

Robuste Verlustfunktionen und Evaluierung von Volatilitätsvorhersagen mit realisierten Volatilitäten

Bachelorarbeit
am Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg

vorgelegt von
Karolin Frohnapfel

Datum: 5. März 2021
Betreuer: Prof. Dr. Hajo Holzmann

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen	3
3	Strikt konsistente Verlustfunktionen	5
3.1	Konsistenz und Elizitierbarkeit	5
3.2	Nicht-Elizitierbarkeit der Varianz	7
3.3	Elizitierbarkeit des Erwartungswertes	8
3.4	Osbands Prinzip	9
3.5	Konsistente Verlustfunktionen zum Erwartungswert	11
3.6	Beispiele	11
3.7	Zwischenfazit	12
4	Robuste Verlustfunktionen	13
4.1	Robustheit	13
4.2	Robustheit der Bregman-Funktionen	14
5	Modellierung der Stellvertreter	16
5.1	Stetige Modellierung	16
5.2	Diskrete Modellierung	18
6	Komparatives Backtesting	20
6.1	Dominanz	20
6.2	Zentraler Grenzwertsatz für abhängige Zufallsvariablen	22
6.3	Diebold-Mariano Test	24
7	Numerische Anwendung	26
7.1	Datengrundlage	26
7.2	Drei-Zonen-Ansatz	27
7.3	Testergebnisse	27
7.4	Interpretation	31
8	Fazit und Ausblick	33
9	Anhang: R-Code	35
10	Literaturverzeichnis	39

1 Einleitung

Wenn es darum geht, konkurrierende Prognosen zu vergleichen, werden typischerweise Verlustfunktionen verwendet, wie etwa der quadratische oder absolute Verlust. Um gute Prognosen zu erstellen, ist es wichtig, dass die zur Evaluierung verwendete Verlustfunktion bereits ex-ante bekannt ist und vorgegeben wird. Andernfalls können die Prognosen sehr stark voneinander abweichen. Die Alternative besteht darin, ein elizitierbares Funktional vorzugeben. Dann können für die Prognosen konsistente und strikt konsistente Verlustfunktionen verwendet werden, die sicherstellen, dass eine optimale Prognose mit minimalem erwartetem Verlust erkannt wird. Ein Problem entsteht dann, wenn es sich bei der Zielgröße um ein latentes Risikomaß handelt, also um eine Größe, die selbst ex-post nicht exakt bestimmt werden kann. Die Volatilität zählt zu den bekanntesten Beispielen für latente Risikomaße, was die Evaluierung von Volatilitätsprognosen verkompliziert. Möchte man zwei konkurrierende Prognosen bewerten, so kann man diese nicht einfach über den erwarteten Verlust mit der tatsächlich eingetretenen Volatilität vergleichen, da diese nicht bekannt ist. Um dieses Problem zu lösen, greift man auf bedingt erwartungstreue Stellvertreter, wie etwa die quadratischen Renditen oder die realisierten Volatilitäten, zurück. Die Verwendung solcher Stellvertreter kann allerdings dazu führen, dass das Ranking, welches sich aus den erwarteten Verlusten der konkurrierenden Prognosen ergibt, von dem tatsächlichen Ranking abweicht. Dann würde eventuell die Beste der Prognosen nicht erkannt und eine ihr Unterlegene als die optimale Prognose angenommen werden. Es muss also sichergestellt werden, dass eine Verlustfunktion dieses Ranking erhält, selbst unter Verwendung eines Stellvertreters. Die Verlustfunktionen, die diese Anforderung erfüllen, werden als robuste Verlustfunktionen bezeichnet. Nur unter Verwendung solcher robuster Verlustfunktionen können sinnvolle Tests zur Evaluierung von Volatilitätsprognosen durchgeführt werden. Das Ziel dieser Arbeit ist es unter anderem, diese robusten Verlustfunktionen für die Volatilität zu identifizieren. Darüber hinaus werden die beiden eben genannten Stellvertreter, die quadratischen Renditen und die realisierten Volatilitäten, näher betrachtet und es wird erörtert, welcher der beiden am geeignetsten ist, um Volatilitätsprognosen in einem komparativen Backtest zu evaluieren.

Wir werden dafür folgendermaßen vorgehen: Nachdem wir in Kapitel 2 die Volatilität in einer geeigneten Umgebung definiert und einige grundlegende Konzepte aufgeführt haben, werden wir uns in Kapitel 3 an den Arbeiten von Gneiting (2009) und Brehmer (2017) orientieren und untersuchen, inwiefern die bedingte Varianz elizitierbar ist, inwiefern es also strikt konsistente Verlustfunktionen gibt. Dabei werden wir feststellen, dass dies grundsätzlich erst einmal nicht der Fall ist. Wenn wir allerdings zentrierte Renditen annehmen, so erhalten wir über das zweite Moment den Erwartungswert als elizitierbares Funktional. Anschließend werden wir dazu konsistente und strikt konsistente Verlust-

funktionen herleiten und feststellen, dass die QLIKE-Verlustfunktion und der quadratische Verlust dazugehören. In Kapitel 4 werden wir dann als Nächstes die Problematik der Latenz thematisieren und in diesem Zusammenhang zeigen, dass die zuvor hergeleiteten konsistenten und strikt konsistenten Verlustfunktionen bereits robust sind und somit für die Evaluierung von Volatilitätsprognosen mittels Stellvertretern verwendet werden können. Denn unter Verwendung robuster Verlustfunktionen können wir Stellvertreter anstelle der tatsächlichen bedingten Varianz einsetzen und erhalten dennoch das gleiche Ranking, das sich aus den erwarteten Verlusten der konkurrierenden Prognosen ergibt. Dabei richten wir uns hauptsächlich nach der Arbeit von Patton (2006). Im darauf folgenden Kapitel 5 wenden wir uns dann den Stellvertretern zu und beziehen uns hier neben Patton (2006) auf Andersen und Benzoni (2018). Wir konzentrieren uns dabei auf die quadratischen Renditen und die realisierten Volatilitäten. Nachdem wir diese modelliert haben, werden wir in Kapitel 6 einen komparativen Backtest, konkret einen Diebold-Mariano Test, konstruieren, mit dem immer zwei konkurrierende Prognosen der Volatilität miteinander verglichen werden können. Solche Tests werden beispielsweise von Nolde und Ziegel (2017) oder Brehmer (2017) vorgeschlagen. Dazu nutzen wir aus, dass strikt konsistente Verlustfunktionen ebenfalls ordnungssensitiv sind, da wir dann das Konzept der Dominanz verwenden können, um die Prognosen zu vergleichen. Diesen Test werden wir dann nutzen, um zu prüfen, inwiefern sich die Verwendung realisierter Volatilitäten im Vergleich zu den quadratischen Renditen positiv auf dessen Aussagekraft auswirkt. Dazu werden wir im letzten Kapitel eine Datensimulation durchführen, mehrere Prognosen erstellen und diese dann mit Hilfe dieses Tests unter Verwendung beider Stellvertreter evaluieren.

2 Grundlagen

Wir werden damit beginnen, die Rahmenbedingungen dieser Arbeit zu stecken, indem wir die Volatilität in einer passenden Umgebung definieren und grundlegende Konzepte wie Verlustfunktionen und die optimale Prognose nach Bayes thematisieren. Im Zentrum steht die Volatilität, die grundsätzlich erst einmal die Standardabweichung einer Veränderung beschreibt. Ist man beispielsweise daran interessiert, ein Asset wie etwa eine Aktie oder eine Anleihe zu bewerten, so ist es naheliegend, dessen Wertänderung, also die Änderung der Renditen, im Zeitverlauf zu betrachten, um Aussagen über den zukünftigen Verlauf treffen zu können. Die Volatilität leistet genau das, indem sie die auf die Vergangenheit bedingte Standardabweichung der Renditen angibt.

Wir betrachten dafür einen Beobachtungsraum $\mathcal{O} \subseteq \mathbb{R}$, der die Renditen r_t und alle Ereignisse, die für die Volatilität σ_t eintreten können, beinhaltet. Sei nun \mathcal{A} eine σ -Algebra auf \mathcal{O} und \mathcal{F} eine Filtration unter \mathcal{A} . Darauf aufbauend definieren wir \mathcal{P} als eine Klasse von Verteilungsfunktionen auf der σ -Algebra \mathcal{A} . In dieser Umgebung definiert Patton (2006) die Volatilität wie folgt:

Definition 2.1. (Volatilität) [vgl. Patton (2006), S.4]

Sei $r_t \in \mathcal{O}$ eine \mathcal{F}_t -messbare Zufallsvariable und \mathcal{F}_{t-1} mit $\sigma(r_{t-j}, j \geq 1) \subseteq \mathcal{F}_{t-1}$ die Informationen bis zum Zeitpunkt $t - 1$ mit $t \in \mathbb{N}$. Dann heißt σ_t mit $\sigma_t^2 = \text{Var}(r_t | \mathcal{F}_{t-1})$ Volatilität.

Im weiteren Verlauf dieser Arbeit wird sich herausstellen, dass es sinnvoll ist, zentrierte Renditen anzunehmen. Dann gilt $\mathbb{E}[r_t | \mathcal{F}_{t-1}] = 0$ und wir erhalten die bedingte Varianz der Renditen über das zweite Moment: $\sigma_t^2 = \mathbb{E}[r_t^2 | \mathcal{F}_{t-1}]$. Das Funktional von Interesse verlagert sich somit von der bedingten Varianz auf den bedingte Erwartungswert.

In den weiteren Ausführungen werden wir uns nur auf σ_t^2 , die bedingte Varianz der Renditen, beziehen, da diese nach Definition (2.1) in direktem Zusammenhang zur Volatilität σ_t steht.

Um nun die bedingte Varianz σ_t^2 vorherzusagen, folgen wir wieder der Notation von Patton (2006) und verwenden $h_t \in \mathcal{H}$ für eine Prognose der bedingten Varianz σ_t^2 zum Zeitpunkt $t \in \mathbb{N}$ und $h_{i,t} \in \mathcal{H}$ mit $i \in \mathbb{N}$ für mehrere Prognosen. Sei dabei $\mathcal{H} \subseteq \mathbb{R}_+$, wobei wir \mathbb{R}_+ als Notation für die nicht-negativen reellen Zahlen verwenden.

Wenn nun eine oder mehrere Prognosen vorliegen, so können Verlustfunktionen genutzt werden, um die Abweichung dieser Prognosen von der tatsächlich eingetretenen Beobachtung zu prüfen. Patton (2006) definiert Verlustfunktionen folgendermaßen:

Definition 2.2. (Verlustfunktion) [vgl. Patton (2006), S.4]

Eine messbare Funktion $L : \mathcal{H} \times \mathcal{O} \rightarrow \mathbb{R}_+$ heißt Verlustfunktion, wobei $L(h, y)$ den Verlust darstellt, falls eine Prognose $h \in \mathcal{H}$ verwendet wird und die Beobachtung $y \in \mathcal{O}$ eintritt.

Über die Abstandsbeurteilung hinaus, werden Verlustfunktionen verwendet, um den erwarteten Verlust $\mathbb{E}[L(h_t, \sigma_t^2)]$ einer Prognose h_t bei tatsächlich eintretender bedingter Varianz der Renditen σ_t^2 zum Zeitpunkt $t \in \mathbb{N}$ zu bestimmen. Dieser erwartete Verlust kommt in vielen Tests zum Vergleich konkurrierender Prognosen eines Risikomaßes zum Einsatz. In Kapitel 6 werden wir einen Diebold-Mariano Test konstruieren, der den erwarteten Verlust verwendet, um die Bessere zweier Volatilitätsprognosen zu identifizieren. Hierbei ist die Wahl der Verlustfunktion wesentlich, da sich die erwarteten Verluste einer Prognose h_t voneinander unterscheiden, je nachdem, welche Verlustfunktion verwendet wird. Einige häufig verwendete Verlustfunktionen sind hier einmal aufgeführt:

quadratischer Verlust	$L(x, y) = (x - y)^2$
absoluter Verlust	$L(x, y) = x - y $
absoluter prozentualer Verlust	$L(x, y) = \left \frac{x-y}{y} \right $
relativer Verlust	$L(x, y) = \left \frac{x-y}{x} \right $
QLIKE-Verlustfunktion	$L(x, y) = \log(x) + \frac{y}{x}$

Tabelle 2.1: Auswahl häufig verwendeter Verlustfunktionen

In unseren Ausführungen richten wir uns nach Gneiting (2009), der Verlustfunktionen als negativ orientiert definiert. Je kleiner also der Wert der Verlustfunktion, umso geringer ist die Abweichung zur Beobachtung und umso besser ist die eingesetzte Prognose. Verlustfunktionen können aber nicht nur dazu verwendet werden, Prognosen zu evaluieren, sondern auch, um eine Prognose zu erstellen. Auf dem erwarteten Verlust aufbauend, definiert Gneiting (2009) die optimale Prognose nach Bayes folgendermaßen:

Definition 2.3. (*optimale Prognose*) [vgl. Gneiting (2009), Kapitel 2.1]

Die optimale Prognose unter der Verteilung $P \in \mathcal{P}$ für eine zukünftige Beobachtung $Y \in \mathcal{O}$ ist die Bayes-Regel:

$$h_t^* := \arg \min_{h \in \mathcal{H}} \mathbb{E}_P[L(h, Y)]. \tag{2.1}$$

Im Allgemeinen ist die Vorgehensweise in der Volatilitätsprognose die Folgende: Eine Verlustfunktion L wird vorgegeben und damit lässt sich dann über Bayes (2.1) die optimale Prognose h_t^* bestimmen. Die Vorgabe der Verlustfunktion ist dabei essentiell, da unterschiedliche Verlustfunktionen in unterschiedlichen Ergebnissen des erwarteten Verlustes und damit der optimalen Prognose nach Bayes resultieren. Soll allerdings lediglich das Funktional, in diesem Fall die bedingte Varianz, vorgegeben werden, so entsteht daraus die Notwendigkeit der Definition sinnvoller Eigenschaften, anhand derer Verlustfunktionen klassifiziert werden können. Die Menge der Verlustfunktionen sollte eingegrenzt werden auf solche, die in der Evaluierung eine optimale Prognose erkennen. Die Frage nach geeigneten Verlustfunktionen wird von verschiedenen Autoren wie etwa Gneiting (2009) oder Patton (2006) diskutiert. Ziel dieser Arbeit ist es unter anderem, geeignete Verlustfunktionen für die Volatilität zu identifizieren. Dafür werden wir die optimale Prognose - die tatsächlich eintretende Volatilität - als bekannt voraussetzen und dann die Verlustfunktionen identifizieren, die diese auch als optimal erkennen.

3 Strikt konsistente Verlustfunktionen

Dieses Kapitel orientiert sich hauptsächlich an der Arbeit von Gneiting (2009), der zwar nicht konkret auf die Volatilität eingeht, sondern eher allgemein verschiedene Funktionale betrachtet, aus dessen Ausführungen aber die Volatilität als Spezialfall sehr gut abgeleitet werden kann.

3.1 Konsistenz und Elizitierbarkeit

Ein häufig verwendeter Ansatz bei der Erstellung von Prognosen ist der, eine Verlustfunktion ex-ante vorzugeben und anschließend die dazu optimale Prognose über die Bayes Regel (2.1) zu ermitteln. Die Alternative, auf welche hier eingegangen wird, ist die Vorgabe eines bestimmten Funktionals, während die Wahl der Verlustfunktion offen bleibt. Wie bereits zuvor erwähnt, kann das zu unterschiedlichen Ergebnissen für die optimale Prognose führen, je nachdem, welche Verlustfunktion für die Prognose verwendet wird. Konsistente Verlustfunktionen garantieren in diesem Zusammenhang, dass wir den tatsächlichen Wert des Funktionals als optimale Prognose erhalten. In unserem Fall geben wir dafür die Varianz als Funktional vor, um die Volatilität nach Definition (2.1) zu erhalten. Das Ziel ist es, dazu konsistente Verlustfunktionen zu identifizieren, um diese dann zur Evaluierung von Volatilitätsprognosen verwenden zu können. Gneiting (2009) definiert Konsistenz und strikte Konsistenz folgendermaßen:

Definition 3.1. (Konsistenz)[vgl. Gneiting (2009), Definition 2.1]

Sei $T : \mathcal{P} \rightarrow \mathcal{H}$, $P \mapsto T(P)$ ein Funktional mit $\mathcal{H} \subseteq \mathbb{R}$. Eine Verlustfunktion L ist konsistent für ein Funktional T bzgl. einer Klasse \mathcal{P} von Wahrscheinlichkeits-Verteilungen, falls

$$\mathbb{E}_P[L(T(P), Y)] \leq \mathbb{E}_P[L(h, Y)] \quad (3.1)$$

für alle $P \in \mathcal{P}$ und alle $h \in \mathcal{H}$ gilt.

Definition 3.2. (strikte Konsistenz)[vgl. Gneiting (2009), Definition 2.1]

Eine Verlustfunktion L ist strikt konsistent für ein Funktional T bzgl. einer Klasse \mathcal{P} von Wahrscheinlichkeits-Verteilungen, falls sie konsistent ist und Gleichheit in (3.1) impliziert, dass $h = T(P)$.

Die Zufallsvariable $Y \in \mathcal{O}$ steht hier und im weiteren Verlauf der Arbeit für die Beobachtungen, die tatsächlich eintreten. Im konkreten Fall dieser Arbeit ist das vorgegebene Funktional T die Varianz der Renditen σ_t^2 , h eine beliebige Prognose für σ_t^2 und $T(P)$ bezeichnet die tatsächliche Varianz der Renditen σ_t^2 . Wird nun h mittels des erwarteten

Verlustes mit dem tatsächlichen Wert des Funktional $T(P)$ verglichen, so führt eine konsistente Verlustfunktion L nach (3.1) dazu, dass der erwartete Verlust des tatsächlichen Funktional, verglichen mit jeder anderen Prognose, am geringsten ausfällt. Wie zuvor bereits erwähnt, gehen wir von negativ orientierten Verlustfunktionen aus. Je kleiner der erwartete Verlust ist, umso besser ist die Prognose. Verwenden wir also eine konsistente Verlustfunktion, so erkennen wir die tatsächliche Varianz der Renditen σ_t^2 als optimal an. Für strikt konsistente Verlustfunktionen gilt sogar, dass das Minimum im erwarteten Verlust eindeutig ist. Das bedeutet, dass bei der Verwendung einer strikt konsistenten Verlustfunktion die tatsächliche Varianz der Renditen die eindeutige Lösung des Optimierungsproblems (2.1) ist. Bevor wir allerdings mit strikt konsistenten Verlustfunktionen rechnen können, müssen wir zuvor sicherstellen, dass ebensolche zu unserem Funktional existieren, was uns direkt zur Elizitierbarkeit führt, die Gneiting (2009) wie folgt definiert:

Definition 3.3. (*Elizitierbarkeit*) [vgl. Gneiting (2009), Definition 2.5]

Ein Funktional T ist elizitierbar bzgl. einer Klasse \mathcal{P} , falls eine strikt konsistente Verlustfunktion zu T bzgl. \mathcal{P} existiert.

Wenn wir also ein elizitierbares Funktional verwenden, stellen wir sicher, dass es strikt konsistente Verlustfunktionen gibt, die eine optimale Prognose dieses Funktional erkennen. Damit reicht es für eine Prognose aus, dieses Funktional ohne eine festgelegte Verlustfunktion vorzugeben. Der Prognostiker kann dann unter den konsistenten Verlustfunktionen wählen, um eine Prognose zu erstellen.

Das besondere Interesse, mit dem sich Gneiting (2009) den konsistenten Verlustfunktionen zuwendet, wird durch den folgenden Satz explizit deutlich, dessen Inhalt bereits angedeutet wurde. Er zeigt die direkte Verbindung auf, die zwischen dem Finden einer optimalen Prognose und der Bewertung von Prognosen mittels konsistenter Verlustfunktionen besteht.

Satz 3.4. [vgl. Gneiting (2009), Theorem 2.2]

Eine Verlustfunktion L ist konsistent für T bzgl. \mathcal{P} genau dann, wenn für alle $P \in \mathcal{P}$ gilt, dass $h = T(P)$ eine optimale Punktprognose unter L ist.

Beweis. Sei L eine Verlustfunktion für das Funktional T . Mit der Definition der optimalen Prognose (2.1) gilt:

$$\begin{aligned} T(P) = \arg \min_{h \in \mathcal{H}} \mathbb{E}_P[L(h, Y)] &\Leftrightarrow \mathbb{E}_P[L(T(P), Y)] \leq \mathbb{E}_P[L(h, Y)] \quad \forall h \in \mathcal{H} \\ &\Leftrightarrow L \text{ ist konsistent für } T \text{ bzgl. } \mathcal{P}. \end{aligned}$$

□

Die Klasse der Verlustfunktionen, die konsistent zu einem Funktional T sind, ist also identisch zu der Klasse der Verlustfunktionen, für die $T(P)$ die optimale Prognose ist. Dieser Satz begründet, weswegen es wünschenswert ist, konsistente Verlustfunktionen zu verwenden.

3.2 Nicht-Elizitierbarkeit der Varianz

Bevor wir damit beginnen, strikt konsistente Verlustfunktionen herzuleiten, sollte allerdings zuerst die Frage geklärt werden, ob dies überhaupt möglich ist, ob also die Varianz elizitierbar ist. Brehmer (2017) führt mit dem folgenden Satz eine notwendige Bedingung für die Elizitierbarkeit eines Funktional ein, die er anschließend nutzt, um zu zeigen, dass die Varianz nicht elizitierbar ist.

Satz 3.5. [vgl. Brehmer (2017), Prop. 1.16]

Sei \mathcal{P} eine konvexe Familie von Verteilungen und T ein Funktional. Dann gilt: Ist T elizitierbar, so hat T konvexe Höhenlinien. Das bedeutet: Sind $P_1, P_2 \in \mathcal{P}$ mit $T(P_1) = T(P_2)$, so gilt $T(\lambda P_1 + (1 - \lambda)P_2) = T(P_1)$ für alle $\lambda \in (0, 1)$.

Beweis. [vgl. Holzmann (2019), Beweis Satz 18.8]

Sei $a = T(P_1) =: r$.

Setze $P_\lambda := \lambda P_1 + (1 - \lambda)P_2$ für $\lambda \in (0, 1)$ beliebig. Da T elizitierbar ist, existiert eine für T strikt konsistente Verlustfunktion L . Damit gilt dann:

$$\begin{aligned} \mathbb{E}_{P_\lambda}[L(a, Y)] &= \int_{\mathcal{O}} L(a, y) dP_\lambda(y) \\ &= \lambda \int_{\mathcal{O}} L(a, y) dP_1(y) + (1 - \lambda) \int_{\mathcal{O}} L(a, y) dP_2(y) \\ &> \lambda \int_{\mathcal{O}} L(r, y) dP_1(y) + (1 - \lambda) \int_{\mathcal{O}} L(r, y) dP_2(y) \\ &= \mathbb{E}_{P_\lambda}[L(r, Y)]. \end{aligned}$$

Für die Ungleichung wurde die strikte Konsistenz von L genutzt. Gleichheit gilt genau dann, wenn $T(P_\lambda) = a = r = T(P_1)$ und ansonsten gilt die strikte Ungleichung. Da λ beliebig gewählt war, folgt die Behauptung. \square

Damit haben wir eine notwendige Bedingung für Elizitierbarkeit bewiesen und können diese nun auf die Varianz anwenden.

Folgerung 3.6. [vgl. Brehmer (2017), Beispiel 1.18]

Da die Varianz keine konvexen Höhenlinien hat, ist sie nicht elizitierbar.

Um diese Folgerung zu beweisen, werden wir ein Gegenbeispiel verwenden, für das die oben hergeleitete notwendige Bedingung aus Satz (3.5) nicht gilt.

Beweis. [vgl. Holzmann (2019)]

Seien $X_1 \sim \mathcal{N}(0, 1)$ mit Verteilung P_1 und $X_2 \sim \mathcal{N}(1, 1)$ mit Verteilung P_2 Zufallsvariablen. Sei weiter $X_\lambda \stackrel{d}{=} P_\lambda$ wie in Satz (3.5) mit $\lambda = \frac{1}{2}$ eine Zufallsvariable. Dann ist $P_\lambda = \frac{P_1 + P_2}{2}$ und für den Erwartungswert gilt: $\mathbb{E}_{P_\lambda}[X_\lambda] = \frac{1}{2}(\mathbb{E}_{P_1}[X_1] + \mathbb{E}_{P_2}[X_2]) = \frac{1}{2}$.

Für die Varianz gilt dann: $Var_{P_1}(X_1) = Var_{P_2}(X_2) = 1$, aber

$$\begin{aligned} Var_{P_\lambda}(X_\lambda) &= \mathbb{E}_{P_\lambda}[X_\lambda^2] - \mathbb{E}_{P_\lambda}[X_\lambda]^2 \\ &= \frac{1}{2}(\mathbb{E}_{P_1}[X_1^2] + \mathbb{E}_{P_2}[X_2^2]) - \frac{1}{4} \\ &= \frac{1}{2}(Var_{P_1}(X_1) + \mathbb{E}_{P_1}[X_1]^2 + Var_{P_2}(X_2) + \mathbb{E}_{P_2}[X_2]^2) - \frac{1}{4} \\ &= \frac{5}{4} \neq 1. \end{aligned}$$

Dies widerspricht der notwendigen Bedingung aus Satz (3.5) und damit ist die Varianz nicht elizitierbar. □

Brehmer (2017) zeigt in einem allgemeineren Beweis [vgl. Brehmer (2017) Beispiel 1.18], dass diese notwendige Bedingung nicht erfüllt ist, sobald \mathcal{P} Verteilungen P_1 und P_2 enthält, die gleiche Varianz, aber unterschiedliche Erwartungswerte haben. Elizitierbarkeit ist hier also abhängig von der Wahl der Klasse \mathcal{P} . Wir werden als Nächstes zeigen, dass der Erwartungswert bzgl. der Klasse \mathcal{P} der Verteilungsfunktionen mit endlichem 1. Moment elizitierbar ist. Das können wir anschließend für die Varianz nutzen, indem wir \mathcal{P} auf die Klasse aller zentrierten Verteilungen mit endlichem 2. Moment einschränken. Dann entspricht die Varianz dem 2. Moment und ist elizitierbar. [vgl. Brehmer (2017), Beispiel 1.20]

3.3 Elizitierbarkeit des Erwartungswertes

Nachdem wir nun festgestellt haben, dass die Varianz nicht elizitierbar ist, wenden wir uns einem anderen Funktional zu, dem Erwartungswert. Wir werden zeigen, dass der Erwartungswert elizitierbar ist, indem wir Beispiele für konsistente und strikte konsistente Verlustfunktionen finden.

Dafür führt Gneiting (2009) zunächst einige Annahmen ein, auf denen aufbauend er den darauf folgenden Satz formuliert:

S_0 : $L(x, y) \geq 0$ mit Gleichheit, falls $x = y$

S_1 : $L(x, y)$ ist stetig in x

S_2 : Die partielle Ableitung $\frac{\partial L(x, y)}{\partial x}$ existiert und ist stetig in x , falls $x \neq y$

Satz 3.7. [vgl. Gneiting (2009), Theorem 3.1]

Sei \mathcal{P} die Klasse der Verteilungen mit kompaktem Träger auf einem Intervall $I \subseteq \mathbb{R}$ mit endlichem 1. Moment und $\Phi : I \rightarrow \mathbb{R}$ eine Funktion mit 1. Ableitung Φ' . Sei L eine Verlustfunktion, die S_0 , S_1 und S_2 auf $\mathcal{D} = I \times I$ erfüllt und die Form

$$L(x, y) = \Phi(y) - \Phi(x) - \Phi'(x)(y - x) \tag{3.2}$$

hat. Dann gilt:

- Ist Φ konvex, so ist L konsistent für den Erwartungswert bzgl. \mathcal{P} .
- Ist Φ strikt konvex, so ist L strikt konsistent für den Erwartungswert bzgl. \mathcal{P} eingeschränkt auf die Verteilungen, für die $\mathbb{E}_P[Y] < \infty$ und $\mathbb{E}_P[\Phi(Y)] < \infty$ existieren.

Beweis. (i) [vgl. Brehmer (2017), Beispiel 1.8, S.3f.]

Seien L eine Verlustfunktion der Form (3.2) mit Φ konvex, $t := T(P) = \mathbb{E}_P[Y]$ das Erwartungswert-Funktional und $x \in \mathcal{D}$.

Durch Einsetzen erhalten wir:

$$\begin{aligned}
 & \mathbb{E}_P[L(x, Y)] - \mathbb{E}_P[L(t, Y)] \\
 &= \mathbb{E}_P[\Phi(Y) - \Phi(x) - \Phi'(x)(Y - x)] - \mathbb{E}_P[\Phi(Y) - \Phi(t) - \Phi'(t)(Y - t)] \\
 &= \Phi(t) - \Phi(x) - \Phi'(x)\mathbb{E}_P[Y - x] + \Phi'(t)\mathbb{E}_P[Y - t] && \text{(Linearität)} \\
 &= \Phi(t) - \Phi(x) - \Phi'(x)\mathbb{E}_P[Y - x] && (t = \mathbb{E}_P[Y]) \\
 &= \Phi(t) - \Phi(x) - \Phi'(x)(t - x) && (t = \mathbb{E}_P[Y]) \\
 &\geq 0.
 \end{aligned}$$

Die Ungleichheit gilt aufgrund der Konvexität von Φ . Damit folgt unter der Bedingung des endlichen 1. Moments, dass $\mathbb{E}_P[L(x, Y)] \geq \mathbb{E}_P[L(t, Y)] \forall x \in \mathcal{D}$ und damit ist gezeigt, dass L eine konsistente Verlustfunktion für den Erwartungswert ist.

(ii) [vgl. Holzmann (2019), Beweis zu Satz 18.5]

Sei nun Φ strikt konvex. Wir setzen wieder ein und erhalten für $x \neq t := T(P) = \mathbb{E}_P[Y]$ wie in Teil (i):

$$\mathbb{E}_P[L(x, Y)] - \mathbb{E}_P[L(t, Y)] = \Phi(t) - \Phi(x) - \Phi'(x)(t - x) > 0.$$

Die strikte Ungleichung folgt aus der strikten Konvexität von Φ und Gleichheit gilt genau dann, wenn $x = t = T(P)$, sodass nun die strikte Konsistenz der Verlustfunktion L gefolgert werden kann. □

Damit haben wir die Elizitierbarkeit des Erwartungswertes bzgl. der Klasse \mathcal{P} der Verteilungen mit kompaktem Träger auf $I \subseteq \mathbb{R}$ mit $\mathbb{E}_P[Y] < \infty$ und $\mathbb{E}_P[\Phi(Y)] < \infty$ gezeigt. Das können wir nun für die Volatilität verwenden, denn nehmen wir Zentriertheit der Renditen an, so gilt: $\sigma_t^2 = \mathbb{E}[r_t^2 | \mathcal{F}_{t-1}]$. Damit ist es also möglich, unter gewissen Einschränkungen strikt konsistente Verlustfunktionen für σ_t^2 zu finden, obwohl die Varianz nicht elizitierbar ist.

3.4 Osbands Prinzip

Nachdem wir nun gezeigt haben, dass es für den Erwartungswert konsistente und strikt konsistente Verlustfunktionen gibt und dieser daher elizitierbar ist, stellt sich die Frage,

wie es möglich ist, solche Verlustfunktionen zu finden. Wir werden in diesem Teilkapitel Osbands Prinzip als eine mögliche Herangehensweise an dieses Problem betrachten. Dabei werden wir nur heuristisch die wesentlichen Punkte darstellen und verweisen für Beweise und eine ausführliche Herleitung auf Gneiting (2009) oder Steinwart et al. (2014), die sich ausführlich mit der Thematik befassen und sich dabei im Wesentlichen auf Osband (1985) beziehen.

Grundlegend für Osbands Prinzip sind Identifikationsfunktionen, welche Gneiting (2009) folgendermaßen definiert.

Definition 3.8. (*Identifikationsfunktion*) [vgl. Gneiting (2009) S.17]

Eine Identifikationsfunktion ist eine Funktion $V : \mathcal{H} \times \mathcal{O} \rightarrow \mathbb{R}$ für die gilt:

$$\mathbb{E}_P[V(x, Y)] = 0 \Leftrightarrow x \in T(P) \text{ und } V(x, y) = 0 \Leftrightarrow x = y.$$

Eine Identifikationsfunktion für den Erwartungswert ist beispielsweise $V(x, y) = x - y$. Diese werden wir später verwenden, um konsistente Verlustfunktionen zum Erwartungswert zu finden. In diesem Abschnitt wollen wir erst einmal eine allgemeine Herangehensweise zur Identifikation konsistenter Verlustfunktionen darstellen. Dazu betrachten wir wieder den erwarteten Verlust $\mathbb{E}_P[L(h, Y)]$ für eine Prognose h , eine konsistente Verlustfunktion L und eine Zufallsvariable Y . Wir nehmen an, dass $Y \sim \text{Ber}(p)$ mit $\mathbb{P}(Y = a) = p$ und $\mathbb{P}(Y = b) = 1 - p$. Dann gilt für den erwarteten Verlust, dass $\mathbb{E}_P[L(h, Y)] = pL(h, a) + (1 - p)L(h, b)$ und dessen Ableitung nach h ergibt sich als

$$\frac{\partial \mathbb{E}_P[L(h, Y)]}{\partial h} = pL_{(1)}(h, a) + (1 - p)L_{(1)}(h, b). \quad (3.3)$$

Dabei bezeichnet $L_{(1)}(h, a)$ die partielle Ableitung von L nach h . Da L konsistent ist, hat der erwartete Verlust sein Minimum bei $h = x \in T(P)$ und wir erhalten 0 für die Ableitung in (3.3). Für den Erwartungswert der Identifikationsfunktion ergibt sich eine ähnliche Gleichung:

$$\mathbb{E}_P[V(h, Y)] = pV(h, a) + (1 - p)V(h, b). \quad (3.4)$$

Auch hier erhalten wir per Definition 0, falls wir $h = x \in T(P)$ setzen. Durch Auflösen nach p und Gleichsetzen der beide Gleichungen (3.3) und (3.4) erhalten wir:

$$\frac{L_{(1)}(x, a)}{V(x, a)} = \frac{L_{(1)}(x, b)}{V(x, b)}.$$

Da a und b beliebig gewählt waren, ist die Funktion $\frac{L_{(1)}(x, y)}{V(x, y)}$ unabhängig von y und wir können festhalten, dass

$$L_{(1)}(x, y) = h(t)V(x, y). \quad (3.5)$$

Ist also eine Identifikationsfunktion zu einem Funktional bekannt, so kann durch Integrieren eine konsistente Verlustfunktion gefunden werden.

3.5 Konsistente Verlustfunktionen zum Erwartungswert

Das im vorherigen Teilkapitel hergeleitete Osbands Prinzip können wir nun verwenden, um eine allgemeine Form konsistenter und strikt konsistenter Verlustfunktionen für den Erwartungswert zu bestimmen. Tatsächlich ist die Klasse der Verlustfunktionen der Form (3.2), mit der wir uns bereits zuvor beschäftigt haben, schon eine vollständige Klasse aller zum Erwartungswert konsistenten Verlustfunktionen.

Satz 3.9. [vgl. Gneiting (2009), Theorem 3.1, S.18]

Sei \mathcal{P} die Klasse der Verteilungen mit kompaktem Träger auf $I \subseteq \mathbb{R}$ und endlichem 1. Moment. Sei L eine Verlustfunktion, die S_0 , S_1 und S_2 auf $\mathcal{D} = I \times I$ erfüllt. Ist L konsistent für den Erwartungswert bzgl. \mathcal{P} , so hat L die Form (3.2) mit Φ konvex.

Beweis. Für den Beweis verwenden wir Osbands Prinzip mit der Identifikationsfunktion $V(x, y) = x - y$. Nach (3.5) gilt: $\frac{\partial L(x, y)}{\partial x} = h(x)(x - y)$ für eine Funktion h . Mittels partieller Integration ergibt sich:

$$\begin{aligned} L(x, y) &= \int h(x)(x - y)dx \\ &= [H(x)(x - y)] - \int H(x)dx \\ &= H(x)(x - y) - \tilde{H}(x) + \tilde{H}(y). \end{aligned}$$

Wenn wir jetzt $\Phi = \tilde{H}$ setzen, erhalten wir für L die Form (3.2), und da per Definition gilt, dass die Verlustfunktion $L(x, y) \geq 0$ ist, muss Φ konvex sein. □

Die hier hergeleiteten Verlustfunktionen heißen auch Bregman-Funktionen. Für sie gilt generell, dass die optimale Prognose unter einer solchen Bregman-Funktion der Erwartungswert der prognostizierten Verteilung ist [vgl. Gneiting (2009), Kapitel 3.1].

3.6 Beispiele

Da wir nun eine allgemeine Form konsistenter Verlustfunktionen für den Erwartungswert hergeleitet haben, können wir diese verwenden, um beispielhaft zwei mögliche Verlustfunktionen zu identifizieren.

(i) Setzen wir $\Phi(x) = \frac{1}{2}x^2$, so erhalten wir

$$L(x, y) = \frac{1}{2}y^2 - \frac{1}{2}x^2 - x(y - x) = \frac{1}{2}(x - y)^2$$

als Verlustfunktion. Da $\Phi(x) = \frac{1}{2}x^2$ strikt konvex ist, ist L eine strikt konsistente Verlustfunktion. Bis auf Konstanten erkennen wir hier den quadratischen Verlust, und da Konstanten eine Verlustfunktion nicht weiter beeinflussen, können wir diesen als strikt konsistent identifizieren.

(ii) Nutzen wir die strikt konvexe Funktion $\Phi(x) = -\log(x)$, so erhalten wir:

$$L(x, y) = -\log(y) + \log(x) + \frac{1}{x}(y - x) = \log(x) + \frac{y}{x} - (\log(y) + 1).$$

Auch hier liegt bis auf Konstanten eine bereits bekannte Verlustfunktion vor. Die QLIKE-Verlustfunktion ist also ebenfalls strikt konsistent für den Erwartungswert.

Diese beiden Verlustfunktionen werden wir in einem späteren Kapitel zur Evaluierung von Volatilitätsprognosen verwenden und überprüfen, ob sich trotz der Tatsache, dass beide strikt konsistent sind, Unterschiede ergeben.

3.7 Zwischenfazit

Wie anfangs bereits erwähnt, gibt es zwei mögliche Vorgehensweisen bei der Prognose der Volatilität: Entweder die Verlustfunktion wird im Vorhinein festgelegt und mittels Bayes kann dann die optimale Prognose ermittelt werden, oder es wird ein elizitierbares Funktional festgelegt und Prognosen können über dazu konsistente Verlustfunktionen erstellt und bewertet werden.

Wir haben festgestellt, dass die Varianz nicht elizitierbar ist und haben daher die Renditen als zentriert angenommen, um zum Erwartungswert als Ziel-Funktional überzugehen. Für den Erwartungswert haben wir anschließend die Bregman-Funktionen der Form (3.2) als konsistente Verlustfunktionen identifiziert. Wenn wir also unter der Annahme zentrierter Renditen die Varianz als elizitierbares Funktional vorgeben, können wir deren Prognosen mittels einer solchen konsistenten Verlustfunktion bewerten. Wir haben gesehen, dass dafür beispielsweise der quadratische Verlust oder die QLIKE-Verlustfunktion in Frage kommen. Im folgenden Kapitel werden wir nun eine weitere wichtige Eigenschaft der Bregman-Verlustfunktionen zeigen: die Robustheit, die speziell bei der Prognose der bedingten Varianz von Bedeutung ist.

4 Robuste Verlustfunktionen

Das folgende Kapitel orientiert sich maßgeblich an der Arbeit von Patton (2006). Patton wendet sich einem weiteren Konzept zu: Den robusten Verlustfunktionen. Auch hierbei liegt der Fokus wieder darauf, Verlustfunktionen zu finden, die die tatsächliche Varianz der Renditen als optimale Prognose erkennen, jedoch wird darüber hinaus die Latenz-Problematik, die insbesondere auch die Volatilität betrifft, mit einbezogen.

4.1 Robustheit

Eine große Herausforderung bei der Erstellung und der Evaluierung von Prognosen für die Volatilität σ_t ist deren Latenz. Die Volatilität ist ein latentes Risikomaß, was bedeutet, dass sie selbst ex-post nicht beobachtbar ist. Sollen also zwei Prognosen $h_{1,t}$ und $h_{2,t}$ über den erwarteten Verlust $\mathbb{E}[L(h_{i,t}, \sigma_t^2)]$ verglichen werden, so kann dafür nicht die tatsächliche bedingte Varianz der Renditen σ_t^2 verwendet werden, da sie nicht bekannt ist. Für die Prognose wird daher ein bedingt erwartungstreuer Stellvertreter $\hat{\sigma}_t^2$ anstelle der tatsächlichen bedingten Varianz eingesetzt. Im Fokus steht dann $\mathbb{E}[L(h_{i,t}, \hat{\sigma}_t^2)]$. Häufig verwendete Stellvertreter sind beispielsweise:

- die quadratischen Renditen r_t^2 ,
- die intra-daily range RG_t ,
- die realisierten Volatilitäten RV_t .

In einem späteren Kapitel werden wir uns besonders mit den quadratischen Renditen und den realisierten Volatilitäten beschäftigen und herausarbeiten, welcher der beiden Stellvertreter sich besser eignet, um Vorhersagen der Volatilität zu vergleichen. In diesem Kapitel werden wir uns allerdings zunächst allgemein auf einen beliebigen Stellvertreter $\hat{\sigma}_t^2$ beziehen, um das Konzept der Robustheit zu verdeutlichen.

Das Hauptproblem in der Verwendung solcher Stellvertreter sind Verzerrungen im erwarteten Verlust. Soll also auf Basis von $\mathbb{E}[L(h_{i,t}, \hat{\sigma}_t^2)]$ eine optimale Prognose nach Bayes (2.1) erstellt werden, so kann es sein, dass sich unterschiedliche Ergebnisse einstellen, je nachdem, ob $\hat{\sigma}_t^2$ oder σ_t^2 verwendet werden. Genauso kann es bei der Evaluierung von Prognosen passieren, dass eine Prognose als optimal angenommen wird, obwohl sie einer anderen Prognose unterlegen ist, wenn Stellvertreter verwendet werden. Die Bedingung der bedingten Erwartungstreue des Stellvertreters reicht dann nicht aus. Benötigt wird also eine Einschränkung der Verlustfunktionen auf solche, die trotz der

Verwendung eines Stellvertreters zum selben Ergebnis für die optimale Prognose gelangen. Daher wird im Folgenden das Konzept der robusten Verlustfunktionen vorgestellt, die Patton (2006) folgendermaßen definiert:

Definition 4.1. (Robustheit)[vgl. Patton (2006), Definition 1, S.5]

Eine Verlustfunktion L ist robust, falls das Ranking zweier Volatilitätsprognosen $h_{1,t}$ und $h_{2,t}$ im erwarteten Verlust gleich ist, unabhängig davon, ob die wahre bedingte Varianz σ_t^2 oder ein erwartungstreuer Stellvertreter $\hat{\sigma}_t^2$ genutzt wird. Also wenn gilt:

$$\mathbb{E}[L(h_{1,t}, \sigma_t^2)] \geq \mathbb{E}[L(h_{2,t}, \sigma_t^2)] \Leftrightarrow \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2)] \geq \mathbb{E}[L(h_{2,t}, \hat{\sigma}_t^2)]. \quad (4.1)$$

Anzumerken ist hier, dass ein gleichbleibendes Ranking nicht bedeutet, dass auch der erwartete Verlust identisch bleibt. Nutzt man einen Stellvertreter, so wird der erwartete Verlust größer ausfallen als unter Verwendung der tatsächlichen bedingten Varianz [vgl. Patton (2006), Kapitel 2]. Dies ist allerdings irrelevant, wenn es lediglich darum geht, die Bessere zweier konkurrierender Prognosen $h_{1,t}$ und $h_{2,t}$ zu identifizieren, was wiederum durch den Erhalt des Rankings gewährleistet ist.

4.2 Robustheit der Bregman-Funktionen

Wir betrachten noch einmal die in Kapitel 3 hergeleiteten Bregman-Funktionen der Form (3.2) und werden zeigen, dass diese bereits robust sind. Dabei ist es irrelevant, ob Φ konvex oder strikt konvex ist. Der folgende Satz gilt sowohl für die konsistenten als auch für die strikt konsistenten Bregman-Funktionen. Eine gekürzte Version der Proposition 2 aus Patton (2006) lautet:

Satz 4.2. [vgl. Patton (2006), Proposition 2]

Eine Verlustfunktion der Form (3.2) ist robust.

Beweis. Angenommen, die Verlustfunktion L habe die Form (3.2) mit Φ konvex. Seien σ_t^2 die tatsächliche Varianz der Renditen, $\hat{\sigma}_t^2$ ein Stellvertreter für σ_t^2 und $h_{1,t}$ und $h_{2,t}$ zwei Prognosen, die verglichen werden sollen.

Mit der Linearität des Erwartungswertes und (3.2) für L eingesetzt gilt:

$$\begin{aligned} \mathbb{E}[L(h_{1,t}, \sigma_t^2)|\mathcal{F}_{t-1}] - \mathbb{E}[L(h_{2,t}, \sigma_t^2)|\mathcal{F}_{t-1}] &= \mathbb{E}[\Phi(h_{1,t}) - \Phi(h_{2,t})|\mathcal{F}_{t-1}] \\ &\quad - \mathbb{E}[\Phi'(h_{1,t})h_{1,t} + \Phi'(h_{2,t})h_{2,t}|\mathcal{F}_{t-1}] \\ &\quad + \mathbb{E}[\sigma_t^2(\Phi'(h_{1,t}) - \Phi'(h_{2,t}))|\mathcal{F}_{t-1}]. \end{aligned}$$

Da nur der letzte Summand von σ_t^2 abhängt, betrachten wir im Weiteren nur diesen letzten Teil. Die beiden ersten Summanden sind identisch, unabhängig davon, ob σ_t^2 oder einen Stellvertreter $\hat{\sigma}_t^2$ eingesetzt wird. Es gilt:

$$\begin{aligned} \mathbb{E}[\sigma_t^2(\Phi'(h_{1,t}) - \Phi'(h_{2,t}))|\mathcal{F}_{t-1}] &= (\Phi'(h_{1,t}) - \Phi'(h_{2,t}))\mathbb{E}[\sigma_t^2|\mathcal{F}_{t-1}] && \text{(Messbarkeit)} \\ &= (\Phi'(h_{1,t}) - \Phi'(h_{2,t}))\mathbb{E}[\hat{\sigma}_t^2|\mathcal{F}_{t-1}] && \text{(Erw'treue)} \\ &= \mathbb{E}[\hat{\sigma}_t^2(\Phi'(h_{1,t}) - \Phi'(h_{2,t}))|\mathcal{F}_{t-1}]. && \text{(Messbarkeit)} \end{aligned}$$

Damit folgt, dass

$$\mathbb{E}[L(h_{1,t}, \sigma_t^2) | \mathcal{F}_{t-1}] - \mathbb{E}[L(h_{2,t}, \sigma_t^2) | \mathcal{F}_{t-1}] = \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2) | \mathcal{F}_{t-1}] - \mathbb{E}[L(h_{2,t}, \hat{\sigma}_t^2) | \mathcal{F}_{t-1}].$$

Mit Erwartungswertbildung und dem Satz der totalen Erwartung ergibt sich dann, dass

$$\mathbb{E}[L(h_{1,t}, \sigma_t^2)] - \mathbb{E}[L(h_{2,t}, \sigma_t^2)] = \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2)] - \mathbb{E}[L(h_{2,t}, \hat{\sigma}_t^2)],$$

womit Robustheit per Definition nach (4.1) folgt. □

Da der Beweis unabhängig davon, ob Φ konvex oder strikt konvex ist, gleichermaßen funktioniert, können wir folgern, dass sowohl die konsistenten als auch die strikt konsistenten Bregman-Funktionen, die in Kapitel 3 hergeleitet wurden, robust sind. Das bedeutet, dass sie verwendet werden können, auch wenn anstelle von σ_t^2 ein Stellvertreter $\hat{\sigma}_t^2$ eingesetzt wird, weil auch dann noch die Bessere zweier Prognosen erkannt wird, wenn diese über den erwarteten Verlust verglichen werden. Damit erfüllen insbesondere auch die QLIKE-Verlustfunktion und der quadratische Verlust die Anforderung in Definition (4.1) und sind robust. Wir können sie also für die Evaluierung von Volatilitätsprognosen verwenden, was wir in Kapitel 7 umsetzen werden.

5 Modellierung der Stellvertreter

Wie im vorherigen Kapitel angekündigt, werden wir uns nun auf die Stellvertreter für σ_t^2 konzentrieren. Wir haben festgestellt, dass das Ranking trotz Verwendung eines Stellvertreters erhalten bleibt, wenn wir eine der Bregman-Funktionen verwenden, da diese robust sind. Das ist bereits ausreichend, um eine bessere Prognose im Vergleich erkennen zu können. Die Frage, die allerdings noch offen bleibt, ist die, welcher Stellvertreter am geeignetsten ist. Inwiefern können wir durch die Wahl des Stellvertreters beeinflussen, wie gut eine bessere Prognose erkannt wird? Dafür betrachten wir zwei Stellvertreter: die quadratischen Renditen r_t^2 , die meist standardmäßig verwendet werden, und die realisierten Volatilitäten RV_t , die beispielsweise von Patton (2006) favorisiert werden. In diesem Kapitel werden wir beide zunächst modellieren, um sie dann in Kapitel 7 im Test einzusetzen und damit zu überprüfen, inwiefern die Wahl eines bestimmten Stellvertreters Einfluss auf die Testergebnisse haben kann.

5.1 Stetige Modellierung

In diesem Teilkapitel werden wir die quadratischen Renditen und die realisierten Volatilitäten unter der Annahme modellieren, dass die bedingte Volatilität ein stetiger Prozess ist. Dafür richten wir uns maßgeblich nach Andersen und Benzoni (2018).

Sei $[0, T]$ mit $T > 0$ ein Zeitintervall, in dem die Preise P_t eines Assets zu $m \cdot T$ äquidistanten Zeitpunkten $t \in \{0, \frac{1}{m}, \dots, T - \frac{1}{m}, T\}$ mit $m > 0$ gemessen werden. Man könnte beispielsweise ein Jahr mit $T = 253$ Handelstagen¹ betrachten, in dem an jedem Handelstag $m = 78$ Beobachtungen der Preise P_t gemacht werden. Das entspricht 5-minütigen Beobachtungen an der New York Stock Exchange (NYSE)². Über $s(t) := \log(P_t)$ kann nun der logarithmische Preis des Assets zum Zeitpunkt t und über $r(t, k) := s(t) - s(t - k)$ die logarithmische Rendite des Assets für das Teilintervall $(t - k, t]$ berechnet werden. Für das Teilintervall $(t - k, t]$ gilt, dass $0 \leq t - k < t \leq T$ und $k = \frac{j}{m}$ mit $j > 0$ ist. Von Interesse für uns ist das Einheitsintervall $(t - 1, t]$ mit $k = 1$. In unserem Beispiel wäre das ein Handelstag, für den die Volatilität gemessen werden soll.

In einem stetigen Finanzmarkt modellieren wir die logarithmischen Renditen über den Itô-Prozess

$$ds(t) = \mu(t)dt + \sigma(t)dW(t), \quad 0 \leq t \leq T, \quad (5.1)$$

¹Das Jahr 2020 hatte 253 Handelstage für Aktien, siehe NYSE (2021)

²An der NYSE wird von 09:30 Uhr bis 16:00 gehandelt. siehe NYSE (2021)

wobei W die Standard Brownsche Bewegung ist. Der Drift $\mu(t)$ kennzeichnet den bedingten Erwartungswert der log>Returns und die Diffusion $\sigma(t)$ steht für die Volatilität der log>Returns. Beides sind prävisible Prozesse, wobei $\mu(t)$ von endlicher Variation und $\sigma(t) > 0$ und quadrat-integrierbar ist. Aus Arbitragegründen muss $s(t)$ ein Semi-Martingal sein. Nehmen wir darüber hinaus, wie in den vorherigen Kapiteln, zentrierte Renditen an, so ist $\mu(t) = 0$ und $s(t)$ ein Martingal wegen

$$\begin{aligned} 0 &= \mathbb{E}[r(t, k) | \mathcal{F}_{t-k}] \\ &= \mathbb{E}[s(t) - s(t - k) | \mathcal{F}_{t-k}] \\ &= \mathbb{E}[s(t) | \mathcal{F}_{t-k}] - s(t - k). \end{aligned}$$

Betrachten wir nun noch einmal die logarithmischen Renditen, so erhalten wir mit dem Itô-Prozess (5.1)

$$r(t, k) = s(t) - s(t - k) = \int_{t-k}^t \mu(\tau) d\tau + \int_{t-k}^t \sigma(\tau) dW(\tau) \quad (5.2)$$

und für die quadratische Variation gilt entsprechend

$$QV(t, k) = \int_{t-k}^t \sigma(\tau)^2 d\tau.$$

Wenn wir die logarithmischen Renditen in (5.2) quadrieren, so erhalten wir die quadratischen Renditen $r(t, k)^2$, den ersten Stellvertreter, den wir für den Vergleich verwenden. Um nun von hier zur Modellierung der realisierten Volatilitäten zu gelangen, nehmen wir wieder das Intervall $(t - k, t]$, für welches die Volatilität bestimmt werden soll, in den Fokus und formulieren darauf eine Partition $\{t - k + \frac{j}{m}, j = 1, \dots, m \cdot k\}$. Zu diesen $m \cdot k$ äquidistanten Zeitpunkten werden die Renditen gemessen. Betrachten wir an dieser Stelle wieder den Fall $k = 1$, so soll die Volatilität für das Intervall $(t - 1, t]$ gemessen werden, was beispielsweise einen Handelstag darstellen könnte. Innerhalb dieses Handelstages liegen m äquidistante Beobachtungen der Rendite vor, welche quadriert und aufsummiert werden, um die realisierten Volatilitäten zu erhalten. Andersen und Benzoni (2018) definieren die realisierten Volatilitäten des Intervalls $(t - k, t]$ als empirisches Maß für die quadratische Variation wie folgt:

$$RV(t, k; m) = \sum_{j=1}^{m \cdot k} r\left(t - k + \frac{j}{m}, \frac{1}{m}\right)^2. \quad (5.3)$$

Sind der Drift μ und die Volatilität σ konstant, so entspricht die bedingte Varianz der Renditen der quadratischen Variation der logarithmischen Preise. Wenn aber die Volatilität $\sigma(t)$ stochastisch ist, so muss klar zwischen den beiden unterschieden werden. Für eine bestimmte Zeitperiode gibt die bedingte Varianz die ex-ante Erwartung und die quadratische Variation die tatsächliche ex-post Realisierung der Volatilität an. Die Semimartingal-Theorie garantiert allerdings, dass gilt:

$$RV(t, k; m) \xrightarrow{\mathbb{P}} QV(t, k), \text{ für } m \rightarrow \infty.$$

Wird also die Stichprobenfrequenz erhöht, so konvergieren die realisierten Volatilitäten in Wahrscheinlichkeit gegen die quadratische Variation, was sie zu einem konsistenten Schätzer für die quadratische Variation macht.

Wir werden nun noch die bedingte Erwartungstreue der realisierten Volatilitäten zeigen. Andersen und Benzoni (2018) geben für diese Aussage keinen Beweis an, daher haben wir diesen hier ergänzt. Wir verwenden dafür, dass unter Annahme der zentrierten Renditen für die bedingte Varianz wieder Folgendes gilt:

$$\text{Var}(r(t, k) | \mathcal{F}_{t-k}) = \mathbb{E}[r(t, k)^2 | \mathcal{F}_{t-k}].$$

Darüber hinaus kommen die Itô-Isometrie und die Linearität des Integrals zum Einsatz. Für $F \in \mathcal{F}_{t-k}$ gilt:

$$\begin{aligned} \mathbb{E}[\mathbb{1}_F RV(t, k; m)] &= \int_F \sum_{j=1}^{m \cdot k} \left(\int_{t-k+\frac{j}{m}-\frac{1}{m}}^{t-k+\frac{j}{m}} \sigma(\tau) dW(\tau) \right)^2 d\mathbb{P} \\ &= \sum_{j=1}^{m \cdot k} \int_F \left(\int_{t-k+\frac{j}{m}-\frac{1}{m}}^{t-k+\frac{j}{m}} \sigma^2(\tau) d\tau \right) d\mathbb{P} && \text{(Linearität, Itô-Isometrie)} \\ &= \int_F \left(\int_{t-k}^t \sigma^2(\tau) d\tau \right) d\mathbb{P} && \text{(Linearität)} \\ &= \int_F \left(\int_{t-k}^t \sigma(\tau) dW(\tau) \right)^2 d\mathbb{P} && \text{(Itô-Isometrie)} \\ &= \mathbb{E}[\mathbb{1}_F r(t, k)^2] \\ &= \mathbb{E}[\mathbb{1}_F \mathbb{E}[r(t, k)^2 | \mathcal{F}_{t-k}]] \\ &= \mathbb{E}[\mathbb{1}_F \text{Var}(r(t, k) | \mathcal{F}_{t-k})]. \end{aligned}$$

Da außerdem gilt, dass $\text{Var}(r(t, k) | \mathcal{F}_{t-k})$ integrierbar und \mathcal{F}_{t-k} -messbar ist, folgt, dass

$$\mathbb{E}[RV(t, k; m) | \mathcal{F}_{t-k}] = \text{Var}(r(t, k) | \mathcal{F}_{t-k}).$$

Damit haben wir gezeigt, dass die realisierten Volatilitäten (5.3) bedingt erwartungstreue Schätzer für die bedingte Varianz der Renditen sind und somit als Stellvertreter verwendet werden können.

5.2 Diskrete Modellierung

Eine ähnliche Herangehensweise verfolgt Patton [vgl. Patton (2006), Kapitel 2.2]. Er geht in diesem Setting von zentrierten Renditen aus und nimmt außerdem an, dass die bedingte Volatilität innerhalb eines Zeitintervalls $(t - k, t]$ konstant ist, so dass gilt:

$$\sigma(\tau) = \sigma_t \quad \forall \tau \in (t - k, t]. \quad (5.4)$$

Damit erhält er für die logarithmischen Renditen die folgende Gleichung:

$$r(t, k) = \int_{t-k}^t \sigma(\tau) dW(\tau) = \sigma_t \int_{t-k}^t dW(\tau).$$

Da W die Standard Brown'sche Bewegung ist, ergibt sich aufgrund der Verteilung der Inkremente von W dann folgende Verteilung für die Renditen:

$$r(t, k) \sim \text{u.i.v. } \mathcal{N}(0, \sigma_t^2 k).$$

Für die innerhalb des Intervalls $(t-k, t]$ beobachteten Renditen, welche für die realisierten Volatilitäten verwendet werden, gilt dann insbesondere folgende Verteilung:

$$r\left(t-k + \frac{j}{m}, \frac{1}{m}\right) \sim \text{u.i.v. } \mathcal{N}\left(0, \frac{\sigma_{t-k+\frac{j}{m}}^2}{m}\right). \quad (5.5)$$

Auch hier ergeben sich die quadratischen Renditen $r(t, k)^2$ wieder durch Quadrieren und die realisierten Volatilitäten wie in (5.3). Zusammen mit der Zentriertheit der Renditen und (5.5) können wir nun die Erwartungstreue der realisierten Volatilitäten folgern:

$$\begin{aligned} \mathbb{E}[RV(t, k; m)] &= \sum_{j=1}^{m \cdot k} \mathbb{E}\left[r\left(t-k + \frac{j}{m}, \frac{1}{m}\right)^2\right] \\ &= \sum_{j=1}^{m \cdot k} \left(\text{Var}\left(r\left(t-k + \frac{j}{m}, \frac{1}{m}\right)\right) + \mathbb{E}\left[r\left(t-k + \frac{j}{m}, \frac{1}{m}\right)\right]^2 \right) \\ &= \sum_{j=1}^{m \cdot k} \frac{\sigma_{t-k+\frac{j}{m}}^2}{m} \quad \text{nach (5.5)} \\ &= \sum_{j=1}^{m \cdot k} \frac{\sigma_t^2}{m} \quad \text{nach (5.4)} \\ &= \sigma_t^2 k. \end{aligned}$$

Für das Einheitsintervall mit $k = 1$, was für uns das Intervall von Interesse darstellt, erhalten wir hier die bedingte Varianz.

In den folgenden Kapiteln werden wir mit der eben formulierten diskreten Modellierung von Patton (2006) arbeiten und speziell den Fall $k = 1$ betrachten. Wir werden die Notationen $r_t^2 := r(t, 1)^2$ und $RV_t := RV(t, 1; m)$ für die quadratischen Renditen und die realisierten Volatilitäten verwenden.

6 Komparatives Backtesting

Wir wollen nun einen Test konstruieren, mit dem wir verschiedene Prognosen der Volatilität vergleichen können. Dafür richten wir uns in diesem Kapitel hauptsächlich nach den Arbeiten von Nolde und Ziegel (2017) und Brehmer (2017). Wir konzentrieren uns auf komparatives Backtesting, da dieses im Gegensatz zu traditionellem Backtesting die Möglichkeit bietet, verschiedene Prognosen zu vergleichen. Bei einem traditionellen Backtest würde jede Prognose einzeln betrachtet und anhand verschiedener Optimalitätskriterien entweder als korrekt oder nicht korrekt eingestuft werden, was in unserem Kontext nicht weiter hilfreich wäre. Wir möchten zwei konkurrierende Prognosen $h_{1,t}$ und $h_{2,t}$ vergleichen, um die Bessere der beiden zu identifizieren. Dazu überprüfen wir mit einem komparativen Backtest folgende Null-Hypothesen:

H_0^- : Die Prognose $h_{1,t}$ ist mindestens so gut wie $h_{2,t}$.

H_0^+ : Die Prognose $h_{1,t}$ ist maximal so gut wie $h_{2,t}$.

Die Prognose $h_{2,t}$ wird dabei als eine Art Richtgröße verwendet, mit der $h_{1,t}$ verglichen wird. Falls die Hypothese H_0^- abgelehnt wird, so ist $h_{1,t}$ schlechter als $h_{2,t}$ und falls H_0^+ abgelehnt wird, so ist $h_{1,t}$ besser als $h_{2,t}$. Den Fehler erster Art, dass eine der beiden Hypothesen abgelehnt wird, obwohl sie wahr ist, kann so direkt über das Signifikanzniveau η gesteuert werden.

Wesentlich für einen komparativen Backtest ist ein elizitierbares Risikomaß, um mit strikt konsistenten Verlustfunktionen arbeiten zu können. Da wir bereits festgestellt haben, dass die bedingte Varianz σ_t^2 unter der Bedingung zentrierter Renditen elizitierbar ist, können wir nun einen solchen Test für Prognosen der Volatilität konstruieren.

6.1 Dominanz

Für ihren Test nutzen Nolde und Ziegel (2017) das Konzept der Dominanz, um eine der beiden Prognosen $h_{1,t}$ und $h_{2,t}$ als die Bessere zu identifizieren. Um aber Dominanz sinnvoll nutzen zu können, werden Verlustfunktionen benötigt, die ordnungssensitiv¹ sind.

Definition 6.1. (*Ordnungssensitivität*²)[vgl. Nolde und Ziegel (2017), Seite 12]
Seien $h_{1,t}$ und $h_{2,t}$ mit $\sigma_t^2 < h_{1,t} < h_{2,t}$ oder $\sigma_t^2 > h_{1,t} > h_{2,t}$ Prognosen für die bedingte Varianz der Renditen σ_t^2 , L eine Verlustfunktion und $\hat{\sigma}_t^2$ ein erwartungstreuer Stellvertreter. Dann heißt L ordnungssensitiv, falls $\mathbb{E}[L(\sigma_t^2, \hat{\sigma}_t^2)] < \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2)] < \mathbb{E}[L(h_{2,t}, \hat{\sigma}_t^2)]$.

¹Im Original: “order sensitive“, hier eigene Übersetzung

²Im Original: “order sensitivity“, hier eigene Übersetzung

Für ordnungssensitive Verlustfunktionen gilt also, dass der erwartete Verlust für eine Prognose umso geringer ausfällt, je näher sie an der tatsächlichen Varianz der Renditen σ_t^2 liegt. Wir werden an dieser Stelle mit den strikt konsistenten Verlustfunktionen der Form (3.2) weiterarbeiten und zeigen, dass diese bereits ordnungssensitiv sind.

Satz 6.2. [vgl. Nolde und Ziegel (2017), S.12]

Zum Erwartungswert strikt konsistente Verlustfunktionen sind ordnungssensitiv.

Ein Beweis für diese Aussage wird nicht angegeben, daher ergänzen wir diesen hier.

Beweis. Sei L eine zum Erwartungswert strikt konsistente Verlustfunktion. Damit hat sie die Form (3.2) mit Φ strikt konvex. Seien weiterhin $h_{1,t}$ und $h_{2,t}$ mit $\sigma_t^2 < h_{1,t} < h_{2,t}$ oder $\sigma_t^2 > h_{1,t} > h_{2,t}$ zwei Prognosen für die bedingte Varianz σ_t^2 und $\hat{\sigma}_t^2$ ein erwartungstreuer Stellvertreter für σ_t^2 .

(i) Mit der strikten Konsistenz von L gilt bereits, dass $\mathbb{E}[L(\sigma_t^2, \hat{\sigma}_t^2)] < \mathbb{E}[L(h_{2,t}, \hat{\sigma}_t^2)]$ und $\mathbb{E}[L(\sigma_t^2, \hat{\sigma}_t^2)] < \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2)]$.

(ii) Um die verbleibende Ungleichung zu zeigen, nutzen wir die Erwartungstreue des Stellvertreters $\hat{\sigma}_t^2$ und die strikte Konvexität von Φ .

$$\begin{aligned} & \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2)] - \mathbb{E}[L(h_{2,t}, \hat{\sigma}_t^2)] \\ &= \mathbb{E}[\Phi(\hat{\sigma}_t^2) - \Phi(h_{1,t}) - \Phi'(h_{1,t})(\hat{\sigma}_t^2 - h_{1,t})] \\ & \quad - \mathbb{E}[\Phi(\hat{\sigma}_t^2) - \Phi(h_{2,t}) - \Phi'(h_{2,t})(\hat{\sigma}_t^2 - h_{2,t})] \\ &= \Phi(h_{2,t}) - \Phi(h_{1,t}) - \Phi'(h_{1,t})(\sigma_t^2 - h_{1,t}) + \Phi'(h_{2,t})(\sigma_t^2 - h_{2,t}) \\ &< -\Phi'(h_{2,t})(h_{1,t} - h_{2,t}) - \Phi'(h_{1,t})(\sigma_t^2 - h_{1,t}) + \Phi'(h_{2,t})(\sigma_t^2 - h_{2,t}) \\ &= (\Phi'(h_{2,t}) - \Phi'(h_{1,t}))(\sigma_t^2 - h_{1,t}) \\ &< 0 \end{aligned}$$

Im letzten Schritt wird ausgenutzt, dass Φ strikt konvex und die Ableitung Φ' damit streng monoton steigend ist. Die Ungleichung gilt dann sowohl für $\sigma_t^2 < h_{1,t} < h_{2,t}$ als auch für $\sigma_t^2 > h_{1,t} > h_{2,t}$. □

Für Verlustfunktionen, die ordnungssensitiv sind, können wir nun sinnvoll das Konzept der Dominanz erarbeiten. Die folgende Definition der L -Dominanz von Nolde und Ziegel (2017) haben wir für unsere Zwecke abgewandelt.

Definition 6.3. (L -Dominanz)[vgl. Nolde und Ziegel (2017), Definition 4]

Sei L eine konsistente Verlustfunktion für die Varianz der Renditen σ_t^2 . Sei weiterhin $\hat{\sigma}_t^2$ ein erwartungstreuer Stellvertreter für σ_t^2 . Dann L -dominiert eine Prognose $h_{1,t}$ eine andere Prognose $h_{2,t}$ im Durchschnitt, falls

$$\mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)] \leq 0 \quad \forall t \in \mathbb{N}.$$

$h_{1,t}$ L -dominiert $h_{2,t}$ bedingt, falls

$$\mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2) | \mathcal{F}_{t-1}] \leq 0 \text{ fast sicher } \forall t \in \mathbb{N}.$$

Bemerkung 6.4. *Aus der bedingten L -Dominanz folgt direkt auch L -Dominanz im Durchschnitt. Ersteres ist daher der stärkere Dominanz-Begriff.*

Im weiteren Teil der Arbeit und vor allem für unseren Backtest werden wir uns dennoch lediglich auf L -Dominanz im Durchschnitt konzentrieren, da bedingte L -Dominanz zu unklaren Ergebnissen führen kann [vgl. Nolde und Ziegel (2017), S.12f.].

Nach der Definition der L -Dominanz würde es ausreichen, eine konsistente Verlustfunktion L zu verwenden, aber da wir bereits festgestellt haben, dass die Klasse der strikt konsistenten Verlustfunktionen der Form (3.2) mit Φ strikt konvex auch ordnungssensitiv sind, werden wir speziell eine strikt konsistente Verlustfunktion verwenden. Das ermöglicht uns einen sinnvollen Vergleich mehrerer Volatilitätsprognosen mit Hilfe von Dominanz.

Auf L -Dominanz aufbauend definiert Ehm et al. (2016) Dominanz dann folgendermaßen, wobei wir auch hier die Definition für unsere Zwecke angepasst haben.

Definition 6.5. *(Dominanz)[vgl. Ehm et al. (2016), Definition 2 b]*

Seien $h_{1,t}$ und $h_{2,t}$ zwei Prognosen für die Varianz der Renditen σ_t^2 und sei $\hat{\sigma}_t^2$ ein erwartungstreuer Stellvertreter für σ_t^2 . Dann wird $h_{2,t}$ von $h_{1,t}$ dominiert, falls $h_{2,t}$ von $h_{1,t}$ für alle zu σ_t^2 konsistenten Verlustfunktionen L L -dominiert wird.

6.2 Zentraler Grenzwertsatz für abhängige Zufallsvariablen

Um in einem nächsten Schritt einen Test auf Dominanz konstruieren zu können und insbesondere um eine geeignete Teststatistik zu finden, wird eine Erweiterung des zentralen Grenzwertsatzes für Folgen abhängiger Zufallsvariablen benötigt. Dazu bezieht sich Brehmer (2017) weitgehend auf Durrett (2005), der sich in [vgl. Durrett (2005), Kapitel 7.c] speziell diesem Konzept widmet. Um die Voraussetzungen für seine Erweiterung des zentralen Grenzwertsatzes zu erfüllen, muss eine Folge von Zufallsvariablen unter anderem strikt stationär und stark mischend sein. Strikte Stationarität definiert Brehmer (2017) folgendermaßen:

Definition 6.6. *(strikt stationär)[vgl. Brehmer (2017), Definition 2.7]*

Eine Folge von Zufallsvariablen $(Y_t)_{t \in \mathbb{Z}}$ heißt strikt stationär, falls für jedes $k \in \mathbb{N}$, $l \in \mathbb{Z}$ und $t_1, \dots, t_k \in \mathbb{Z}$ gilt, dass die Verteilungen von $(Y_{t_1}, \dots, Y_{t_k})$ und $(Y_{t_1+l}, \dots, Y_{t_k+l})$ identisch sind.

Aus strikter Stationarität einer Folge von Zufallsvariablen $(Y_t)_{t \in \mathbb{Z}}$ folgt insbesondere auch, dass Y_t identisch verteilt sind für alle $t \in \mathbb{N}$. Da aber Unabhängigkeit der Zufallsvariablen hier nicht vorausgesetzt werden kann, kann der zentrale Grenzwertsatz nach Lindeberg-Lévy nicht verwendet werden. Falls allerdings eine stark mischende Folge von Zufallsvariablen vorliegt, kann dennoch ein ähnlicher Satz formuliert werden.

Definition 6.7. (*stark mischend*³) [vgl. Brehmer (2017), Definition 2.8]

Sei $(Y_t)_{t \in \mathbb{Z}}$ eine Folge von Zufallsvariablen und seien außerdem $\mathcal{A}_t^\infty := \sigma(Y_t, Y_{t+1}, \dots)$ und $\mathcal{A}_{-\infty}^t := \sigma(Y_t, Y_{t-1}, \dots)$ σ -Algebren. Für $k \in \mathbb{N}$ heißt

$$\alpha(k) := \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \mid t \in \mathbb{Z}, A \in \mathcal{A}_{-\infty}^t, B \in \mathcal{A}_{t+k}^\infty\}$$

starker Mischungskoeffizient⁴ von $(Y_t)_{t \in \mathbb{Z}}$ für das Zeitfenster k .

$(Y_t)_{t \in \mathbb{Z}}$ heißt stark mischend, falls $\alpha(k) \rightarrow 0$ für $k \rightarrow \infty$.

Der starke Mischungskoeffizient $\alpha(k)$ stellt ein Maß für das Abhängigkeitsverhältnis der beiden σ -Algebren $\mathcal{A}_{-\infty}^t$ und \mathcal{A}_{t+k}^∞ dar, wobei $\alpha(k) = 0$ Unabhängigkeit impliziert.

Beide Definitionen kommen nun in folgendem Satz zur Anwendung, in dem ein zentraler Grenzwertsatz für stationäre und stark mischende Folgen von Zufallsvariablen formuliert wird.

Satz 6.8. [vgl. Brehmer (2017), Theorem 2.9]

Sei $(Y_t)_{t \in \mathbb{Z}}$ eine stationäre und stark mischende Folge von Zufallsvariablen mit starkem Mischungskoeffizient α und $\mathbb{E}[Y_0] = 0$. Angenommen, es existiert ein $\delta > 0$, sodass $\mathbb{E}[|Y_t|^{2+\delta}] < \infty$ und $\sum_{k=1}^\infty \alpha(k)^{\delta/(2+\delta)} < \infty$. Dann gilt für

$$\sigma_Y^2 := \mathbb{E}[Y_0^2] + 2 \sum_{t=1}^\infty \mathbb{E}[Y_0 Y_t] < \infty,$$

falls $\sigma_Y^2 > 0$ ist:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n Y_t \xrightarrow{d} \mathcal{N}(0, \sigma_Y^2) \text{ für } n \rightarrow \infty.$$

Für den Beweis verweist Brehmer (2017) auf Durrett (2005). Der eben formulierte Satz ist eine Kombination aus [Durrett (2005), Theorem 7.8] und der nachfolgenden Bemerkung.

Mit einem konsistenten Schätzer $\tilde{\sigma}_n^2$ für σ_Y^2 folgert Brehmer (2017) dann mit dem Lemma von Slutsky, dass für eine Folge von Zufallsvariablen $(Y_t)_{t \in \mathbb{Z}}$, die die Anforderungen im vorigen Satz erfüllt, gilt:

$$\frac{1}{\sqrt{\tilde{\sigma}_n^2}} \frac{1}{\sqrt{n}} \sum_{t=1}^n Y_t \xrightarrow{d} \mathcal{N}(0, 1) \text{ für } n \rightarrow \infty. \quad (6.1)$$

Diese Konvergenz können wir nun im folgenden Teilkapitel nutzen, um eine Teststatistik für einen asymptotischen Hypothesentest zu konstruieren.

³Im Original: “strongly mixing“

⁴Im Original: “strong mixing coefficient“

6.3 Diebold-Mariano Test

Mit dem Konzept der Dominanz können wir nun die beiden Null-Hypothesen H_0^- und H_0^+ konkreter ausformulieren. Es ist naheliegend, eine L -dominante Prognose als die Bessere anzunehmen, wenn eine Verlustfunktion verwendet wird, die strikt konsistent und damit ordnungssensitiv ist. Nolde und Ziegel (2017) definieren dafür das Limit

$$\lambda := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)] \in [-\infty, +\infty]$$

und nehmen an, dass es existiert. Per Definition der Dominanz folgt direkt, dass $\lambda \leq 0$ ist, falls $h_{2,t}$ von $h_{1,t}$ im Durchschnitt L -dominiert wird. Für die Umkehrung dieser Implikation benötigen wir zusätzlich strikte Stationarität, denn, falls die Folge der Zufallsvariablen $(L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2))_{t \in \mathbb{N}}$ strikt stationär ist, so sind diese insbesondere identisch verteilt und es gilt:

$$\begin{aligned} 0 \geq \lambda &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)] \text{ für } t \in \mathbb{N} \text{ beliebig} \\ &= \mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)] \text{ für } t \in \mathbb{N} \text{ beliebig.} \end{aligned}$$

Somit impliziert $\lambda \leq 0$ dann auch L -Dominanz im Durchschnitt und wir erhalten folgende Null-Hypothesen:

$$H_0^- : \lambda \leq 0,$$

$$H_0^+ : \lambda \geq 0.$$

Unter H_0^- L -dominiert die Prognose $h_{1,t}$ die als Richtgröße fungierende Prognose $h_{2,t}$, ist also mindestens so gut wie $h_{2,t}$. Für H_0^+ gilt umgekehrt, dass $h_{1,t}$ von $h_{2,t}$ L -dominiert wird und daher ist $h_{1,t}$ nur maximal so gut wie $h_{2,t}$. Wir haben also das Konzept der Dominanz genutzt, um Hypothesen eines komparativen Backtests zu konstruieren. In diesem Setting betrachten sowohl Nolde und Ziegel (2017) als auch Brehmer (2017) vorerst die Teststatistik

$$\Delta_n \bar{L} := \frac{1}{n} \sum_{t=1}^n (L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)).$$

Nach dem starken Gesetz der großen Zahlen konvergiert $\Delta_n \bar{L}$ für $n \rightarrow \infty$ fast sicher gegen $\mathbb{E}[L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)]$ für ein $t \in \mathbb{N}$ und entspricht damit unserem Dominanz-Argument für oder gegen die Prognose $h_{1,t}$.

Außerdem entspricht $\mathbb{E}[\Delta_n \bar{L}]$ asymptotisch dem Limit λ , falls n groß genug gewählt wird, womit wir folgende Umformulierung unserer Null-Hypothesen erhalten:

$$H_0^- : \mathbb{E}[\Delta_n \bar{L}] \leq 0,$$

$$H_0^+ : \mathbb{E}[\Delta_n \bar{L}] \geq 0.$$

Tests für H_0^- und H_0^+ , die auf einer passend skalierten Version von $\Delta_n \bar{L}$ basieren, sind sogenannte Diebold-Mariano Tests. Auch wir werden $\Delta_n \bar{L}$ skalieren und einen solchen Test konstruieren.

An dieser Stelle können wir die Ergebnisse des vorigen Teilkapitels anwenden, denn falls die Folge der Zufallsvariablen $(D_t)_{t \in \mathbb{N}}$ mit $D_t := L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)$ die Anforderungen aus Satz (6.8) erfüllt und $\tilde{\sigma}_n^2$ ein konsistenter Schätzer für σ_D^2 ist, so gilt nach (6.1), dass

$$\sqrt{n} \frac{\Delta_n \bar{L}}{\tilde{\sigma}_n} = \frac{1}{\sqrt{\tilde{\sigma}_n^2}} \frac{1}{\sqrt{n}} \sum_{t=1}^n D_t \xrightarrow{d} \mathcal{N}(0, 1) \text{ für } n \rightarrow \infty.$$

Wenn wir also die Teststatistik

$$T := \sqrt{n} \frac{\Delta_n \bar{L}}{\tilde{\sigma}_n}$$

verwenden, so erhalten wir einen asymptotischen Level- η -Test. Bei einem Test für die Null-Hypothese H_0^+ wird diese abgelehnt, wenn $\Phi(T) \leq \eta$ und bei einem Test für die Null-Hypothese H_0^- wird diese abgelehnt, wenn $1 - \Phi(T) \leq \eta$ ist, wobei das Level $\eta \in (0, 1)$ im Vorhinein festgelegt wird. An dieser Stelle steht Φ als Notation für die Verteilungsfunktion der Standardnormalverteilung.

Was jetzt noch fehlt, ist ein konsistenter Schätzer $\tilde{\sigma}_n^2$ für σ_D^2 . Wir greifen dafür auf den von Gneiting und Ranjan (2008) vorgeschlagenen Schätzer mit $k = 0$ zurück:

$$\begin{aligned} \tilde{\sigma}_n^2 &= \frac{1}{n} \sum_{j=-1}^1 \sum_{t=1}^{n-|j|} [(L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)) \\ &\quad \cdot (L(h_{1,t+|j|}, \hat{\sigma}_{t+|j|}^2) - L(h_{2,t+|j|}, \hat{\sigma}_{t+|j|}^2))] \\ &= \frac{1}{n} \sum_{t=1}^n (L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2))^2 \\ &\quad + \frac{2}{n} \sum_{t=1}^{n-1} [(L(h_{1,t}, \hat{\sigma}_t^2) - L(h_{2,t}, \hat{\sigma}_t^2)) \\ &\quad \cdot (L(h_{1,t+1}, \hat{\sigma}_{t+1}^2) - L(h_{2,t+1}, \hat{\sigma}_{t+1}^2))] . \end{aligned}$$

Damit haben wir die Konstruktion unseres komparativen Backtests für die bedingte Varianz abgeschlossen und können diesen nun verwenden, um mehrere konkurrierende Volatilitätsprognosen zu evaluieren.

7 Numerische Anwendung

In diesem Kapitel werden wir nun mit dem eben entwickelten Diebold-Mariano Test überprüfen, inwiefern sich die Verwendung realisierter Volatilitäten als Stellvertreter auf die Aussagekraft des Tests auswirkt. Dazu werden wir die quadratischen Renditen r_t^2 und die realisierten Volatilitäten RV_t als Stellvertreter einsetzen und den Test für mehrere Prognosen der bedingten Varianz durchführen. Als Verlustfunktionen nutzen wir den quadratischen Verlust und die QLIKE-Verlustfunktion, da diese strikt konsistent und somit auch robust sind. Wir werden den Test dann jeweils für verschiedene Signifikanzniveaus durchführen und so versuchen, Rückschlüsse ziehen zu können.

7.1 Datengrundlage

Die Grundlage für die Tests bildet eine in \mathbf{R} simulierte Zeitreihe. Wir simulieren zunächst eine Reihe standardisierter intraday Renditen $\epsilon_{t,i}$ mit

$$\epsilon_{t,i} \sim \mathcal{N}\left(0, \frac{1}{m}\right), \text{ mit } i = 1, \dots, m \text{ und } t \in [0, T].$$

Dabei gehen wir von $m = 100$ und $T = 6100$ aus. Durch Aufsummieren der $\epsilon_{t,i}$ ergeben sich die standardisierten, täglichen Renditen ϵ_t , die wir dann nutzen, um mit den Parametern $a_0 = 0,02$, $a_1 = 0,08$ und $b = 0,85$ eine $GARCH(1,1)$ -Zeitreihe der täglichen Renditen r_t zu erzeugen. Dafür gehen wir iterativ vor und berechnen für jeden Zeitpunkt $t \in [0, T]$ die bedingte Varianz und die täglichen Renditen mit

$$\begin{aligned}\sigma_t^2 &= a_0 + a_1 r_{t-1}^2 + b \sigma_{t-1}^2 \text{ und} \\ r_t &= \sigma_t \epsilon_t.\end{aligned}$$

Um Stationarität zu erhalten, löschen wir die ersten 100 Realisierungen unserer Zeitreihe und erhalten damit eine Zeitreihe mit 6000 Datenpunkten. Für diese Zeitreihe erstellen wir nun mithilfe des Paketes `rugarch` und insbesondere der Funktion `ugarchroll()` Prognosen der bedingten Varianz mittels rollierendem Zeitfenster für $GARCH(1,1)$ -, $ARCH(1)$ -, $ARCH(2)$ - und $ARCH(7)$ -Zeitreihen. Dazu verwenden wir ein Zeitfenster von 2000 Datenpunkten und treffen jeweils Einschnitt-Vorhersagen. Diese vergleichen wir dann anschließend, mittels des komparativen Backtests aus dem vorherigen Kapitel, miteinander und mit der tatsächlichen bedingten Varianz σ_t^2 unserer simulierten Zeitreihe r_t . Um die Güte des Tests zu erhöhen, werden 500 Durchgänge des Tests berechnet und die Teststatistiken anschließend gemittelt. Wir erhalten eine 5×5 -Matrix der Teststatistiken, welche wir dann mit dem im nächsten Teilkapitel beschriebenen Drei-Zonen-Ansatz markieren.

7.2 Drei-Zonen-Ansatz

Für die Auswertung der Tests verwenden wir den Drei-Zonen-Ansatz von Fissler et al. (2016), die diesen für Backtests des Expected Shortfall nutzen. Wie weiter oben bereits erklärt, gehen wir davon aus, dass die Prognose $h_{1,t}$ schlechter ist als $h_{2,t}$, falls wir die Nullhypothese H_0^- ablehnen, und wir nehmen an, dass $h_{1,t}$ besser ist als $h_{2,t}$, falls wir die Nullhypothese H_0^+ ablehnen. Können wir keine der beiden Hypothesen ablehnen, so können wir keine Aussage über die beiden Prognosen treffen. Daraus ergeben sich drei Zonen, die Fissler et al. (2016) mit einem Ampelsystem markieren. Dieses bezieht sich auf die Prognose $h_{1,t}$, die mit der als Richtgröße fungierenden Prognose $h_{2,t}$ verglichen wird. Lehnen wir H_0^+ ab, so wird der Test intuitiv grün markiert, um zu signalisieren, dass $h_{1,t}$ besser ist als $h_{2,t}$. Lehnen wir dagegen H_0^- ab, so wird der Test rot markiert. Und falls keine Hypothese abgelehnt werden kann, so ist die Markierung gelb.

Zone	Hypothesen	Ergebnis
● rot	H_0^- wird abgelehnt	$h_{1,t}$ ist schlechter als $h_{2,t}$
● gelb	keine Hypothese wird abgelehnt	keine Aussage
● grün	H_0^+ wird abgelehnt	$h_{1,t}$ ist besser als $h_{2,t}$

Tabelle 7.1: Drei-Zonen-Ansatz nach Fissler et al. (2016)

Mit diesen Markierungen können wir sehr übersichtliche Plots erstellen, wie etwa Abbildung (7.1) im folgenden Teilkapitel. Je nachdem, welchen Wert die Teststatistik hat, wird das entsprechende Feld der Matrix eingefärbt.

7.3 Testergebnisse

Die folgenden Abbildungen zeigen die Testergebnisse, wie wir sie oben bereits beschrieben haben. Wir erhalten verschiedene Ergebnisse, je nachdem, welchen Stellvertreter und welche Verlustfunktion wir verwenden. Und um zusätzlich die Aussagekraft der Tests besser beurteilen zu können, betrachten wir diese jeweils für die Signifikanzniveaus $\eta = 0,005$, $\eta = 0,05$ und $\eta = 0,1$.

Auf der x -Achse einer einzelnen Matrix ist die Prognose $h_{2,t}$ und auf der y -Achse die Prognose $h_{1,t}$ dargestellt. Der Bezeichner 'opt' steht für die tatsächliche bedingte Varianz σ_t^2 , die die optimale Prognose darstellt. In diesem speziellen Fall kennen wir den Wert der tatsächlichen bedingten Varianz, da der Datensatz vollständig simuliert ist. Der Wert innerhalb eines Matrix-Eintrages gibt den Wert der jeweiligen Teststatistik T wieder. Dieser wird mit den dem Signifikanzniveau η entsprechenden Quantilen der Standardnormalverteilung verglichen und nach dem Drei-Zonen-Ansatz rot, grün oder gelb eingefärbt. Für $T \geq \Phi^{-1}(1 - \frac{\eta}{2})$ erhalten wir eine rote Markierung, für $T \leq \Phi^{-1}(\frac{\eta}{2})$ eine grüne und falls $\Phi^{-1}(\frac{\eta}{2}) < T < \Phi^{-1}(1 - \frac{\eta}{2})$ ist, so wird das Feld gelb gefärbt. Hierbei verwenden wir Φ wieder für die Verteilungsfunktion der Standardnormalverteilung.

Wir betrachten zunächst einmal beispielhaft die erste Matrix oben rechts in Abbildung

(7.1). Hier werden die Testergebnisse unter Verwendung der quadratischen Verlustfunktion, einem Signifikanzniveau von $\eta = 0,05$ und der realisierten Volatilitäten RV_t als Stellvertreter dargestellt. Der Test mit der tatsächlichen bedingten Varianz σ_t^2 als $h_{1,t}$ und der auf der $ARCH(7)$ -Zeitreihe beruhenden Prognose als $h_{2,t}$ liefert die Teststatistik $T = -13,333$. Wir haben die Ergebnisse an dieser Stelle der Übersichtlichkeit halber auf drei Nachkommastellen gerundet. Das entsprechende Feld in der Matrix ist grün markiert, da $T \leq \Phi^{-1}(0,025)$. Das bedeutet, dass wir die Hypothese H_0^+ ablehnen können und damit nehmen wir an, dass die tatsächliche bedingte Varianz besser ist, als die $ARCH(7)$ -Prognose.

Der entsprechende Test in der Matrix oben links unterscheidet sich dazu nur durch die Wahl des Stellvertreters. Hier werden die quadratischen Renditen r_t^2 anstelle der realisierten Volatilitäten RV_t verwendet. Betrachten wir hier das äquivalente Feld in der Matrix, so können wir ablesen, dass die Teststatistik hier $T = -2,901$ ist, also im Betrag deutlich kleiner. Auch hier ist das Feld wieder grün markiert, da $T \leq \Phi^{-1}(0,025)$ und wir nehmen an, dass die tatsächliche bedingte Varianz die Bessere der beiden Prognosen ist.

Das Ergebnis des Tests ist in diesem konkreten Fall also identisch, unabhängig davon, welcher Stellvertreter eingesetzt wird, lediglich die Teststatistiken unterscheiden sich in ihrem Betrag voneinander.



Abbildung 7.1: Test unter Verwendung der quadratischen Verlustfunktion



Abbildung 7.2: Test unter Verwendung der QLIKE-Verlustfunktion

7.4 Interpretation

Abbildung (7.1) zeigt die Ergebnisse des komparativen Backtests unter Verwendung des quadratischen Verlustes als Verlustfunktion. Betrachten wir die obersten beiden Matrizen, also die Testergebnisse für das Signifikanzniveau $\eta = 0,05$, so fällt auf, dass der Test unter Verwendung der quadratischen Renditen als Stellvertreter (oben links) mehr Felder gelb markiert hat, als der Test unter Verwendung der realisierten Volatilitäten (oben rechts). Bei drei der zehn Tests kann links keine Aussage getroffen werden, wohingegen rechts bei allen Tests eine der Hypothesen abgelehnt werden kann. Führen wir den Test beispielsweise für die tatsächliche bedingte Varianz der Renditen σ_t^2 und die $GARCH(1,1)$ -Prognose durch, so können wir keine Aussage treffen, wenn wir die quadratischen Renditen r_t^2 als Stellvertreter verwenden, da die Teststatistik mit $T = -1,137$ im Betrag zu gering ausfällt. Wenn wir allerdings die realisierten Volatilitäten RV_t einsetzen, so ist die Teststatistik mit $T = -7,176$ im Betrag groß genug und wir können annehmen, dass die $GARCH(1,1)$ -Prognose der tatsächlichen bedingten Varianz σ_t^2 unterlegen ist. Es kann insgesamt bei einer größeren Anzahl an Tests keine der beiden Nullhypothesen H_0^- und H_0^+ abgelehnt werden, wenn wir die quadratischen Renditen als Stellvertreter verwenden. Darüber hinaus sind die Werte der einzelnen Teststatistiken unter Verwendung der realisierten Volatilitäten im Betrag ohne Ausnahme größer, was dazu führt, dass in allen Fällen eine Aussage getroffen werden kann.

Die beiden Matrizen in zweiter Reihe zeigen die Testergebnisse, wenn wir das Signifikanzniveau und somit auch die Wahrscheinlichkeit für den Fehler erster Art auf $\eta = 0,1$ erhöhen. Der Unterschied in der Aussagekraft wird geringer, da wir bei Verwendung der quadratischen Renditen r_t^2 bei einem der vorher gelb markierten Tests nun eine Aussage treffen können. Nichtsdestotrotz bleibt der Unterschied in allen anderen Fällen bestehen. Verringern wir dagegen das Signifikanzniveau stark auf $\eta = 0,005$, so wird der Unterschied noch deutlicher, wobei die Markierungen der Tests mit den realisierten Volatilitäten selbst dann unverändert bleiben. Die entsprechenden Tests für das Signifikanzniveau $\eta = 0,005$ werden in den unteren beiden Matrizen dargestellt.

Verwenden wir also den quadratischen Verlust als Verlustfunktion, so führt die Verwendung der realisierten Volatilitäten anstelle der quadratischen Renditen als Stellvertreter dazu, dass die Aussagekraft des Tests stark erhöht wird und wir selbst bei einem sehr geringen Signifikanzniveau Hypothesen ablehnen und Aussagen treffen können.

Zum Vergleich haben wir identische Tests ebenfalls für die QLIKE-Verlustfunktion durchgeführt. Die Ergebnisse werden in Abbildung (7.2) dargestellt. Auch hier ist ein klarer Unterschied hinsichtlich der Aussagekraft erkennbar, obwohl dieser nicht so groß ist wie in der ersten Testreihe. Die Verwendung der realisierten Volatilitäten als Stellvertreter führt dazu, dass bei einer größeren Anzahl an Tests Hypothesen abgelehnt werden können und damit eine Aussage getroffen werden kann. Auch hier sind die absoluten Beträge der Teststatistiken ausnahmslos größer, wenn wir die realisierten Volatilitäten verwenden.

Eine Erhöhung des Signifikanzniveaus auf $\eta = 0,1$ hat keine Auswirkungen in beiden Fällen, wohingegen eine Senkung auf $\eta = 0,005$ die Aussagekraft in beiden Fällen vermindert. Der Unterschied zwischen den beiden Stellvertretern bleibt aber bestehen und

wird sogar noch verstärkt. Somit können wir auch hier folgern, dass die realisierten Volatilitäten im Vergleich zu den quadratischen Renditen als Stellvertreter stark zur Verbesserung der Testergebnisse, auch unter Verwendung der QLIKE-Verlustfunktion, führen.

Zusätzlich zu den Unterschieden, die sich durch die verschiedenen Stellvertreter ergeben, sind auch Unterschiede zu erkennen, je nachdem, welche Verlustfunktion zum Einsatz kommt. Beide sind strikt konsistent und damit auch robust und ordnungssensitiv und können daher prinzipiell für den Vergleich von Volatilitätsprognosen mittels Stellvertretern verwendet werden. Betrachten wir nur die Tests, bei denen die quadratischen Renditen als Stellvertreter verwendet wurden, also nur die linken Matrizen in Abbildung (7.1) und Abbildung (7.2), so scheint die QLIKE-Verlustfunktion besser geeignet, um Volatilitätsprognosen zu evaluieren. Die Teststatistiken fallen insgesamt im Betrag größer aus und dadurch kann in mehr Fällen eine der Hypothesen abgelehnt werden. Im Fall der realisierten Volatilitäten ergibt sich allerdings ein etwas anderes Bild. In sechs der zehn Tests führt die Verwendung der QLIKE-Verlustfunktion zu im Betrag größer ausfallenden Teststatistiken. In den übrigen Fällen verhält es sich umgekehrt. Dies ist beispielsweise beim Vergleich der $ARCH(1)$ - und der $ARCH(2)$ -Prognosen der Fall. Für die Teststatistik ergeben sich die Werte $T = 3,055$ für den quadratischen Verlust und $T = 2,246$ für die QLIKE-Verlustfunktion. An dieser Stelle hat das sogar zur Folge, dass bei einem Signifikanzniveau von $\eta = 0,005$ keine der beiden Hypothesen mehr abgelehnt werden kann, falls die QLIKE-Verlustfunktion verwendet wird. In den übrigen Fällen ergeben sich keine Auswirkungen auf die Markierungen. Insgesamt können wir also keine eindeutige Aussage über die Eignung der beiden Verlustfunktionen zur Evaluierung von Volatilitätsprognosen formulieren, allerdings scheint die QLIKE-Verlustfunktion in den meisten Fällen bessere Ergebnisse zu liefern.

8 Fazit und Ausblick

Das Ziel der Arbeit war die Evaluierung von Volatilitätsvorhersagen unter Verwendung robuster Verlustfunktionen und realisierter Volatilitäten. Zu Beginn haben wir gezeigt, dass σ_t^2 , die bedingte Varianz der Renditen, unter Annahme der Zentriertheit der Renditen r_t elizitierbar ist. Somit konnten wir dazu konsistente und strikt konsistente Verlustfunktionen herleiten, die den Vorteil haben, dass die optimale Prognose, in unserem Fall die tatsächliche bedingte Varianz σ_t^2 , als optimal erkannt wird. Darüber hinaus haben wir gezeigt, dass diese konsistenten Verlustfunktionen auch robust sind und somit grundsätzlich für die Evaluierung von Prognosen für ein latentes Risikomaß, wie etwa die Volatilität, mittels Stellvertretern verwendet werden können. Wir haben speziell die quadratischen Renditen r_t^2 und die realisierten Volatilitäten RV_t als Stellvertreter betrachtet. Dabei war das Ziel, zu untersuchen, ob es sich positiv auf die Evaluierung von Volatilitätsprognosen auswirkt, wenn die realisierten Volatilitäten anstelle der häufig verwendeten quadratischen Renditen genutzt werden. Dazu haben wir einen komparativen Backtest, konkret einen Diebold-Mariano Test konstruiert, der zwei konkurrierende Prognosen vergleicht und überprüft, welche der beiden die Bessere ist.

Diesen Test haben wir anschließend auf einer simulierten Zeitreihe von Renditen durchgeführt und dafür die beiden Stellvertreter RV_t und r_t^2 sowie die QLIKE-Verlustfunktion und den quadratischen Verlust als Verlustfunktionen verwendet. Dabei ist aufgefallen, dass die Teststatistiken in allen Fällen im Betrag größer ausfallen, wenn die realisierten Volatilitäten verwendet werden. Das hat zur Folge, dass der Test insgesamt aussagekräftiger wird und wir in einigen Fällen eine Aussage treffen können, während wir bei Verwendung der quadratischen Renditen keine treffen können. Im Vergleich der beiden Stellvertreter führen die realisierten Volatilitäten also insgesamt zu besseren Ergebnissen. Der Vergleich der Verlustfunktionen zeigt, dass die QLIKE-Verlustfunktion in den meisten Fällen geeigneter erscheint. Jedoch bestätigt sich dies nicht in allen Fällen und somit ergibt sich keine einheitliche Schlussfolgerung.

An dieser Stelle ist allerdings anzumerken, dass die hier erzielten Ergebnisse keinesfalls statistisch signifikant sind. Wir haben lediglich den quadratischen Verlust und die QLIKE-Verlustfunktion als Repräsentanten der Bregman-Funktionen verwendet und können daher keine allgemein gültige Aussage für alle konsistenten Verlustfunktionen treffen. Darüber hinaus haben wir die Menge der zu evaluierenden Prognosen sehr stark eingegrenzt und abgesehen von der tatsächlichen, bedingten Varianz lediglich vier weitere Prognosen betrachtet. Das führt wiederum dazu, dass die Resultate nicht stellvertretend für die Evaluierung beliebiger Volatilitätsprognosen stehen können. Es müssten also weitere und vor allem signifikantere Tests durchgeführt werden, in denen die Auswahl der Prognosen und der Verlustfunktionen noch weiter variiert werden würde, um aussagekräftigere Ergebnisse zu erhalten. Zusätzlich dazu basieren unsere Tests auf einer

simulierten Zeitreihe und die Anwendung und Auswirkung auf realen Daten bleibt in dieser Arbeit ungetestet.

Trotz allem geben uns die hier erzielten Ergebnisse einen Einblick in die Möglichkeiten, die die Verwendung realisierter Volatilitäten bieten könnte. Die Verbesserung, die die Verwendung realisierter Volatilitäten bereits in unserem Setting bewirkt hat, ist erheblich. Sollten sich die Resultate in einer signifikanteren Umgebung gleichermaßen einstellen, so würde die Umstellung auf die realisierten Volatilitäten als Stellvertreter starke, positive Auswirkungen auf die Evaluierung von Volatilitätsprognosen mit sich bringen. Es ist vorstellbar, dass die realisierten Volatilitäten dann in der Praxis zunehmend Anwendung finden und die derzeit viel verwendeten quadratischen Renditen als Stellvertreter ablösen. Jedoch ist noch anzumerken, dass die Anwendung sehr stark davon abhängt, inwiefern die für die Berechnung der realisierten Volatilitäten verwendeten intraday Renditen verfügbar sind.

Obwohl diese Arbeit keinen signifikanten Beweis liefern konnte und weitere wissenschaftliche Betrachtungen nötig sein werden, bleibt insgesamt festzuhalten, dass die realisierten Volatilitäten scheinbar eine Verbesserung gegenüber der quadratischen Renditen im Bezug auf die Evaluierung von Volatilitätsprognosen darstellen.

9 Anhang: R-Code

```

1 library(rugarch)
2 library('plot.matrix')
3 #####
4 ##### Parameter festlegen
5 #####
6 m <- 100 # Number of intraday returns
7 T <- 6100 # Number of days for Simulation
8 #####
9 ##### Verlustfunktionen
10 #####
11 MSE <- function(estimate, proxy){
12   return((estimate-proxy)^2)
13 }
14 QLIKE <- function(estimate, proxy){
15   return(log(estimate)+(proxy/estimate))
16 }
17 #####
18 ##### Diebold Mariano Teststatistik
19 #####
20 DM.Test <- function(h_1, h_2, proxy, scoreFunction){
21   score.diff <- scoreFunction(h_1, proxy) - scoreFunction(h_2, proxy)
22   n <- length(score.diff)
23   sigma.DM <- sqrt(mean(score.diff^2)+2*mean(score.diff[-n]*score.diff[-1]))
24   if(sigma.DM == 0){ return(0) }
25   testresult <- sqrt(n)*mean(score.diff)/sigma.DM
26   return(testresult)
27 }
28 #####
29 ##### Matrix Plot
30 #####
31 MatrixPlot <- function(matrix, eta, proxy){
32   par(mar=c(5.5,1,0.5,6.3)) # Abstand zum Rand (unten, links, oben, rechts)
33   rownames(matrix) <- c("opt", "GARCH(1,1)", "ARCH(1)", "ARCH(2)", "ARCH(7)")
34   colnames(matrix) <- c("opt", "GARCH(1,1)", "ARCH(1)", "ARCH(2)", "ARCH(7)")
35   plot(matrix, col=c('green','yellow','red'),
36        breaks=c(-100,qnorm(eta/2), qnorm(1-(eta/2)),100), #Intervalle f r die Farben
37        digits=3, #Nachkommastellen in der Matrix
38        cex=0.9, #Schriftgr e in der Matrix
39        xlab="", ylab="",
40        axis.row = NULL,#list(side=2, las=1),
41        axis.col = NULL,#3, #Achsenbeschriftung auf Seite 3
42        cex.axis=0.9, #Schriftgr e der Achsenbeschriftung
43        key = NULL, #keine Legende
44        main = ""
45   )
46   mtext(bquote(eta == .(eta)~", Stellvertreter"~"~.(proxy)), side = 1, line = 1.1,
47        cex = 0.9)
48 }
49 #####
50 ##### Beginne Loop
51 #####
52 n.loop <- 500 #Anzahl der Schleifendurchl ufe
53 RV.MSE.array <- array(0,c(5,5,n.loop))
54 R2.MSE.array <- array(0,c(5,5,n.loop))
55 RV.QLIKE.array <- array(0,c(5,5,n.loop))
56 R2.QLIKE.array <- array(0,c(5,5,n.loop))

```

```

56 for (k in 1:n.loop) {
57   #####
58   ##### simulierte Zeitreihe der Renditen
59   #####
60   set.seed(k)
61   Epsilon.intraday <- matrix(rnorm(T*m, mean = 0, sd = sqrt(1/m)), nrow = m)
62   # Standardisierte Renditen pro Tag
63   Epsilon.daily <- colSums(Epsilon.intraday)
64
65   # Simuliere bed. Varianz und Renditen
66   a0 <- 0.02
67   a1 <- 0.08
68   b <- 0.85
69   sigma.squared <- c(a0/(1-a1-b))
70   returns <- c(0)
71   for (t in 2:T) {
72     # bed. Varianz pro Tag
73     sigma.squared <- c(sigma.squared, a0 + a1*(returns^2)[t-1] + b*sigma.squared[t-1])
74     # t gleiche Renditen
75     returns <- c(returns, (sqrt(sigma.squared))[t]*Epsilon.daily[t])
76   }
77
78   # L schie die ersten 100 Simulationen, damit wir Stationarit t erhalten
79   returns <- returns[-c(1:100)]
80   Epsilon.daily <- Epsilon.daily[-c(1:100)]
81   sigma.squared <- sigma.squared[-c(1:100)]
82   Epsilon.intraday <- Epsilon.intraday[,-c(1:100)]
83
84   # Simuliere intraday Returns
85   returns.intraday <- Epsilon.intraday %*% diag(sqrt(sigma.squared))
86   #####
87   ##### Stellvertreter
88   #####
89   # 1. Stellvertreter: quadratische Renditen
90   squared.returns <- returns^2
91
92   # 2. Stellvertreter: realisierte Volatilit t
93   RV <- colSums(returns.intraday^2)
94
95   #####
96   ##### erster Sch tzer GARCH(1,1)
97   #####
98   # Modellspezifikationen
99   garch.spec <- ugarchspec(variance.model=list(model="fGARCH",garchOrder=c(1,1),
100     submodel="GARCH"),
101     mean.model=list(armaOrder=c(0,0),include.mean=FALSE))
102   # rollierendes Fenster mit GARCH(1,1) Modell
103   garch.roll <- ugarchroll(spec = garch.spec, data = returns,
104     n.ahead = 1, n.start = 2000,
105     refit.every = 10, refit.window = "moving",
106     window.size = 2000, calculate.Var=FALSE)
107   garch.estimate <- garch.roll@forecast$density$Sigma^2
108   #####
109   ##### zweiter Sch tzer Arch(1)
110   #####
111   arch1.spec <- ugarchspec(variance.model=list(model="fGARCH",garchOrder=c(1,0),
112     submodel="GARCH"),
113     mean.model=list(armaOrder=c(0,0),include.mean=FALSE))
114   arch1.roll <- ugarchroll(spec = arch1.spec, data = returns,
115     n.ahead = 1, n.start = 2000,
116     refit.every = 10, refit.window = "moving",
117     window.size = 2000, calculate.Var=FALSE)
118   arch1.estimate <- arch1.roll@forecast$density$Sigma^2
119   #####
120   ##### dritter Sch tzer Arch(2)
121   #####

```

```

120 arch2.spec <- ugarchspec(variance.model=list(model="fGARCH",garchOrder=c(2,0),
121                               submodel="GARCH"),
122                               mean.model=list(armaOrder=c(0,0),include.mean=FALSE))
123 arch2.roll <- ugarchroll(spec = arch2.spec, data = returns,
124                               n.ahead = 1, n.start = 2000,
125                               refit.every = 10, refit.window = "moving",
126                               window.size = 2000, calculate.VaR=FALSE)
127 arch2.estimate <- arch2.roll@forecast$density$Sigma^2
128 ##### vierter Sch tzer Arch(7)
129 #####
130 arch7.spec <- ugarchspec(variance.model=list(model="fGARCH",garchOrder=c(7,0),
131                               submodel="GARCH"),
132                               mean.model=list(armaOrder=c(0,0),include.mean=FALSE))
133 arch7.roll <- ugarchroll(spec = arch7.spec, data = returns,
134                               n.ahead = 1, n.start = 2000,
135                               refit.every = 10, refit.window = "moving",
136                               window.size = 2000, calculate.VaR=FALSE)
137 arch7.estimate <- arch7.roll@forecast$density$Sigma^2
138 ##### passe Sch tzer und Stellvertreter f r Test an
139 #####
140 sigma.squared.true <- sigma.squared[2001:6000]
141 squared.returns.test <- squared.returns[2001:6000]
142 RV.test <- RV[2001:6000]
143 estimates <- cbind(sigma.squared.true, garch.estimate, arch1.estimate, arch2.estimate
144                               , arch7.estimate)
145 ##### Test mit quadratischen Renditen
146 #####
147 result.squared.returns <- c()
148 for (i in 1:5) {
149   for (j in 1:5) {
150     result.squared.returns <- append(result.squared.returns,
151                                       DM.Test(h_1 = estimates[,i],
152                                               h_2 = estimates[,j],
153                                               proxy = squared.returns.test,
154                                               scoreFunction = MSE))
155   }
156 }
157 R2.MSE.array[, ,k] <- matrix(result.squared.returns, byrow = TRUE, nrow = 5)
158
159 result.squared.returns <- c()
160 for (i in 1:5) {
161   for (j in 1:5) {
162     result.squared.returns <- append(result.squared.returns,
163                                       DM.Test(h_1 = estimates[,i],
164                                               h_2 = estimates[,j],
165                                               proxy = squared.returns.test,
166                                               scoreFunction = QLIKE))
167   }
168 }
169 R2.QLIKE.array[, ,k] <- matrix(result.squared.returns, byrow = TRUE, nrow = 5)
170 ##### Test mit realisierten Volatilit ten
171 #####
172 result.RV <- c()
173 for (i in 1:5) {
174   for (j in 1:5) {
175     result.RV <- append(result.RV, DM.Test(h_1 = estimates[,i],
176                                           h_2 = estimates[,j],
177                                           proxy = RV.test,
178                                           scoreFunction = MSE))
179   }
180 }
181
182 RV.MSE.array[, ,k] <- matrix(result.RV, byrow = TRUE, nrow = 5)

```

```

183
184 result.RV <- c()
185 for (i in 1:5) {
186   for (j in 1:5) {
187     result.RV <- append(result.RV, DM.Test(h_1 = estimates[,i],
188                                           h_2 = estimates[,j],
189                                           proxy = RV.test,
190                                           scoreFunction = QLIKE))
191   }
192 }
193 RV.QLIKE.array[, ,k] <- matrix(result.RV, byrow = TRUE, nrow = 5)
194 }# Ende der gesamten For-Schleife (Beginn Zeile 56)
195
196 #####
197 ##### Ermittlele Mittelwert der Testergebnisse
198 #####
199 #MSE mit realisierter Volatilität
200 RV.matrix.MSE <- RV.MSE.array[, ,1]
201 for (i in 2:n.loop) {
202   RV.matrix.MSE <- RV.matrix.MSE + RV.MSE.array[, ,i]
203 }
204 RV.matrix.MSE <- RV.matrix.MSE*(1/n.loop)
205
206 #QLIKE mit realisierter Volatilität
207 RV.matrix.QLIKE <- RV.QLIKE.array[, ,1]
208 for (i in 2:n.loop) {
209   RV.matrix.QLIKE <- RV.matrix.QLIKE + RV.QLIKE.array[, ,i]
210 }
211 RV.matrix.QLIKE <- RV.matrix.QLIKE*(1/n.loop)
212
213 #MSE mit quadratischen Renditen
214 R2.matrix.MSE <- R2.MSE.array[, ,1]
215 for (i in 2:n.loop) {
216   R2.matrix.MSE <- R2.matrix.MSE + R2.MSE.array[, ,i]
217 }
218 R2.matrix.MSE <- R2.matrix.MSE*(1/n.loop)
219
220 #QLIKE mit quadratischen Renditen
221 R2.matrix.QLIKE <- R2.QLIKE.array[, ,1]
222 for (i in 2:n.loop) {
223   R2.matrix.QLIKE <- R2.matrix.QLIKE + R2.QLIKE.array[, ,i]
224 }
225 R2.matrix.QLIKE <- R2.matrix.QLIKE*(1/n.loop)
226
227 #####
228 ##### Plots
229 #####
230 par(mfrow=c(3,2))
231 MatrixPlot(matrix = R2.matrix.QLIKE, eta = 0.05, proxy = "quad. Renditen")
232 mtext(bquote(h["2,t"]), side=3, line=3, cex = 0.9)
233 mtext(bquote(h["1,t"]), side=2, line=4.8, las=1, cex = 0.9)
234 MatrixPlot(matrix = RV.matrix.QLIKE, eta = 0.05, proxy = "real. Volatilität")
235 MatrixPlot(matrix = R2.matrix.QLIKE, eta = 0.1, proxy = "quad. Renditen")
236 MatrixPlot(matrix = RV.matrix.QLIKE, eta = 0.1, proxy = "real. Volatilität")
237 MatrixPlot(matrix = R2.matrix.QLIKE, eta = 0.005, proxy = "quad. Renditen")
238 MatrixPlot(matrix = RV.matrix.QLIKE, eta = 0.005, proxy = "real. Volatilität")

```

10 Literaturverzeichnis

- Andersen, T. G. und Benzoni, L. (2018). Realized Volatility. *Federal Reserve Bank of Chicago*.
- Brehmer, J. R. (2017). Elicitability and its Application in Risk Management. *University of Mannheim, School of Business Informatics and Mathematics*.
- Durrett, R. (2005). *Probability: Theory and Examples*, Band third edition. Duxbury advanced series. Thomson Brooks/Cole.
- Ehm, W., Gneiting, T., Jordan, A., und Krüger, F. (2016). Of quantiles and expectiles: Consistent scoring functions, choquet representations and forecast rankings. *Journal of the Royal Statistical Society, Series B*, 78:505–562.
- Fissler, T., Ziegel, J. F., und Gneiting, T. (2016). Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting. *Risk Magazine*, January 2016:58–61.
- Gneiting, T. (2009). Evaluating Point Forecasts. *Institut für angewandte Mathematik Universität Heidelberg*.
- Gneiting, T. und Ranjan, R. (2008). Comparing Density Forecasts Using Threshold and Quantile Weighted Scoring Rules. *Department of Statistics, University of Washington, Technical Report no. 533*.
- Holzmann, H. (2019). Scoring Funktionen und Parameterelicitierbarkeit, Skript Mathematische Statistik.
- Nolde, N. und Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The annals of applied statistics*, 11(4):1833–1874.
- NYSE (2021). Holidays & Trading Hours. Website. Online erhältlich unter <https://www.nyse.com/markets/hours-calendars>; abgerufen am 21. Februar 2021.
- Osband, K. H. (1985). Providing Incentives for Better Cost Forecasting. *PhD thesis, University of California, Berkeley*.
- Patton, A. J. (2006). Volatility Forecast Comparison Using Imperfect Volatility Proxy. *London School of Economics*.
- Steinwart, I., Pasin, C., Williamson, R., und Zhang, S. (2014). Elicitation and Identification of Properties. *JMLR: Workshop and Conference Proceedings*, 35:1–45.

Eidesstattliche Erklärung

Hiermit versichere ich, Karolin Heike Anna Frohnapfel, Matrikelnummer 3074985, dass ich diese Bachelorarbeit mit dem Titel „Robuste Verlustfunktionen und Evaluierung von Volatilitätsvorhersagen mit realisierten Volatilitäten“ eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

5.3.2021, Marburg

Ort, Datum



Karolin Frohnapfel