

Bachelorarbeit

Tail-Risikomaße: Eigenschaften und Backtesting

Oliver Mühlbach
16. September 2025

Philipps-Universität Marburg
Fachbereich Mathematik und Informatik
Prof. Dr. H. Holzmann

Inhaltsverzeichnis

1	Einführung	1
2	Risikomaße	2
2.1	Definition und Eigenschaften	2
2.2	Beispiele für Risikomaße	3
3	Tail-Risikomaße	5
3.1	Definitionen	5
3.2	Beziehung zwischen Tail-Risikomaß und Erzeuger	10
4	Identifizierbarkeit und Elizitierbarkeit	16
4.1	Identifizierbarkeit	16
4.2	Elizitierbarkeit	19
5	Backtesting	25
5.1	Komparatives Backtesting	27
5.2	Traditionelles Backtesting	27
6	Numerische Studie	28
6.1	Scores und Diebold-Mariano Tests	28
6.2	Kalibrierungstests	31
7	Fazit	33
A	Zeitreihenanalyse	34
A.1	AR-GARCH-Modelle	34
A.2	Vorhersagen im AR-GARCH-Modell	37

1 Einführung

Die quantitative Bewertung finanzieller Risiken ist ein zentraler Bestandteil moderner Risikosteuerung. Besonders Banken und Versicherungen sind darauf angewiesen, potenzielle Verluste zuverlässig abzuschätzen, um regulatorische Vorgaben einzuhalten und ihre Stabilität zu sichern. Solche Verluste werden als Zufallsvariablen modelliert und mit Hilfe spezifischer Risikomaße quantifiziert.

In Kapitel 2 führen wir Risikomaße ein, diskutieren zentrale Eigenschaften wie Kohärenz und Konvexität und stellen einige Beispiele dar. Anschließend betrachten wir in Kapitel 3 die sogenannten Tail-Risikomaße, welche insbesondere den Randbereich der Verlustverteilungen erfassen. Wir erläutern die Zusammenhänge zwischen allgemeinen Risikomaßen und Tail-Risikomaßen und stellen zentrale Beziehungen zwischen beiden Konzepten dar.

In Kapitel 4 widmen wir uns den Konzepten der Identifizierbarkeit und Elizitierbarkeit. Diese bilden die theoretische Grundlage für die spätere statistische Überprüfung von Risikomaßen. Ein Schwerpunkt liegt hierbei auf den Resultaten von [Fissler et al., 2024], insbesondere Satz 4.1.7 und Satz 4.2.9, welche aufzeigen, dass unter bestimmten Voraussetzungen für jedes identifizierbare bzw. elizitierbare Risikomaß auch das zugehörige Tail-Risikomaß identifizierbar bzw. elizitierbar ist.

Darauf aufbauend behandeln wir in Kapitel 5 die theoretischen Grundlagen des komparativen und traditionellen Backtestings. Diese werden anschließend in Kapitel 6 praktisch angewandt. Dort analysieren wir mit Hilfe geeigneter Scoring- und Identifikationsfunktionen aus Kapitel 4 die Prognosegüte verschiedener Schätzmethoden. Neben einem Vergleich der Methoden mittels Diebold-Mariano-Test untersuchen wir auch, ob die Vorhersagen im Durchschnitt konsistent sind oder ob systematische Abweichungen auftreten.

Ziel dieser Arbeit ist es somit, die theoretischen Grundlagen von Risikomaßen und Tail-Risikomaßen mit Konzepten wie Identifizierbarkeit und Elizitierbarkeit zu verbinden und diese im Rahmen von Backtesting-Verfahren sowohl theoretisch als auch empirisch zu untersuchen.

2 Risikomaße

Bemerkung 2.0.1. Angenommen, wir betrachten eine Vermögensposition mit einem zeitabhängigen Wert. In zukünftigen Zeitpunkten kann diese Position zufälligen Schwankungen unterliegen, die sowohl Gewinne als auch Verluste verursachen können. Aus Sicht des Risikomanagements sind jedoch vor allem potenzielle Verluste relevant. Wir beschäftigen uns daher ausschließlich mit diesen und bezeichnen den Verlust mit L .

Dabei handelt es sich um eine messbare Funktion bzw. um eine Zufallsvariable $L : \Omega \rightarrow \mathbb{R}$ auf einem messbaren Raum (Ω, \mathcal{A}) . Die Unsicherheit über den tatsächlichen Verlust wird durch ein Wahrscheinlichkeitsmaß \mathbb{P} auf (Ω, \mathcal{A}) modelliert. Die daraus resultierende Verteilung von L bezeichnen wir als *Verlustverteilung*. (Positive Werte von L entsprechen dabei negativen Auswirkungen auf die Vermögensposition.)

Ein *Risikomaß* ist eine Abbildung ρ , die einem zufälligen Verlust L eine reelle Zahl $\rho(L)$ zuordnet. Diese Zahl liefert eine quantitative Bewertung des mit dem Verlust verbundenen Risikos: Je größer $\rho(L)$, desto riskanter wird die Position eingeschätzt. Man kann $\rho(L)$ auch als Kapitalreserve interpretieren, die ein Unternehmen vorhalten muss, um potenzielle Verluste auszugleichen.

2.1 Definition und Eigenschaften

In diesem Abschnitt wollen wir uns mit der grundlegenden Definition von Risikomaßen und deren Eigenschaften beschäftigen.

(Wir orientieren uns an der Darstellung aus [Holzmann, 2023], [Liu and Wang, 2021] und [Fissler et al., 2024].)

Sei \mathcal{X}^q die Menge aller Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit endlichem q -ten Moment, $q \in [0, \infty)$. Sei weiter \mathcal{X}^∞ die Menge der fast-überall beschränkten Zufallsvariablen. Ein positiver Wert von $X \in \mathcal{X}^0$ steht also für einen finanziellen Verlust. Weiter sei \mathcal{M}^0 die Menge aller Borel-messbaren Verteilungsfunktionen auf \mathbb{R} . \mathcal{M}^q ist die Menge aller Verteilungsfunktionen auf \mathbb{R} mit endlichem q -ten Moment, $q \in [0, \infty)$ und \mathcal{M}^∞ die Menge aller Verteilungen mit kompaktem Träger.

Definition 2.1.1 (Risikomaß). Sei \mathcal{L} ein konvexer Kegel, der \mathcal{X}^∞ enthält. Es gilt also für $\lambda > 0, L, L_1, L_2 \in \mathcal{L}$, dass $\lambda L \in \mathcal{L}$ und $L_1 + L_2 \in \mathcal{L}$. Ein *Risikomaß* ist eine Abbildung

$$\rho : \mathcal{L} \rightarrow \mathbb{R}$$

Wir können $\rho(L)$ als Reserve interpretieren, um L auszugleichen.

Definition 2.1.2 (Eigenschaften von Risikomaßen). Für $L_1, L_2 \in \mathcal{L}$ heißt ein Risikomaß $\rho : \mathcal{L} \rightarrow \mathbb{R}$:

1. *positiv homogen*, falls für $\lambda \geq 0$ gilt, dass $\rho(\lambda L) = \lambda \rho(L)$,
2. *translationsinvariant*, falls für $\mu \in \mathbb{R}$ gilt, dass $\rho(\mu + L) = \mu + \rho(L)$,
3. *monoton*, falls für $L_1 \leq L_2$ gilt, dass $\rho(L_1) \leq \rho(L_2)$,

4. *konvex*, falls für $\lambda \in [0, 1]$ gilt, dass $\rho(\lambda L_1 + (1 - \lambda)L_2) \leq \lambda\rho(L_1) + (1 - \lambda)\rho(L_2)$,
5. *subadditiv*, falls $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$ und
6. *komonoton additiv*, falls $\rho(L_1 + L_2) = \rho(L_1) + \rho(L_2)$ gilt, wenn L_1 und L_2 komonoton sind. Zwei Zufallsvariablen L_1 und L_2 heißen *komonoton*, wenn es eine Menge $\Omega_0 \in \mathcal{A}$ mit $\mathbb{P}(\Omega_0) = 1$ gibt so, dass für alle $\omega, \omega' \in \Omega_0$ gilt: $(L_1(\omega) - L_1(\omega'))(L_2(\omega) - L_2(\omega')) \geq 0$.

Bemerkung 2.1.3. Die Eigenschaften *positive Homogenität* und *Konvexität* implizieren *Subadditivität*.

Für weitere Eigenschaften benötigen wir folgende Konzepte:

Definition 2.1.4. Für $L_1, L_2 \in \mathcal{L}^0$ gilt, dass L_1 kleiner als L_2 in *stochastischer Ordnung* ist (Notation: $L_1 \prec_{st} L_2$), genau dann, wenn $\mathbb{E}[f(L_1)] \leq \mathbb{E}[f(L_2)]$ für alle monoton wachsenden Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ gilt und der Erwartungswert existiert.

Bemerkung 2.1.5. Äquivalent zur Bedingung für stochastische Ordnung $L_1 \prec_{st} L_2$ ist Folgendes:

Für alle $t \in \mathbb{R}$ gilt

$$\begin{aligned} \mathbb{P}(L_1 \geq t) &\leq \mathbb{P}(L_2 \geq t) \\ \Leftrightarrow F_{L_1}(t) &\geq F_{L_2}(t), \quad \text{falls } F_{L_1} \text{ und } F_{L_2} \text{ stetig.} \end{aligned}$$

Definition 2.1.6. Für $L_1, L_2 \in \mathcal{L}^1$ gilt, dass L_1 kleiner als L_2 in *konvexer Ordnung* ist (Notation: $L_1 \prec_{cx} L_2$), falls $\mathbb{E}[\phi(L_1)] \leq \mathbb{E}[\phi(L_2)]$ für alle monoton wachsenden, konvexen Funktionen $\phi : \mathbb{R} \rightarrow \mathbb{R}$ und zusätzlich $\mathbb{E}[L_1] = \mathbb{E}[L_2]$ ist.

Dies führt zu folgenden weiteren Eigenschaften für Risikomaße:

Definition 2.1.7 (Weiterführung von (2.1.2)). Für $L_1, L_2 \in \mathcal{L}$ heißt ein Risikomaß $\rho : \mathcal{L} \rightarrow \mathbb{R}$:

7. *\prec_{st} -monoton*, falls für $L_1 \prec_{st} L_2$ gilt, dass $\rho(L_1) \leq \rho(L_2)$,
8. *\prec_{cx} -monoton*, falls für $L_1 \prec_{cx} L_2$ gilt, dass $\rho(L_1) \leq \rho(L_2)$.

Definition 2.1.8. Ein *kohärentes Risikomaß* erfüllt die Eigenschaften 1., 2., 3. und 4. (5.), ein *konvexes Risikomaß* erfüllt die Eigenschaften 2., 3. und 4. und ein *monetäres Risikomaß* erfüllt die Eigenschaften 2. und 3. Man beachte hierbei, dass mit einem konvexen Risikomaß nicht allein gemeint ist, dass das Risikomaß konvex ist.

2.2 Beispiele für Risikomaße

Die zwei wohl bekanntesten Risikomaße sind der *Value at Risk* (VaR) und der *Expected Shortfall* (ES). Sie kommen in der Praxis vor allem in Banken und Versicherungen zum Einsatz.

Definition 2.2.1 (Value at Risk). Der *Value at Risk* $VaR_p(L)$ ist der kleinstmögliche Verlustbetrag $l \in [-\infty, \infty]$, der vom Verlust L mit gegebener Wahrscheinlichkeit von höchstens $1 - p$ überschritten wird:

$$\begin{aligned} VaR_p(L) &= \inf\{l \in \mathbb{R} \mid \mathbb{P}(L > l) \leq 1 - p\} \\ &= \inf\{l \in \mathbb{R} \mid \mathbb{P}(L \leq l) \geq p\} \\ &= \inf\{l \in \mathbb{R} \mid F_L(l) \geq p\} \\ &=: F_L^{-1}(p) \end{aligned}$$

Der VaR_p ist also genau die Quantilsfunktion der Verteilungsfunktion F_L .

Definition 2.2.2. Für ein Konfidenzniveau p unterscheidet man zwischen dem *Links-* und *Rechtsquantil*:

$$\begin{aligned} VaR_p^+(L) &= \inf\{l \in \mathbb{R} \mid F_L(l) > p\} = F_L^{-1}(p+), \\ VaR_p^-(L) &= \inf\{l \in \mathbb{R} \mid F_L(l) \geq p\} = F_L^{-1}(p) \end{aligned}$$

Das p -Quantil einer Verteilungsfunktion ist dann das Intervall

$$Q_p(L) = [VaR_p^-(L), VaR_p^+(L)].$$

Bemerkung 2.2.3. Falls die Funktion F_L^{-1} im Punkt p stetig ist, sind die beiden Grenzen gleich. Es gilt dann also $VaR_p^-(L) = VaR_p^+(L)$.

Bemerkung 2.2.4. Der Value-at-Risk ist ein monetäres Risikomaß.

Definition 2.2.5 (Expected Shortfall). Der *Expected Shortfall* $ES_p(L)$ einer Vermögensposition ist definiert als

$$ES_p(L) = \frac{1}{1-p} \int_p^1 VaR_u^+(L) du.$$

Man kann den *Expected Shortfall* $ES_p(L)$ als den erwarteten Verlust bei Überschreitung des VaRs interpretieren (Siehe [Holzmann, 2023]).

Bemerkung 2.2.6. Der Expected Shortfall $ES_p(L)$ ist unabhängig von der Wahl der Quantilgrenze, selbst wenn F_L^{-1} im Punkt p nicht stetig ist, da $VaR_p^+(L) \neq VaR_p^-(L)$ nur in höchstens abzählbar vielen $p \in [0, 1]$ sein kann.

Bemerkung 2.2.7. Der Expected Shortfall ist ein kohärentes Risikomaß.

Ein weiteres Risikomaß ist die Standardabweichung $\sigma(L)$, die sowohl positive als auch negative Abweichungen vom Erwartungswert gleichmäßig berücksichtigt.

Definition 2.2.8 (Standardabweichung). Das Risikomaß der Standardabweichung ist für $\beta > 0$ definiert als:

$$SD_\beta(L) = \mathbb{E}[L] + \beta \sqrt{var(L)}, \quad L \in \mathcal{L}^2. \quad (\text{SD})$$

Definition 2.2.9 (Verteilungsinvarianz). Wenn $L, X \in \mathcal{L}$ und $L \stackrel{d}{=} X$, also $F_L(x) = F_X(x) \forall x \in \mathbb{R}$, dann ist $\rho(L) = \rho(X)$. Der Wert von ρ hängt also nur von der Verteilung der Zufallsvariablen ab.

Mit dieser Eigenschaft können wir nun das vorher definierte Risikomaß analog als Abbildung von den Verteilungsfunktionen auf die reellen Zahlen interpretieren:

Bemerkung 2.2.10. Sei $\mathcal{M} := \{F_L \in \mathcal{M}^0 \mid L \in \mathcal{L}\}$. Dann ist

$$\hat{\rho} : \mathcal{M} \rightarrow \mathbb{R}, \quad \hat{\rho}(F) := \rho(L) \quad \text{für } L \in \mathcal{L} \text{ mit } F_L = F.$$

Wir werden später ρ und $\hat{\rho}$ nicht mehr unterscheiden und nur noch ρ nennen.

3 Tail-Risikomaße

3.1 Definitionen

In Kapitel 2.1 haben wir definiert, was ein Risikomaß ist und welche grundlegenden Eigenschaften es erfüllen kann. Nun betrachten wir Verlustfunktionen und Risikomaße im Bereich oberhalb eines vorgegebenen Konfidenzniveaus $p \in (0, 1)$. Wir bezeichnen dies als das *Tail-Risiko*.

Darunter verstehen wir das Verhalten der Verlustfunktion L in den Ausprägungen, die mit einer Wahrscheinlichkeit von höchstens $1 - p$ eintreten, also im rechten Rand („Tail“) der Verteilung. In diesem Abschnitt wollen wir zuerst die Definitionen hinter Tail-Risikomaßen erläutern und uns danach mit Eigenschaften und Zusammenhängen zwischen Risikomaßen und Tail-Risikomaßen beschäftigen.

Definition 3.1.1. Für $p \in [0, 1]$ ist L_p das *p-Tail-Risiko* (*p-Tail-Verlust*) von L hinter dem p -quantil:

$$L_p := F_L^{-1}(p + (1 - p)U_L)$$

U_L ($:= F_L(L)$) ist hierbei wie in [Liu and Wang, 2021] die Uniform verteilte Zufallsvariable auf $[0, 1]$ so, dass $F_L^{-1}(U_L) = L$ fast sicher für alle L gilt.

Definition 3.1.2. Sei $p \in [0, 1]$. Die Verteilungsfunktion F_{L_p} von L_p heißt *p-Tail-Verteilungsfunktion* und lässt sich darstellen als:

$$F_{L_p}(l) = \mathbb{P}(L_p \leq l) = \mathbb{P}(L \leq l \mid U_L \geq p) = \frac{(\mathbb{P}(L \leq l) - p)_+}{1 - p} = \frac{(F(l) - p)_+}{1 - p}, \quad l \in \mathbb{R}.$$

Bemerkung 3.1.3. Es gilt

$$\begin{aligned} \mathbb{P}(L \leq l, U_L > p) &= \mathbb{P}(U_L \leq F_L(l), U_L > p) \\ &= \mathbb{P}(p < U_L \leq F_L(l)) \\ &= \max(0, F_L(l) - p) = (\mathbb{P}(L \leq l) - p)_+. \end{aligned}$$

Beispiel 3.1.4 (Standardnormalverteilung). Sei $X \sim \mathcal{N}(0, 1)$ eine standardnormalverteilte Zufallsvariable mit der Verteilungsfunktion

$$F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt$$

und $p = 0,95$. Dann ist $\Phi^{-1}(p) \approx 1,645$ und daher ist die p -Tail-Verteilungsfunktion gegeben durch

$$F_{X_p}(x) = \begin{cases} 0 & , \text{ falls } x \leq 1,645 \\ \frac{\Phi(x) - 0,95}{0,05} & , \text{ sonst.} \end{cases}$$

(Siehe Abbildung 1.)

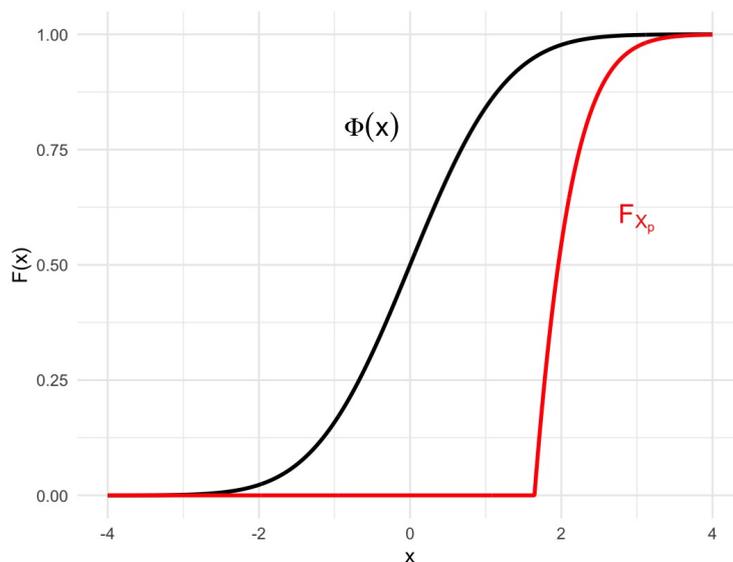


Abbildung 1: Vergleich von Standardnormalverteilung und p -Tail-Verteilungsfunktion

Beispiel 3.1.5 (t-Verteilung). Seien die Zufallsvariablen Y_0, Y_1, \dots, Y_k stochastisch unabhängig und je $\mathcal{N}(0, 1)$ -verteilt. Dann ist die durch

$$X_k := \frac{Y_0}{\sqrt{\frac{1}{k} \sum_{j=1}^k Y_j^2}}$$

erzeugte Zufallsvariable X_k mit Dichte

$$f_k(x) = \frac{1}{\sqrt{\pi k}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}},$$

wobei

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

(siehe Abbildung 2), eine (*Student'sche*) *t*-Verteilung mit *k* Freiheitsgraden. Die Verteilungsfunktion von X_k ist dann gegeben durch

$$F_k(x) = \int_{-\infty}^x f_k(x) dx.$$

Sei wieder $p = 0,95$. Für $k = 1$ ist $F_1^{-1}(p) \approx 6,3138$ und für $k = 5$ ist $F_5^{-1}(p) \approx 2,0150$ und daher ist die *p*-Tail-Verteilungsfunktion gegeben durch

$$F_{X_p,1}(x) = \begin{cases} 0 & , \text{ falls } x \leq 6,3138 \\ \frac{F_1(x) - 0,95}{0,05} & , \text{ sonst} \end{cases}$$

und

$$F_{X_p,5}(x) = \begin{cases} 0 & , \text{ falls } x \leq 2,0150 \\ \frac{F_5(x) - 0,95}{0,05} & , \text{ sonst.} \end{cases}$$

(Siehe Abbildung 3.)

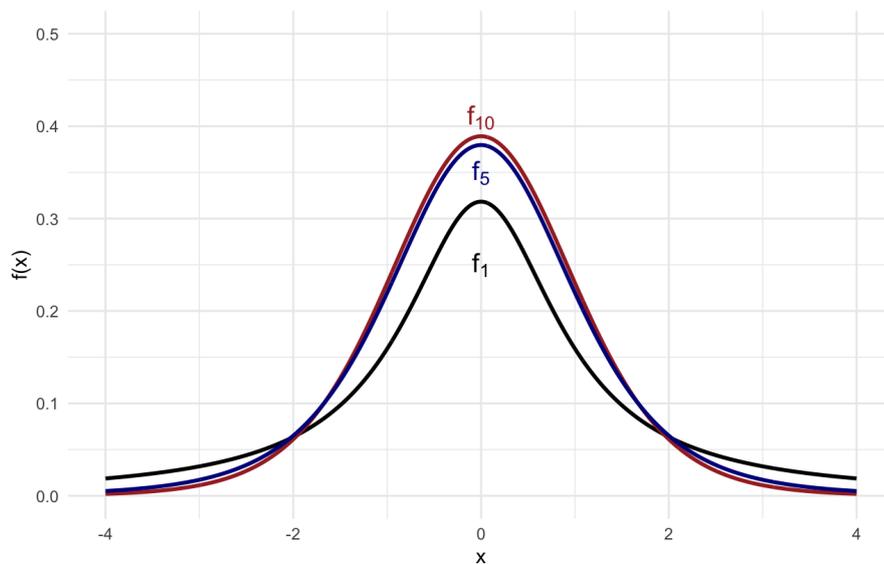


Abbildung 2: Dichte der *t*-Verteilung mit $k = 1, 5, 10$ Freiheitsgraden

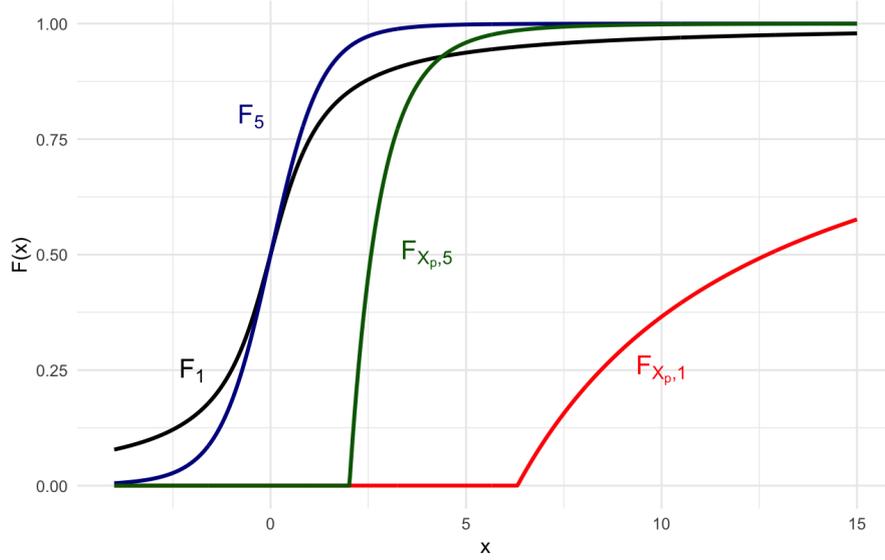


Abbildung 3: Vergleich von t-Verteilungen mit $k = 1, 5$ Freiheitsgraden und p -Tail-Verteilungsfunktionen

Auf Basis der bisherigen Eigenschaften können wir nun die Definition aus [Liu and Wang, 2021] für Tail-Risikomaße angeben.

Definition 3.1.6. Für $p \in (0, 1)$ ist ρ ein p -Tail-Risikomaß, falls $\rho(X) = \rho(Y)$ für alle $X, Y \in \mathcal{L}$ mit $X_p \stackrel{d}{=} Y_p$ ist. Ein Risikomaß ρ heißt *Tail-relevant*, falls es ein p -Tail-Risikomaß für ein $p \in (0, 1)$ ist. Wir bezeichnen ein p -Tail-Risikomaß einfachheitshalber als *Tail-Risikomaß*.

Bemerkung 3.1.7. Es fällt auf, dass die vorher gezeigten Risikomaße Var_p^+ und ES_p bereits Tail-Risikomaße sind. Dies liegt an Folgendem:

Seien $X, Y \in \mathcal{L}$ zwei Verlustfunktionen mit gleichverteiltem p -Tail-Verlust $X_p \stackrel{d}{=} Y_p$. Dann gilt:

$$Var_p^+(X) = \text{ess-inf}(X_p) = \text{ess-inf}(Y_p) = Var_p^+(Y) \quad \text{und} \\ ES_p(X) = \mathbb{E}[X_p] = \mathbb{E}[Y_p] = ES_p(Y).$$

Die Standardabweichung aus (SD) sowie der Erwartungswert sind hingegen keine Tail-Risikomaße, da sie nicht ausschließlich den Randbereich, sondern die gesamte Verteilung von L berücksichtigen.

Als Nächstes wollen wir auf die Eigenschaft hinarbeiten, dass Tail-Risikomaße von Risikomaßen eindeutig erzeugt werden. Dies führt zu folgender Definition:

Definition 3.1.8 (p -Erzeuger von ρ). Sei ρ^* ein verteilungsinvariantes Risikomaß auf \mathcal{L} und $p \in (0, 1)$. Das dazugehörige p -Tail-Risikomaß ρ ist definiert durch

$$\rho(L) := \rho^*(L_p), \quad L \in \mathcal{L}. \quad (1)$$

Falls dies gilt, so heißt ρ das von ρ^* erzeugte p -Tail-Risikomaß und ρ^* heißt der p -Erzeuger von ρ .

Wir werden im Folgenden die Eindeutigkeit und Existenz des Erzeugers von Tail-Risikomaßen behandeln. Mit anderen Worten: Ein Risikomaß ist genau dann ein Tail-Risikomaß, falls es durch ein Risikomaß erzeugt wurde.

Zunächst werden wir sehen, dass wir für jedes p -Tail-Risikomaß einen p -Erzeuger finden können.

Definition 3.1.9. Sei \mathcal{L}^* die Menge aller fast überall nach unten beschränkten Zufallsvariablen:

$$\mathcal{L}^* := \{L \in \mathcal{L} : \text{ess-inf}(L) > -\infty\}.$$

Sei weiter für jedes $L \in \mathcal{L}^*$ und $p \in (0, 1)$ $L^{(p)}$ eine Zufallsvariable mit Verteilungsfunktion

$$\mathbb{P}(L^{(p)} \leq l) = p\mathbb{1}_{\{l \geq \text{ess-inf}(L)\}} + (1-p)\mathbb{P}(L \leq l), \quad l \in \mathbb{R}$$

Bemerkung 3.1.10. Für die Zufallsvariable $L^{(p)}$ gilt:

$$L^{(p)} \stackrel{d}{=} \text{ess-inf}(L)B + L(1-B)$$

Für $B \sim \text{Bern}(p)$ unabhängig von L . Das heißt $\mathbb{P}(B = 1) = p$ und $\mathbb{P}(B = 0) = 1 - p$.

Bemerkung 3.1.11. $(L^{(p)})_p$ hat die gleiche Verteilung wie L :

$$\begin{aligned} \mathbb{P}((L^{(p)})_p \leq l) &= \frac{(\mathbb{P}(L^{(p)} \leq l) - p)_+}{1-p} \\ &= \frac{((p\mathbb{1}_{\{l \geq \text{ess-inf}(L)\}} + (1-p)\mathbb{P}(L \leq l)) - p)_+}{1-p} \end{aligned}$$

Falls $l \leq \text{ess-inf}(L)$ gilt, ist der Zähler 0. Wir betrachten daher nur die Menge $\{l \geq \text{ess-inf}(L)\}$:

$$\begin{aligned} \mathbb{P}((L^{(p)})_p \leq l) &= \frac{p + (1-p)\mathbb{P}(L \leq l) - p}{1-p} \\ &= \mathbb{P}(L \leq l) \end{aligned}$$

Dies führt uns zu Proposition 3.1 in [Liu and Wang, 2021]:

Proposition 3.1.12. Ein p -Erzeuger ρ^* von einem p -Tail-Risikomaß ρ ist eindeutig auf \mathcal{L}^* und gegeben durch

$$\rho^*(L) = \rho(L^{(p)}). \quad (2)$$

Beweis.

$$\rho^*(L) = \rho^*((L^{(p)})_p) = \rho(L^{(p)}).$$

□

Bemerkung 3.1.13. ρ^* ist nur auf \mathcal{L}^* eindeutig, da L_p für $p \in (0, 1)$ immer nach unten beschränkt ist.

Da wir nun ρ genau dann kennen, wenn wir ρ^* kennen, erhalten wir Folgendes:

Definition 3.1.14. Sei $p \in (0, 1)$. Das Paar (ρ, ρ^*) heißt p -Tail-Paar, falls ρ^* Verteilungsinvariant ist und ρ von ρ^* erzeugt wurde.

3.2 Beziehung zwischen Tail-Risikomaß und Erzeuger

Als nächstes wollen wir uns mit Eigenschaften beschäftigen, die vom Erzeuger an das Tail-Risikomaß (oder andersherum) vererbt werden.

Lemma 3.2.1. *Für eine Konstante $c \in \mathbb{R}$ gilt, dass $\rho(c) = \rho^*(c)$.*

Beweis. Für die Zufallsvariable $L = c$ gilt auch $L_p = c$. □

Lemma 3.2.2. *Für ein monotonen Risikomaß ρ^* gilt $\rho(L) \geq \rho^*(L)$ für alle $L \in \mathcal{L}^*$.*

Beweis. Es gilt $L_p \geq L$ und daher $\rho(L) = \rho^*(L_p) \geq \rho^*(L)$. □

Weitere Eigenschaften werden in Satz 1 aus [Liu and Wang, 2021] behandelt:

Satz 3.2.3. *Sei $p \in (0, 1)$ und (ρ, ρ^*) ein p -Tail-Paar. Es gilt:*

(i) *Wenn ρ^* monoton (bzw. translationsinvariant, positiv homogen oder komonoton additiv) ist, dann gilt dies auch für ρ .*

(ii) *Wenn ρ^* ein kohärentes (bzw. konvexes) Risikomaß ist, dann gilt dies auch für ρ .*

Bemerkung 3.2.4. Teil (i) von Satz 3.2.3 impliziert: Wenn ρ^* ein monetäres Risikomaß ist, dann gilt dies auch für ρ .

Satz 3.2.5. *Sei $p \in (0, 1)$ und (ρ, ρ^*) ein p -Tail-Paar. Es gilt:*

(i) *Wenn ρ monoton (bzw. translationsinvariant, positiv homogen oder komonoton additiv) ist, dann gilt dies auch für ρ^* .*

(ii) *Wenn ρ subadditiv (bzw. konvex oder \prec_{cx} -monoton) ist, dann gilt dies auch für ρ^* .*

Bemerkung 3.2.6. Satz 3.2.5 impliziert: Wenn ρ ein kohärentes (bzw. konvexes oder monetäres) Risikomaß ist, dann gilt dies auch für ρ^* .

Zunächst wollen wir mit einem Gegenbeispiel zeigen, dass die Rückrichtung von (ii) in Satz 3.2.5 im Allgemeinen nicht gilt.

Beispiel 3.2.7. Sei das Risikomaß der Standardabweichung definiert wie in (SD):

$$\rho^*(L) = SD_\beta(L) = \mathbb{E}[L] + \beta\sqrt{\text{var}(L)}, \quad L \in \mathcal{L}^2.$$

Dass dieses Risikomaß *translationsinvariant, konvex, positiv homogen, subadditiv* und \prec_{cx} -*monoton* ist, folgt direkt aus den Eigenschaften des Erwartungswerts und der Varianz. Betrachtet man nun für $p \in (0, 1)$ das zugehörige p -Tail-Risikomaß ρ , das durch SD_β erzeugt wurde, erhält man:

$$\rho(L) = \mathbb{E}(L_p) + \beta\sqrt{\text{var}(L_p)} = \rho^*(L_p) \quad L \in \mathcal{L}^2.$$

Seien nun X und Y unabhängig identisch verteilt mit $\mathbb{P}(X = -1) = p$ und $\mathbb{P}(X = 0) = 1 - p$. Dann ist $\mathbb{E}[X] = -p$ und $\text{Var}(X) = p(1 - p)$.

Sei weiter $Z := X + Y$. Dann ist $\mathbb{P}(Z = 0) = (1 - p)^2$ und

$$\mathbb{P}(Z_p = 0) = \mathbb{P}(Z_p \leq 0) - \mathbb{P}(Z_p \leq -1) = \frac{(1 - p)^2}{1 - p} = 1 - p.$$

Hieraus folgt, dass Z_p nicht konstant ist. Also gilt $\text{var}(Z_p) > 0$.

Daher gilt für $\beta > 0$:

$$\rho(X + Y) = \rho(Z) = \mathbb{E}[Z_p] + \beta \sqrt{\text{var}(Z_p)} > 0.$$

Es gilt $X_p = Y_p = 0$ und daher auch $\rho(X) = \rho(Y) = 0$. Also kann ρ nicht subadditiv sein, da $\rho(X + Y) > \rho(X) + \rho(Y)$. Dementsprechend kann ρ auch nicht konvex sein. Außerdem gilt $X + Y \prec_{cx} 2X$ (siehe Theorem 3.5 aus [Rüschendorf, 2013]) und $\rho(X + Y) > \rho(2X) = 0$.

Für den Beweis von Satz 3.2.3 und Satz 3.2.5 sind im folgendem Lemma noch einige Relationen zwischen X , X_p und $X^{(p)}$ dargestellt.

Lemma 3.2.8. *Sei $p \in (0, 1)$.*

(i) *Für $L \in \mathcal{L}^*$ gilt: $(X^{(p)})_p \stackrel{d}{=} X$.*

(ii) *Für $L, J \in \mathcal{L}$ gilt: Falls $L \prec_{st} J$, dann ist $L_p \prec_{st} J_p$. Für $L, J \in \mathcal{L}^*$ gilt: Falls $L \prec_{st} J$, dann ist $L^{(p)} \prec_{st} J^{(p)}$.*

(iii) *Für $L \in \mathcal{L}$ gilt: $L \prec_{st} L_p$.*

Beweis von Lemma 3.2.8. Zu (i): Diese Eigenschaft wurde bereits in Bemerkung 3.1.11 gezeigt.

Zu (ii):

Es gilt $L \prec_{st} J \Leftrightarrow F_L^{-1}(u) \leq F_J^{-1}(u)$ für fast alle $u \in (0, 1)$.

Da für $U \sim \mathcal{U}(0, 1)$ gilt, dass $p + (1 - p)U \sim \mathcal{U}(p, 1) \subset (p, 1)$, muss auch

$$F_L^{-1}(p + (1 - p)U) \leq F_J^{-1}(p + (1 - p)U)$$

punktweise gelten.

Dies ist wieder äquivalent zu $L_p \prec_{st} J_p$.

Weiter ist $L^{(p)} \stackrel{d}{=} \text{ess-inf}(L)B + L(1 - B)$ mit $B \sim \text{Bern}(p)$. Falls $B = 1$, dann ist

$$L^{(p)} = \text{ess-inf}(L) \leq \text{ess-inf}(J) = J^{(p)}$$

und falls $B = 0$, dann ist

$$L^{(p)} = L \leq J = J^{(p)},$$

da

$$\mathbb{P}(L \geq b) \leq \mathbb{P}(J \geq b) \quad \forall b \in \mathbb{R}$$

gilt.

Zu (iii):

Es ist

$$L = F_L^{-1}(U) \quad \text{und} \quad L_p = F_L^{-1}(p + (1-p)U).$$

Da $U \leq p + (1-p)U$ für alle $p \in (0, 1)$ gilt und F^{-1} monoton wachsend ist, gilt auch

$$F_L^{-1}(U) \leq F_L^{-1}(p + (1-p)U)$$

und somit $L \prec_{st} L_p$. □

Lemma 3.2.9. *Ein Risikomaß ρ ist genau dann monoton, wenn es \prec_{st} -monoton ist.*

Beweis von Lemma 3.2.9. “ \Leftarrow ” Angenommen ρ ist ein \prec_{st} -monotones Risikomaß. Falls $L \leq J$ ist, dann folgt direkt $\mathbb{P}(L \leq t) \geq \mathbb{P}(J \leq t)$. Also gilt auch $L \prec_{st} J$ und damit $\rho(L) \leq \rho(J)$.

“ \Rightarrow ” Sei ρ ein monotonen Risikomaß und $L \prec_{st} J$. Dann existieren zwei Zufallsvariablen \hat{L}, \hat{J} mit $\hat{L} \stackrel{d}{=} L$ und $\hat{J} \stackrel{d}{=} J$ mit $\mathbb{P}(\hat{L} \leq \hat{J}) = 1$ [Shaked and Shanthikumar, 2007].

Daraus folgt

$$\rho(L) = \rho(\hat{L}) \leq \rho(\hat{J}) = \rho(J),$$

da ρ verteilungsinvariant ist. □

Beweis von Satz 3.2.3. Für den Beweis werden wir (1) häufig verwenden.

Zu (i):

(Monotonie)

Sei ρ^* \prec_{st} -monoton und $L \prec_{st} J$. Nach Lemma 3.2.8 gilt auch $L_p \prec_{st} J_p$ und daher auch $\rho^*(L_p) \prec_{st} \rho^*(J_p)$. Daraus folgt:

$$\rho(L) = \rho^*(L_p) \prec_{st} \rho^*(J_p) = \rho(J).$$

Somit ist $\rho \prec_{st}$ -monoton und wegen Lemma 3.2.9 auch monoton.

(Translationsinvarianz)

Sei ρ^* translationsinvariant und $\mu \in \mathbb{R}$. Es gilt:

$$\begin{aligned} \mathbb{P}((L + \mu)_p \leq l) &= \frac{(\mathbb{P}(L + \mu \leq l) - p)_+}{1 - p} \\ &= \frac{(\mathbb{P}(L \leq l - \mu) - p)_+}{1 - p} \\ &= \mathbb{P}(L_p \leq l - \mu) = \mathbb{P}(L_p + \mu \leq l). \end{aligned}$$

Daher ist $(L + \mu)_p \stackrel{d}{=} L_p + \mu$. Es folgt

$$\rho(L + \mu) = \rho^*((L + \mu)_p) = \rho^*(L_p + \mu) = \rho^*(L_p) + \mu = \rho(L) + \mu.$$

(Positive Homogenität)

Sei ρ^* positiv homogen und $\lambda \geq 0$. Für $\lambda > 0$ gilt:

$$\begin{aligned}\mathbb{P}((\lambda L)_p \leq l) &= \frac{(\mathbb{P}(\lambda L \leq l) - p)_+}{1 - p} \\ &= \frac{(\mathbb{P}(L \leq \frac{l}{\lambda}) - p)_+}{1 - p} \\ &= \mathbb{P}(L_p \leq \frac{l}{\lambda}) = \mathbb{P}(\lambda L_p \leq l).\end{aligned}$$

Daher ist $(\lambda L)_p \stackrel{d}{=} \lambda L_p$. Es folgt

$$\rho(\lambda L) = \rho^*((\lambda L)_p) = \rho^*(\lambda L_p) = \lambda \rho^*(L_p) = \lambda \rho(L).$$

(Komonotone Additivität)

Sei ρ^* komonoton additiv und $L, J \in \mathcal{L}$ komonoton. Dann gilt für $U \sim U[0, 1]$:

$$L + J \stackrel{d}{=} F_{L+J}^{-1}(U) \stackrel{d}{=} F_L^{-1}(U) + F_J^{-1}(U)$$

und

$$(L + J)_p \stackrel{d}{=} F_{L+J}^{-1}(p + (1 - p)U) \stackrel{d}{=} F_L^{-1}(p + (1 - p)U) + F_J^{-1}(p + (1 - p)U).$$

Es folgt, dass

$$\begin{aligned}\rho(L + J) &= \rho^*((L + J)_p) \\ &= \rho^*(F_L^{-1}(p + (1 - p)U) + F_J^{-1}(p + (1 - p)U)) \\ &= \rho^*(L_p + J_p) = \rho^*(L_p) + \rho^*(J_p) \\ &= \rho(L) + \rho(J).\end{aligned}$$

Zu (ii):

Für den Beweis verwenden wir Theorem 3 aus [Liu and Wang, 2021], das sich mit Worst-Case-Szenarien für verschiedenen Verlustfunktionen befasst.

Dabei wird die Menge aller möglichen aggregierten Verluste betrachtet, deren Verteilungen vorgegeben sind. Diese sogenannte Aggregationsmenge ist definiert als

$$\mathcal{S}_n(F_1, \dots, F_n) := \{L_1 + \dots + L_n : L_i \in \mathcal{L}, L_i \sim F_i, i = 1, \dots, n\},$$

also die Menge aller Summen von Zufallsvariablen, deren jeweilige Verteilung F_i entspricht. Für das Worst-Case-Szenario berechnet man

$$\sup\{\rho(S) : S \in \mathcal{S}_n(F_1, \dots, F_n)\}.$$

Satz 3 aus [Liu and Wang, 2021] behauptet Folgendes:

Sei $p \in (0, 1)$ und (ρ, ρ^*) ein p -Tail-Paar von monotonen Risikomaßen. Für univariate Verteilungsfunktionen F_1, \dots, F_n gilt:

$$\sup\{\rho(S) : S \in \mathcal{S}_n(F_1, \dots, F_n)\} = \sup\{\rho^*(S) : S \in \mathcal{S}_n(F_{1p}, \dots, F_{np})\}.$$

Das heißt das Worst-Case-Szenario eines monotonen p -Tail-Risikomaßes ρ über alle möglichen aggregierte Verluste ist gleich dem Worst-Case-Szenarios des p -Erzeugers ρ^* über alle möglichen aggregierten p -Tail-Verluste.

Sei nun ρ^* ein konvexes Risikomaß (monoton, translationsinvariant und konvex). Für $L, J \in \mathcal{L}$ und $\lambda \in [0, 1]$ gilt:

$$\begin{aligned} \rho(\lambda L + (1 - \lambda)J) &\leq \sup\{\rho(\lambda X + (1 - \lambda)Y) : X, Y \in \mathcal{L}, X \sim F_L, Y \sim F_J\} \\ &\stackrel{\text{Theorem 3}}{=} \sup\{\rho^*(\lambda Z + (1 - \lambda)W) : Z, W \in \mathcal{L}, Z \sim F_{Lp}, W \sim F_{Jp}\} \\ &\stackrel{\rho^* \text{ konvex}}{\leq} \sup\{\lambda \rho^*(Z) + (1 - \lambda)\rho^*(W) : Z, W \in \mathcal{L}, Z \sim F_{Lp}, W \sim F_{Jp}\} \\ &= \lambda \sup_{Z \sim F_{Lp}} \{\rho^*(Z)\} + (1 - \lambda) \sup_{W \sim F_{Jp}} \{\rho^*(W)\} \\ &= \lambda \rho^*(L_p) + (1 - \lambda)\rho^*(J_p) \\ &= \lambda \rho(L) + (1 - \lambda)\rho(J) \end{aligned}$$

Daher gilt: Wenn ρ^* ein konvexes Risikomaß ist, dann gilt dies auch für ρ .

Falls ρ^* zusätzlich noch positiv homogen ist, gilt wegen (i), dass dann auch ρ positiv homogen ist. Daher gilt: Wenn ρ^* ein kohärentes Risikomaß ist, dann gilt dies auch für ρ . \square

Beweis von Satz 3.2.5. Für den Beweis werden wir (2) häufig verwenden.

Zu (i):

(Monotonie)

Sei ρ \prec_{st} -monoton und $L \prec_{st} J$. Nach Lemma 3.2.8 gilt auch $L^{(p)} \prec_{st} J^{(p)}$ und daher auch $\rho(L^{(p)}) \prec_{st} \rho(J^{(p)})$. Daraus folgt:

$$\rho^*(L) = \rho(L^{(p)}) \prec_{st} \rho(J^{(p)}) = \rho^*(J).$$

Somit ist ρ^* \prec_{st} -monoton und wegen Lemma 3.2.9 auch monoton.

(Translationsinvarianz)

Sei ρ translationsinvariant und $\mu \in \mathbb{R}$. Es gilt nach Bemerkung 3.1.10:

$$\begin{aligned} (L + \mu)^{(p)} &\stackrel{d}{=} \text{ess-inf}(L + \mu)B + (L + \mu)(1 - B) \\ &= \text{ess-inf}(L)B + \mu B + L(1 - B) + \mu(1 - B) \\ &= \text{ess-inf}(L)B + L(1 - B) + \mu \\ &\stackrel{d}{=} L^{(p)} + \mu, \end{aligned}$$

für $B \sim \text{Bern}(p)$ unabhängig von L . Daher gilt

$$\rho^*(L + \mu) = \rho((L + \mu)^{(p)}) = \rho(L^{(p)} + \mu) = \rho(L^{(p)}) + \mu = \rho^*(L) + \mu.$$

(Positive Homogenität)

Sei ρ positiv homogen und $\lambda \geq 0$. Es gilt nach Bemerkung 3.1.10:

$$\begin{aligned} (\lambda L)^{(p)} &\stackrel{d}{=} \text{ess-inf}(\lambda L)B + \lambda L(1 - B) \\ &= \lambda \text{ess-inf}(L)B + \lambda L(1 - B) \\ &= \lambda(\text{ess-inf}(L)B + L(1 - B)) \\ &\stackrel{d}{=} \lambda L^{(p)}, \end{aligned}$$

für $B \sim \text{Bern}(p)$ unabhängig von L . Daher gilt

$$\rho^*(\lambda L) = \rho((\lambda L)^{(p)}) = \rho(\lambda L^{(p)}) = \lambda \rho(L^{(p)}) = \lambda \rho^*(L).$$

(Komonotone Additivität)

Sei ρ komonoton additiv und $L, J \in \mathcal{L}^*$ komonoton. Sei $B \sim \text{Bern}(p)$ unabhängig von L und J und $l = \text{ess-inf}(L)$ und $j = \text{ess-inf}(J)$. Dann gilt:

$$(L + J)^{(p)} \stackrel{d}{=} \text{ess-inf}(L + J)B + (L + J)(1 - B) = lB + L(1 - B) + jB + J(1 - B)$$

Es sind $lB + L(1 - B)$ und $jB + J(1 - B)$ ebenfalls komonoton und daher gilt:

$$\begin{aligned} \rho^*(L + J) &= \rho((L + J)^{(p)}) = \rho(lB + L(1 - B) + jB + J(1 - B)) \\ &= \rho(lB + L(1 - B)) + \rho(jB + J(1 - B)) \\ &= \rho(L^{(p)}) + \rho(J^{(p)}) \\ &= \rho^*(L) + \rho^*(J). \end{aligned}$$

Zu (ii):

Sei wieder $B \sim \text{Bern}(p)$. Für $Z \in \mathcal{L}^*$ unabhängig von B und $t \leq \text{ess-inf}(Z)$ gilt:

$$(tB + Z(1 - B))_p \stackrel{d}{=} (\text{ess-inf}(Z)B + Z(1 - B))_p \stackrel{d}{=} Z. \quad (*)$$

Sei weiter $L, J \in \mathcal{L}^*$ unabhängig von B und $l = \text{ess-inf}(L)$, $j = \text{ess-inf}(J)$, $z = \text{ess-inf}(L + J)$ und $w = \min\{l, j\}$.

(Subadditivität)

Sei ρ subadditiv. Es gilt $z \geq l + j$ und daher wegen (*):

$$\begin{aligned} \rho^*(L + J) &= \rho^*((L + J)^{(p)})_p = \rho((L + J)^{(p)}) \\ &= \rho(zB + L(1 - B) + J(1 - B)) \\ &= \rho((l + j)B + L(1 - B) + J(1 - B)) \\ &\leq \rho(lB + L(1 - B)) + \rho(jB + J(1 - B)) \\ &= \rho^*(L) + \rho^*(J). \end{aligned}$$

(Konvexität)

Sei ρ konvex und $\lambda \in [0, 1]$.

$$\begin{aligned}
\rho^*(\lambda L + (1 - \lambda)J) &= \rho((\lambda L + (1 - \lambda)J)^{(p)}) \\
&= \rho((\lambda l + (1 - \lambda)j)B + \lambda L(1 - B) + (1 - \lambda)J(1 - B)) \\
&= \rho(\lambda lB + \lambda L(1 - B) + (1 - \lambda)jB + (1 - \lambda)J(1 - B)) \\
&= \rho(\lambda(lB + L(1 - B)) + (1 - \lambda)(jB + J(1 - B))) \\
&\leq \lambda\rho(lB + L(1 - B)) + (1 - \lambda)\rho(jB + J(1 - B)) \\
&= \lambda\rho^*(L) + (1 - \lambda)\rho^*(J).
\end{aligned}$$

(\prec_{cx} -Monotonie)

Sei ρ \prec_{cx} -monoton und $L \prec_{cx} J$. Dann gilt $l \leq j$ und $wB + L(1 - B) \prec_{cx} wB + J(1 - B) \prec_{cx} jB + J(1 - B)$ und daher wegen (*):

$$\begin{aligned}
\rho^*(L) &= \rho(L^{(p)}) \\
&= \rho(lB + L(1 - B)) \\
&= \rho(wB + L(1 - B)) \\
&\leq \rho(wB + J(1 - B)) \\
&\leq \rho(jB + J(1 - B)) = \rho^*(J).
\end{aligned}$$

□

4 Identifizierbarkeit und Elizitierbarkeit

Bemerkung 4.0.1. Identifizierbarkeit und Elizitierbarkeit sind zwei zentrale Eigenschaften von Risikomaßen, die es ermöglichen, die Qualität und Prognosegüte ihrer Schätzungen systematisch zu bewerten. Sie bilden die theoretische Grundlage für die Konstruktion von Teststatistiken und Verfahren des Backtestings, da sie durch Identifikations- und Scoring-Funktionen eine objektive Überprüfung sowie den Vergleich verschiedener Vorhersagen erlauben. Ohne diese Eigenschaften wäre eine konsistente und faire Bewertung von Risikomaßen nicht möglich. Im Folgenden werden die Begriffe präzise definiert und einige wichtige Zusammenhänge von Tail-Risikomaßen und Erzeuger erläutert.

4.1 Identifizierbarkeit

Definition 4.1.1. Sei $A \subseteq \mathbb{R}^k$ der Aktionsraum mit möglichen Vorhersagen und $O \subseteq \mathbb{R}$ (messbar) der Beobachtungsraum. Sei \mathcal{M} die Menge aller Verteilungen auf $(O, \mathcal{B}(O))$. Ein Funktional $T : \mathcal{M} \rightarrow A$ heißt Parameter-Funktion (oder Risikomaß) und weist jeder Verteilung $F \in \mathcal{M}$ einen Wert im Aktionsraum $T(F) \in A$ zu.

Definition 4.1.2 (Identifikationsfunktion). Eine Funktion $V : A \times O \rightarrow \mathbb{R}$, für die $V(a, \cdot)$ messbar für alle $a \in A$ ist und

$$\mathbb{E}_F[|V(a, L)|] := \int_O |V(a, l)| dF(l) < \infty \quad \forall F \in \mathcal{M}$$

heißt \mathcal{M} -Identifikationsfunktion für T , falls für alle $F \in \mathcal{M}$ gilt, dass

$$T(F) \subseteq \left\{ a \in A : \int_{\mathcal{O}} V(a, l) dF(l) = 0 \right\}.$$

Falls Gleichheit gilt, ist V eine *strikte* \mathcal{M} -Identifikationsfunktion für T . Falls für T eine strikte \mathcal{M} -Identifikationsfunktion existiert, heißt T \mathcal{M} -identifizierbar.

Bemerkung 4.1.3. Der Erwartungswert der Identifikationsfunktion ist genau dann null, wenn die Vorhersage $a \in A$ dem wahren Wert des Risikomaßes $T(F)$ unter der Verteilung F entspricht.

Beispiel 4.1.4 (Erwartungswert). Sei $T(F) = \mathbb{E}_F[L]$. Dann ist $V(a, l) = a - l$, $a, l \in \mathbb{R}$ eine strikte \mathcal{M}^1 -Identifikationsfunktion für T , denn

$$\begin{aligned} \int_{\mathbb{R}} V(a, l) dF(l) &= \int_{\mathbb{R}} a - l dF(l) \\ &= a \int_{\mathbb{R}} 1 dF(l) - \int_{\mathbb{R}} l dF(l) \\ &= a - \mathbb{E}_F[L] \end{aligned}$$

ist genau dann 0, wenn $a = \mathbb{E}_F[L]$. Somit gilt für alle $F \in \mathcal{M}^1$

$$T(F) = \left\{ a \in \mathbb{R} : \int_{\mathbb{R}} V(a, l) dF(l) = 0 \right\}.$$

\mathbb{E}_F ist also \mathcal{M}^1 -identifizierbar.

Definition 4.1.5. Wir definieren folgende Menge:

$$\mathcal{M}_{(p)} := \{F \in \mathcal{M}^0 : F \text{ stetig in } F^{-1}(p)\}, \quad p \in (0, 1).$$

Falls wir die Menge \mathcal{M} nicht weiter definieren, gilt die Behauptung für jede Untermenge der Borel-messbaren Verteilungsfunktionen auf \mathbb{R} . Also für alle $\mathcal{M} \subseteq \mathcal{M}^0$.

Beispiel 4.1.6 (p-Quantil). Sei $p \in (0, 1)$ und $T(F) = Q_p(F)$. Dann ist T identifizierbar auf

$$\mathcal{M}_{(p)} = \{F \in \mathcal{M}^0 : F \text{ stetig in } F^{-1}(p)\}.$$

$V(a, l) = \mathbb{1}_{\{l \leq a\}} - p$, $a, l \in \mathbb{R}$ ist dann eine strikte $\mathcal{M}_{(p)}$ -Identifikationsfunktion, denn

$$\begin{aligned} \int_{\mathbb{R}} V(a, l) dF(l) &= \int_{\mathbb{R}} (\mathbb{1}_{l \leq a} - p) dF(l) \\ &= \int_{-\infty}^a 1 dF(l) - p \int_{\mathbb{R}} 1 dF(l) \\ &= F(a) - p \end{aligned}$$

ist genau dann 0, wenn $a \in Q_p(F)$. Somit gilt für alle $F \in \mathcal{M}_{(p)}$

$$T(F) = \left\{ a \in \mathbb{R} : \int_{\mathbb{R}} V(a, l) dF(l) = 0 \right\}.$$

Das Quantil $Q_p(F)$ ist also $\mathcal{M}_{(p)}$ -identifizierbar.

Satz 4.1.7 (Satz 4.3 aus [Fissler et al., 2024]). Sei (ρ, ρ^*) ein p -Tail-Paar für ein $p \in (0, 1)$. Wenn $\rho^* : \mathcal{M} \rightarrow \mathbb{R}$ \mathcal{M} -identifizierbar mit strikter Identifikationsfunktion $V^* : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ist, dann ist (Q_p, ρ) $\mathcal{M}_{(p)} \cap \mathcal{M}$ -identifizierbar mit der strikten $\mathcal{M}_{(p)} \cap \mathcal{M}$ -Identifikationsfunktion

$$V(v, a, l) = \begin{pmatrix} \mathbb{1}_{\{l \leq v\}} - p \\ \mathbb{1}_{\{l > v\}} V^*(a, l) + (\mathbb{1}_{\{l \leq v\}} - p) V^*(a, v) \end{pmatrix}, \quad v, a, l \in \mathbb{R}.$$

Beweis. Sei $F \in \mathcal{M}_{(p)} \cap \mathcal{M}$. Nach Beispiel 4.1.6 gilt:

$$\int_{\mathbb{R}} (\mathbb{1}_{\{l \leq v\}} - p) dF(l) = 0 \quad \Leftrightarrow \quad v \in Q_p(F).$$

Sei nun $v \in Q_p(F)$ fest. Wegen $F_p(l) = \frac{(F(l)-p)_+}{1-p}$, $F_p \ll F$ (F_p absolut stetig bzgl. F) und dem Satz von Radon-Nikodým gilt

$$dF_p(l) = \frac{1}{1-p} \mathbb{1}_{\{F(l) > p\}} dF(l). \quad (4.1)$$

Daraus folgt:

$$\begin{aligned} & \int_{\mathbb{R}} (\mathbb{1}_{\{l > v\}} V^*(a, l) + (\mathbb{1}_{\{l \leq v\}} - p) V^*(a, v)) dF(l) \\ &= \int_{\mathbb{R}} \mathbb{1}_{\{l > v\}} V^*(a, l) dF(l) + \int_{\mathbb{R}} (\mathbb{1}_{\{l \leq v\}} - p) V^*(a, v) dF(l) \\ &= \int_{\mathbb{R}} \mathbb{1}_{\{l > v\}} V^*(a, l) dF(l) + V^*(a, v) \int_{\mathbb{R}} \mathbb{1}_{\{l \leq v\}} dF(l) - p V^*(a, v) \int_{\mathbb{R}} 1 dF(l) \\ &= \int_{\mathbb{R}} \mathbb{1}_{\{l > v\}} V^*(a, l) dF(l) + V^*(a, v) F(v) - p V^*(a, v) \\ &\stackrel{F(v)=p}{=} \int_{\mathbb{R}} \mathbb{1}_{\{l > v\}} V^*(a, l) dF(l) \\ &= (1-p) \int_{\mathbb{R}} V^*(a, l) \frac{1}{1-p} \mathbb{1}_{\{l > v\}} dF(l) \\ &\stackrel{(4.1)}{=} (1-p) \int_{\mathbb{R}} V^*(a, l) dF_p(l). \end{aligned}$$

Da V^* eine strikte Identifikationsfunktion für ρ^* ist, gilt

$$\rho^*(F_p) = \left\{ a \in \mathbb{R} : \int_{\mathbb{R}} V^*(a, l) dF_p(l) = 0 \right\}$$

und somit

$$\int_{\mathbb{R}} V^*(a, l) dF_p(l) = 0 \quad \Leftrightarrow \quad a = \rho^*(F_p) = \rho(F).$$

Also gilt

$$(Q_p, \rho)(F) = \left\{ (v, a) \in \mathbb{R} \times \mathbb{R} : \int_{\mathbb{R}} V(v, a, l) dF(l) = 0 \right\}$$

und somit folgt die Behauptung. \square

Beispiel 4.1.8. Sei (Q_p, ES_p) ein Paar von Risikomaßen. Der Erzeuger von ES_p ist nach Bemerkung 3.1.7 der Erwartungswert \mathbb{E} , der nach Beispiel 4.1.4 die strikte \mathcal{M}^1 -Identifikationsfunktion $V^*(a, l) = a - l$ hat. Das Paar (Q_p, ES_p) ist $\mathcal{M}_{(p)} \cap \mathcal{M}^1$ identifizierbar mit der strikten Identifikationsfunktion

$$\begin{aligned} V(v, a, l) &= \begin{pmatrix} \mathbb{1}_{\{l \leq v\}} - p \\ \mathbb{1}_{\{l > v\}}(a - l) + (\mathbb{1}_{\{l \leq v\}} - p)(a - v) \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{1}_{\{l \leq v\}} - p \\ a(\mathbb{1}_{\{l > v\}} + (\mathbb{1}_{\{l \leq v\}} - p)) - [\mathbb{1}_{\{l > v\}}l + (\mathbb{1}_{\{l \leq v\}} - p)v] \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{1}_{\{l \leq v\}} - p \\ a(1 - p) - [\mathbb{1}_{\{l > v\}}l + (\mathbb{1}_{\{l \leq v\}} - p)v] \end{pmatrix}. \end{aligned}$$

Wegen des ersten Eintrags der Identifikationsfunktion $V(v, a, l)$ muss wieder $v \in Q_p(F)$ gelten. Betrachtet man nun den zweiten Eintrag, so fällt auf:

$$\begin{aligned} &\int_{\mathbb{R}} a(1 - p) - [\mathbb{1}_{\{l > v\}}l + (\mathbb{1}_{\{l \leq v\}} - p)v] dF(y) = 0 \\ \Leftrightarrow &a - \int_{\mathbb{R}} \frac{1}{1 - p} [\mathbb{1}_{\{l > v\}}l + (\mathbb{1}_{\{l \leq v\}} - p)v] dF(y) = 0 \\ \Leftrightarrow &a - \frac{1}{1 - p} \left(\int_{l > v} l dF(l) + v \left(\int_{l \leq v} 1 dF(l) - p \right) \right) = 0 \\ \Leftrightarrow &a - \frac{1}{1 - p} \left(\int_{l > v} l dF(l) + v(F(v) - p) \right) = 0 \\ \stackrel{F(v)=p}{\Leftrightarrow} &a - \frac{1}{1 - p} \int_{l > v} l dF(l) = 0 \end{aligned}$$

und

$$\frac{1}{1 - p} \int_{l > v} l dF(l) = ES_p(F).$$

Daher ist

$$\int_{\mathbb{R}} V(v, a, l) dF(l) = 0$$

genau dann, wenn $v \in Q_p(F)$ und $a = ES_p(F)$ und somit ist $V(v, a, l)$ eine strikte $\mathcal{M}_{(p)} \cap \mathcal{M}^1$ -Identifikationsfunktion für (Q_p, ES_p) .

4.2 Elizitierbarkeit

Bemerkung 4.2.1. Wir verwenden die Notation aus Kapitel 4.1. $A \subseteq \mathbb{R}^k$ ist der Aktionsraum, $O \subseteq \mathbb{R}$ (messbar) der Beobachtungsraum und \mathcal{M} die Menge aller Verteilungen auf $(O, \mathcal{B}(O))$. Ein Funktional $T : \mathcal{M} \rightarrow A$ weist jeder Verteilung $F \in \mathcal{M}$ einen Wert im Aktionsraum $T(F) \in A$ zu.

Definition 4.2.2 (Scoring-Funktion). Eine Funktion $S : A \times O \rightarrow \mathbb{R}$, für die $S(a, \cdot)$ messbar für alle $a \in A$ ist und

$$\mathbb{E}_F[|S(a, L)|] := \int_O |S(a, L)| dF(l) < \infty \quad \forall F \in \mathcal{M}$$

heißt \mathcal{M} -Scoring-Funktion. Falls für alle $F \in \mathcal{M}$

$$T(F) \subseteq \arg \min_{a \in A} \int_O S(a, l) dF(l)$$

bzw.

$$\mathbb{E}_F[S(T(F), L)] \leq \mathbb{E}_F[S(a, L)], \quad \forall a \in A,$$

gilt, dann heißt S eine *konsistente* \mathcal{M} -Scoring-Funktion für T . Falls für alle $F \in \mathcal{M}$

$$T(F) = \arg \min_{a \in A} \int_O S(a, l) dF(l)$$

bzw.

$$\mathbb{E}_F[S(T(F), L)] < \mathbb{E}_F[S(a, L)], \quad a \neq T(F)$$

gilt, dann heißt S eine *strikt-konsistente* \mathcal{M} -Scoring-Funktion. Falls für T eine strikt-konsistente \mathcal{M} -Scoring-Funktion existiert, heißt T *\mathcal{M} -Elizitierbar*.

Beispiel 4.2.3 (Erwartungswert). Sei I ein Intervall, $\phi : I \rightarrow \mathbb{R}$ eine strikt konvexe differenzierbare Funktion und $F \in \mathcal{M}^1$. Für $T(F) = \mathbb{E}_F[L]$ ist

$$S(a, l) = \phi(l) - \phi(a) - \phi'(a)(l - a), \quad a, l \in I$$

eine strikt-konsistente \mathcal{M}^1 -Scoring-Funktion für T , denn für $T(F) \neq a$ gilt:

$$\begin{aligned} & \mathbb{E}_F[S(a, L)] - \mathbb{E}_F[S(T(F), L)] \\ &= \mathbb{E}_F[\phi(L) - \phi(a) - \phi'(a)(L - a)] - \mathbb{E}_F[\phi(L) - \phi(T(F)) - \phi'(T(F))(L - T(F))] \\ &= -\phi(a) - \phi'(a)(\mathbb{E}_F[L] - a) + \phi(T(F)) - \phi'(T(F))(\mathbb{E}_F[L] - T(F)) \\ &= \phi(T(F)) - \phi(a) - \phi'(a)(T(F) - a). \end{aligned}$$

Da ϕ streng konvex ist, gilt

$$\phi(T(F)) > \phi(a) - \phi'(a)(T(F) - a)$$

und somit

$$\mathbb{E}_F[S(a, L)] - \mathbb{E}_F[S(T(F), L)] > 0.$$

Bemerkung 4.2.4. Für $\phi(x) = x^2$ ist $S(a, l)$ eine strikt-konsistente \mathcal{M}^2 -Scoring-Funktion, die auch als quadratischer Fehler bekannt ist:

$$S(a, l) = l^2 - a^2 - 2a(l - a) = (l - a)^2$$

Bemerkung 4.2.5. Die Scoring-Funktion für das p -Quantil, später auch bekannt als pinball-loss Funktion, wurde erstmals 1978 in [Koenker and Bassett, 1978] erwähnt. In dieser Publikation wurde eine spezielle asymmetrische Verlustfunktion eingeführt, die es erlaubte, beliebige Quantile eines Regressionsmodells zu schätzen. Die Funktion ist so gestaltet, dass sie positive und negative Abweichungen unterschiedlich gewichtet und damit direkt beim gewünschten Quantil minimiert wird. Den Namen erhielt die Verlustfunktion aufgrund ihres Aussehens, was an einen Flipperautomaten (pinball) erinnert.

Satz 4.2.6 (p-Quantil, Satz 4.54 aus [Holzmann, 2023]). Sei $p \in (0, 1)$ und $T(F) = Q_p(F)$ und \mathcal{M}_c^{sm} die Menge aller stetigen Verteilungsfunktionen, die streng monoton wachsend auf der Menge $[\inf\{x \mid F(x) > 0, \sup\{x \mid F(x) < 1\}]$ sind. Dann ist T elizitierbar auf \mathcal{M}_c^{sm} mit der strikt-konsistenten \mathcal{M}_c^{sm} -Scoring-Funktion

$$\begin{aligned} S(a, l) &= \begin{cases} p \cdot (l - a), & \text{falls } l > a \\ (1 - p)(a - l), & \text{falls } l \leq a \end{cases} \\ &= (\mathbb{1}_{\{l \leq a\}} - p)(a - l) \\ &= (1 - p)(a - l)_+ + p(l - a)_+, \quad a, l \in \mathbb{R}. \end{aligned}$$

Beweis. Um dies zu zeigen, benutzen wir die partielle Integration für Lebesgue-Stieltjes Integrale wie in [Holzmann, 2023]: Falls $F, G : \mathbb{R} \rightarrow \mathbb{R}$ rechtsstetig, streng monoton wachsend mit endlichem Grenzwert gegen $\pm\infty$ und eine der beiden Funktionen absolut stetig ist, dann gilt auf dem kompakten Intervall $[a, b]$:

$$\int_{[a,b]} G(x) dF(x) = G(b)F(b) - G(a)F(a) - \int_{[a,b]} F(x) dG(x). \quad (4.2)$$

(Siehe [Hewitt, 1960].)

Für unsere Scoring-Funktion gilt:

$$\mathbb{E}_F[S(a, l)] = (1 - p) \int_{-\infty}^a (a - l) dF(l) + p \int_a^{\infty} (l - a) dF(l).$$

Für $G(l) = l$ und $c < a$ können wir (4.2) nutzen:

$$\begin{aligned} \int_c^a (a - l) dF(l) &= a \int_c^a 1 dF(l) - \int_c^a l dF(l) \\ &= a(F(a) - F(c)) - (aF(a) - cF(c) - \int_c^a F(l) dl) \\ &= -(aF(c) - cF(c)) + \int_c^a F(l) dl. \end{aligned}$$

Nun gilt $\lim_{c \rightarrow -\infty} aF(c) = 0$, und da für hinreichend kleine c die Ungleichung $|L| \geq |c|$ gilt, folgt:

$$\begin{aligned} &|c| \cdot \mathbb{1}_{L \leq c} \leq |L| \cdot \mathbb{1}_{L \leq c} \\ \Leftrightarrow &|c| \cdot \mathbb{P}(L \leq c) \leq \mathbb{E}_F[\mathbb{1}_{L \leq c} \cdot |L|] \\ \Leftrightarrow &|c| \cdot F(c) \leq \mathbb{E}_F[\mathbb{1}_{L \leq c} \cdot |L|] \end{aligned} \quad (4.3)$$

und

$$\mathbb{E}_F[\mathbb{1}_{L \leq c} \cdot |L|] \rightarrow 0, \quad c \rightarrow -\infty,$$

wegen $\mathbb{E}_F[|L|] < \infty$ und dem Satz der majorisierten Konvergenz.

Dadurch erhalten wir:

$$\int_{-\infty}^a (a - l) dF(l) = \int_{-\infty}^a F(l) dl$$

Weil $F(l) = 1 - (1 - F(l))$ ist, können wir nun $dF(l) = d(-(1 - F(l)))$ setzen und erhalten

$$\begin{aligned}
\int_a^c (l - a) dF(l) &= \int_a^c (l - a) d(-(1 - F(l))) \\
&= - \int_a^c (l - a) d(1 - F(l)) \\
&= - \left(\int_a^c l d(1 - F(l)) - \int_a^c a d(1 - F(l)) \right) \\
&= - \left(c(1 - F(c)) - a(1 - F(a)) - \int_a^c (1 - F(l)) dl \right. \\
&\quad \left. - a((1 - F(c)) - (1 - F(a))) \right) \\
&= - c(1 - F(c)) + a(1 - F(a)) + \int_a^c (1 - F(l)) dl \\
&\quad + a(1 - F(c)) - a(1 - F(a)).
\end{aligned}$$

Da nun $a(1 - F(c)) \rightarrow 0, c \rightarrow \infty$ und $|c|(1 - F(c)) \rightarrow 0, c \rightarrow \infty$ analog zu (4.3) gilt, erhalten wir

$$\int_a^\infty (l - a) dF(l) = \int_a^\infty (1 - F(l)) dl.$$

Sei nun $a > Q_p(F) = F^{-1}(p) := q$. Dann gilt

$$\begin{aligned}
&\mathbb{E}_F[S(a, l)] - \mathbb{E}_F[S(T(F), l)] \\
&= (1 - p) \left(\int_{-\infty}^a F(l) dl - \int_{-\infty}^q F(l) dl \right) + p \left(\int_a^\infty (1 - F(l)) dl - \int_q^\infty (1 - F(l)) dl \right) \\
&= (1 - p) \int_q^a F(l) dl - p \int_q^a (1 - F(l)) dl \\
&= \int_q^a ((1 - p)F(l) - p(1 - F(l))) dl > 0,
\end{aligned}$$

da für alle $l \in (q, a]$ gilt, dass $F(l) - p > 0$ ist.

Sei nun $a < Q_p(F) := q$. Dann gilt

$$\begin{aligned}
&\mathbb{E}_F[S(a, l)] - \mathbb{E}_F[S(T(F), l)] \\
&= (1 - p) \left(\int_{-\infty}^a F(l) dl - \int_{-\infty}^q F(l) dl \right) + p \left(\int_a^\infty (1 - F(l)) dl - \int_q^\infty (1 - F(l)) dl \right) \\
&= \int_a^q (-(1 - p)F(l) + p(1 - F(l))) dl > 0,
\end{aligned}$$

da für alle $l \in [a, q)$ gilt, dass $-F(l) + p > 0$ ist.

Somit ist $S(a, l)$ eine strikt-konsistente \mathcal{M}_c^{sm} -Scoring-Funktion für $T(F) = Q_p(F)$. \square

Bemerkung 4.2.7. Im Allgemeinen ist das Quantil Q_p auf ganz \mathcal{M}^0 elizitierbar mit der strikt-konsistenten \mathcal{M}^0 -Scoring-Funktion

$$S(a, l) = \mathbb{1}_{\{l > a\}}g(l) + (\mathbb{1}_{\{l \leq a\}} - p)g(a) + h(l), \quad a, l \in \mathbb{R}, \quad (4.4)$$

wobei g streng monoton wachsend und $h(l)$ eine Funktion ist (Siehe [Fissler et al., 2024]). Durch $g(x) = x$ und $h(l) = (1 - p)l$ erhält man dann genau die Scoring-Funktion in Satz 4.2.6.

Definition 4.2.8 (Bedingte Elizitierbarkeit). Ein Funktional $T : \mathcal{M} \rightarrow A$ ist \mathcal{M} -bedingt elizitierbar mit einem Funktional $T' : \mathcal{M} \rightarrow A'$, falls T' \mathcal{M} -elizitierbar ist und T $\mathcal{M}(r, T')$ -elizitierbar ist für jedes $r \in A'$, wobei

$$\mathcal{M}(r, T') = \{F \in \mathcal{M} : r \in T'(F)\}.$$

Satz 4.2.9 (Satz 5.3 aus [Fissler et al., 2024]). Sei (ρ, ρ^*) ein p -Tail-Paar für ein $p \in (0, 1)$. Wenn $\rho^* : \mathcal{M} \rightarrow \mathbb{R}$ \mathcal{M} -Elizitierbar mit der strikt-konsistenten \mathcal{M} -Scoring-Funktion $S^* : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ist, dann gilt Folgendes:

(i) ρ ist \mathcal{M} -bedingt elizitierbar mit Q_p . Für jedes $v \in \mathbb{R}$ ist die Funktion

$$S_v(a, l) = \mathbb{1}_{\{l > v\}} S^*(a, l) + (\mathbb{1}_{\{l \leq v\}} - p) S^*(a, v), \quad a, l \in \mathbb{R}$$

eine strikt-konsistente $\mathcal{M}(v, Q_p)$ -Scoring-Funktion für ρ , wobei

$$\mathcal{M}(v, Q_p) = \{F \in \mathcal{M} : v \in Q_p(F)\}.$$

(ii) Wenn für alle $a \in \mathbb{R}$ gilt, dass $S^*(a, l)$ streng monoton wachsend in l ist, dann ist (Q_p, ρ) \mathcal{M} -elizitierbar mit der strikt-konsistenten \mathcal{M} -Scoring-Funktion

$$S(v, a, l) = \mathbb{1}_{\{l > v\}} S^*(a, l) + (\mathbb{1}_{\{l \leq v\}} - p) S^*(a, v), \quad v, a, l \in \mathbb{R}.$$

Beweis. Zu (i):

Sei $F \in \mathcal{M}$ und $v \in Q_p(F)$. Es gilt

$$\begin{aligned} & \int_{\mathbb{R}} S_v(a, l) dF(l) \\ &= \int_{\mathbb{R}} (\mathbb{1}_{\{l > v\}} S^*(a, l)) dF(l) + \int_{\mathbb{R}} ((\mathbb{1}_{\{l \leq v\}} - p) S^*(a, v)) dF(l) \\ &= \int_v^{\infty} S^*(a, l) dF(l) + S^*(a, v) \left(\int_{-\infty}^v 1 dF(l) - \int_{\mathbb{R}} p dF(l) \right) \\ &= \int_v^{\infty} S^*(a, l) dF(l) + S^*(a, v)(p - p) \\ &= \int_v^{\infty} S^*(a, l) dF(l). \end{aligned}$$

Da $F(v) = p$ ist, gilt

$$d(F(l) - p)_+ = \begin{cases} 0, & l < v \\ dF(l), & l \geq v \end{cases}$$

und somit

$$\int_{\mathbb{R}} S_v(a, l) dF(l) = \int_{\mathbb{R}} S^*(a, l) d(F(l) - p)_+.$$

Da sich die Minimierungsstelle des Erwartungswerts nicht ändert, wenn wir das Maß $d(F(l) - p)_+$ normieren, erhalten wir

$$\begin{aligned} \arg \min_{a \in \mathbb{R}} \int_{\mathbb{R}} S_v(a, l) dF(l) &= \arg \min_{a \in \mathbb{R}} \int_{\mathbb{R}} S^*(a, l) d\left(\frac{(F(l) - p)_+}{1 - p}\right) \\ &= \arg \min_{a \in \mathbb{R}} \int_{\mathbb{R}} S^*(a, l) dF_p(l) \\ &= \rho^*(F_p) = \rho(F) \end{aligned}$$

Zu (ii):

Da $S^*(a, l)$ streng monoton wachsend in l für alle $a \in \mathbb{R}$ ist, ist die Funktion

$$S(v, l) = \mathbb{1}_{\{l > v\}} S^*(a, l) + (\mathbb{1}_{\{l \leq v\}} - p) S^*(a, v) = S(v, a, l), \quad v, l \in \mathbb{R}$$

wegen Bemerkung 4.2.7 strikt-konsistent für Q_p für alle $a \in \mathbb{R}$. Also gilt für alle $a \in \mathbb{R}$:

$$Q_p(F) = \arg \min_{v \in \mathbb{R}} \int_{\mathbb{R}} S(v, a, l) dF(l).$$

Da nun $v \in Q_p(F)$ ist, gilt

$$\arg \min_{a \in \mathbb{R}} \int_{\mathbb{R}} S(v, a, l) dF(l) \stackrel{(i)}{=} \rho^*(F_p) = \rho(F).$$

Somit gilt die Behauptung. □

Bemerkung 4.2.10. Die gemeinsame Elizitierbarkeit von (Q_p, ES_p) unter gewissen Voraussetzungen wurde bereits durch [Acerbi and Szekely, 2014] gezeigt. Später konnte das Ergebnis dann verallgemeinert werden.

Beispiel 4.2.11. Sei $\rho^* = \mathbb{E}$ der Erwartungswert, $\phi(x) = x^2$ und $\rho = ES_p$ der Expected Shortfall. Nach Beispiel 4.2.3 ist der Erwartungswert \mathbb{E} elizitierbar mit der strikt-konsistenten \mathcal{M}^2 -Scoring-Funktion

$$S(a, l) = (l - a)^2 \quad a, l \in I,$$

wobei ϕ strikt konvex und differenzierbar ist.

Nach Satz 4.2.9 ist ES_p \mathcal{M}^2 -bedingt elizitierbar mit Q_p . Für $v \in \mathbb{R}$ ist

$$S_v(a, l) = \mathbb{1}_{\{l > v\}} (l - a)^2 + (\mathbb{1}_{\{l \leq v\}} - p) (v - a)^2, \quad a, l \in \mathbb{R} \quad (4.5)$$

die strikt-konsistente $\mathcal{M}^2(v, Q_p)$ -Scoring-Funktion für ES_p , wobei

$$\mathcal{M}^2(v, Q_p) = \{F \in \mathcal{M}^2 : v \in Q_p(F)\}.$$

Beispiel 4.2.12 (τ -Expectile). Sei $\tau \in (0, 1)$. Das τ -Expectile ρ_τ ist für $F \in \mathcal{M}^1$ definiert als die eindeutige Lösung $t = \rho_\tau(F)$ der Gleichung

$$\tau \int_t^\infty (l - t) dF(l) = (1 - \tau) \int_{-\infty}^t (t - l) dF(l).$$

Der Punkt t ist also genau dann ein Expectile, wenn die gewichtete Summe der Abweichungen links von t gleich der gewichteten Summe der Abweichungen rechts von t ist. Für $\tau = \frac{1}{2}$ ist das τ -Expectile genau der Erwartungswert.

Lemma 4.2.13 (Lemma 5.5. aus [Fissler et al., 2024]). *Angenommen, $S^* : \mathbb{R}^2 \rightarrow \mathbb{R}$ ist partiell differenzierbar bzgl. des zweiten Arguments. Dann gilt: Es gibt eine differenzierbare Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ so, dass $l \mapsto S^*(a, l) + g(l)$ monoton wachsend für alle $a \in \mathbb{R}$ ist, genau dann wenn es eine Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$ gibt so, dass*

$$h(l) \leq \partial_l S^*(a, l), \quad \text{für alle } a, l \in \mathbb{R}.$$

Beweis. Siehe [Fissler et al., 2024]. □

Proposition 4.2.14. *Sei (ρ, ρ^*) ein p -Tail-Paar und ρ^* das τ -Expectile mit $\tau \in (0, 1)$ und der \mathcal{M}^1 -Scoring-Funktion*

$$S^*(a, l) = |\mathbb{1}_{\{l \leq a\}} - \tau|(\phi(l) - \phi(a) + \phi'(a)(a - l)) + h(l),$$

wobei ϕ strikt konvex mit Ableitung $|\phi'| < C$ für ein $C > 0$ und h eine beliebige Funktion ist. Dann ist das Paar (Q_p, ρ) \mathcal{M}^1 -elizitierbar mit der strikt-konsistenten Scoring-Funktion:

$$S(v, a, l) = \mathbb{1}_{\{l > v\}} S^*(a, l) + (\mathbb{1}_{\{l \leq v\}} - p) S^*(a, v), \quad v, a, l \in \mathbb{R}.$$

Beweis. Jede Scoring-Funktion für ρ^* ist von der Form

$$S^*(a, l) = |\mathbb{1}_{\{l \leq a\}} - \tau|(\phi(l) - \phi(a) + \phi'(a)(a - l)) + h(l),$$

wobei ϕ strikt konvex mit Ableitung ϕ' und h eine beliebige Funktion ist [Gneiting, 2010]. Die partielle Ableitung für $h = 0$ und $l \neq a$ lautet

$$\partial_l S^*(a, l) = |\mathbb{1}_{\{l \leq a\}} - \tau|(\phi'(l) - \phi'(a)).$$

Für $l = a$ existieren die links- und rechtsseitige Ableitungen und sind beide 0. Wähle nun $\phi(x) = \frac{x^2}{1+|x|}$. Dann gilt für ein $C > 0$, dass $|\phi'(x)| = \frac{x(2+|x|)}{(1+|x|)^2} < C$ und somit auch

$$|\partial_l S^*(a, l)| < 2C.$$

Nun folgt aus Lemma 4.2.13, dass eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ existiert so, dass $l \mapsto S^*(a, l) + g(l)$ monoton wachsend ist. Aus Satz 4.2.9 lässt sich dann die strikt-konsistente \mathcal{M}^1 -Scoring-Funktion für (Q_p, ρ) konstruieren. □

5 Backtesting

Bemerkung 5.0.1. Backtesting bezieht sich auf die Überprüfung, ob zuvor festgelegte Risikomaße mit den tatsächlich realisierten Verlusten übereinstimmen. Dabei unterscheidet man zwei grundlegende Ansätze des Backtestings:

1. *Komparatives Backtesting*: In diesem Ansatz werden verschiedene Prognosemethoden miteinander verglichen, um zu bewerten, welche Methode bessere Schätzungen liefert. Ein Beispiel hierfür ist der Vergleich zwischen standardisierten Ansätzen (z.B. regulatorisch vorgegebenen Verfahren) und individuell entwickelten Methoden.
2. *Traditionelles Backtesting*: Hier wird getestet, ob die Prognosen der Risikomaße mit den tatsächlich eingetretenen Verlusten konsistent sind. Beim Value at Risk wird also zum Beispiel überprüft, ob die Anzahl und Schwere der Überschreitungen (sogenannte Violations) mit der erwarteten Häufigkeit übereinstimmt.

In diesem Abschnitt beschäftigen wir uns mit dem Backtesting von Zeitreihen. Dabei kommen insbesondere die in Abschnitt Kapitel 4.1 und Kapitel 4.2 eingeführten Identifikations- und Scoringfunktionen zum Einsatz.

Bemerkung 5.0.2 (Prognose für Risikomaße). Angenommen, wir betrachten nun eine Zeitreihe von Verlusten $(L_t)_{t \in \mathbb{N}}$ modelliert als

$$L_t = \mu_t + \sigma_t Z_t,$$

wobei $(Z_t)_{t \in \mathbb{N}}$ eine Folge von unabhängig identisch verteilten Zufallsvariablen mit $\mathbb{E}[Z_t] = 0$ und $\mathbb{E}[Z_t^2] = 1$ und μ_t, σ_t messbar bzgl. \mathcal{F}_{t-1} , also der Informationen zum Zeitpunkt $t - 1$, ist. Da wir versuchen realitätsnah zu sein, verwenden wir für die zeitdynamischen Werte von Verlusten hierzu das AR(1)-GARCH(1,1) Modell (siehe Anhang A.1). Sei weiter ρ ein Risikomaß. Dann ist die Prognose von ρ für den nächsten Zeitschritt

$$\rho(F(L_t | \mathcal{F}_{t-1})) = \mu_t + \sigma_t \rho(F(Z)), \quad (5.1)$$

wobei Z eine Zufallsvariable mit der gleichen Verteilung wie die (Z_t) (siehe [Nolde and Ziegel, 2017]). Die Prognose von $\rho(F(L_t | \mathcal{F}_{t-1}))$ erfolgt dann in zwei Schritten: Zuerst schätzt man die Parameter des AR(1)-GARCH(1,1) Modells durch das Maximum-Likelihood Verfahren unter Annahme der Standardnormalverteilung von Z . Danach schätzt man $\rho(F(Z))$ auf Basis der standardisierten Residuen $\hat{z}_t = \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}$. Wir benutzen dafür die *voll parametrische Schätzung*, bei der man eine Verteilung der Innovationen $(Z_t)_{t \in \mathbb{N}}$ annimmt. Beispiele hierfür sind meistens die Normal-, (Student'sche) t- und Schiefe-t-Verteilung (skew-t-Verteilung). Die Parameter der angenommenen Verteilung schätzt man dann durch die Residuen mit beispielsweise dem Maximum-Likelihood Verfahren. In unserer numerischen Studie in Kapitel 6 verwenden wir für die Schätzung der Parameter die Funktion 'ugarchfit' und für die Prognosen die Funktion 'ugarchforecast' aus dem 'rugarch' Paket. Die Funktion 'ugarchfit' schätzt die Parameter des spezifizierten AR(1)-GARCH(1,1)-Modells über das Maximum-Likelihood-Verfahren. Dabei werden die Likelihood-Gleichungen für die gewählte Innovationsverteilung (z. B. Normal-, t- oder skew-t-Verteilung) numerisch maximiert. Das Ergebnis sind Schätzer für die AR-Parameter, die GARCH-Parameter sowie die Parameter der Innovationsverteilung. Auf Basis dieser geschätzten Modellparameter erzeugt die Funktion 'ugarchforecast' bedingte Prognosen für den Erwartungswert und die Volatilität der Zeitreihe in zukünftigen Zeitpunkten. Diese Vorhersagen sind wiederum die Grundlage für die Berechnung der Risikomaße nach Gleichung (5.1).

5.1 Komparatives Backtesting

Bemerkung 5.1.1 (Diebold-Mariano Test). Um verschiedene Prognosen anhand der Scoring-Funktionen vergleichen zu können verwendet man häufig den Diebold-Mariano Test, der den Durchschnitt der Differenzen der Scores zu jedem Zeitpunkt t betrachtet:

$$W_n = \frac{1}{n} \sum_{k=1}^n (S(a_{t,1}, L_t) - S(a_{t,2}, L_t)).$$

$(a_{t,1})$ und $(a_{t,2})$ stehen dabei für zwei Verschiedene Prognosen, die zu jedem Zeitpunkt t Verglichen werden sollen. Dabei gibt es folgende Hypothesen:

H_0 : Die Vorhersagen $a_{t,1}$ sind genauso gut wie $a_{t,2}$.

H_1 : Die Vorhersagen $a_{t,1}$ unterscheiden sich von $a_{t,2}$.

Falls W_n negativ ist, sind die $(a_{t,1})$ besser als die $(a_{t,2})$. Ist W_n hingegen positiv, so sind die $(a_{t,2})$ besser als die $(a_{t,1})$. Das schwache Gesetz der großen Zahlen sagt uns:

$$W_n \rightarrow \lambda \in \mathbb{R}, \quad n \rightarrow \infty \quad \text{fast sicher,}$$

wobei $\lambda = \mathbb{E}[S(a_{t,1}, L_t) - S(a_{t,2}, L_t)]$.

Der Zentrale Grenzwertsatz sagt uns, dass W_n für große n asymptotisch um λ normalverteilt ist:

$$\sqrt{n}(W_n - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau^2), \quad n \rightarrow \infty.$$

Der Parameter τ^2 lässt sich schätzen durch:

$$\hat{\tau}^2 = \hat{\tau}^2(h) = \hat{\gamma}_0 = \widehat{\text{Var}}(d_t), \quad h = 1,$$

wobei $d_t := S(a_{t,1}, L_t) - S(a_{t,2}, L_t)$.

Die Teststatistik lautet daher:

$$DM = \frac{W_n}{\sqrt{\hat{\tau}^2/n}}.$$

Diese ist unter der Nullhypothese ($\lambda = 0$) standardnormalverteilt.

5.2 Traditionelles Backtesting

Bemerkung 5.2.1. Sei $\Theta = (\rho_1, \dots, \rho_k)$ ist ein identifizierbares Funktional mit Identifikationsfunktion V für alle $F \in \mathcal{M}$ und unser Prognoseverfahren ist das aus Bemerkung 5.0.2.

Definition 5.2.2. Eine Folge von Vorhersagen $(a_t)_{t \in \mathbb{N}}$ heißt *im Durchschnitt kalibriert* für Θ , falls

$$\mathbb{E}[V(a_t, L_t)] = 0 \quad \text{für alle } t \in \mathbb{N}.$$

Die Folge heißt *im Durchschnitt super-kalibriert (sub-kalibriert)* für Θ , falls $\mathbb{E}[(V(a_t, L_t))] \geq 0$ ($\mathbb{E}[(V(a_t, L_t))] \leq 0$) komponentenweise für alle $t \in \mathbb{N}$ gilt.

Die Folge heißt *bedingt kalibriert* für Θ , falls

$$\mathbb{E}[(V(a_t, L_t)|\mathcal{F}_{t-1})] = 0 \quad \text{fast sicher für alle } t \in \mathbb{N}.$$

Bedingt super-kalibriert und *bedingt sub-kalibriert* für Θ ist analog definiert wie oben.

Bemerkung 5.2.3. Wir werden, wie in [Nolde and Ziegel, 2017], die Nullhypothese

$$H_0 : \text{Die Folge von Vorhersagen } (a_t)_{t \in \mathbb{N}} \text{ ist bedingt kalibriert für } \Theta. \quad (5.2)$$

testen. Da wir in diesem Kapitel vor allem die Anwendung der in Kapitel 4.1 vorgestellten Identifikationsfunktionen demonstrieren wollen, werden wir einfachheitshalber wieder nur Ein-Schritt Prognosen betrachten und die sogenannten *simplen* bedingten Kalibrierungstests durchführen. Die Voraussetzung für die Nullhypothese $\mathbb{E}[V(a_t, L_t) | \mathcal{F}_{t-1}] = 0$ fast sicher ist äquivalent dazu, dass $\mathbb{E}[h_t^\top V(a_t, L_t)] = 0$ für alle \mathcal{F}_{t-1} messbaren \mathbb{R}^k -wertigen Funktionen h_t ist. Wir betrachten eine \mathcal{F} -messbare Folge $(H_t)_{t \in \mathbb{N}}$ von $q \times k$ -Matrizen H_t , die die sogenannten *Test-Funktionen* sind, um einen *Wald-Test* mit der Statistik T_1 zu konstruieren:

$$T_1 = n \left(\frac{1}{n} \sum_{t=1}^n H_t V(a_t, L_t) \right)^\top \hat{\Omega}_n^{-1} \left(\frac{1}{n} \sum_{t=1}^n H_t V(a_t, L_t) \right), \quad (5.3)$$

wobei

$$\hat{\Omega}_n = \frac{1}{n} \sum_{t=1}^n (H_t V(a_t, L_t))(H_t V(a_t, L_t))^\top$$

ein Schätzer der Varianz des Vektors $H_t V(a_t, L_t)$ ist. Idealerweise wählt man q so, dass die Reihen von $H_t \mathcal{F}_{t-1}$ erzeugen. Im simplen Kalibrierungstest setzen wir $H_t = 1$.

6 Numerische Studie

Bemerkung 6.0.1. Die Zeitreihe $\{L_t\}_{t \in \mathbb{N}}$ für die Analyse ist durch einen $AR(1)$ - $GARCH(1, 1)$ Prozess mit folgenden Parametern generiert:

$$\begin{aligned} L_t &= 0,1 \cdot L_{t-1} + u_t \\ u_t &= \sigma_t \cdot Z_t \\ \sigma_t^2 &= 0,01 + 0,1 \cdot u_{t-1}^2 + 0,85 \cdot \sigma_{t-1}^2, \end{aligned}$$

wobei die Innovationen $\{Z_t\}_{t \in \mathbb{N}}$ standardnormalverteilt sind.

6.1 Scores und Diebold-Mariano Tests

Bemerkung 6.1.1. Beim komparativen Backtesting vergleichen wir Vorhersagen nach der voll parametrischen Schätzung mit einer Vorhersage, die die wahren Parameter unserer Zeitreihe kennt (opt). Bei unserer voll parametrischen Schätzung nehmen wir an, dass die zugrundeliegende Zeitreihe einem $AR(1)$ - $GARCH(1, 1)$ Prozess folgt und die Innovationen entweder normalverteilt, student-t-verteilt oder skewed-t-verteilt sind. Wir schätzen die Parameter mit einem bewegenden Fenster von 500 Werten. Es wird also der 501. Wert anhand der ersten 500 Werte geschätzt, der 502. wird anhand des 2. bis 501. geschätzt und so weiter. Insgesamt hat unsere Zeitreihe eine Größe von 2500. Für den Vergleich nutzen wir dann die Scoring-Funktionen aus Kapitel 4.2 und den Diebold-Mariano Test. Je kleiner der Wert der Scoring-Funktion für die Vorhersage im Durchschnitt ist, desto besser ist die Vorhersage.

Tabelle 6.1: Auflistung durchschnittlicher Werte und Scores des Value at Risk und Expected Shortfall mit vollparametrischer Schätzung (Innovationsverteilung normal, student-t oder skew-t) und der Schätzung mit vorgegebenen Parameter. Die Spalten \overline{VaR}_p und \overline{ES}_p zeigen jeweils die durchschnittlichen Werte des jeweiligen Risikomaßes, während $\frac{1}{1-p}\overline{S}$ die durchschnittlichen - mit $\frac{1}{1-p}$ skalierten - Scores aus (4.4) und (4.5) angeben.

p	Methode	\overline{VaR}_p	$\frac{1}{1-p}\overline{S}$ [(4.4)]	\overline{ES}_p	$\frac{1}{1-p}\overline{S}$ [(4.5)]
0,99	norm	0,9967	1,2100	1,1407	0,0212
	std	0,9989	1,2100	1,1433	0,0215
	sstd	1,0003	1,2109	1,1449	0,0218
	opt	1,0036	1,1790	1,1497	0,0144
0,95	norm	4,75	0,8995	0,8846	0,0366
	std	4,60	0,9000	0,8866	0,0366
	sstd	4,70	0,9009	0,8878	0,0365
	opt	4,45	0,8876	0,8900	0,0334
0,90	norm	9,95	0,7553	0,7537	0,0407
	std	9,95	0,7558	0,7554	0,0407
	sstd	9,85	0,7561	0,7565	0,0409
	opt	9,80	0,7493	0,7573	0,0373

Bemerkung 6.1.2. Wir sehen in Tabelle 6.1, dass es insgesamt keine großen Abweichungen zwischen den verschiedenen Schätzmethode gibt. Wie erwartet liefert die optimale Schätzmethode die niedrigsten Scores. Dahinter folgt die vollparametrische Schätzung mit normalverteilten Innovationen, was zu erwarten war, da der zugrunde liegende Prozess tatsächlich normalverteilte Innovationen aufweist. Die Schätzungen mit angenommener t- und skew-t-Verteilung schneiden etwas schlechter ab.

Beispiel 6.1.3 (Diebold-Mariano Test). Wir wollen nun den Diebold-Mariano Test aus Bemerkung 5.1.1 auf unsere Scores für den VaR und den ES anwenden und damit die Schätzmethoden vergleichen und auf signifikante Unterschiede überprüfen. Dazu vergleichen wir die Schätzung mit den angenommenen normalverteilten Innovationen mit den jeweils anderen Schätzmethoden auf dem $p = 0,95$ Niveau. Unsere Nullhypothese lautet dabei:

$$H_0 : \lambda = 0,$$

während unsere Alternative

$$H_1 : \lambda \neq 0$$

ist. Wir führen diesen beidseitigen Test auf dem Signifikanzniveau von 5% im Statistik Programm R durch. Die Ergebnisse sind in Tabelle 6.2 zusammengefasst.

Tabelle 6.2: Teststatistiken und zugehörige p -Werte des Diebold-Mariano Tests. Wir vergleichen hierbei alle Methoden mit der Schätzmethode mit angenommener Normalverteilung. Die fettgedruckten p -Werte liegen unter dem Konfidenzniveau von 5%. In diesen Fällen würde man die Nullhypothese ablehnen.

p	Vergleich von	VaR-Score		ES-Score	
		DM	p -Wert	DM	p -Wert
0.99	norm-std	-0.0365	0.9709	1.3767	0.1688
	norm-sstd	-0.4483	0.6540	1.0720	0.2838
	norm-opt	1.5948	0.1109	0.1751	0.8610
0.95	norm-std	-1.3550	0.1756	0.2836	0.7768
	norm-sstd	-2.3315	0.0198	-0.4355	0.6633
	norm-opt	2.1478	0.0319	1.6283	0.1036
0.90	norm-std	-1.7820	0.0749	0.3103	0.7563
	norm-sstd	-1.9260	0.0542	0.6749	0.4998
	norm-opt	1.9199	0.0550	0.3905	0.6962

Bemerkung 6.1.4. Es fällt auf: Für $p = 0,99$ ergeben sich keine signifikanten Unterschiede zwischen den Schätzmethoden. Zwar zeigt die Teststatistik, dass die optimale Methode beim VaR leicht besser abschneidet als die Normalverteilung, die p -Werte liegen jedoch deutlich oberhalb des Signifikanzniveaus. Auch beim Expected Shortfall sind die Abweichungen zwischen den Modellen gering. Hier bemerken wir allerdings, dass die Methode mit angenommener Normalverteilung am schlechtesten abschneidet. Für $p = 0,95$ ist die Teststatistik beim Vergleich von normalverteilten und t - bzw. skew- t -verteilten Innovationen negativ, was auf eine bessere Performance der Schätzung mit Normalverteilung hindeutet. Signifikant ablehnen kann man die Nullhypothese jedoch nur beim Vergleich der normal- gegenüber der skew- t -Verteilung ($p = 0,0198$) und der optimalen Methode ($p = 0,0319$). Im ersten Fall würde man die Methode mit Normalverteilung und im zweiten Fall die optimale Methode präferieren. Für den Expected Shortfall ergeben sich beim Vergleich der Schätzmethoden hingegen keine signifikanten Unterschiede. Weder zwischen der Schätzung mit normalverteilten und der mit skew- t -verteilten Innovationen noch zwischen der normalverteilten Schätzung und der optimalen Methode ist die Nullhypothese signifikant ablehnbar. Für $p = 0,9$ ergibt sich für den VaR ein ähnliches Bild wie für $p = 0,95$. Allerdings liegen unsere p -Werte hier knapp über dem 5%-Niveau, weshalb wir keine signifikanten Unterschiede feststellen können. Beim Expected Shortfall erhalten wir die gleiche Tendenz wie bei $p = 0,99$. Allerdings sind die p -Werte sehr hoch, sodass sich auch hier keine signifikanten Unterschiede ergeben. Dass beim ES generell keine signifikanten Unterschiede feststellbar sind, könnte an der gemeinsamen Scoringfunktion aus (4.5) liegen, die den Expected Shortfall schwerer gewichtet als den VaR.

Insgesamt zeigt sich, dass die Normalverteilung als zugrundeliegende Innovationsverteilung nicht durchgängig am besten abschneidet, obwohl sie im Simulationsmodell tatsächlich korrekt spezifiziert ist. Der Grund liegt darin, dass die Schätzmethoden auf endlichen Stich-

proben basieren und die Anpassung in den Tails für das Backtesting entscheidend ist, da mit dem VaR und ES ausschließlich Tail-Risikomaße betrachtet werden. Die verwendeten Scoring-Funktionen bestrafen Unterschätzungen der Tail-Risikomaße stark, was zu einem erhöhten Score führt. Modelle mit schwereren Tails (wie t- oder skew-t-Verteilungen) reagieren robuster auf Extremwerte und liefern deshalb in endlichen Stichproben oft bessere Ergebnisse. Dadurch kann es vorkommen, dass ein vermeintlich falsch spezifiziertes Modell im Backtesting besser abschneidet als das tatsächlich richtige Modell. Zusammenfassend schneiden alle Methoden im Vergleich zur Normalverteilung ähnlich gut ab. Signifikante Unterschiede ergeben sich lediglich beim VaR zwischen der Normal- und der skew-t-Verteilung bzw. zwischen der Normalverteilung und der optimalen Methode.

6.2 Kalibrierungstests

Bemerkung 6.2.1. Zur Varianzschätzung verwenden wir einen sogenannten *Heteroskedasticity and Autocorrelation Consistent* (HAC)-Schätzer. Dieser berücksichtigt, dass die Sequenz der Identifikationsfunktionen im Zeitreihenrahmen sowohl heteroskedastisch als auch autokorreliert sein kann. Würde man lediglich das Kreuzprodukt der Identifikationsfunktionen verwenden, entspräche dies der Annahme von unkorrelierten Residuen. Diese Annahme ist in unserem Kontext nicht realistisch, weshalb die Schätzung der Kovarianzmatrix über einen HAC-Ansatz notwendig ist. Ein zentrales Element dabei ist die sogenannte Bandbreite, die angibt, wie viele Autokovarianzen in die Schätzung einbezogen werden. Eine zu kleine Bandbreite vernachlässigt Abhängigkeiten, während eine zu große Bandbreite die Schätzung unnötig verrauschen lässt. In der praktischen Umsetzung in R wird hierfür die Funktion `'vcovHAC'` aus dem Paket `'sandwich'` verwendet. Die Bandbreite kann z. B. mit der Funktion `'bwAndrews'` automatisch gewählt werden. Dadurch wird die Kovarianzmatrix an die zeitabhängige Struktur der Daten angepasst und in die Berechnung der Wald-Teststatistik einbezogen.

Bemerkung 6.2.2. Unter der Nullhypothese (5.2) gilt dann: $T_1 \xrightarrow{d} \chi_q^2$. Wir können also ein Signifikanzniveau α festlegen so, dass für $T_1 > \chi_{q,\alpha}^2$ H_0 verworfen wird. $\chi_{q,\alpha}^2$ ist dabei das α -Quantil der χ^2 -Verteilung mit q Freiheitsgraden. Analog geht man für die Identifikationsfunktion des Paares (VaR_p, ES_p) vor.

Bemerkung 6.2.3. Die Ergebnisse aus den durchgeführten Tests (Tabelle 6.3) zeigen, dass durchweg keine signifikanten Abweichungen festgestellt werden können. Sowohl für den VaR als auch für den ES liegen die Teststatistiken T_1 deutlich unter den jeweils kritischen Werten der zugehörigen χ_1^2 Verteilung. Dies bedeutet, dass die Nullhypothese (5.2), die besagt, dass die jeweiligen Verfahren korrekt kalibriert sind, in keinem Fall verworfen werden kann. Es gibt also keinen statistischen Hinweis darauf, dass die Schätzungen von VaR und ES nicht adäquat sind. Die durchgeführten Tests liefern somit konsistent die Aussage, dass die Methoden keine systematischen Abweichungen zeigen und dass die Kalibrierung für die betrachteten Risikomaße als gelungen gelten kann.

Ein Grund hierfür könnte sein, dass die t-Verteilung (und die skew-t-) Verteilung mit hohem Freiheitsgrad (und Schiefeparameter nahe bei 1) gegen die Normalverteilung konvergieren. Die Schätzverfahren mit angenommener t- (bzw. skew-t-) Verteilung schätzen

in unserem Fall tatsächlich für die t-Verteilung den Freiheitsgrad $\nu \approx 92,9$ und für die skew-t-Verteilung den Freiheitsgrad $\nu \approx 58,9$ mit Schiefeparameter $\gamma \approx 0,99$.

Tabelle 6.3: Ergebnis-Tabelle der Wald-Teststatistik und die dazugehörigen p -Werte für VaR und ES. Die Spalte „ p “ gibt das Konfidenzniveau für die Risikomaße an, die Spalte „Methode“ die verwendete Methode und die Spalten „ T_1 “ und „ p -Wert“ die Teststatistik und den zugehörigen p -Werte aus (5.3) bzw. Bemerkung 6.2.2 für das jeweilige Risikomaß.

p	Methode	VaR		ES	
		T_1	p -Wert	T_1	p -Wert
0.99	norm	1.3539	0.2446	1.2245	0.2685
	std	1.3539	0.2446	1.2244	0.2685
	sstd	1.3539	0.2446	1.2228	0.2688
	opt	0.9753	0.3234	0.8759	0.3493
0.95	norm	0.2754	0.5997	0.2700	0.6033
	std	0.7296	0.3930	0.6981	0.4034
	sstd	0.4012	0.5265	0.3904	0.5321
	opt	1.4218	0.2331	1.4891	0.2224
0.90	norm	0.0056	0.9405	0.0058	0.9391
	std	0.0056	0.9405	0.0058	0.9391
	sstd	0.0507	0.8219	0.0531	0.8178
	opt	0.0913	0.7626	0.1001	0.7517

7 Fazit

In dieser Arbeit wurde untersucht, wie die Prognose von Risikomaßen mit Hilfe der Konzepte Identifizierbarkeit und Elizitierbarkeit analysiert und bewertet werden kann. Zunächst wurden die theoretischen Grundlagen von Risikomaßen und Tail-Risikomaßen sowie deren Eigenschaften dargestellt. Aufbauend darauf wurde gezeigt, wie sich mit Hilfe der Resultate aus [Fissler et al., 2024] die Beziehung zwischen allgemeinen Risikomaßen und deren Tail-Risikomaßen formal begründen lässt. Ein zentrales Ergebnis dabei ist, dass für ein identifizierbares bzw. elizitierbares Risikomaß auch das zugehörige Tail-Risikomaß identifizierbar bzw. elizitierbar zusammen mit dem Quantil ist.

Die anschließende numerische Studie verdeutlichte die praktische Anwendbarkeit der theoretischen Konzepte. Mit Hilfe von Scoring-Funktionen und dem Diebold-Mariano-Test konnten Unterschiede in der Prognosegüte verschiedener Schätzmethode aufgezigt werden. Darüber hinaus gaben die Identifikationsfunktionen Aufschluss darüber, ob die Vorhersagen insgesamt konsistent waren oder signifikante Abweichungen auftraten. Die wichtigsten Ergebnisse lassen sich wie folgt zusammenfassen:

1. Beim VaR schneidet die Schätzung mit vorgegebenen Parametern (opt) am Besten und die Schätzung mit skew-t-Verteilung am schlechtesten ab. Zwischen Schätzung mit Normal- und t-Verteilung gibt es keine signifikanten Unterschiede.
2. Beim ES gibt es keine signifikanten Unterschiede zwischen den verschiedenen Schätzmethoden.
3. Im traditionellen Backtesting kann die Nullhypothese, die besagt, dass die Vorhersagen kalibriert sind, in keinem Fall verworfen werden.

Trotz dieser Erkenntnisse müssen wir Einschränkungen beachten und die Ergebnisse kritisch betrachten. So wurde zum Beispiel der AR(1)-GARCH(1,1) Prozess nur mit normalverteilten Innovationen durchgeführt, was die guten Schätzungen der Methoden mit t- und skew-t-Verteilung erklärt. Außerdem betrachten wir in unserer Datenanalyse eine endliche Zeitreihe mit 2500 Datenpunkten. Für zukünftige Arbeiten würde sich anbieten die Zeitreihe zu vergrößern, was den Schätzungen mehr Stabilität geben würde. Außerdem könnte man die Innovationsverteilung anpassen und das rolling Window von 500 zu verkleinern und/oder zu vergrößern um weitere Aussagen für die verschiedenen Schätzmethoden treffen zu können.

A Zeitreihenanalyse

A.1 AR-GARCH-Modelle

Definition A.1.1 (Zeitreihe). Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine *reelle Zeitreihe* ist eine Familie $\mathcal{X} = (X_t)_{t \in T}$ von reellen Zufallsvariablen $X_t : \Omega \rightarrow \mathbb{R}$ mit Indexmenge $T = \mathbb{N}, T = \mathbb{N}_0$ oder auch $T = \mathbb{Z}$. Sie heißt

- *zentriert*, falls $\mathbb{E}[X_t] = 0$ für alle $t \in T$ ist,
- *quadrat-integrierbar*, falls $\mathbb{E}[X_t^2] < \infty$ für alle $t \in T$ ist.

Definition A.1.2 (Stationarität). Eine reelle Zeitreihe $\mathcal{X} = (X_t)_{t \in T}$ heißt

- *stationär*, wenn für alle $n, h \in \mathbb{N}$ und $t_1 < \dots < t_n$ mit $t_j \in T$ gilt, dass

$$(X_{t_1}, \dots, X_{t_n})^\top \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h})^\top.$$

- *schwach stationär*, wenn
 - \mathcal{X} quadrat-integrierbar ist,
 - $\mathbb{E}[X_t] = \mu$ für alle $t \in T$ für ein $\mu \in \mathbb{R}$ gilt,
 - für alle $h \in \mathbb{N}$ und $t, s \in T$ gilt, dass

$$\text{Cov}(X_{t+h}, X_{s+h}) = \text{Cov}(X_t, X_s).$$

Definition A.1.3 (AR(p) Modell). Eine stationäre Zeitreihe (X_t) folgt einem *autoregressiven Modell der Ordnung p* (kurz: AR(p)), wenn es Konstanten $\phi_1, \dots, \phi_p \in \mathbb{R}, \phi_p \neq 0, \mu \in \mathbb{R}$ und unkorrelierte Zufallsvariablen (e_t) gibt mit $\mathbb{E}[e_t^2] < \infty$ und $\mathbb{E}[e_t] = 0$ so, dass

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + e_t, \quad \forall t \in T.$$

Dies lässt sich umschreiben zu:

$$\begin{aligned} X_t &= \phi_1 X_{t-1} - \phi_1 \mu + \dots + \phi_p X_{t-p} - \phi_p \mu + e_t + \mu \\ &= \mu(1 - \phi_1 - \dots - \phi_p) + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + e_t \\ &= \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + e_t, \end{aligned}$$

wobei $\phi_0 = \mu(1 - \phi_1 - \dots - \phi_p)$.

Beispiel A.1.4 (AR(1)). Wir betrachten für eine Zeitreihe $(X_t)_{t \in T}, T = \mathbb{Z}$ die Darstellung

$$X_t = \phi X_{t-1} + e_t, \quad \phi \in \mathbb{R},$$

wobei die (e_t) unabhängig identisch Verteilt sind mit $\mathbb{E}[e_t] = 0$ und $\mathbb{E}[e_t^2] < \infty$.

Fall 1: $|\phi| < 1$.

Durch Induktion über $k \geq 1$ erhält man

$$\begin{aligned} X_t &= \phi(\phi X_{t-2} + e_{t-1}) + e_t \\ &= \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j e_{t-j}. \end{aligned}$$

Wenn X_t stationär ist, also den gleichen Erwartungswert für alle $t \in T$ hat, gilt wegen $|\phi| < 1$, dass $\phi^k X_{t-k} \rightarrow 0$ für $k \rightarrow \infty$. Daraus folgt

$$X_t = \sum_{j=0}^{\infty} \phi^j e_{t-j}. \quad (\text{A.1})$$

Andersrum bildet (A.1) eine stationäre Lösung für die Gleichung $X_t = \phi X_{t-1} + e_t$, die eindeutig bestimmt ist für gegebene e_t .

Fall 2: $\phi = 1$.

Es gilt $X_t = X_{t-1} + e_t$ und somit

$$X_{t+h} = X_t + e_{t+1} + \dots + e_{t+h}, \quad h \in \mathbb{Z}.$$

Daher gilt auch

$$\begin{aligned} X_{t+h} - X_t &= X_0 + e_1 + \dots + e_{t+h} - (X_0 + e_1 + \dots + e_t) \\ &= e_{t+1} + \dots + e_{t+h}. \end{aligned}$$

Somit gilt für $\sigma_e^2 > 0$:

$$\mathbb{E}[(X_{t+h} - X_t)^2] = \sum_{j=1}^h \mathbb{E}[e_{t+j}^2] = h\sigma_e^2 \rightarrow \infty, \quad h \rightarrow \infty$$

und daher existiert keine stationäre Lösung. Analog geht man für $\phi = -1$ vor.

Fall 3: $\phi > 1$.

Die Gleichung aus (A.1) gilt offensichtlich nicht mehr. Man könnte schreiben:

$$\begin{aligned} X_t &= X_t + \phi^{-1}e_{t+1} - \phi^{-1}e_{t+1} \\ &= \phi^{-1}X_{t+1} - \phi^{-1}e_{t+1} \\ &= \phi^{-k}X_{t+k} - \sum_{j=1}^k \phi^{-j}e_{t+j} \\ &\rightarrow -\sum_{j=1}^{\infty} \phi^{-j}e_{t+h}, \quad k \rightarrow \infty, \end{aligned}$$

da $\phi > 1$. Allerdings würde die Lösung dann von den zukünftigen Ereignissen e_{t+j} abhängen.

Nun betrachten wir Zeitreihen, bei denen die Varianz nicht konstant bleibt, sondern sich im Zeitverlauf verändert. Dieses Verhalten bezeichnet man als Heteroskedastizität. Dazu führen wir das GARCH-Modell ein.

Definition A.1.5. Eine stationäre Zeitreihe $(X_t)_{t \in \mathbb{Z}}$ heißt ARCH(∞) Zeitreihe (*autoregressive conditionally heteroskedastic*), wenn für unabhängige und identisch verteilte Zufallsvariablen $(e_t)_{t \in \mathbb{Z}}$ mit $\mathbb{E}[e_t] = 0$, $\mathbb{E}[e_t^2] = 1$, sowie für eine stationäre, positive Zeitreihe $(\sigma_t)_{t \in \mathbb{Z}}$ gilt, dass

$$X_t = \sigma_t e_t,$$

$$\sigma_t \text{ ist messbar bzgl. } \mathcal{F}_{t-1} := \sigma\{e_s, s \leq t-1\}.$$

Die Zufallsvariablen (e_t) heißen Innovationen vom ARCH Modell und (σ_t) die Volatilität zum Zeitpunkt t .

Definition A.1.6. Wenn $(X_t)_{t \in \mathbb{Z}}$ eine ARCH(∞) Zeitreihe bzgl. Innovationen $(e_t)_{t \in \mathbb{Z}}$ ist, dann heißt (X_t) eine GARCH(p, q) (generalized-ARCH) Zeitreihe, wenn es Konstanten $a_0 > 0, a_1, \dots, a_p, b_1, \dots, b_q \geq 0$ gibt so, dass (σ_t^2) die Rekursion

$$\sigma_t^2 = a_0 + \sum_{i=1}^p a_i X_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2$$

erfüllt.

Bemerkung A.1.7 (AR-GARCH-Modelle). Wenn man modellieren möchte, wie der Erwartungswert und gleichzeitig die Varianz von bisherigen Informationen abhängen, verwendet man AR Modelle kombiniert mit GARCH Modellen. Für \tilde{p}, p, q ist ein AR(\tilde{p})-GARCH(p, q) Modell gegeben durch

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_{\tilde{p}} X_{t-\tilde{p}} + u_t,$$

$$u_t = \sigma_t e_t,$$

$$\sigma_t^2 = a_0 + a_1 u_{t-1}^2 + \dots + a_p u_{t-p}^2 + b_1 \sigma_{t-1}^2 + \dots + b_q \sigma_{t-q}^2,$$

wobei $\phi_0, \dots, \phi_{\tilde{p}} \in \mathbb{R}$, $a_0, \dots, a_p, b_1, \dots, b_q \geq 0$ und (e_t) die unabhängig, identisch verteilten Zufallsvariablen mit $\mathbb{E}[e_t] = 0$ und $\text{Var}(e_t) = 1$ sind.

Bemerkung A.1.8 (AR(1)-GARCH(1,1) Modell). Das AR(1)-GARCH(1, 1) Modell hat die Form

$$X_t = \phi_0 + \phi_1 X_{t-1} + u_t,$$

$$u_t = \sigma_t e_t,$$

$$\sigma_t^2 = a_0 + a_1 u_{t-1}^2 + b_1 \sigma_{t-1}^2,$$

wobei $\phi_0 = \mu(1 - \phi_1)$ ist.

In diesem Modell ist μ der Durchschnittswert der Zeitreihe. ϕ_1 gibt an, wie stark die Zufallsvariable X_t vom Wert des Vortages der Zufallsvariable X_{t-1} abhängt. a_0 ist das Grundniveau der Varianz und a_1 gibt an wie groß der Einfluss des letzten Schocks ist, also wie hoch die Volatilität nach großen Veränderungen von u_t ist, während b_1 festlegt, wie lange hohe Volatilität nachwirkt.

A.2 Vorhersagen im AR-GARCH-Modell

Bemerkung A.2.1. (Schätzung der Parameter) Die Schätzung der Parameter $a_0 > 0, a_1, b_1 \geq 0$ im AR(1)-GARCH(1, 1) Modell erfolgt durch das Gaußsche-Maximum-Likelihood Verfahren. Also das Maximum-Likelihood Verfahren unter der Annahme, dass die Innovationen normalverteilt sind (Siehe [Straumann, 2005]).

Bemerkung A.2.2. (Schätzung der Volatilität) Angenommen wir beobachten X_1, \dots, X_n einer GARCH(p, q) Zeitreihe mit bekannten Parametern $a_0, \dots, a_p, b_1, \dots, b_q$. Die Volatilitäten $\sigma_1, \dots, \sigma_n$ sind messbar bzgl. der σ -Algebren \mathcal{F}_n , die alle Informationen bis zum Zeitpunkt n enthalten. Das ist allerdings nur der Fall in dem wir eine unendliche Vergangenheit haben. In der Praxis schätzt man die Volatilitäten durch die Beobachtungen:

$$\hat{\sigma}_t^2 = \begin{cases} a_0 + \sum_{i=1}^{\min(t-1, p)} a_i X_{t-i}^2 + \sum_{j=1}^q b_j \hat{\sigma}_{t-1}^2, & t \geq 1, \\ \frac{a_0}{1 - b_1 - \dots - b_q - a_1 - \dots - a_p}, & t = 1 - q, \dots, 0. \end{cases}$$

Bemerkung A.2.3 (GARCH-Residuen). Die GARCH-Residuen sind auf Basis der geschätzten Volatilitäten von der Form

$$\hat{e}_t = \frac{X_t}{\hat{\sigma}_t}, \quad t = 1, \dots, n$$

und schätzen die Innovationen e_t eines GARCH Modells.

Literatur

- [Acerbi and Szekely, 2014] Acerbi, C. and Szekely, B. (2014). Back-testing expected shortfall. *Risk*, 27:76–81.
- [Fissler et al., 2024] Fissler, T., Liu, F., Wang, R., and Wei, L. (2024). Elicitability and identifiability of tail risk measures.
- [Gneiting, 2010] Gneiting, T. (2010). Making and evaluating point forecasts.
- [Hewitt, 1960] Hewitt, E. (1960). Integration by parts for stieltjes integrals. *The American Mathematical Monthly*, 67(5):419–423.
- [Holzmann, 2023] Holzmann, H. (2023). Skript zur Vorlesung Quantitative Risk Management. Vorlesungsskript, Philipps-Universität Marburg.
- [Koenker and Bassett, 1978] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.
- [Liu and Wang, 2021] Liu, F. and Wang, R. (2021). A Theory for Measures of Tail Risk. *Mathematics of Operations Research*, 46(3):1109–1128.
- [Nolde and Ziegel, 2017] Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation.
- [Rüschendorf, 2013] Rüschendorf, L. (2013). *Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.
- [Shaked and Shanthikumar, 2007] Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Series in Statistics. Springer, New York.
- [Straumann, 2005] Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics. Springer Berlin, Heidelberg.