lag-Länge ist  $2\sqrt{n}$ , wobei  $n$  die Länge des Datensatzes ist. Abbildung 4.3 zeigt den acf-plot der Log-Renditen (oben) für die Lag-Längen 30 und  $2\sqrt{1256} = 70.88018$  im Vergleich zum acf-plot der Residuen  $z_t$  (unten) für die gleichen Lag-Längen. In diesem Vergleich ist zu erkennen, dass die Residuen eher in dem gekennzeichnetem Intervall liegen, als die Log-Renditen.

Mit diesem acf-plot kann aber nur die Unkorreliertheit der Daten überprüft werden. Um festzustellen, ob die Daten als unabhängig voneinander angesehen werden können, wird der acf-plot der quadrierten Werte betrachtet. Abbildung 4.4 zeigt die Ergebnisse wieder für die Lag-Längen 30 und  $2\sqrt{1256} = 70.88018$ . Es ist zu beobachten, dass die Log-Renditen für beide Lag-Längen eindeutig nicht als unabhängig angesehen werden können. Für die Residuen hingegen liegt die Autokorrelation sehr nah bei Null. Es kann also aufgrund des acf-Plots von annähernd unabhängigen Residuen ausgegangen werden.

### Ljung-Box-Test

Zusätzlich zum acf-plot kann mit einem Ljung-Box-Test die Unabhängigkeitsannahme überprüft werden. Dazu wird ein Ljung-Box-Test auf die Residuen angewendet. Mit ihm lässt sich signifikant nachweisen, dass eine Beobachtung nicht von einer iid Folge stammt. Dazu wird die Teststatistik

$$Q = n(n+2) \sum_{h=1}^L \frac{1}{n-h} \hat{\rho}(h)^2$$

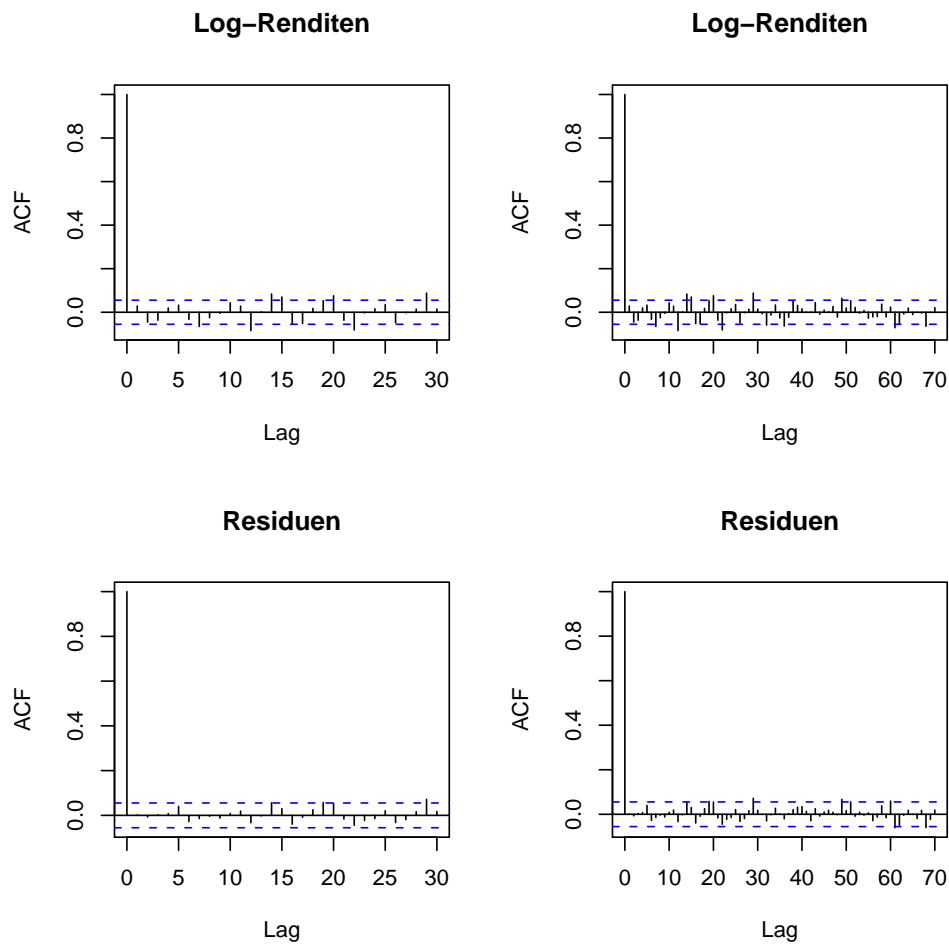


Abbildung 4.3: acf-plot der Log-Renditen (oben) mit Lag=30 (links) und Lag=70.88018 (rechts) im Vergleich zum acf-plot der Residuen (unten) mit Lag=30 (links) und Lag=70.88018 (rechts)

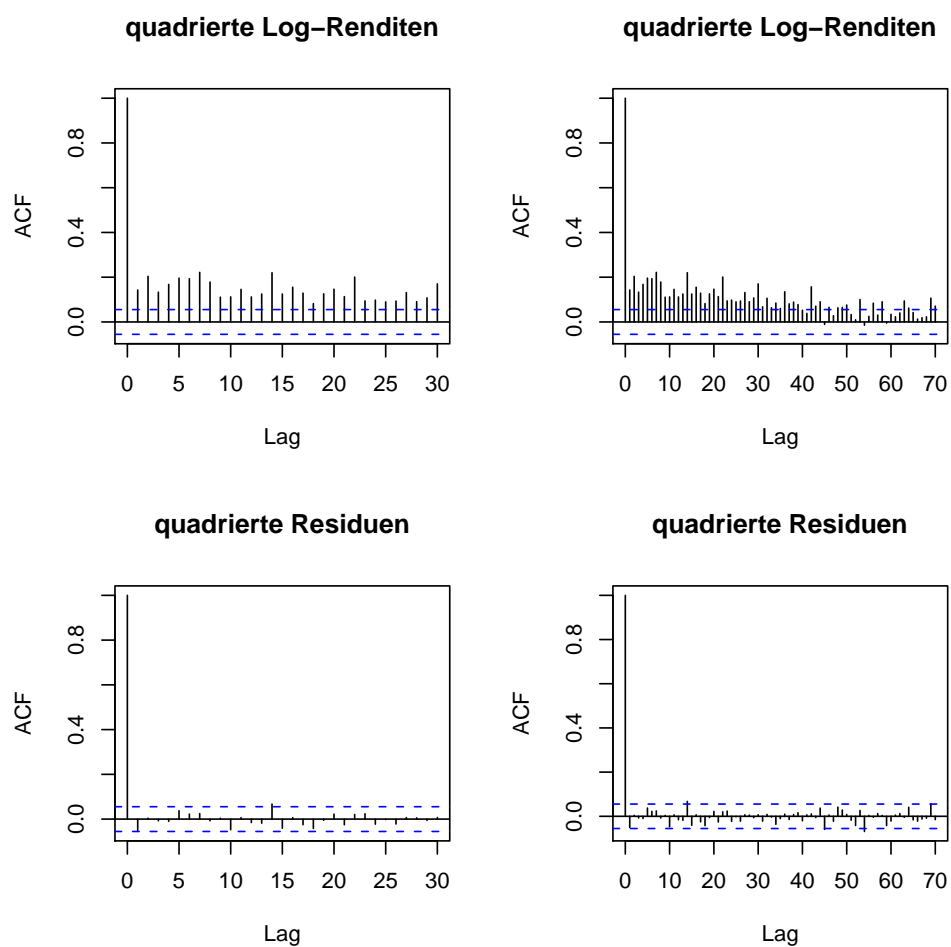


Abbildung 4.4: acf-plot der quadrierten Log-Renditen (oben) mit Lag=30 (links) und Lag=70.88018 (rechts) im Vergleich zum acf-plot der quadrierten Residuen (unten) mit Lag=30 (links) und Lag=70.88018 (rechts)

Lag-Länge	p-Wert
50	0.4669
$2\sqrt{(1256)}$	0.2648
100	0.2293
150	0.2347
200	0.4448
250	0.3721
300	0.606
400	0.719
500	0.5396
600	0.8524
700	0.818
800	0.7433
900	0.7905
1000	0.8687
1100	0.8644
1200	0.9118

Tabelle 4.2: Ljung-Box-Test Ergebnisse für verschiedene Lag-Längen

bestimmt.  $L$  ist dabei die maximale Lag-Länge. Tabelle 4.2 zeigt für verschiedene Lag-Längen den p-Wert, der von R ausgegeben wird. Es ist zu sehen, dass der p-Wert für keine Lag-Länge kleiner als 0.05 ist. (Es wurden noch weitere Lag-Längen eingesetzt, und auch hier wurden keine p-Werte kleiner als 0.05 ausgegeben.) Zu den üblichen Niveaus kann somit die Nullhypothese der Unabhängigkeit nicht abgelehnt werden.

Insgesamt kann also, aufgrund der acf-Plots und der Ljung-Box-Tests, davon ausgegangen werden, dass die Residuen unabhängig voneinander sind und die Anpassung gut gelungen ist.

Weiterhin muss die Stationarität geprüft werden. Da

$$\alpha_1 + \beta = 0.09199 + 0.9 = 0.99199 < 1$$

gilt, kann davon ausgegangen werden, dass die Anpassung strikt stationär ist, obwohl der Wert sehr nah an der zulässigen oberen Schranke liegt.

### 4.2.2 Anwendung der POT-Methode auf die Residuen

Auf die Residuen wird jetzt die POT-Methode für iid Zufallsvariablen angewendet.

## QQ-Normal Plot

Der QQ-Normal-Plot (Abbildung 4.5) zeigt, dass auch die Residuen heavier tailed sind als eine Normalverteilung, da die Punkte an den Rändern wieder deutlich von der Linie abweichen. Es ist also auch für die Residuen sinnvoll eine GPD-Verteilung anzupassen.

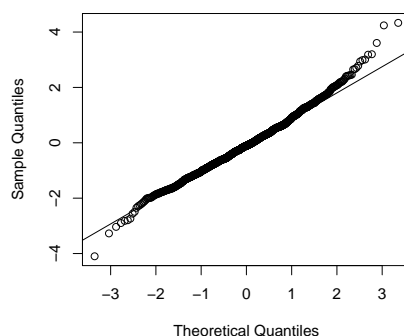


Abbildung 4.5: QQ-Normal Plot der Residuen

## mrl-Plot

Mit Hilfe des Mean Residual Life Plots der Residuen wird wie in Kapitel 2.1.1 ein Threshold ausgewählt. Abbildung 4.6 zeigt den Mean Residual Life Plot und Abbildung 4.7 zeigt wieder einen Ausschnitt daraus. Aus der Graphik könnte man die Schwellenwerte 2.45 (gelbe Linie) oder 1.8 (blaue Linie) als geeignete Werte entnehmen. Diese Werte liegen allerdings nicht in dem von den 90%- und 95%-Quantilen aufgespannten Intervall (1.258764, 1.619105) sondern darüber und bieten somit zu wenige Überschreitungen. Ein besserer Threshold ist 1.45 (grüne Linie) oder 1.3 (rote Linie). Für diese Werte erhält man 82 bzw. 111 Überschreitungen und sie liegen beide in dem von den 90%- und 95%-Quantilen aufgespannten Intervall. Für die weiteren Berechnungen wurde der Threshold 1.3 ausgewählt.

Alternativ wurde ein tcplot betrachtet, aber dieser führt zu einem ähnlichen Threshold und wird deshalb hier nicht weiter aufgeführt.

## Maximum-Likelihood-Schätzung

Nachdem der Threshold gewählt wurde, können wie in Abschnitt 2.2 durch eine Maximum-Likelihood-Schätzung Schätzer für  $\xi$  und  $\beta$  berechnet werden. R gibt folgende Werte aus:





Abbildung 4.6: Mean Residual Life Plot der Residuen

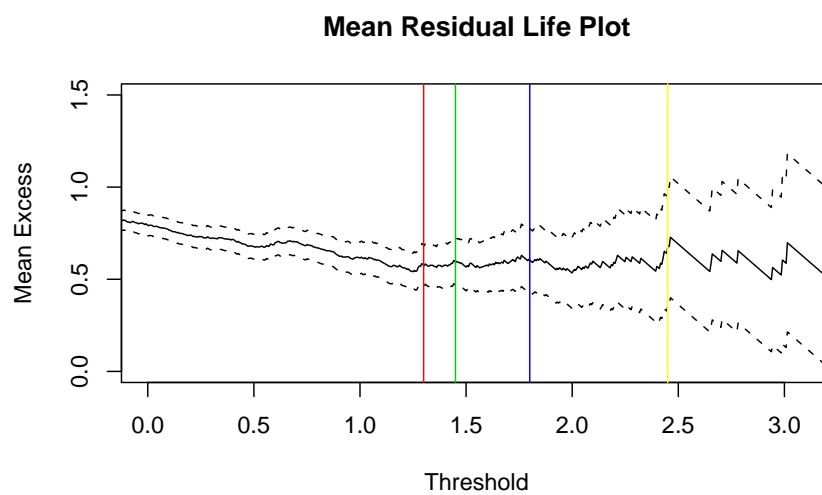


Abbildung 4.7: Ausschnitt aus dem Mean Residual Life Plot der Residuen

Threshold:	1.3	
Anzahl der Exzesse:	111	
Parameterschätzer:	$\xi$	0.01012838
	$\beta$	0.57002697
Standardabweichung der Schätzer:	$\xi$	0.10093950
	$\beta$	0.07898203
Varianz-Kovarianzmatrix:	0.01018878	-0.005755520
	-0.00575552	0.006238161

Als Schätzer für  $\xi$  wurde der Wert 0.01012838 ermittelt und als Schätzer für  $\beta$  0.57002697. Der Threshold 1.3 ist automatisch wieder der Schätzer für  $\mu$ . Leider ist die Standardabweichung des Schätzers für  $\xi$  sehr hoch. Dies lässt sich jedoch nicht immer vermeiden, da der Schätzer für  $\xi$  sehr klein ist. Darüber hinaus ergibt die Überprüfung der Anpassung, die als nächstes folgt, ein sehr gutes Ergebnis. Daher kann davon ausgegangen werden, dass die Maximum-Likelihood Schätzung trotzdem gut gelungen ist.

### 4.2.3 Überprüfung der Anpassung

#### Histogramm

Die Anpassung wird durch ein Histogramm überprüft. Abbildung 4.8 zeigt das Histogramm der Residuen mit der Dichte der geschätzten Verteilung (links) und die empirische Verteilungsfunktion der Residuen mit der geschätzten Verteilungsfunktion (rechts). In beiden Abbildungen ist zu sehen, dass die Verteilung sehr gut an die Daten angepasst ist.

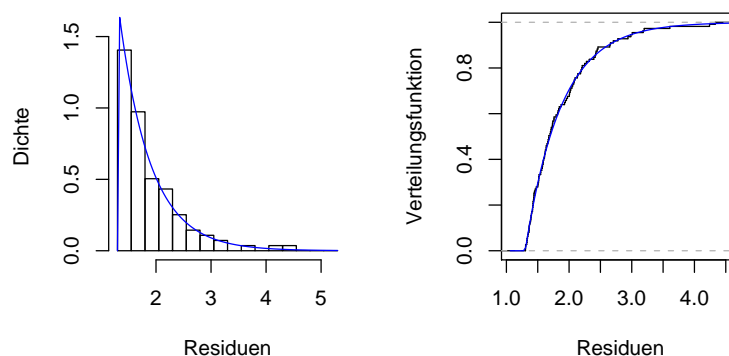


Abbildung 4.8: Histogramm der Residuen mit der Dichte der geschätzten Verteilung (links) und empirische Verteilungsfunktion der Residuen mit der geschätzten Verteilungsfunktion (rechts).

## Kolmogorov-Smirnov-Test

Ein Kolmogorov-Smirnov-Test mit einem p-Wert von 0.9994 bestätigt dieses Ergebnis.

### 4.2.4 Value at Risk der Residuen

Mit der angepassten Verteilung kann nun, wie in Kapitel 2.4, mit der Formel

$$\hat{z}_q = u + \frac{\hat{\beta}}{\hat{\xi}} \left( \left( \frac{n}{k} (1 - q) \right)^{-\hat{\xi}} - 1 \right)$$

der Value at Risk der Residuen berechnet werden. In Tabelle 4.3 sind die Ergebnisse für verschiedene Konfidenzintervalle aufgelistet.

$q$	$\hat{z}_q$
95%	1.625611
99%	2.555904
99.9%	3.913498
99.99%	5.303125

Tabelle 4.3: Value at Risk der Residuen unter Annahme GPD verteilter Residuen (für verschiedene Konfidenzintervalle).

Der Value at Risk hängt wieder stark vom Konfidenzintervall ab. Für  $q = 95\%$  beträgt er nur 1.625611 und für  $q = 99.99\%$  mit 5.303125 mehr als das Dreifache. Diese Werte sind aber noch nicht sehr aussagekräftig, da sie „nur“ ein Zwischenergebnis für die Schätzung des VaR der gesamten Zeitreihe sind.

Zum Vergleich wurde zusätzlich der Value at Risk der Residuen unter Annahme einer Normalverteilung berechnet. Die Ergebnisse sind in Tabelle 4.4 aufgelistet. Auch diese Werte sind noch nicht sehr aussagekräftig. Sie werden aber für einen

$q$	$\hat{z}_q$
95%	1.577590
99%	2.259068
99.9%	3.022934
99.99%	3.651703

Tabelle 4.4: Value at Risk der Residuen unter Annahme normal verteilter Residuen (für verschiedene Konfidenzintervalle).

späteren Vergleich der Methoden benötigt. Im Vergleich der beiden Tabellen (4.3 und 4.4) ist aber, wie schon in Kapitel 2, zu sehen, dass der VaR für eine GPD Verteilung teilweise deutlich größer ist, als der VaR einer Normalverteilung.

### 4.2.5 Expected Shortfall der Residuen

Schließlich wird mit der Formel

$$E[Z|Z > \hat{z}_q] = \frac{\hat{z}_q}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1 - \hat{\xi}},$$

wie in Kapitel 2.4, der ES der Residuen berechnet. Die Ergebnisse für verschiedene Konfidenzintervalle sind in Tabelle 4.5 aufgelistet. Auch hier hängen die Werte stark vom Konfidenzintervall ab.

$q$	$E[Z Z > \hat{z}_q]$
95%	2.204803
99%	3.144614
99.9%	4.516099
99.99%	5.919945

Tabelle 4.5: Expected Shortfall der Residuen unter Annahme GPD verteilter Residuen (für verschiedene Konfidenzintervalle).

Tabelle 4.6 zeigt den ES der Residuen unter der Annahme, dass die Residuen normalverteilt sind. Insgesamt ist beim ES dasselbe Verhalten wie beim VaR zu beobachten, er variiert sehr stark bei Veränderung des Konfidenzintervalls und der ES unter Annahme normalverteilter Daten ist immer kleiner als der ES unter Annahme GPD verteilter Daten.

$q$	$E[Z Z > \hat{z}_q]$
95%	1.995439
99%	2.597926
99.9%	3.299785
99.99%	3.891160

Tabelle 4.6: Expected Shortfall der Residuen unter Annahme normalverteilter Residuen (für verschiedene Konfidenzintervalle).

## 4.3 Bedingter Value at Risk

Der bedingte Value at Risk kann jetzt durch Zusammensetzen der Ergebnisse berechnet werden. Es gilt nach (3.11) aus Kapitel 3.2

$$\hat{x}_q^t = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \hat{z}_q,$$

wobei  $\hat{\mu}_{t+1} = \hat{\phi}X_t$  und  $\hat{\sigma}_{t+1}^2 = \hat{\alpha}_0 + \hat{\alpha}_1\epsilon_t^2 + \hat{\beta}\hat{\sigma}_t^2$  und  $\hat{z}_q$  ist der VaR der Residuen für das Konfidenzintervall  $q$ , der aus Tabelle 4.3 entnommen werden kann. Da es sich nun um einen bedingten VaR, also ein dynamisches Verfahren handelt, erhält man jetzt keine Zahl mehr, sondern eine Funktion.

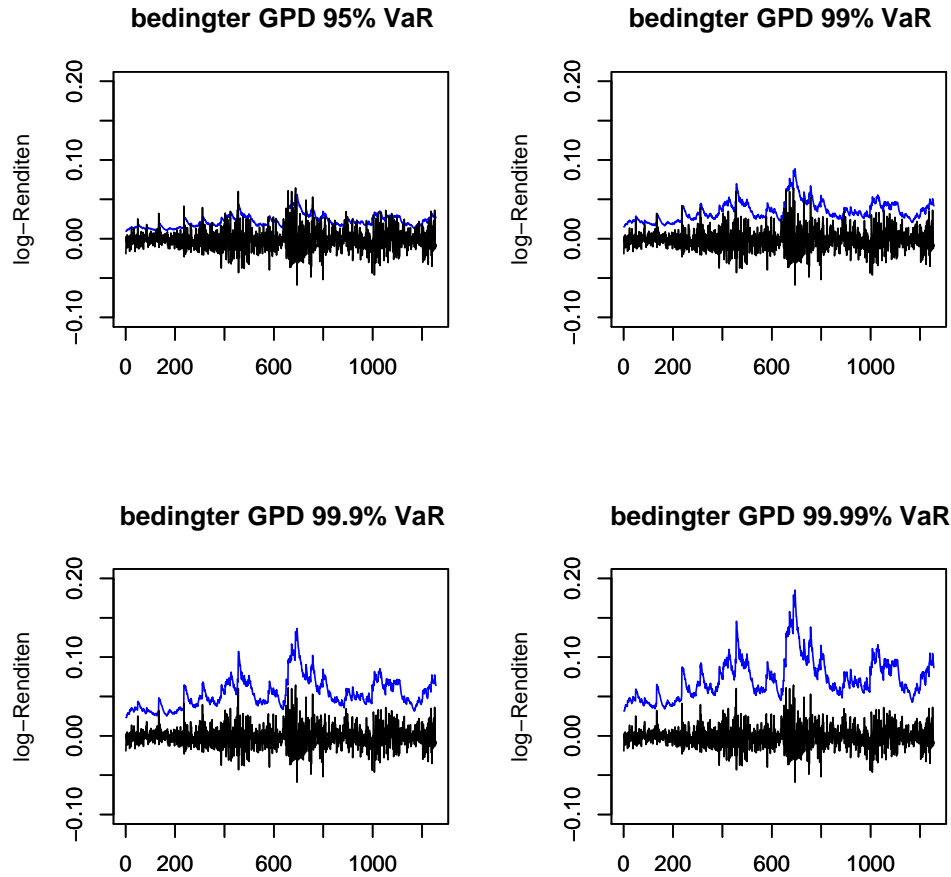


Abbildung 4.9: Log-Renditen mit den Schätzern für den bedingten VaR unter Annahme GPD verteilter Residuen für verschiedene Konfidenzintervalle.

Abbildung 4.9 zeigt den geschätzten bedingten Value at Risk für die verschiedenen Konfidenzintervalle zusammen mit den Log-Renditen. Es ist zu sehen, dass sich der VaR sehr gut an die Log-Renditen anpasst und sehr schnell auf Volatilitätsveränderungen reagiert. Für ein Konfidenzintervall von 95% gibt es noch einige sogenannte *Störungen*, d.h. Werte, für die der geschätzte VaR kleiner ist als die Log-Rendite, also  $\hat{x}_q^t < x_{t+1}$ . Aber für ein Konfidenzintervall von 99.9% ist der VaR schon sehr „sicher“. Er weist nur 2 Störungen auf. Für ein Konfidenzintervall von 99.99% gibt es sogar gar keine Störungen mehr.

## Vergleich GPD und Normalverteilung

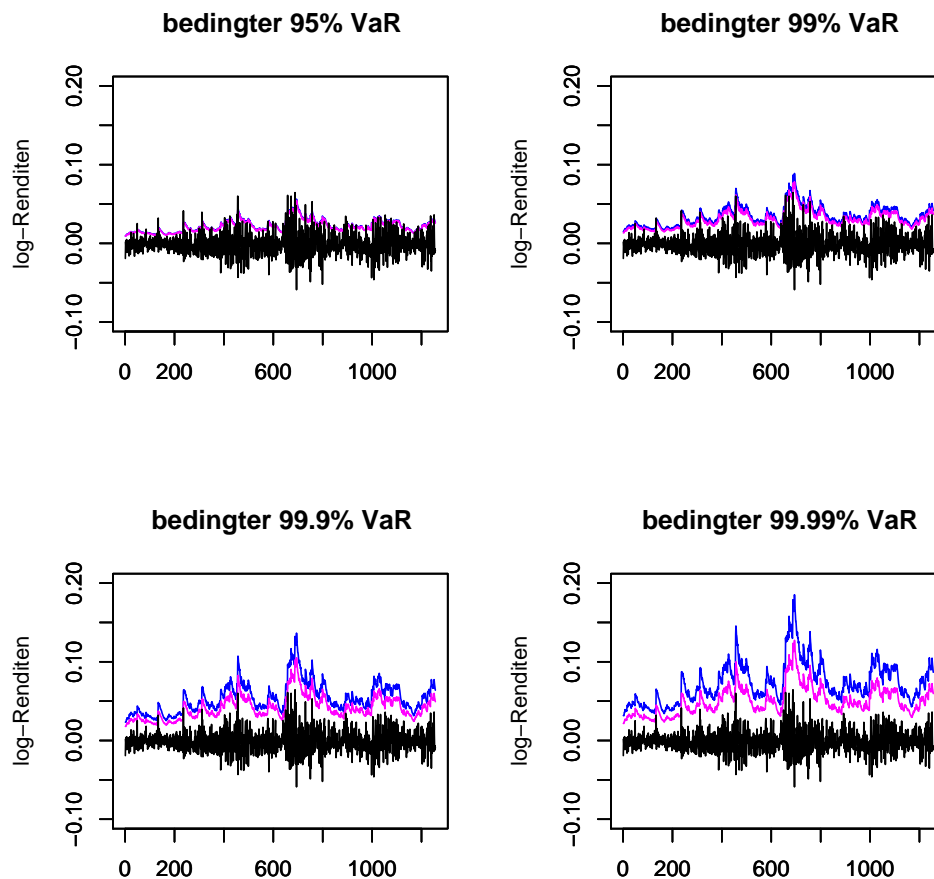


Abbildung 4.10: Log-Renditen mit dem bedingten VaR unter Annahme GPD verteilter Residuen (blau) und dem bedingten VaR unter Annahme normalverteilter Residuen (pink) (für verschiedene Konfidenzintervalle).

Abbildung 4.10 zeigt für die verschiedenen Konfidenzintervalle die Log-Renditen mit dem bedingten VaR unter der Annahme, dass die Residuen GPD-verteilt sind (kurz: bedingter GPD VaR) (blau) und den bedingten VaR unter der Annahme normalverteilter Residuen (kurz: bedingter normal VaR) (pink). Beide VaR-Funktionen liefern gute Ergebnisse. Für ein Konfidenzintervall von 95% liegen die Graphen fast genau übereinander, es gibt also keine großen Unterschiede. Bei einem Konfidenzintervall von 99.9% oder 99.99% ist schon ein deutlicherer Unterschied zu erkennen. Hier liegt der bedingte GPD VaR immer leicht über dem bedingten normal VaR, d.h. er ist etwas konservativer. Beim 99.99% normal VaR treten allerdings noch Störungen (2) auf, was beim 99.99% GPD VaR nicht der

Fall ist. Beide VaR-Funktionen reagieren aber sehr gut auf die Volatilitätsschwankungen.

### Vergleich bedingter und unbedingter VaR

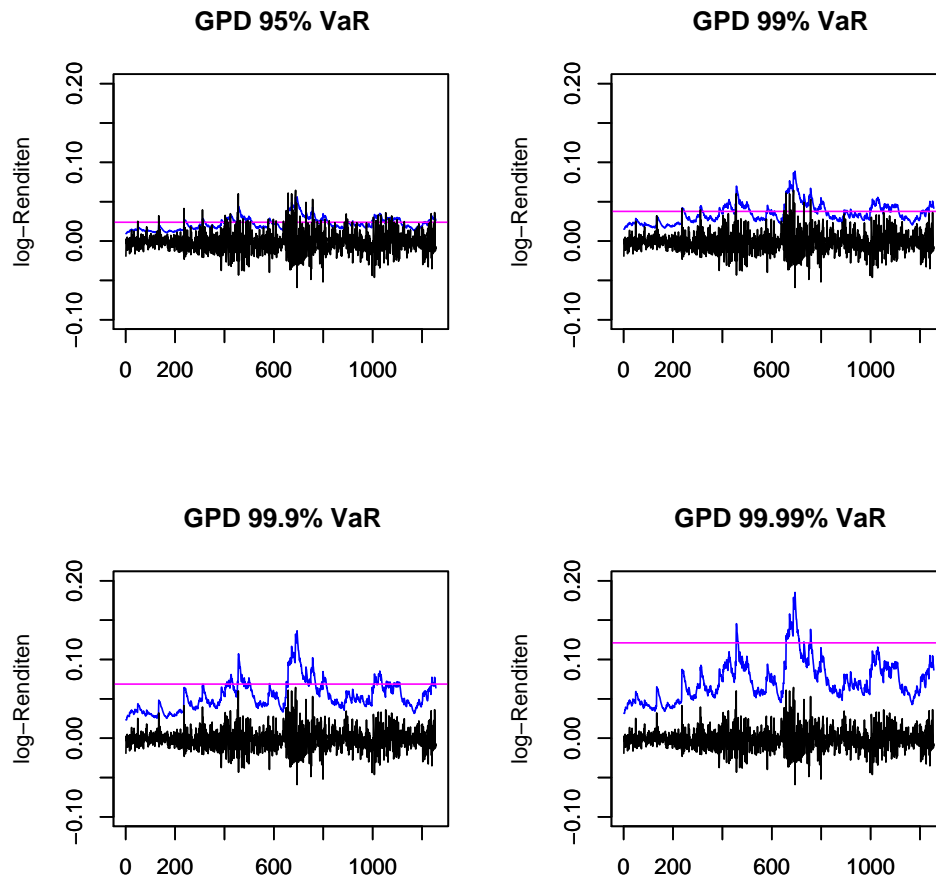


Abbildung 4.11: Log-Renditen mit dem bedingten VaR unter Annahme GPD verteilter Residuen (blau) und dem unbedingten VaR unter Annahme GPD verteilter Zufallsvariablen (aus Kapitel 2) (pink).

In Abbildung 4.11 ist der bedingte GPD VaR (blau) im Vergleich zum unbedingten GPD VaR (aus Kapitel 2.4) (pink) für verschiedene Konfidenzintervalle in die Graphik der Log-Renditen eingezeichnet. Gerade bei einem Konfidenzintervall von 99% ist sehr gut zu sehen, dass der bedingte und der unbedingte VaR gleich oft Störungen aufweisen. Aber wenn eine Störung auftritt, überschreitet die Log-Rendite den unbedingten VaR teilweise schon sehr deutlich. Der bedingte VaR wird hingegen immer nur ganz leicht überschritten. Bei einem Konfidenzintervall

von 99.99% ist sehr gut zu sehen, dass der unbedingte VaR weit über den Daten liegt. Der bedingte VaR liegt zwar auch immer über den Log-Renditen, aber er liegt die meiste Zeit deutlich unter dem unbedingten VaR und beschreibt das Risiko großer Verluste in diesem Fall besser.

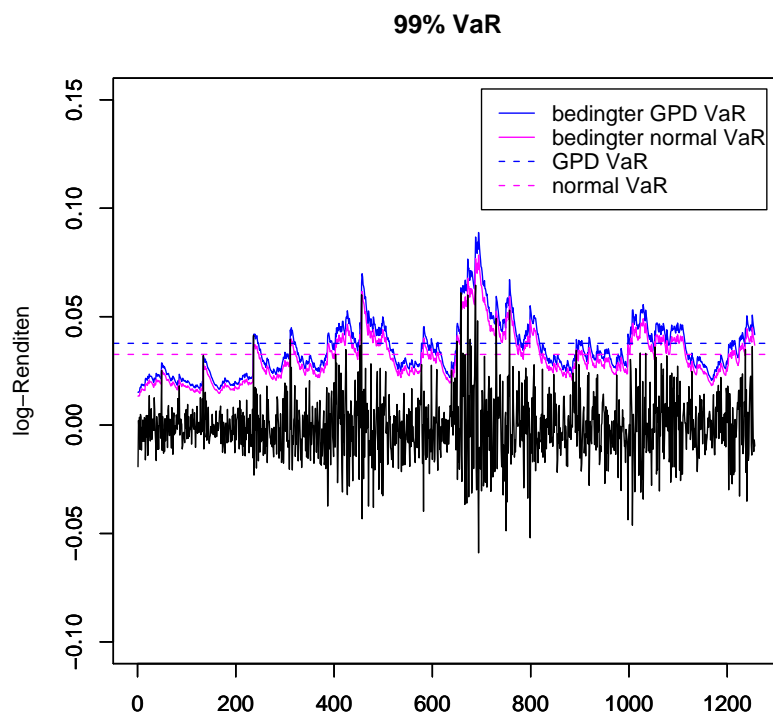


Abbildung 4.12: Log-Renditen mit dem bedingten und unbedingten VaR unter Annahme GPD verteilter und normalverteilter Residuen bzw. Zufallsvariablen.

Abbildung 4.12 zeigt noch einmal speziell für ein 99%-Konfidenzintervall den bedingten GPD VaR (blaue durchgezogene Kurve), den bedingten normal VaR (pinke durchgezogene Kurve), den unbedingten GPD VaR (blaue gestrichelte Linie) und den unbedingten normal VaR (pinke gestrichelte Linie) im Vergleich. Am besten schneidet in diesem Vergleich der bedingte GPD VaR ab. Er hat zwar genauso viele Störungen (nämlich 12) wie der unbedingte GPD VaR, aber die Log-Renditen überschreiten diesen VaR immer nur leicht. Der bedingte normal VaR zeigt mit 20 Störungen eine deutlich schlechtere Risikoabschätzung und der unbedingte normal VaR schneidet mit 23 Störungen am schlechtesten ab. Insgesamt ist aus diesem Ergebnis festzuhalten, dass der VaR unter Annahme einer GPD Verteilung im bedingten und unbedingten Fall ein besseres Ergebnis liefert als der VaR unter Annahme einer Normalverteilung.



## 4.4 Bedingter Expected Shortfall

Im Folgenden wird der bedingte Expected Shortfall genauer betrachtet. Dieser ergibt sich durch Einsetzen der Ergebnisse in

$$\hat{S}_q^t = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} E[Z \mid Z > \hat{z}_q],$$

wobei  $\hat{\mu}_{t+1} = \hat{\phi} X_t$  und  $\hat{\sigma}_{t+1}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \epsilon_t^2 + \hat{\beta} \hat{\sigma}_t^2$  und  $E[Z \mid Z > \hat{z}_q]$  ist der ES der Residuen für das Konfidenzintervall  $q$ , der aus Tabelle 4.5 abgelesen werden kann.

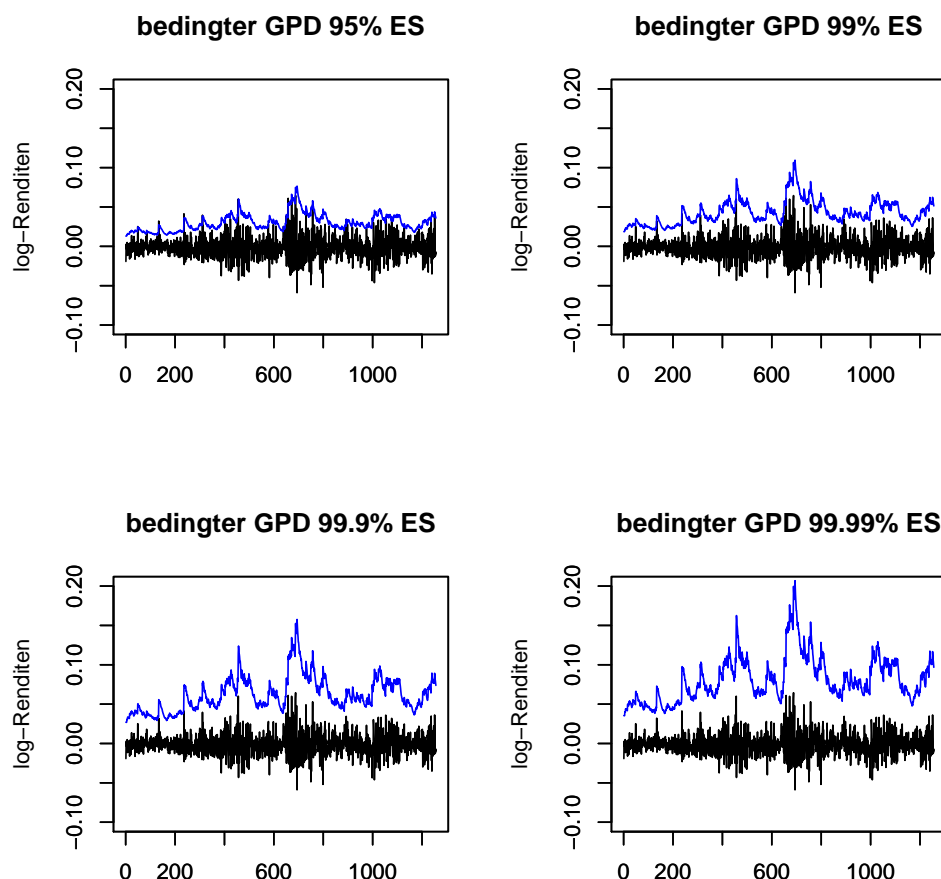


Abbildung 4.13: Log-Renditen mit den Schätzern für den bedingten ES unter Annahme GPD verteilter Residuen für verschiedene Konfidenzintervalle.

Abbildung 4.13 zeigt die Log-Renditen mit dem bedingten ES für verschiedene Konfidenzintervalle. Auch hier ist zu sehen, dass die Abschätzung recht gut gelungen ist und der ES schnell auf Volatilitätsschwankungen reagiert.

## Vergleich GPD und Normalverteilung

Abbildung 4.14 zeigt den bedingten GPD ES (blaue Linie) und den bedingten normal ES (pinke Linie). Im Vergleich der Verteilungen ist dasselbe Verhalten wie beim VaR (Abbildung 4.10) zu beobachten, nur dass der bedingte ES höher über den Log-Renditen liegt als der bedingte VaR. Dies folgt aber aus der Definition des ES.

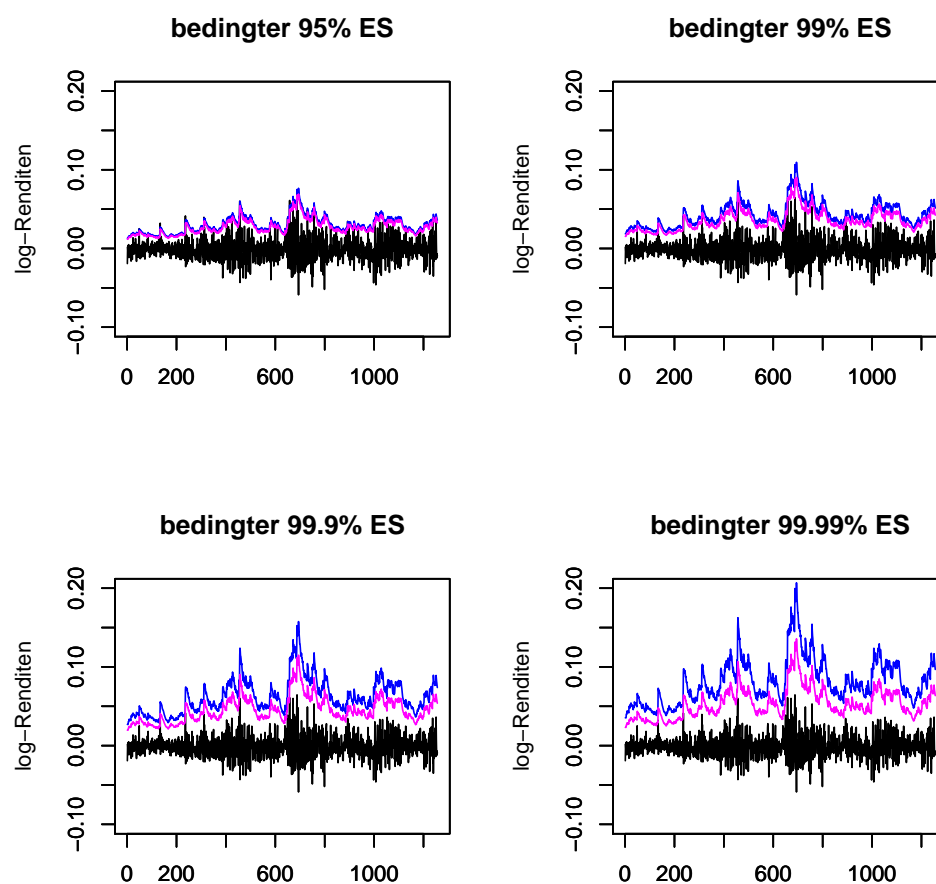


Abbildung 4.14: Log-Renditen mit dem bedingten ES unter Annahme GPD verteilter Residuen (blau) und mit dem bedingten ES unter Annahme normalverteilter Residuen (pink) (für verschiedene Konfidenzintervalle).

## Vergleich VaR und ES

Dass der bedingte ES höher über den Log-Renditen liegt als der bedingte VaR, wird in Abbildung 4.15 verdeutlicht. Denn dort ist der bedingte GPD VaR (blau) und der bedingte GPD ES (pink) für verschiedene Konfidenzintervalle zusammen mit den Log-Renditen abgebildet.

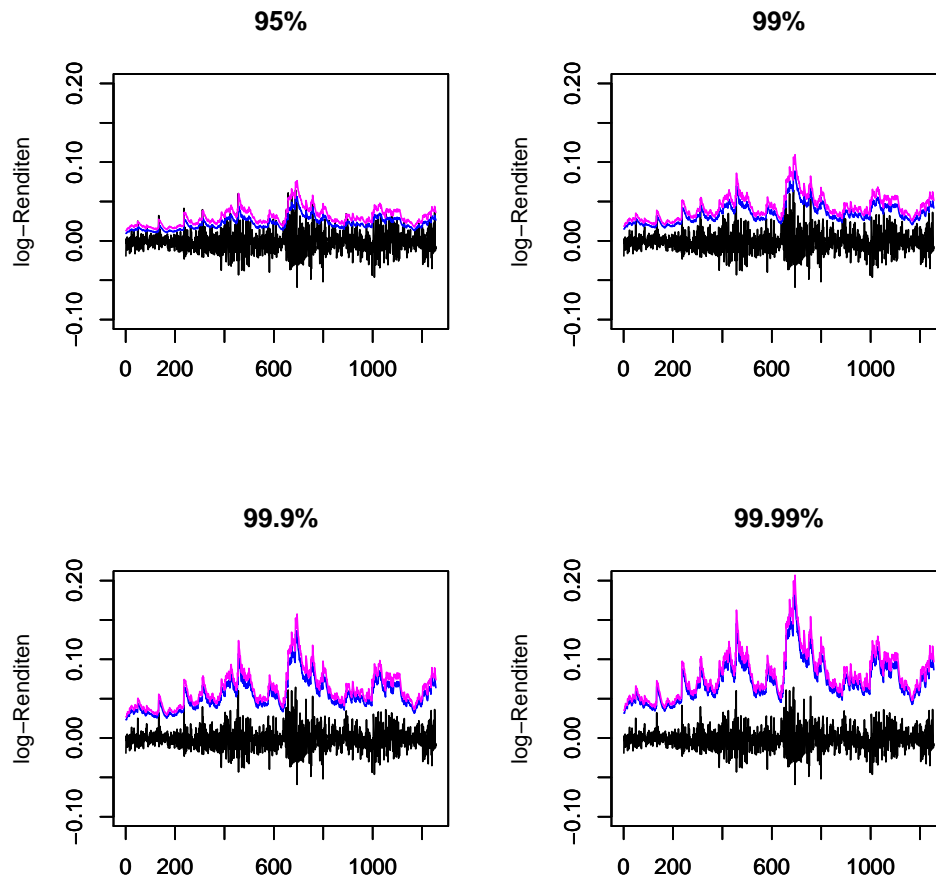


Abbildung 4.15: Log-Renditen mit dem bedingten Value at Risk (blau) und dem bedingten Expected Shortfall (pink) unter Annahme GPD verteilter Residuen (für verschiedene Konfidenzintervalle).

In Abbildung 4.16 wurde speziell für ein Konfidenzintervall von 99% der bedingte VaR (blaue durchgezogene Kurve), der unbedingte VaR (blaue gestrichelte Linie), der bedingte ES (pinke durchgezogene Kurve) und der unbedingte ES (pinke gestrichelte Linie) (alles unter Annahme einer GPD Verteilung) in die Graphik der Log-Renditen eingezeichnet. Hier ist sehr schön zu sehen, dass der bedingte VaR und der bedingte ES sehr nah beieinander liegen. Zwischen dem unbedingten VaR

und dem unbedingten ES liegt ein größerer Abstand. Das liegt an dem schon erwähnten „Phänomen“, dass der unbedingte VaR, wenn er überschritten wird, meist deutlich überschritten wird, was beim bedingten VaR nicht der Fall ist. Da der ES nach Definition der erwartete Verlust, gegeben dass der VaR überschritten wurde ist, liegt dieser im unbedingten Fall deutlich über dem VaR als im bedingten Fall.

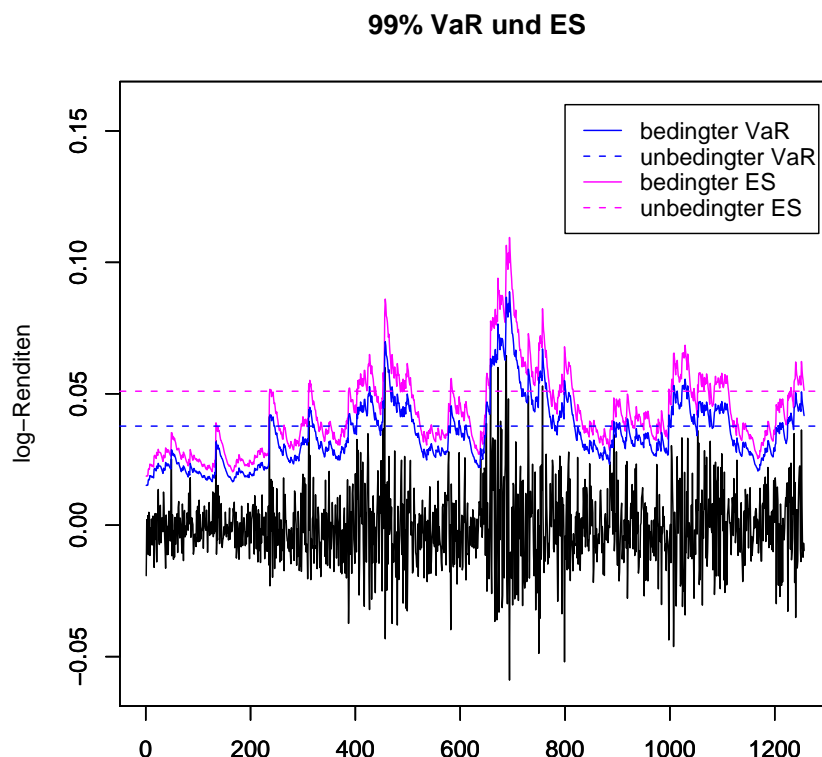


Abbildung 4.16: Log-Renditen mit dem bedingten und unbedingten VaR und ES unter Annahme GPD verteilter Residuen bzw. Zufallsvariablen (für ein Konfidenzintervall von 99%)

## 4.5 Abschließender Vergleich

Insgesamt kann aus den Graphiken gefolgert werden, dass der bedingte GPD VaR das Risiko großer Verluste am besten beschreibt. Um dies zu verdeutlichen wurden mit R die Störungen (also die Werte für die  $x_{t+1} > \hat{x}_q^t$  gilt) für die verschiedenen Methoden und Konfidenzintervalle gezählt und in Tabelle 4.7 aufgelistet. In den

Klammern hinter den Werten sind die gerundeten Prozentzahlen aufgelistet, also der Anteil der Störungen im Vergleich zur Länge des Datensatzes.

Störungen	bedingter GPD VaR	unbedingter GPD VaR	bedingter normal VaR	unbedingter normal VaR
95%	62 (0.0493)	63 (0.0502)	69 (0.0549)	72 (0.0573)
99%	12 (0.0096)	12 (0.0096)	20 (0.0159)	23 (0.0183)
99.9%	2 (0.0016)	0 (0.0000)	5 (0.0040)	8 (0.0063)
99.99%	0 (0.0000)	0 (0.0000)	2 (0.0016)	6 (0.0048)

Tabelle 4.7: Anzahl der Störungen für die verschiedenen Methoden und Konfidenzintervalle. In Klammern steht der Anteil der Störungen im Vergleich zur Länge des Datensatzes.

Im Vergleich der Anzahl der Störungen schneiden der bedingte GPD VaR und der unbedingte GPD VaR am besten ab. Sie liegen, was die Anzahl der Störungen angeht, sehr nah beieinander. Wie in den Abbildungen gesehen, wird der unbedingte VaR bei einer Störung aber im Durchschnitt deutlich mehr überschritten als der bedingte VaR, da er nicht so schnell auf Volatilitätsschwankungen reagiert. Dies wird durch die Betrachtung des ES noch verdeutlicht. Deshalb schneidet der bedingte VaR in diesem Vergleich besser ab, als der unbedingte VaR. Der VaR unter Annahme einer Normalverteilung zeigt ein deutlich schlechteres Ergebnis. Die Anzahl der Störungen liegt in allen Fällen über der Anzahl der Störungen des GPD VaR, wobei der unbedingte normal VaR noch ein schlechteres Ergebnis liefert, als der bedingte normal VaR.

Dieses Ergebnis bestätigt die Aussage, dass eine GPD Verteilung Log-Rendite Daten besser beschreibt als eine Normalverteilung und dass die POT Methode für iid Folgen und für stetige Zeitreihen eine sehr gute Abschätzung der Daten liefert. Insgesamt ist bei dieser Art von Daten die POT Methode für stetige Zeitreihen zu präferieren. Trotz der Abhängigkeitsstruktur der Daten liefert die POT Methode für iid Folgen aber auch noch sehr gute Ergebnisse.



## 5 Alternative Anwendung

Das in Kapitel 3 und 4 vorgestellte Verfahren ist ein gängiges Verfahren um den VaR von stationären Zeitreihen zu schätzen. An dieser Stelle soll aber noch einmal auf die Signifikanz des Parameters  $\phi$  eingegangen werden. Da für diesen Parameter bei dem vorliegenden Beispiel der DAX-Abschlusswerte bei der Pseudo-Maximum-Likelihood Schätzung, in Abschnitt 4.1, ein sehr hoher Signifikanzcode ausgegeben wurde, soll in diesem Kapitel kritisch hinterfragt werden, ob der AR(1)-Prozess in diesem Fall für die Anpassung benötigt wird. Darüber hinaus stellt sich die Frage, ob der AR(1)-Prozess die Log-Renditen gut abbildet, da diese, wie in den acf-plots (Abbildung 4.3) gesehen, annähernd unkorreliert sind und diese Eigenschaft nicht durch den AR-Prozess abgebildet wird. Um diesen Fragen nachzugehen wird in diesem Kapitel „nur“ ein GARCH(1,1)-Prozess an die Log-Renditen angepasst und die Ergebnisse werden mit denen aus Kapitel 4 verglichen.

### 5.1 Anpassung eines GARCH(1,1)-Prozesses

Im Gegensatz zum gängigen Verfahren ist  $\mu_t$  in dieser Anwendung konstant. Als Wert für  $\mu$  wird der Erwartungswert der Log-Renditen verwendet. Dieser beträgt  $-0.0008242146$ . Dementsprechend haben die Log-Renditen die Gestalt

$$X_t = \mu + \epsilon_t,$$

wobei  $\epsilon_t = \sigma_t Z_t$  mit  $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$  wieder ein GARCH(1,1)-Prozess ist. Es wird also ein GARCH(1,1)-Prozess an die Mittelwert-bereinigte Folge  $X_t - \mu$  angepasst.

Die Pseudo-Maximum-Likelihood Schätzung mit R liefert folgende Schätzer:

	Schätzer	Standard-abweichung	t-Wert	Signifikanzcode
$\alpha_0$	0.00000258386	0.000001012	2.554	0.0107 *
$\alpha_1$	0.0944175	0.01727	5.468	0.0000000454 ***
$\beta$	0.89665	0.01785	50.228	$< 2 * 10^{-16}$ ***

Tabelle 5.1: Ergebnisse der Pseudo-Maximum-Likelihood Schätzung

Die Ergebnisse für die Parameter  $\alpha_0$ ,  $\alpha_1$  und  $\beta$  liegen sehr nah an den Werten aus der AR(1)-GARCH(1,1)-Anpassung aus Kapitel 4.1. Auch die Signifikanzcodes für  $\alpha_1$  und  $\beta$  sind wieder sehr gut.

Mit den geschätzten Parametern kann mit

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

wieder rekursiv der Schätzer für  $(\hat{\sigma}_{t-n+1}, \dots, \hat{\sigma}_t)$  berechnet werden. Auch hier wurde Null als Startwert gewählt.

## 5.2 Residuen

Als nächstes werden die Residuen berechnet. Diese erhält man in diesem Fall durch

$$(z_{t-n+1}, \dots, z_t) = \left( \frac{x_{t-n+1} - \mu}{\hat{\sigma}_{t-n+1}}, \dots, \frac{x_t - \mu}{\hat{\sigma}_t} \right).$$

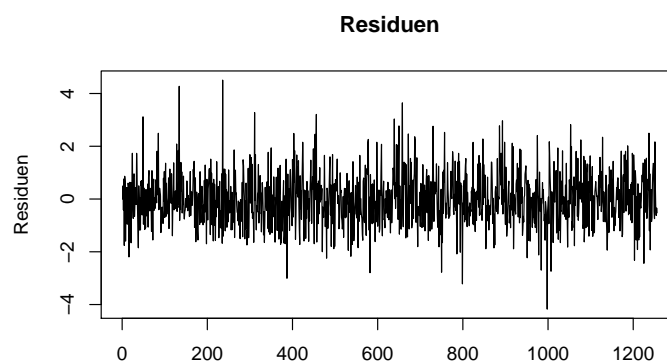


Abbildung 5.1: Residuenplot

Abbildung 5.1 zeigt den Plot der Residuen. Im Vergleich mit dem Plot der Residuen aus Kapitel 4.2 (Abbildung 4.2) sind auf den ersten Blick keine Unterschiede zu erkennen. Selbst wenn beide Graphiken übereinander gelegt werden, sind nur minimale Unterschiede zu beobachten. Im Durchschnitt unterscheiden die Werte sich um 0.06599787 und der maximale Unterschied beträgt 0.1755684.

### 5.2.1 Überprüfung der Unabhängigkeit der Residuen

Die Unabhängigkeit der Residuen wird mit einem acf-plot und einigen Ljung-Box-Tests überprüft. Abbildung 5.2 zeigt die acf-plots der quadrierten Residuen



für die Lag-Längen 30 (links) und  $2\sqrt{1256}$  (rechts). Die Autokorrelationen liegen ungefähr in dem vorgegebenen Intervall. Somit können die Residuen als annähernd unabhängig angenommen werden.

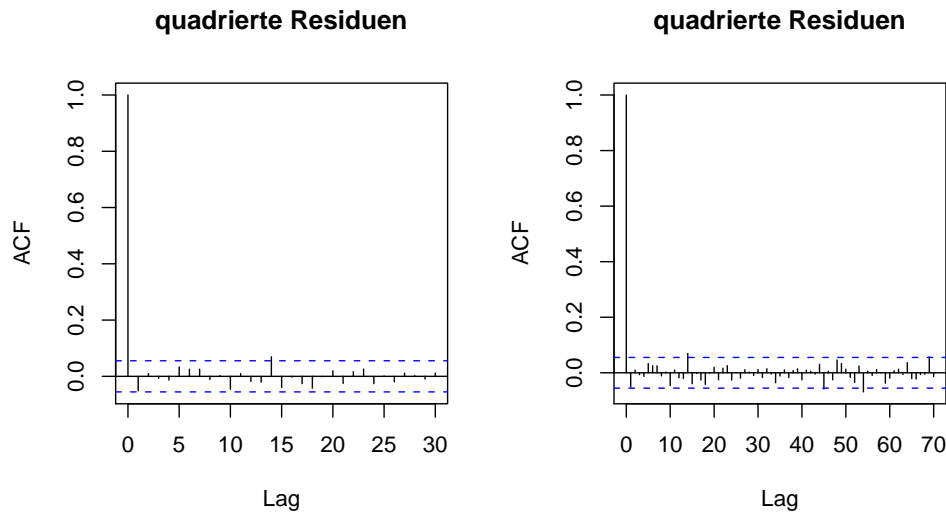


Abbildung 5.2: acf-plot der quadrierten Residuen für die Lag-Längen 30 (links) und  $2\sqrt{1256}$  (rechts).

Zusätzlich wurden wieder einige Ljung-Box-Tests für verschiedene Lag-Längen durchgeführt um die Unabhängigkeit zu überprüfen. Die Ergebnisse sind in Tabelle 5.2 aufgeführt. Auch hier liegen die p-Werte alle über 0.05 und somit kann die Nullhypothese der Unabhängigkeit zu keinem der üblichen Niveaus abgelehnt werden.

Schließlich wird die Stationarität überprüft. Es gilt

$$\alpha_1 + \beta = 0.0944175 + 0.89665 = 0.9910675 < 1,$$

daher kann davon ausgegangen werden, dass die Anpassung strikt stationär ist.

### 5.2.2 Anwendung der POT-Methode auf die Residuen

Der QQ-Normal-Plot (Abbildung 5.3) zeigt wie bei den anderen Anwendungen, dass die Residuen an den Enden deutlich von der Linie abweichen. Eine Normalverteilung würde die Tails der Verteilung also nicht sehr gut beschreiben.

Um einen Threshold zu wählen wird wieder der Mean Residual Life Plot der Residuen ausgegeben. Abbildung 5.4 zeigt einen Ausschnitt aus dem mrl-plot mit den möglichen Thresholds 2.51 (gelbe Linie), 2.155 (blaue Linie), 1.75 (grüne Linie)

Lag-Länge	p-Wert
50	0.4284
$2\sqrt{(1256)}$	0.2404
100	0.2151
150	0.2136
200	0.4121
250	0.3583
300	0.606
400	0.6928
500	0.4941
600	0.8272
700	0.797
800	0.7287
900	0.7791
1000	0.8652
1100	0.8735
1200	0.9254

Tabelle 5.2: Ljung-Box-Test Ergebnisse für verschiedene Lag-Längen

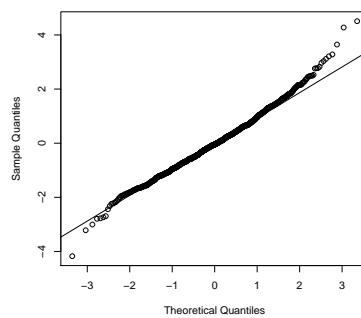


Abbildung 5.3: QQ-Normal-Plot der Residuen

und 1.3 (rote Linie). In dem von den 90%- und 95%-Quantilen aufgespannten Intervall (1.31561, 1.68296) liegt aber nur der Wert 1.3. Dieser wird für die weiteren Berechnungen als Threshold gesetzt.

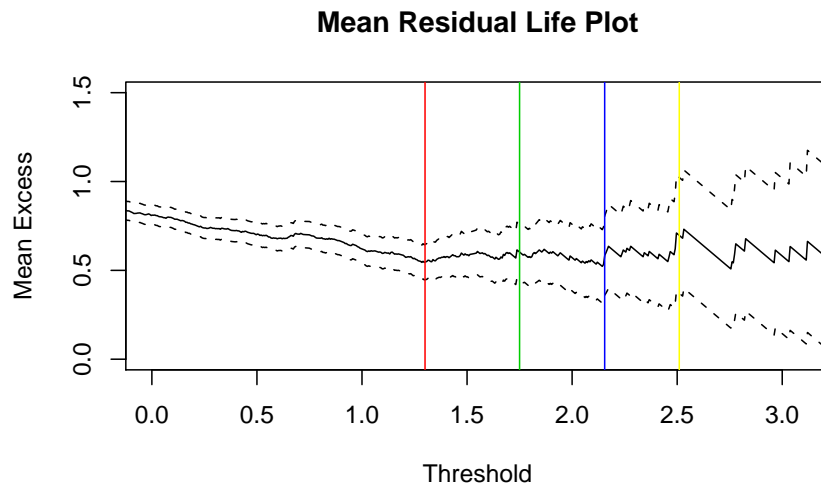


Abbildung 5.4: Ausschnitt aus dem Mean Residual Life Plot der Residuen.

Die Maximum-Likelihood Schätzung liefert folgende Parameterschätzer:

Threshold:	1.3
Anzahl der Exzesse:	132
Parameterschätzer:	$\xi$ 0.0742653 $\beta$ 0.5055251
Standardabweichung der Schätzer:	$\xi$ 0.10378328 $\beta$ 0.06839695
Varianz-Kovarianzmatrix:	0.010770968   -0.005139739 -0.005139739   0.004678143

In diesem Fall ist der  $\xi$ -Parameter deutlich größer als der  $\xi$ -Parameter der GPD-Anpassung aus der herkömmlichen Anwendung in Kapitel 4.2.2. Der  $\beta$ -Parameter ist ungefähr gleich groß. Für die angepasste Verteilung heißt das, dass die Tails der Verteilung in diesem Fall größer sind als die der GPD-Verteilung aus Kapitel 4.2.2. Die Lage und die Streckung ist bei beiden Verteilungen ungefähr gleich.

Die Überprüfung der Anpassung liefert auch für diese Anpassung gute Resultate. Abbildung 5.5 zeigt das Histogramm der Residuen mit der Dichte der geschätzten Verteilung (links) und die empirische Verteilungsfunktion der Residuen mit der geschätzten Verteilungsfunktion (rechts). Es ist zu erkennen, dass die Verteilung gut an die Daten angepasst ist.

Auch ein Kolmogorov-Smirnov-Test liefert einen p-Wert von 0.986 und somit ein

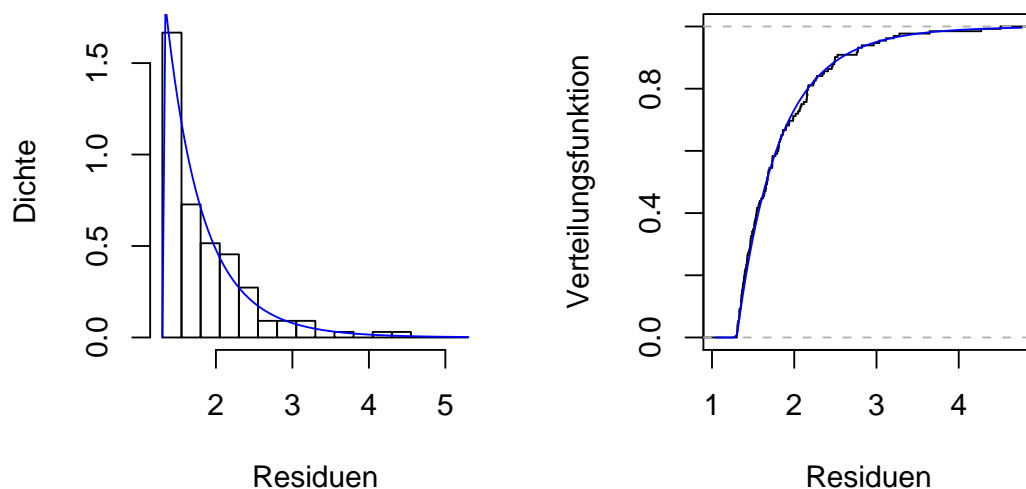


Abbildung 5.5: Histogramm der Residuen mit der Dichte der geschätzten Verteilung (links) und die empirische Verteilungsfunktion der Residuen mit der geschätzten Verteilungsfunktion (rechts).

sehr gutes Ergebnis. Insgesamt ist die Anpassung der Verteilung an die Daten also sehr gut gelungen.

### 5.2.3 VaR und ES der Residuen

Mit der geschätzten Verteilung kann jetzt der Value at Risk und der Expected Shortfall der Residuen berechnet werden. Die Tabellen 5.3 und 5.4 zeigen die Ergebnisse für verschiedene Konfidenzintervalle.

$q$	$\hat{z}_q$
95%	1.686079
99%	2.599328
99.9%	4.111107
99.99%	5.904823

Tabelle 5.3: Value at Risk der Residuen unter Annahme GPD verteilter Residuen (für verschiedene Konfidenzintervalle).

Zusätzlich wurde wieder der VaR und der ES unter der Annahme normalverteilter Residuen berechnet (Tabelle 5.5 und 5.6).

$q$	$E[Z Z > \hat{z}_q]$
95%	2.263132
99%	3.249644
99.9%	4.882703
99.99%	6.820316

Tabelle 5.4: Expected Shortfall der Residuen unter Annahme GPD verteilter Residuen (für verschiedene Konfidenzintervalle).

$q$	$\hat{z}_q$
95%	1.647469
99%	2.330626
99.9%	3.096374
99.99%	3.726692

Tabelle 5.5: Value at Risk der Residuen unter Annahme normalverteilter Residuen (für verschiedene Konfidenzintervalle).

$q$	$E[Z Z > \hat{z}_q]$
95%	2.066348
99%	2.670319
99.9%	3.373907
99.99%	3.966740

Tabelle 5.6: Expected Shortfall der Residuen unter Annahme normalverteilter Residuen (für verschiedene Konfidenzintervalle).

Insgesamt liegt der VaR und der ES der Residuen für diese Anpassung immer leicht über dem VaR und dem ES der Residuen aus Kapitel 4.2.4 und 4.2.5.

### 5.3 Bedingter VaR und ES

Der bedingte VaR kann wie in Kapitel 4.3 durch Zusammensetzen der Ergebnisse berechnet werden. In diesem Fall muss  $\mu_{t+1}$  durch die Konstante  $\mu$  ersetzt werden. Es gilt also

$$\hat{x}_q^t = \mu + \hat{\sigma}_{t+1} \hat{z}_q,$$

wobei  $\hat{\sigma}_{t+1}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \epsilon_t^2 + \hat{\beta} \hat{\sigma}_t^2$  und  $\hat{z}_q$  ist der VaR der Residuen für das Konfidenzintervall  $q$ , der aus Tabelle 5.3 entnommen werden kann.

Ebenso wird der bedingte ES durch Einsetzen der Schätzer für  $\mu$ ,  $\hat{\sigma}_{t+1}$  und  $E[Z|Z > \hat{z}_q]$  in

$$S_q^t = \mu + \hat{\sigma}_{t+1} E[Z|Z > \hat{z}_q]$$

berechnet.

Abbildung 5.6 zeigt für verschiedene Konfidenzintervalle den bedingten GPD VaR der GARCH(1,1)-Anpassung (blau) im Vergleich zum bedingten GPD VaR der AR(1)-GARCH(1,1)-Anpassung (pink) aus Abschnitt 4.3. Für die Konfidenzintervalle 95%, 99% und 99.9% sind im Prinzip keine Unterschiede zu erkennen, die blaue Funktion liegt in diesen Fällen fast genau über der pinken Funktion. Nur bei ganz genauem Hinschauen sind minimale Unterschiede zu erkennen. Bei einem Konfidenzintervall von 99.99% sind hingegen schon etwas deutlichere Unterschiede zu sehen. Dort liegt der bedingte VaR der GARCH-Anpassung immer leicht über dem bedingten VaR der AR-GARCH-Anpassung. Insgesamt ist aber zu sehen, dass aus beiden Methoden ein guter Schätzer resultiert.

Der Vergleich zwischen dem bedingten normal VaR der GARCH(1,1)-Anpassung (blau) und dem bedingten normal VaR der AR(1)-GARCH(1,1)-Anpassung (pink) (wieder für verschiedene Konfidenzintervalle) ist in Abbildung 5.7 zu sehen. In diesem Fall ist nur die blaue Funktion zu erkennen, weil beide VaR-Funktionen genau übereinander liegen, d.h. unter der Annahme normalverteilter Residuen liefern beide Methoden dasselbe Ergebnis.

Der gleiche Vergleich wurde auch für den bedingten Expected Shortfall durchgeführt. Abbildung 5.8 zeigt dazu den bedingten GPD ES der GARCH(1,1)-Anpassung (blau) im Vergleich zum bedingten GPD ES der AR(1)-GARCH(1,1)-Anpassung (pink) für verschiedene Konfidenzintervalle. Hier ist das gleiche Verhalten wie beim VaR zu erkennen. Für die Konfidenzintervalle 95%, 99% und 99.9% ist praktisch kein Unterschied zu sehen. Lediglich bei einem Konfidenzintervall

von 99.99% liegt der bedingte GPD ES der GARCH(1,1)-Anpassung leicht über dem bedingten GPD ES der AR(1)-GARCH(1,1)-Anpassung.

Auch die Betrachtung der ES-Funktionen unter Annahme normalverteilter Residuen liefert dasselbe Ergebnis wie beim Vergleich der VaR-Funktionen unter Annahme normalverteilter Residuen. Abbildung 5.9 zeigt den bedingten normal ES der GARCH(1,1)-Anpassung (blau) und den bedingten normal ES der AR(1)-GARCH(1,1)-Anpassung (pink) (für verschiedene Konfidenzintervalle). Da die beiden Funktionen wieder genau übereinander liegen, ist nur die blaue Funktion zu erkennen.

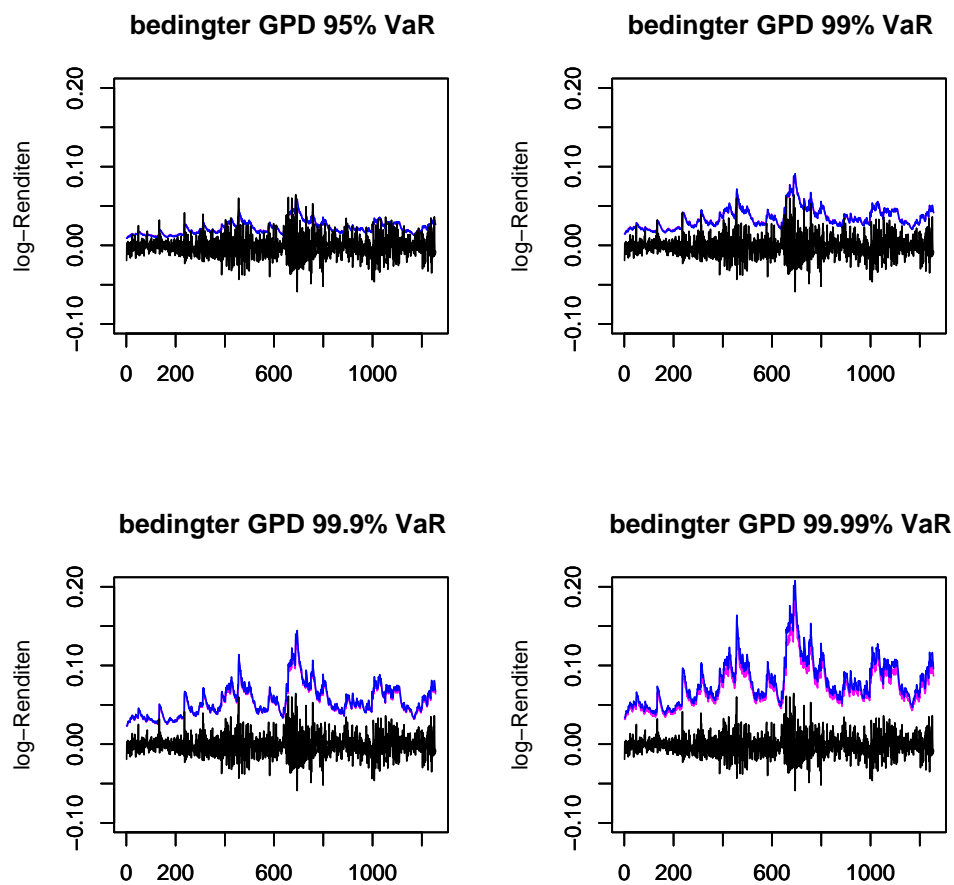


Abbildung 5.6: Log-Renditen mit den Schätzern für den bedingten GPD VaR der AR(1)-GARCH(1,1)-Anpassung (pink) und den bedingten GPD VaR der GARCH(1,1)-Anpassung (blau) (für verschiedene Konfidenzintervalle).

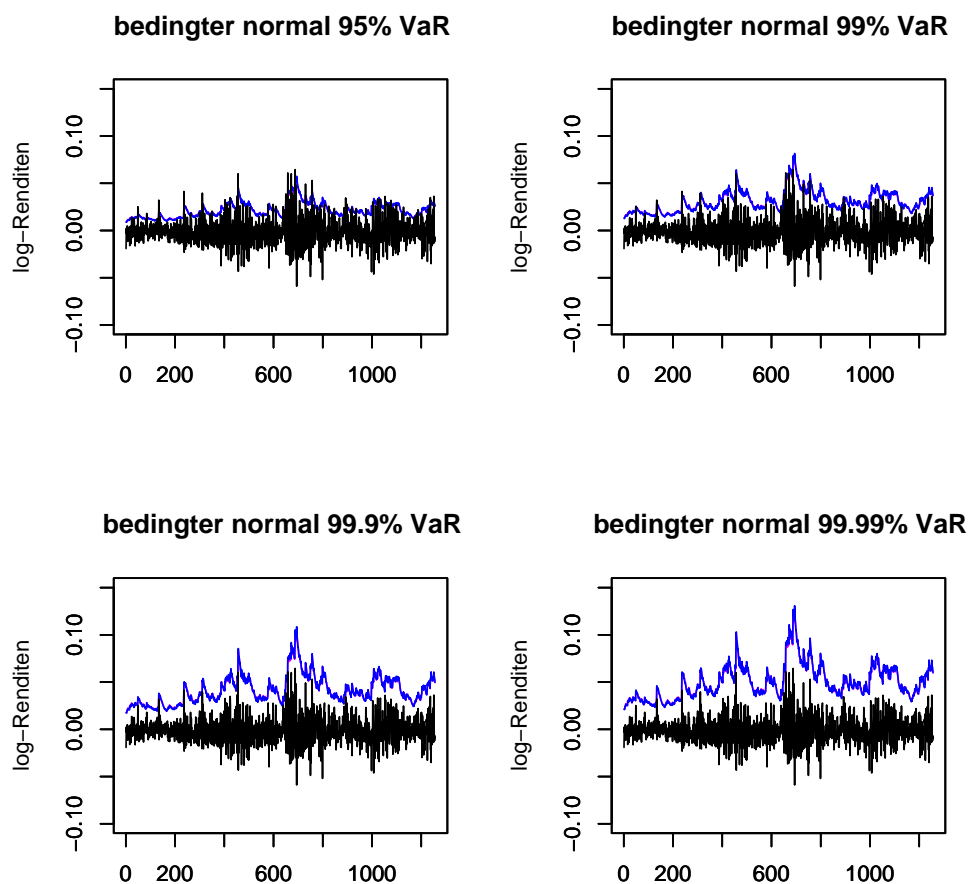


Abbildung 5.7: Log-Renditen mit den Schätzern für den bedingten normal VaR der AR(1)-GARCH(1,1)-Anpassung (pink) und den bedingten normal VaR der GARCH(1,1)-Anpassung (blau) (für verschiedene Konfidenzintervalle).



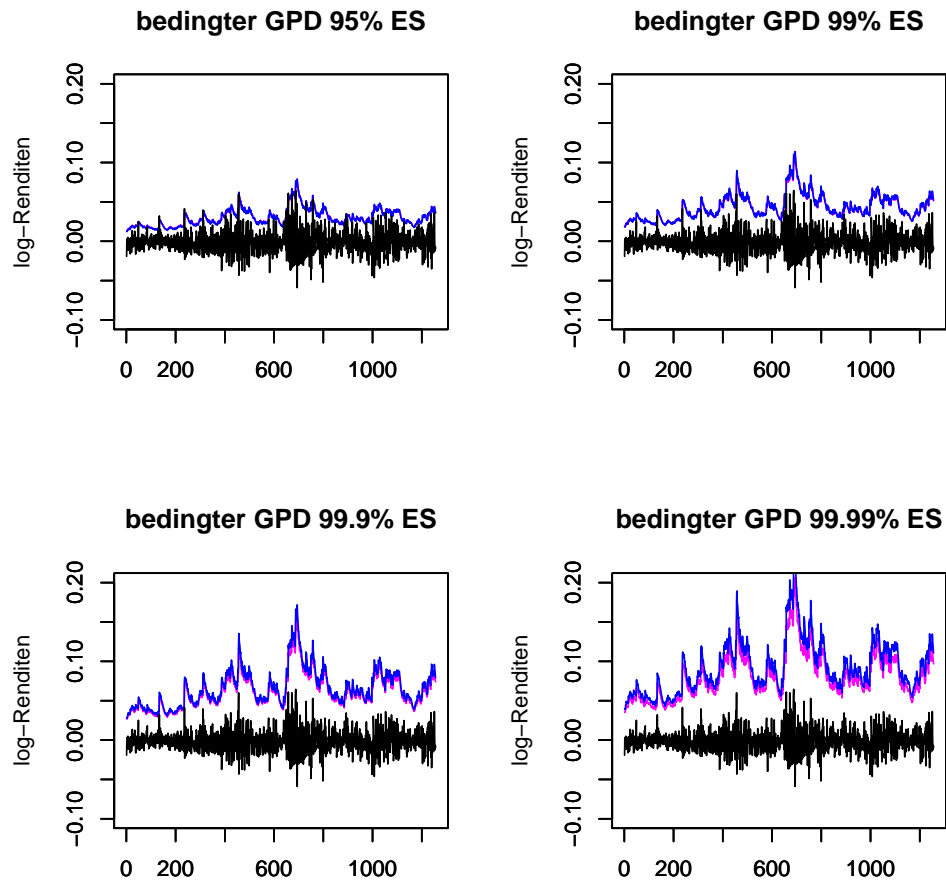


Abbildung 5.8: Log-Renditen mit den Schätzern für den bedingten GPD ES der AR(1)-GARCH(1,1)-Anpassung (pink) und den bedingten GPD ES der GARCH(1,1)-Anpassung (blau) (für verschiedene Konfidenzintervalle).

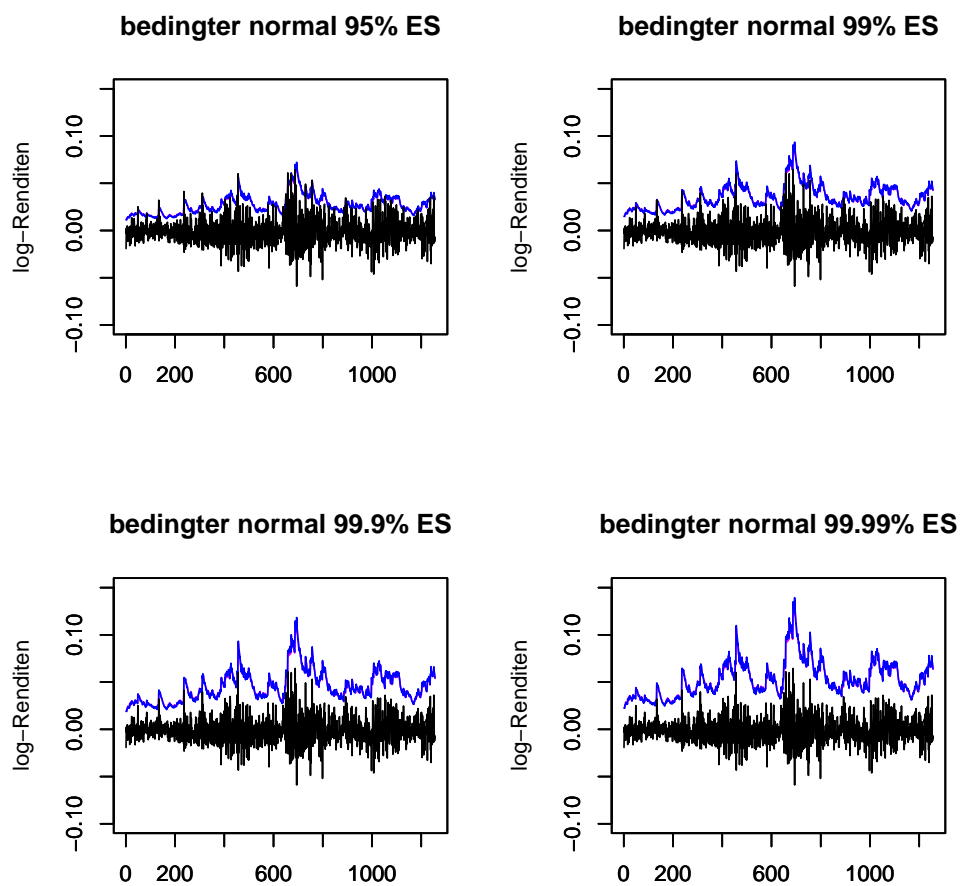


Abbildung 5.9: Log-Renditen mit den Schätzern für den bedingten normal ES der AR(1)-GARCH(1,1)-Anpassung (pink) und den bedingten normal ES der GARCH(1,1)-Anpassung (blau) (für verschiedene Konfidenzintervalle).

Zählt man für die einzelnen Methoden und Konfidenzintervalle wieder die Störungen, also die Werte für die  $x_{t+1} > \hat{x}_q^t$  gilt, erhält man ein ähnliches Ergebnis wie in Abschnitt 4.5. Der bedingte GPD VaR der GARCH(1,1)-Anpassung liefert für jedes Konfidenzintervall exakt die gleiche Anzahl an Störungen wie der bedingte GPD VaR der AR(1)-GARCH(1,1)-Anpassung. Beim bedingten normal VaR der GARCH(1,1)-Anpassung sieht es ähnlich aus, dort tritt bei einigen Konfidenzintervallen lediglich eine Störung mehr als beim bedingten normal VaR der AR(1)-GARCH(1,1)-Anpassung auf.

## 5.4 Abschließender Vergleich

Abschließend soll die Frage geklärt werden, welche Parameter ausschlaggebend sind für die leichten Abweichungen zwischen den Methoden. Wie bereits festgestellt, liegen die Schätzer des GARCH(1,1)-Prozesses, also  $\alpha_0$ ,  $\alpha_1$  und  $\beta$  für beide Methoden sehr nah beieinander. Daher sind die Werte für  $\sigma_t$  im Vergleich auch sehr ähnlich. Eine leichte Abweichung gibt es bei  $\mu_t$  bzw.  $\mu$ , denn beim bedingten VaR der GARCH(1,1)-Anpassung ist  $\mu$  konstant  $-0.0008242146$ . Beim bedingten VaR der AR(1)-GARCH(1,1)-Anpassung hingegen nimmt  $\mu_t$  verschiedene Werte an. Dabei beträgt  $\mu_t$  im Durchschnitt ca.  $-0.00001230768$  und ist somit im Durchschnitt etwas größer als das  $\mu$  der GARCH(1,1)-Anpassung. Der kleinste Wert, den  $\mu_t$  in dieser Anpassung annimmt, ist  $-0.0008807802$ . Dieser Wert liegt nah an  $\mu$ , ist aber noch etwas kleiner. Also ist festzuhalten, dass das  $\mu_t$  für die AR(1)-GARCH(1,1)-Schätzung insgesamt etwas größer ist als das  $\mu$  für die GARCH(1,1)-Schätzung. Dieser Unterschied macht aber im Vergleich des bedingten VaR und ES der Methoden nicht viel aus, da  $\mu$  bzw.  $\mu_t$  in der Formel des bedingten VaR (bzw. des bedingten ES) „nur“ hinzu addiert wird und die Abweichungen im Vergleich zu den anderen Werten minimal sind. Entscheidender ist der Unterschied des VaR und des ES der Residuen. Dieser beträgt z.B. für den 99.99% GPD ES 0.900371 und für den 99.99% GPD VaR 0.601698. Dies sind die größten Unterschiede im Vergleich des VaR und des ES der Residuen. Für diese beiden Maße sind auch in den Abbildungen (Abbildung 5.6 und 5.8) Unterschiede zu erkennen. Minimale Unterschiede sind auch beim bedingten 99.9% GPD VaR und beim bedingten 99.9% GPD ES zu erkennen. Hier beträgt der Unterschied des VaR bzw. des ES der Residuen 0.197609 bzw. 0.3410931. Bei den anderen Vergleichen war kein Unterschied in den Abbildungen des bedingten VaR bzw. des bedingten ES zu erkennen. Der Unterschied des VaR bzw. des ES der Residuen ist in diesen Fällen auch immer kleiner als 0.11. In den meisten Fällen ist er sogar deutlich kleiner als 0.11 und liegt ca. bei 0.07.

Zusammenfassend ist festzuhalten, dass die Berechnung der Residuen und die entsprechende GPD Anpassung mit der schließlich der VaR und der ES der Residuen geschätzt wird, entscheidend für die Bestimmung des bedingten VaR und des be-

dingten ES ist. Die Schätzer für  $\mu_t$  (bzw.  $\mu$ ) und  $\sigma_t$  haben in diesem Vergleich keine große Bedeutung, da sie nur verhältnismäßig kleine Unterschiede aufweisen und wegen ihrer Größe ohnehin nicht wesentlich in die Schätzung eingehen.

Insgesamt kann für diesen Beispieldatensatz festgestellt werden, dass für die Schätzung des bedingten VaR und des bedingten ES auf den AR(1)-Prozess verzichtet werden kann ohne das Ergebnis zu verschlechtern. Da sich das Ergebnis aber durch das Weglassen des AR(1)-Prozesses auch nicht verbessert hat, kann ebenso gut die herkömmliche Methode verwendet werden. Der einzige Vorteil der GARCH(1,1)-Anpassung ist, dass sie etwas weniger zeitaufwändig ist, als die AR(1)-GARCH(1,1)-Anpassung.

# Fazit

Ziel dieser Arbeit war es, die Peak over Threshold Methode mit einer Anwendung vorzustellen und zu überprüfen, ob diese Methode für die Abschätzung von Renditezeitreihen gute bis sehr gute Resultate liefert.

Dazu wurde die Methode in Kapitel 1 zunächst für den einfachsten Fall, also unabhängig und identisch verteilte Zufallsvariablen eingeführt. Es wurde gezeigt, wie ein geeigneter Threshold gewählt werden kann, so dass die Überschreitungen dieses Thresholds nach dem Pickands-Balkema-de Haan Theorem durch eine verallgemeinerte Paretoverteilung modelliert werden können. Darüber hinaus wurde gezeigt, wie die Parameter der verallgemeinerten Paretoverteilung mit Hilfe der Maximum-Likelihood Methode geschätzt werden, und wie damit der Value at Risk und der Expected Shortfall der Daten berechnet werden kann. Die Anwendung der Methode auf die Log-Renditen der DAX-Abschlusswerte vom 01.01.1996 bis zum 31.12.2000 in Kapitel 2 lieferte bereits für diesen Fall sehr gute Ergebnisse. Darüber hinaus wurde, für den iid Fall, gezeigt dass die Anpassung einer heavy tail Verteilung (also in diesem Fall einer verallgemeinerten Paretoverteilung) an die Daten einen deutlich besseren Schätzer für den Value at Risk und den Expected Shortfall liefert, als die Anpassung einer Normalverteilung.

Da Renditezeitreihen häufig eine sehr komplexe Abhängigkeitsstruktur aufweisen, konnten die Ergebnisse noch nicht als zufriedenstellend angesehen werden und es wurde in Kapitel 3 die Peak over Threshold Methode für stationäre Zeitreihen eingeführt. Dazu wurden zunächst einige Grundlagen der Zeitreihenanalyse vorgestellt. Es wurde gezeigt, wie mit Hilfe einer Pseudo-Maximum-Likelihood Schätzung ein  $AR(1)$ -GARCH(1,1)-Prozess an Renditezeitreihen angepasst werden kann und wie mit den Parameterschätzern des  $AR(1)$ -GARCH(1,1)-Prozesses 1-Schritt Vorhersagen für die bedingte Varianz und den bedingten Erwartungswert geschätzt werden können. Darüber hinaus wurde gezeigt, wie mit Hilfe der 1-Schritt Vorhersagen Residuen berechnet werden können, auf die schließlich die Peak over Threshold Methode für unabhängig und identisch verteilte Zufallsvariablen angewendet werden kann. Mit Hilfe des Value at Risk und des Expected Shortfalls der Residuen wurde eine Formel für den bedingten Value at Risk und den bedingten Expected Shortfall der stationären Zeitreihe vorgestellt. Die Anwendung dieser Methode auf die DAX-Abschlusswerte vom 01.01.1996 bis zum 31.12.2000 in Kapitel 4 lieferte auch in diesem Fall sehr gute Ergebnisse. Da aus dieser Anpassung ein „variabler“ Value at Risk resultiert, kann dieser sich schneller an Volatilitätsschwankungen anpassen und somit das Risiko von Kurs-

verlusten besser beschreiben als der statische Value at Risk aus der Peak over Threshold Methode für iid Zufallsvariablen. Zusätzlich wurde zum Vergleich wieder eine Normalverteilung an die Daten angepasst. Auch hier schnitt die Peak over Threshold Methode mit der Anpassung einer verallgemeinerten Paretoverteilung deutlich besser ab.

Weil sich bei der Anpassung des AR(1)-GARCH(1,1)-Prozesses in dem Anwendungsbeispiel herausstellte, dass der AR(1)-Prozess nicht signifikant in die Schätzung eingeht, wurde schließlich in Kapitel 5 noch eine alternative Anwendung der Peak over Threshold Methode für stationäre Zeitreihen vorgestellt, bei der „nur“ ein GARCH(1,1)-Prozess an die Log-Renditen der DAX-Abschlusswerte angepasst wurde. Diese Methode lieferte ein sehr ähnliches Ergebnis wie die herkömmliche Methode. Nur bei der Betrachtung des Value at Risk oder des Expected Shortfalls für ein Konfidenzintervall von 99.99% waren leichte Unterschiede zwischen den Methoden zu erkennen.

Insgesamt ist die vorgestellte Peak over Threshold Methode für stationäre Zeitreihen sehr zu empfehlen. Dabei kann, wenn es die Anpassung zulässt, auf die Verwendung eines autoregressiven Prozesses verzichtet werden ohne das Ergebnis zu verschlechtern. Darüber hinaus liefert die Peak over Threshold Methode für unabhängig und identisch verteilte Zufallsvariablen, trotz ihrer Einfachheit, verhältnismäßig gute Ergebnisse und ist daher, z.B. für einen schnellen Vergleich zweier Investitionen, ebenfalls zu empfehlen.

# Anhang

Für die Anwendung mit R wurden unter anderem folgende Funktionen verwendet:

Aus dem Paket evd:

mrlplot	Mit dieser Funktion wird ein Mean Residual Life Plot erstellt. (Siehe z.B. Abbildung 2.3, S.32)
tcplot	Die Funktion tcplot erstellt einen Threshold Choice Plot. (Siehe z.B. Abbildung 2.5 oder Abbildung 2.6, S.34)
hist	Mit diesem Befehl kann ein Histogramm erzeugt werden. (Siehe z.B. Abbildung 2.7 (links), S.36)
plot.ecdf	ecdf ist die Abkürzung für empirical cumulative distribution function. plot.ecdf zeichnet also die empirische Verteilungsfunktion. (Siehe z.B. Abbildung 2.7 (rechts), S.36)
pgpd	gpd bezeichnet die verallgemeinerte Paretoverteilung (generalized pareto distribution) und das p steht für die Verteilungsfunktion. Mit pgpd kann also die Verteilungsfunktion einer verallgemeinerten Paretoverteilung in eine Graphik eingezeichnet werden. (Siehe z.B. Abbildung 2.7 (rechts), S.36)
dgpd	gpd bezeichnet die verallgemeinerte Paretoverteilung und das d steht für die Dichte. Mit dgpd wird also die Dichte einer verallgemeinerten Paretoverteilung in eine Graphik eingezeichnet. (Siehe z.B. Abbildung 2.7 (links), S.36)

Aus dem Paket fGarch:

garchFit	Mit diesem Befehl kann ein GARCH-Prozess (oder auch ein AR-GARCH-Prozess) an einen Datensatz angepasst werden.
----------	--

Aus dem Paket `evir`:

<code>gpd</code>	Mit diesem Befehl kann eine verallgemeinerte Paretoverteilung an einen Datensatz angepasst werden. In der Standardeinstellung werden die Parameter durch die Maximum-Likelihood-Methode geschätzt.
<code>riskmeasures</code>	Mit <code>riskmeasures</code> kann der Value at Risk und der Expected Shortfall einer mit <code>gpd</code> angepassten verallgemeinerten Paretoverteilung berechnet werden. Dabei können verschiedene Konfidenzintervalle angegeben werden.
<code>tailplot</code>	Diese Funktion erstellt einen <code>tailplot</code> einer mit <code>gpd</code> angepassten verallgemeinerten Paretoverteilung. (Siehe z.B. Abbildung 2.8, S.40)
<code>gpd.q</code>	Mit diesem Befehl wird die Profile Likelihood Kurve für den VaR und ein 95%-Konfidenzintervall für den Schätzer in die <code>tailplot</code> -Graphik eingefügt. (Siehe z.B. Abbildung 2.8, S.40)
<code>gpd.sfall</code>	Mit diesem Befehl wird die Profile Likelihood Kurve für den ES und ein 95%-Konfidenzintervall für den Schätzer in die <code>tailplot</code> -Graphik eingefügt. (Siehe z.B. Abbildung 2.8, S.40)

Aus dem Paket `stats`:

<code>acf</code>	Mit dieser Funktion wird ein Autokorrelationsfunktionsplot erstellt. (Siehe z.B. Abbildung 4.3, S.69)
<code>Box.test</code>	Führt einen Box-Pierce- oder einen Ljung-Box-Test für eine angegebene Lag-Länge durch.
<code>qqnorm</code>	Mit dem Befehl <code>qqnorm</code> wird ein QQ-Normal-Plot erstellt. (siehe z.B. Abbildung 2.2, S.31)
<code>qqline</code>	Mit <code>qqline</code> wird eine Linie in den QQ-Normal-Plot eingefügt um festzustellen, ob die Werte von der Linie abweichen. (siehe z.B. Abbildung 2.2, S.31)
<code>ks.test</code>	Führt einen Kolmogorov-Smirnov-Test für einen angegebenen Datensatz und eine angegebene Verteilung durch.



# Literaturverzeichnis

- [1] ARTZNER, P., DELBAEN, F., EBER, J.-M. & HEATH, D.: *Thinking coherently*. RISK, 10(11):68–71, 1997.
- [2] ARTZNER, P., DELBAEN, F., EBER, J.-M. & HEATH, D.: *Coherent measures of risk*. Math. Fin., 9(3):203–228, 1999.
- [3] BOLLERSLEV, T.: *Generalized autoregressive conditional heteroscedasticity*. Journal of Econometrics, 31:307–327, 1986.
- [4] COLES, S.: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2007.
- [5] EISELE, B.: *Value-at-Risk-basiertes Risikomanagement in Banken - Portefeuilleentscheidungen, Risikokapitalallokation und Risikolimitierung unter Berücksichtigung des Bankenaufsichtsrechts*. Deutscher Universitäts-Verlag, 2004.
- [6] EMBRECHTS, P., KLÜPPELBERG, C. & MIKOSCH, T.: *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.
- [7] ENGLE, R.F.: *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. Econometrica, 50(4):987–1007, 1982.
- [8] FISHER, R. & TIPPET, L.: *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. Proceedings of the Cambridge Philosophical Society 24, 180-190, 1928.
- [9] GOURIÉROUX, C.: *ARCH Models and Financial Applications*. Springer, 1997.
- [10] HANISCH, J.: *Risikomessung mit dem conditional Value-at-Risk - Implikationen für das Entscheidungsverhalten*. Kovač, 2006.
- [11] HOSKING, J.R.M. & WALLIS, J.R.: *Parameter and Quantile Estimation for the Generalized Pareto Distribution*. Technometrics, 29(3):339–349, August 1987.
- [12] JOCKUSCH, A.: *Value-at-risk-Modelle für Aktienportfolios auf der Basis der Varianz-Kovarianz-Methode*. Lang, 2002.
- [13] LÖWE, M.: *Skript zur Extremwerttheorie*, Westfälische Wilhelms-Universität Münster, WS 2008/09.

- [14] MCNEIL, A.J. & FREY, R.: *Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach*, April 2000.
- [15] MCNEIL, A.J. & SALADIN, T.: *The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions*, April 1997.
- [16] MCNEIL, A.J.: *Extreme Value Theory for Risk Managers*, Mai 1999.
- [17] MCNEIL, A.J., FREY, R. & EMBRECHTS, P.: *Quantitative Risk Management - Concepts, Techniques and Tools*. Princeton University Press, 2005.
- [18] MEYER, C.: *Value at Risk für Kreditinstitute*. Deutscher Universitäts-Verlag, 1999.
- [19] REISS, R.-D. & THOMAS, M.: *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and other Fields*. Birkhäuser, 1997.
- [20] ROOTZÉN, H. & TAJVIDI, N.: *Extreme value statistics and wind storm losses: a case study*, Oktober 1995.
- [21] SCHMID, F. & TREDE, M.: *Finanzmarktstatistik*. Springer, 2006.
- [22] TSAY, R.S.: *Analysis of financial time series*. Wiley, 2002.