

Faktorenkörbe

Theresa Hausen

Seminararbeit im Bereich
Anwendung der Finanzmathematik

RWTH Aachen
Institut für Mathematik
Prof. Dr. Stanislaus Maier-Paape

Inhaltsverzeichnis

1	Einleitung	3
2	Aktienselektion nach Volatilität	4
3	Analyse durch POEM	6
3.1	Vorgehensweise	6
3.2	Ergebnisse über Veränderungen der Reihenfolge aller Quantile . .	8
4	Statistische Tests	11
5	Fazit	18
	Literaturverzeichnis	19
A	Quellcode aus R	20

Kapitel 1

Einleitung

Wer einen ersten Versuch der Investition in Aktien betreiben möchte und einen hilfreichen Tipp benötigt, dem könnte man die Theorie „volatility is one of the best predictors of bankruptcy“ (S.20 [Kon16]) aus Pim van Vliets Buch *High returns from low risk* empfehlen. Der Autor zeigt unter Anderem, dass Low-Vola Portfolios besser performen als Portfolios mit hoher Volatilität (vgl. S.36 [Kon16]).

Wäre es deshalb nicht von großem Nutzen, eine Aussage über zukünftige Volatilität von Aktienkörben basierend auf historischer Volatilität treffen zu können? Damit soll sich diese Arbeit beschäftigen.

Zu Beginn wird auf die Arbeit „*Momentum vs. Volatilität*“ von Platen und Haase [Pla19] eingegangen, die als Inspiration dieser Seminararbeit diente. Danach wird die historische Volatilität eines Marktes mithilfe von POEM, einem Programm zum Backtest von Portfolios (vgl. [Bre21]), und R analysiert, um schließlich die These, dass Volatilität prognostizierbar ist, zu prüfen.

Kapitel 2

Aktienselektion nach Volatilität

Da sich die folgenden Kapitel mit der statistischen Verifikation von Aussagen über Volatilität der Arbeit „*Momentum vs. Volatilität*“ von Haase und Platen beschäftigen [Pla19], wird der wesentliche Inhalt zu Beginn thematisiert.

Die von der Vereinigung technischer Analysten Deutschlands ausgezeichnete Arbeit untersucht die Prognostizierbarkeit von Momentum sowie Volatilität. Dazu wurden die 3.000 US-Aktien auf Basis des Russell 3000 Index sowie die 600 größten Aktien aus der Eurozone und Japan nach Marktkapitalisierung jeweils in 14 Gruppen gleicher Größe (300 Aktien) unterteilt. Für die 300 Gruppenmitglieder eines jeden Marktes wurden Ranglisten nach Momentum respektive nach Volatilität erstellt und Dezile zu 30 Aktien gebildet.

Das Ziel ist es, herauszufinden welche Aktienkörbe (30-er Dezile in den Gruppen) am Ende des Quartals immer noch die gleiche Anordnung nach Momentum bzw. Volatilität haben, wie zu Beginn. Dabei werden vor allem jene Aktienkörbe beobachtet, die das höchste Momentum oder die niedrigste Volatilität haben (diese werden als *Gewinner* bezeichnet) und jene Dezile, die ein niedriges Momentum und hohe Volatilität haben, bekannt als *Verlierer* (S.4f [Pla19]).

Die Ergebnisse der Aktienselektion in Bezug auf den ersten 300er-Markt (des Russell 3000 Index) nach Momentum ergeben, dass sich der Leitspruch „Verlierer bleiben Verlierer“ im Gegensatz zu „Gewinner bleiben Gewinner“ bewahrt. Da nur einer von 14 Märkten betrachtet wurde, wird die These auch für die restlichen Dezile untersucht. Das Ergebnis zeigt, dass die Performance-Daten der Gewinner in 67% aller Fälle hinter ihren Equal-Weight-Vergleichsindizes zurückbleiben (S.7 [Pla19]).

Da sich diese Seminararbeit ausschließlich auf die Sortierung der Aktienkörbe nach Volatilität bezieht, werden die Ergebnisse der Aktienselektion nach Volatilität von Haase und Platen kurz erläutert.

In der Zusammenfassung über alle Märkte wurde allgemein gezeigt, dass hochvolatile Aktien ein schlechteres Chance-Risiko-Verhältnis aufweisen, als die Verlierer (Aktien im letzten Momentum Dezil). Des Weiteren stellte man fest,

dass die Aktien des Low-Vola-Dezils über bessere Werte verfügten als die Gewinner, indem es eine viel höhere Überrendite zum jeweiligen Equal-Weight-Vergleichsindex erzielte als jedes Momentum-Dezil. Es wird von einer hohen Prognosequalität von historischer relativer Volatilität in Bezug auf zukünftige relative Volatilität ausgegangen (S.15 [Pla19]). Diese Annahme wird im weiteren Verlauf der Arbeit anhand von historischen Daten statistisch überprüft. Es besteht hohes Interesse daran, diese These zu verifizieren, da sich so verzichtbares Risiko im Anlegen von Kapital vermeiden ließe.

Kapitel 3

Analyse durch POEM

3.1 Vorgehensweise

Mithilfe von POEM werden die Daten aller Aktien des S&P 500 Index über einen Zeitraum von 20 Jahren ausgewertet (Beginn: 22.05.2000, Ende: 20.05.2020), indem jedes Equal Weight Portfolio aus den Aktien eines der gleich gewichteten Volatilitäts-Quantile erzeugt wird. Alle 230 Tage werden Aktienkörbe nach aufsteigender Volatilität erstellt, die insgesamt zwanzig Quantile gleicher Größe ergeben. Zum Zeitpunkt einer jeden Umschichtung befinden sich im ersten Quantil genau die 5% aller Aktien, welche geringste Volatilität vorweisen. Mit jedem Quantil steigt die Volatilität der Aktien, sodass sich die Aktien mit höchster Volatilität im letzten Quantil befinden. In der folgenden Abbildung sind die Equity-Kurven jedes Quantils zu sehen.

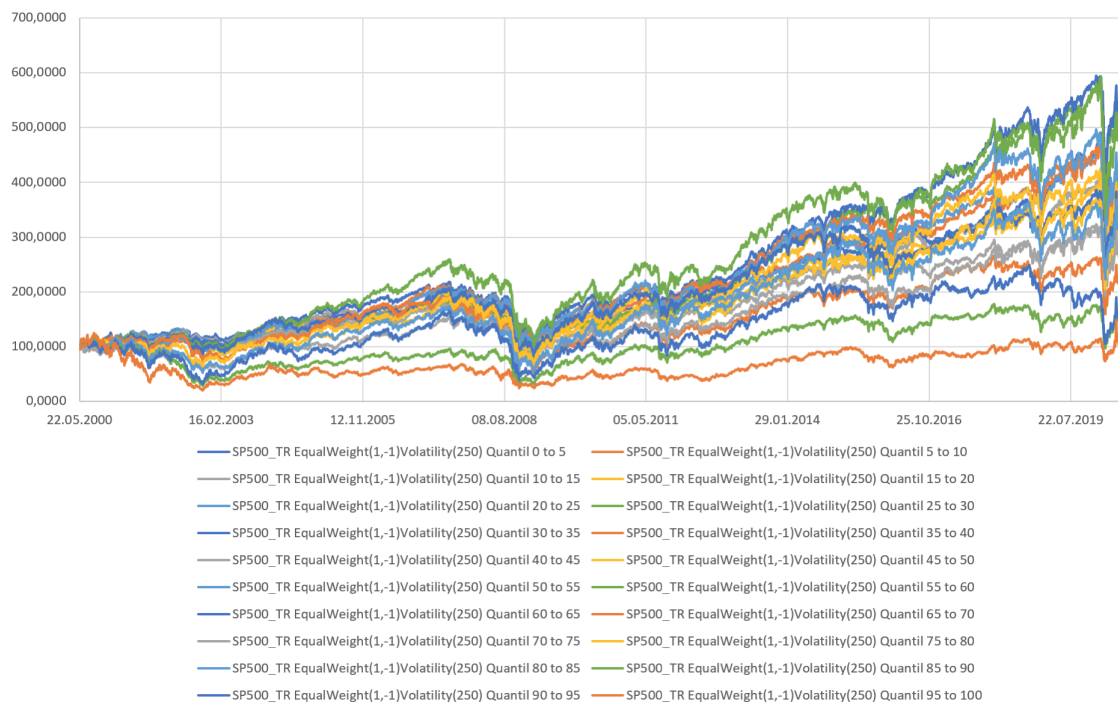


ABBILDUNG 3.1: Equity-Plot aller Quantile

Das Ziel dieser Arbeit ist es, zu zeigen, dass Aktienkörbe mit geringer Volatilität gegenüber den restlichen Aktienkörben, diese über einen bestimmten Zeitraum beibehalten. Weiter soll gezeigt werden, dass sich für risikoreiche Aktien mit vergleichsweise hoher Volatilität jene Schwankung in Zukunft fortsetzen wird.

Um diese Aussage mithilfe von verschiedenen z-Tests zu verifizieren, werden die Quantile jeder Optimierung benötigt. Denn es soll gezeigt werden, dass die Quantile mit minimaler (bzw. maximaler) Volatilität zum Zeitpunkt einer beliebigen Optimierung auch am Ende der Periode (nach 230 Tagen) im Outsample noch immer die Quantile mit minimaler (bzw. maximaler) Volatilität sind. Das bedeutet, wenn die Reihenfolge der Quantile im Outsample der natürlichen Reihenfolge der Quantile entspricht, hat sich das Verhältnis der Volatilitäten aller Aktien untereinander nicht verändert („Low/High Vola bleibt Low/High Vola“).

Definition 3.1. Von nun an wird für jede Optimierung $n \in O := \{1, 2, \dots, 21\}$ ein Tupel aus der Symmetrischen Gruppe mit 20 Elementen

$$\begin{aligned} i_n &\in S_{20}, \text{ für die Reihenfolge der Quantile des Insamples und} \\ o_n &\in S_{20}, \text{ für die Reihenfolge der Quantile des Outsamples} \end{aligned}$$

betrachtet, wobei $n = 1$ den Start-Zeitpunkt der ersten Optimierung der Daten beschreibt, für welchen keine historischen Messdaten existieren und die Volatilität für alle Quantile noch bei 0 liegt. Sind die Quantile im Insample der Periode $n \in O$ in der natürlichen Reihenfolge angeordnet, gilt

$$i_n = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20) =: id,$$

wobei dieses Ereignis auch zum Outsample-Zeitpunkt nicht ausgeschlossen ist.

Man mag sich nun die Frage stellen, wie es möglich ist, dass sich die Permutationen des Insamples i_n mit $n \in O$ von der natürlichen Reihenfolge unterscheiden, da die von POEM vorgenommene Optimierung genau auf dem Merkmal beruht Quantile nach aufsteigender Volatilität zu bilden.

Im Optimierungs-Prozess wird jede Aktie bezüglich ihrer Varianz einzeln betrachtet und anhand dieser Größe in die Aktienkörbe sortiert. Jedoch existiert zwischen den Aktien aufgrund der gleichen Gewichtung eine Kovarianz. Wenn dann die Summe der Aktien im Quantil betrachtet wird, verändert sich somit die Varianz des Aktienkorbs, die damit das auftretende Phänomen einer (möglicherweise) abgeänderten Reihenfolge im Insample erklärt.

Es bietet sich nun die Möglichkeit drei Vergleiche zu unternehmen, indem entweder

1. die Reihenfolge der Quantile des Insamples mit der Reihenfolge im Outsample (i_n versus o_n für $n \in O$) oder

2. die Reihenfolge der Quantile des Insamples mit der natürlichen Reihenfolge (i_n versus id für $n \in O$) oder
3. die Reihenfolge der Quantile des Outsamples mit der natürlichen Reihenfolge (o_n versus id für $n \in O$) gegenübergestellt wird.

Anhand der Daten, die mit Hilfe von POEM berechnet wurden, ist es im Statistik-Programm R möglich, die obigen drei Verfahren über die Reihenfolge der Quantile in i_n und o_n anzuwenden und auszuwerten. Der dafür notwendige Code ist im Anhang zu finden.

3.2 Ergebnisse über Veränderungen der Reihenfolge aller Quantile

Das Vorgehen i_n und o_n für $n \in O$ mit der natürlichen Reihenfolge der Quantile id zu vergleichen wird durch die Berechnung eines Differenzvektors

$$d_{id,i}^{(n)} := id - i_n \quad \text{bzw.} \quad d_{id,o}^{(n)} := id - o_n$$

veranschaulicht. Dieser zeigt, ob das k -te Quantil im Insample bzw. Outsample der n -ten Optimierung vor dem k -ten Quantil der natürlichen Reihenfolge steht ($d_{id,i}^{(n)}(k) > 0$ bzw. $d_{id,o}^{(n)}(k) > 0$), nach dem k -ten Quantil der natürlichen Reihenfolge folgt ($d_{id,i}^{(n)}(k) < 0$ bzw. $d_{id,o}^{(n)}(k) < 0$) oder an der „richtigen“ Stelle, also an k -ter Stelle, steht ($d_{id,i}^{(n)}(k) = 0$ bzw. $d_{id,o}^{(n)}(k) = 0$). Somit gilt für die Differenzvektoren

$$d_{id,i}^{(n)}, d_{id,o}^{(n)} \in \{-19, -18, \dots, -1, 0, 1, \dots, 18, 19\}^{20}$$

für $n \in O$. Je mehr Einträge im Differenzvektor zum Outsample Zeitpunkt gleich 0 sind, desto näher ist man der Annahme, dass sich Volatilität prognostizieren lässt. Demnach erhofft man sich besonders für die Einträge in $d_{id,o}^{(n)}$ betragsmäßig kleine Werte.

Abbildung 3.2 zeigt beispielhaft die Veränderungen in der Reihenfolge des Outsamples am Ende des betrachteten Zeitraums im Vergleich zur (erwünschten) natürlichen Reihenfolge. Der Differenzvektor zum Schluss der letzten Optimierung nach zwanzig Jahren $d_{id,o}^{(21)}$ wird graphisch dargestellt.

Anhand nicht vorhandener Balken zeichnet sich ab, dass die Quantile mit niedriger Volatilität (0% – 10%) und die Quantile mit höchster Volatilität (80% – 100%) und ein Quantil in mittleren Bereich (55% – 60%) in der natürlichen Anordnung liegen.

Die rot gefärbten Balken zeigen welche Quantile in der Reihenfolge des Outsamples hinter ihrem Platz in der natürlichen Reihenfolge liegen, wobei sich anhand der x-Achse, die den Wertebereich des betrachteten Differenzvektors angibt,

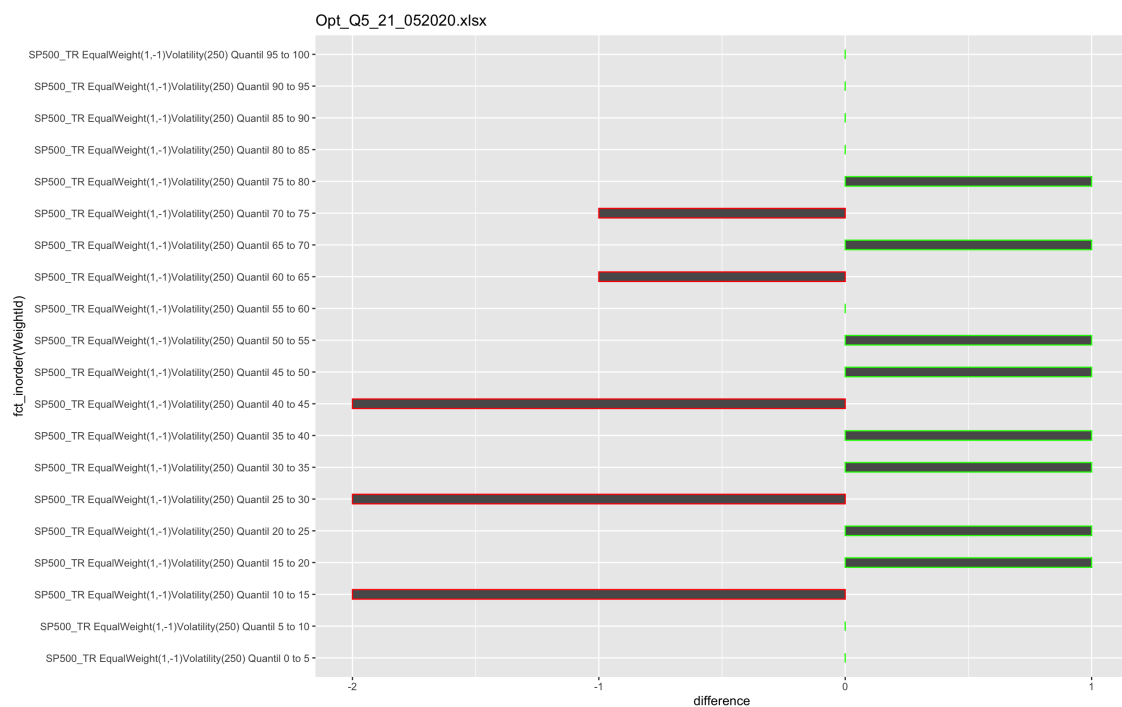


ABBILDUNG 3.2: Differenzvektor nach der letzten Optimierung

ablesen lässt, dass diese um maximal zwei Plätze verschoben sind. Analog dazu folgt die Interpretation der Quantile mit grün eingefärbten Balken, diese befinden sich hier maximal einen Stelle vor ihrer natürlichen Anordnung.

Um einen Eindruck von den gesamten Daten der Optimierungen über die betrachteten zwanzig Jahre zu bekommen, ist es von Interesse den Boxplot, der sich aus dem obigen und den weiteren 20 Diagrammen zusammensetzt, zu betrachten (vgl. Abb. 3.3).

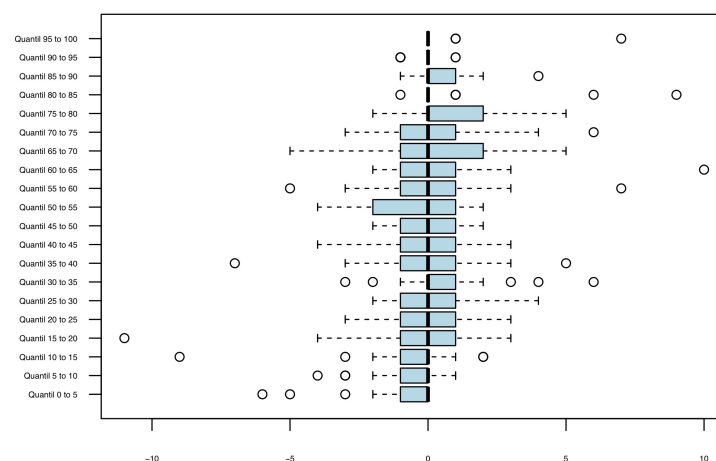


ABBILDUNG 3.3: Boxplot über Differenzen aller Optimierungen

Anhand der Veranschaulichung aller Differenzen ist zu erkennen, dass sich die

Quantile im Median in den meisten Fällen um nur einen Platz gegenüber der natürlichen Reihenfolge verschieben. Besonders die Quantile die „Low-“ bzw. „High-Vola“-Aktienkörbe repräsentieren, liegen mit dem Mittelwert der Entfernung zur „richtigen“ Stelle unter dem arithmetischem Mittel über alle Quantile, welches durch

$$\frac{1}{21} \sum_{n=1}^{21} \frac{1}{20} \sum_{k=1}^{20} |d_{id,o}^{(n)}(k)| = 1.1714$$

gegeben ist. Ohne die Ausreißer wäre der Unterschied zu den Quantilen in der Mitte sogar noch deutlicher, was in Abbildung 3.3 an den längeren gestrichelten Linien in diesem Bereich abgelesen werden kann. Das heißt für alle Jahre des betrachteten Zeitraums, also 230 Tage nach der vorangegangenen Optimierung, unterscheidet sich jedes Quantil im Durchschnitt an weniger als zwei Stellen von dem zugewiesenen Platz in der natürlichen Reihenfolge im Outsample.

Das arithmetischem Mittel \bar{d} über die betragsmäßigen Differenzen der Quantile im Outsample über alle Optimierungen $n \in O$ ist gegeben durch

$$\bar{d} = \frac{1}{21} \sum_{n=1}^{21} (|d_{id,o}(1)|, |d_{id,o}(2)|, \dots, |d_{id,o}(20)|)^{(n)} \\ \approx (0.86, 0.90, 1.19, 1.70, 1.29, 1.10, 1.24, 1.71, 1.33, 0.86, 1.43, 1.71, 1.52, 1.81, 1.33, 1.24, 0.90, 0.62, 0.29, 0.48).$$

Aufgrund der niedrigen Werte für die ersten und letzten drei Einträge in \bar{d} , sollten die Quantile an diesen Stellen genauer studiert werden. Diese Erkenntnis ist besonders von Vorteil, da jene Quantile genau die „Low-“ bzw. „High-Vola“-Aktienkörbe repräsentieren.

Statt die Differenzen an den zwanzig Stellen der Quantile zu betrachten, hat man die Option, die Zahlen an einer bestimmten Stelle $k \in O$ des Vektors $o_n(k)$ zu betrachten. Da die Quantile mit kleinster und größter Volatilität aus genannten Gründen von besonderem Interesse sind, würde es sich anbieten, $o_n(1)$ und $o_n(20)$ für alle $n \in O$ zu betrachten. Diese Quantile werden im folgenden Kapitel näher untersucht. Ferner werden z-Tests auf Basis verschiedener Annahmen durchgeführt, um die Hauptthese dieser Arbeit zu belegen.

Kapitel 4

Statistische Tests

Im letzten Teil der Arbeit wird anhand des Datensatzes, welcher bereits aus Kapitel 3 bekannt ist, die Anordnung der Quantile im Outsample nach jeder Optimierung im betrachteten Zeitraum von zwanzig Jahren analysiert. Ab jetzt beziehen sich alle Werte auf die Stellen der Quantile in o_n , mit $n \in O = \{1, \dots, 21\}$ und nicht mehr auf deren Differenzen zur natürlichen Reihenfolge in $d_{id,o}^{(n)}$.

Um ein allgemeines Bild über die Anordnungen der Quantile am Ende jeder Optimierungsphase zu bekommen, wird zunächst der Mittelwert über alle o_n betrachtet:

$$\bar{o} = \frac{1}{21} \sum_{n=1}^{21} o_n$$

$$\approx (1.9, 2.7, 3.5, 4.7, 5.1, 5.8, 6.4, 8.1, 9.1, 10.2, 11.6, 12.1, 12.6, 13.7, 14.7, 15.4, 16.2, 17.6, 19.1, 19.5).$$

An fast jeder Stelle steht die natürliche Zahl vor dem Komma, die der Nummerierung der Stellen entspricht. Dies kann so interpretiert werden, dass das arithmetische Mittel über die o_n sehr nah an der natürlichen Reihenfolge der Quantile liegt. Das Ziel ist es, diese 'Nähe' durch eine statistische Größe zu quantifizieren.

Zu Beginn wird sich nur auf das Quantil mit kleinster Volatilität im Outsample jeder Optimierung n , also $o_n(1)$ konzentriert. Dazu wird folgende Zufallsvariable definiert.

Definition 4.1. Sei $\Omega := S_{20}$ die Menge der möglichen Ereignisse, dann ist die Zufallsvariable X , die auf die erste Stelle jeder Folge in S_{20} abbildet, definiert durch

$$X : \Omega \rightarrow \{1, 2, \dots, 20\}, \pi \mapsto \pi(1).$$

Die Wahrscheinlichkeitsverteilung von X ist

$$\begin{aligned} P^X(X = t) &= P(\{\pi \in S_{20} : X(\pi) = \pi(1) = t\}) \\ &= \frac{|\{\pi \in S_{20} : X(\pi) = t\}|}{|S_{20}|} \\ &= \frac{19!}{20!} \\ &= \frac{1}{20} \end{aligned}$$

für ein beliebiges $t \in \{1, 2, \dots, 20\}$.

Das arithmetische Mittel über alle Elemente in Ω ist

$$\frac{1}{20!} \sum_{\pi \in S_{20}} \pi = (10.5, 10.5, \dots, 10.5)$$

und der Erwartungswert von X ist somit $\mu_X := EX = 10.5$. Es folgt $P(X \leq t) = \lfloor t \rfloor \cdot \frac{1}{20}$ für $t \in [1, 20]$. Die Wahrscheinlichkeit, dass die erste Stelle einer zufällig verteilten Folge $\pi \in S_{20}$ kleiner als \bar{o}_1 ist, liegt bei

$$P(X \leq \bar{o}(1)) = \lfloor 1.9 \rfloor \cdot \frac{1}{20} = 5\%.$$

Das bedeutet, dass sich der empirische Erwartungswert $\bar{o}(1)$ stark vom Erwartungswert μ_X der Zufallsvariable X unterscheidet. Um eine allgemeine Aussage über die Quantile, die sich auf dem ersten Platz befinden, mithilfe der gegebenen Daten zu erhalten wird die Nullhypothese, dass der empirische Mittelwert $\bar{o}(1)$ größer oder gleich dem erwarteten Mittelwert μ_X ist, aufgestellt:

$$H_0 : \mu_X \leq \bar{o}(1) \quad \longleftrightarrow \quad H_1 : \mu_X > \bar{o}(1).$$

Dem gegenüber steht die Hypothese dass $\bar{o}(1)$ kleiner als μ_X ist. Eine Methode die Nullhypothese zu testen ist der z-Test, auch bekannt unter dem Namen Gauß-Test. Um diesen anwenden zu können, sollten die Beobachtungen unabhängige Zufallsstichproben aus Normalverteilungen mit derselben Varianz und gleichem Erwartungswert der Grundgesamtheit sein. Bei den vorliegenden empirischen Daten handelt es sich um eine Summe gleichverteilter Zufallsstichproben.

Der Zentrale Grenzwertsatz, nach Lindeberg und Lévy besagt, dass eine Summe identisch verteilter Zufallsvariablen X_1, \dots, X_n mit $EX_i = \mu$ und $Var(X_i) = \sigma^2$, für alle i , approximativ normalverteilt ist (vgl. S.316f [Fah06]). Der zentrale Grenzwertsatz gilt in noch wesentlich allgemeineren Varianten, wobei die Zufallsvariablen X_1, \dots, X_n abhängig und verschieden verteilt sein dürfen. Entscheidend ist, dass keine der Zufallsvariablen X_i die restlichen deutlich dominiert (vgl. S.318 [Fah06]). Durch Anwenden des Zentralen Grenzwertsatzes erfüllen die empirischen Messdaten alle Voraussetzungen für den z-Test, welcher

im Folgenden für ein Signifikanzniveau von $\alpha = 5\%$ angewendet wird. Um die obige Nullhypothese anhand der ersten Zeile der Outsample Matrix (hier durch $x1$ repräsentiert) zu testen, wird für den Gauß-Test außerdem der Erwartungswert $\mu_X = 10.5$ und die Standardabweichung $\sigma_X = 5.76$ der Grundgesamtheit benötigt.

One Sample z-test

```
data: x1
z = -6.8686, n = 21.0000, Std. Dev. = 5.7663, Std. Dev. of the sample mean = 1.2583,
p-value = 3.241e-12
alternative hypothesis: true mean is less than 10.5
95 percent confidence interval:
  -Inf 3.926872
sample estimates:
mean of x1
  1.857143
```

Der einseitige z-Test verwirft die Nullhypothese, mit einem sehr kleinen p-Wert von $3.241 \cdot 10^{-12}$. Die Eigenschaft, dass Aktien des S&P 500 Index, die in der Vergangenheit eine relativ niedrige Volatilität vorweisen und diese auch in der Zukunft beibehalten, zeigt eine statistische Signifikanz auf.

Um eine allgemeinere Aussage über eine größere Menge von Aktien mit geringer Volatilität treffen zu können, wird nun nicht nur die erste Stelle der Outsample-Vektoren o_n , sondern die ersten drei Stellen betrachtet. Dazu wird die Zufallsvariable Y benötigt.

Definition 4.2. Sei $\Omega := S_{20}$ die Menge der möglichen Ereignisse, dann ist die Zufallsvariable Y , die auf die Summe der ersten drei Stellen jeder Folge in S_{20} abbildet, definiert durch

$$Y : \Omega \rightarrow \{1, 2, \dots, 20\} : \pi \mapsto \pi(1) + \pi(2) + \pi(3).$$

Der empirische Erwartungswert der Summe der Quantile auf den ersten drei Plätzen ist gegeben durch

$$\overline{o(1) + o(2) + o(3)} = \frac{1}{21} \sum_{n=1}^{21} o_n(1) + o_n(2) + o_n(3) = 8.095$$

und der theoretische Erwartungswert lässt sich anhand von μ_X bestimmen, da gilt

$$\mu_Y = 31.5 = 3 \cdot \mu_X.$$

Das Ergebnis für μ_Y kann mit Hilfe des Codes, welcher im Anhang zu finden ist, verifiziert werden. Des Weiteren folgt damit der Wert $\sigma_Y = 9.45$ der Standardabweichung von Y . Somit sind alle notwendigen Informationen gegeben, um einem weiteren z-Test anwenden zu können, der die folgende Nullhypothese prüft:

$$H_0 : \mu_Y \leq \overline{o(1) + o(2) + o(3)} \quad \longleftrightarrow \quad H_1 : \mu_Y > \overline{o(1) + o(2) + o(3)}.$$

One-sample z-Test

```

data: x
z = -11.353, p-value < 2.2e-16
alternative hypothesis: true mean is less than 31.5
95 percent confidence interval:
    NA 11.48619
sample estimates:
mean of x
  8.095238

```

Analog zum ersten Test wird auch hier die Nullhypothese durch den sehr kleinen p-Wert verworfen. Es darf somit die These H_1 angenommen werden, welche die Behauptung „Low Volatility bleibt Low Volatility“ ([Pla19]) bestätigt.

Da sich die kombinatorische Struktur der Zufallsvariablen X und Z nicht verändert, wenn auf andere Stellen des Vektors $\pi \in S_{20}$ abgebildet wird, kann von gleichem Erwartungswert μ_X bzw. μ_Y und der selben Standardabweichung σ_X bzw. σ_Y ausgegangen werden. Betrachtet man die Quantile mit höchster Volatilität, wählt man die Zufallsvariable X_{high} so, dass auf $\pi(20)$ abgebildet wird. Analog ist das Bild von Y_{high} gegeben durch $\pi(18) + \pi(19) + \pi(20)$, da man eine Aussage über die Quantile auf den letzten Stellen treffen möchte. Die empirischen Erwartungswerte der Zufallsvariablen X_{high} und Y_{high} lassen sich wie $\bar{o}(1)$ und $\bar{o}(1) + \bar{o}(2) + \bar{o}(3)$ berechnen und man erhält

$$\bar{o}(20) = 19.52 \quad \text{und} \quad \overline{o(18) + o(19) + o(20)} = 56.19.$$

Um, zu testen, ob eine statistische Signifikanz in Bezug auf den Erwartungswert der Messdaten vorliegt, wird hier mit einem rechtsseitigem z-Test geprüft. Der Grund dafür ist, dass die empirischen Erwartungswerte viel größer sind als die der Zufallsvariablen X_{high} und Y_{high} . Der entsprechende Code der Gauß-Tests ist im Anhang zu finden. Da die p-Werte bei $2.2 \cdot 10^{-16}$ liegen und damit denen des obigen linksseitigen z-Tests stark ähneln, wird an diesem Punkt nur betont, dass der vorliegende Effekt der Quantile mit geringer Volatilität auch bei jenen mit hoher Volatilität vorliegt. Durch kompilieren des Codes in R sind die Resultate der z-Test vollständig einsehbar. Es gilt auch hier „High Volatility bleibt High Volatility“.

Die obige Analyse bezieht sich nur auf die Aufteilung aller Aktien in zwanzig Aktienkörbe gleicher Größe. Um eine weitere statistische Überprüfung der Aussage „Low Volatility bleibt Low Volatility und High Volatility bleibt High Volatility“ durchführen zu können, werden die Aktien des S&P 500 Index im Folgenden in nur zehn Quantile aufgeteilt. Die Quantile enthalten nun die doppelte Zahl an Aktien (10% Körbe) im Vergleich zu den bisher betrachteten Quantilen. Um ein allgemeines Bild von der Differenz zwischen den Stellen im Outsample und der natürlichen Reihenfolge der Optimierungen im selben betrachteten Zeitraum von zwanzig Jahren zu bekommen, ist der folgende Boxplot in Abbildung 4.1 gegeben.

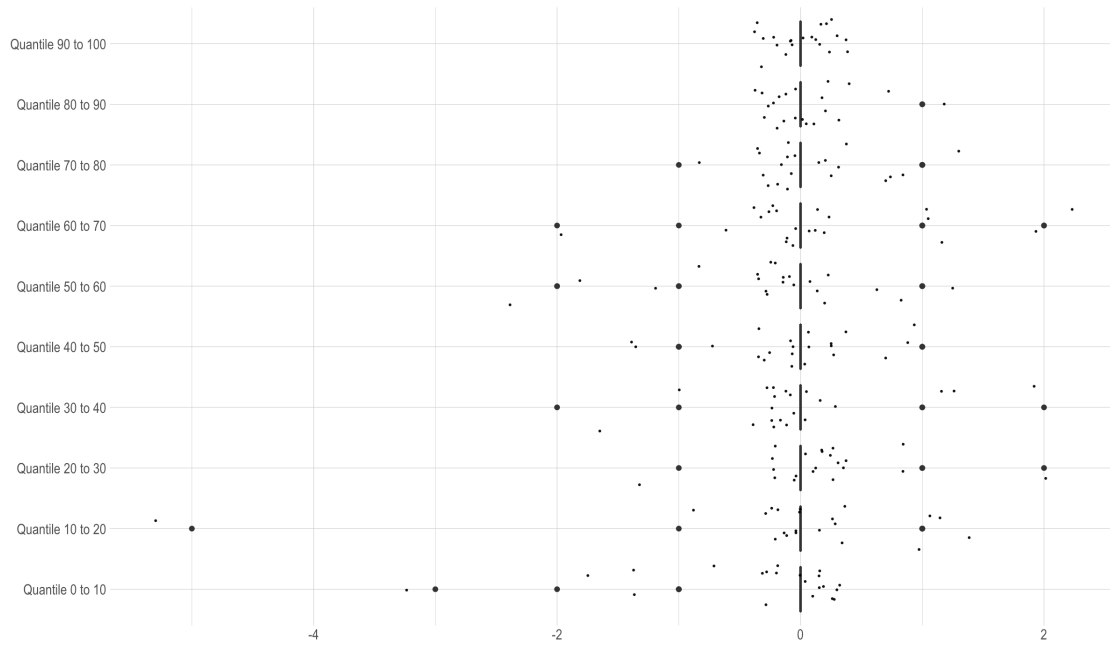


ABBILDUNG 4.1: Boxplot der 10 Quantile über alle Optimierungen

Der Mittelwert der Differenzen ist für jedes Quantil so nah an Null angesiedelt, dass keine Balken zu sehen sind. Dies ist auch daran zu erkennen, dass die Ansammlung der Punkte um die 0 herum liegt. Das bedeutet, dass die Anordnung der Quantile des Outsamples sehr ähnlich zu der natürlichen Reihenfolge ist. Deshalb soll auch hier getestet werden, ob der gleiche Effekt in der Anordnung der Quantile vorliegt, wie bei der Unterteilung des Marktes in zwanzig Quantile. Um das bereits bekannte Verfahren anwenden zu können, wird zunächst die folgende Zufallsvariable definiert.

Definition 4.3. Sei hier $\Omega := S_{10}$ die Menge der möglichen Ereignisse, dann ist die Zufallsvariable Z , die auf die Summe der ersten zwei Stellen jeder Folge in S_{10} abbildet, definiert durch

$$Z : \Omega \rightarrow \{3, 4, \dots, 19\} : \pi \mapsto \pi(1) + \pi(2).$$

Die Zufallsvariable Z ist wie folgt verteilt.

Satz 4.4. Die zugehörige Verteilungsfunktion von Z ist

$$P^Z(Z = t) = \begin{cases} 2 \cdot \lfloor \frac{t-1}{2} \rfloor \cdot \frac{1}{90} & , \text{für } t < 11 \\ 2 \cdot \lfloor \frac{19-t+2}{2} \rfloor \cdot \frac{1}{90} & , \text{sonst} \end{cases}$$

für $t \in \{3, 4, \dots, 19\}$ und es folgt $\mu_Z = 11$.

Beweis. Da nur zwei Einträge des Vektors π betrachtet werden, ergeben sich für die restlichen 8 Stellen 8! Permutations-Möglichkeiten. Es gilt somit

$$|\{\pi \in S_{10} : Z(\pi) = t\}| = |\{i \in \{1, \dots, 10\} : i + j = t, i \neq j, j \in \{1, \dots, 10\}\}| \cdot 8!$$

Für jedes $t \in \{3, 4, \dots, 10\}$ gibt es genau $2 \cdot \lfloor \frac{t-1}{2} \rfloor$ Kombinations-Möglichkeiten, dieses durch einer Summe von zwei verschiedenen natürlichen Zahlen darzustellen. Um eine Summe gleicher Struktur für die restlichen $t \in \{11, \dots, 19\}$ zu konstruieren, sind genau $2 \cdot \lfloor \frac{19-t+2}{2} \rfloor$ Kombinations-Möglichkeiten gegeben. Damit folgt

$$\begin{aligned}
 P^Z(Z = t) &= P(\{\pi \in S_{10} : Z(\pi) = \pi(1) + \pi(2) = t\}) \\
 &= \frac{|\{\pi \in S_{10} : Z(\pi) = t\}|}{|S_{10}|} \\
 &= |\{i \in \{1, \dots, 10\} : i + j = t, i \neq j, j \in \{1, \dots, 10\}\}| \cdot \frac{8!}{10!} \\
 &= \begin{cases} 2 \cdot \lfloor \frac{t-1}{2} \rfloor \cdot \frac{1}{90} & , \text{für } t < 11 \\ 2 \cdot \lfloor \frac{19-t+2}{2} \rfloor \cdot \frac{1}{90} & , \text{sonst.} \end{cases}
 \end{aligned}$$

□

Auch hier liegt der theoretische Erwartungswert weit über dem empirischen Erwartungswert $\overline{o(1) + o(2)} = 3.48$, man beachte, dass $o_n \in S_{10}$ nun nur noch zehn Einträge hat. Es wird die gleiche Nullhypothese wie für den Erwartungswert der Zufallsvariable Y aufgestellt und durch den linksseitigen z-Test geprüft. Auch hier kann die Nullhypothese anhand des niedrigen p-Wertes $2.2 \cdot 10^{-16}$ verworfen werden. Gleiches gilt für die Beobachtung der letzten zwei Stellen des Vektors in S_{10} . Um eine Repetition des Inhalts zu vermeiden, können die genaueren Ergebnisse der genannten z-Tests anhand des Codes im Anhang nachvollzogen werden.

Es ist somit gezeigt worden, dass für relativ geringe bzw. hoch volatile Aktienkörbe eine statistische Signifikanz vorliegt, die dazu führt, dass diese Volatilität auch in Zukunft beibehalten wird. Da bisher nur die Quantile mit Randwerten der gemessenen Volatilitäten betrachtet wurden, wird zum Schluss der Arbeit ein Blick auf die drei Quantile, deren Volatilität sich genau in der Mitte der zwanzig Aktienkörbe befindet, geworfen. Die Anzahl der Optimierungen und der betrachtete Zeitraum ist dazu immer noch der selbe.

Die dafür notwendige Zufallsvariable, die auf die Summe $\pi(9) + \pi(10) + \pi(11)$ abbildet, steht auf Grund von Symmetrie der Symmetrischen Gruppe S_{20} in Bijektion zu der in 4.2 definierten Zufallsvariable Y . Das bedeutet, dass der Erwartungswert und die Standardabweichung durch μ_Y und σ_Y gegeben sind. Der empirische Mittelwert der Outsample-Vektoren $o_n \in S_{20}$ mit $n \in O$ ist

$$\overline{o(9) + o(10) + o(11)} = 30.85.$$

Dieser ist dem theoretischen Mittelwert $\mu_Z = 31.5$ sehr ähnlich. Es ist eine interessante Überlegung, wie nun das Ergebnis eines zweiseitigen Hypothesentests

in Bezug auf die Nullhypothese

$$H_0 : \mu_Z = \overline{o(9) + o(10) + o(11)} \longleftrightarrow H_1 : \mu_Z \neq \overline{o(9) + o(10) + o(11)}.$$

aussehen würde. Ein beidseitiger z-Test in R liefert folgendes Ergebnis:

One-sample z-Test

```
data: x3.2
z = -0.31183, p-value = 0.7552
alternative hypothesis: true mean is not equal to 31.5
95 percent confidence interval:
 26.81657 34.89771
sample estimates:
mean of x
 30.85714
```

Eine Interpretation des vergleichsweise hohen p-Wertes, der bei 75.52% liegt, ergibt, dass H_0 mit einer Wahrscheinlichkeit von 75.52% eingehalten wird. Das bedeutet, für die Quantile, die eine durchschnittliche historische Volatilität haben, wird sich die gemessene Volatilität im Vergleich zu den anderen Quantilen in Zukunft auch in der Mitte bewegen.

Kapitel 5

Fazit

Zusammenfassend konnte die Prognosequalität von historischer relativer Volatilität in Bezug auf zukünftige relative Volatilität gezeigt werden. Obwohl die Analyse nur auf einem Markt basiert, lässt sich der vorliegende Effekt mit gleicher Methode auch für andere Märkte vorweisen.

Die These „a stock that displays a low degree of volatility will probably also remain a low-volatility stock in the future. The same applies to high-volatility stocks.“ (S. 20 [Kon16]), die Pim van Vliet aufstellte, konnte somit verifiziert werden. Darüber hinaus wurde auch gezeigt, dass man bei Aktien mit durchschnittlicher Volatilität davon ausgehen kann, dass diese in der Zukunft beibehalten wird.

Literaturverzeichnis

- [Bre21] René Brenner. *Consistency in Portfolio Optimization: A new Approach using the General Framework of Portfolio Theory*. unveröffentlichte Dissertation. RWTH Aachen University, 2021.
- [Fah06] Iris Pigeot; Gerhard Tutz; Ludwig Fahrmeir. *Statistik: Der Weg zur Datenanalyse*. Fifth. Springer, 2006, S. 614.
- [Kon16] Jan de Koning; Pim van Fliet. *High returns from low risk*. Wiley, 16. Dezember 2016, S. 164.
- [Pla19] Daniel Haase; Dr. Andreas Platen. *Momentum vs. Volatilität*. 2019, S. 43.

Anhang A

Quellcode aus R

Dieses Kapitel beinhaltet den Code, der in R von mir geschrieben wurde um die Analyse an den durch POEM erhaltenen Daten durchzuführen. Dieser wurde auf den Datensatz angewendet, der S&P 500 Index in zwanzig Quantile unterteilt, um die gleichen Funktionen für die Messungen mit zehn Quantilen durchführen zu können, wurde der gleiche Code nur mit anderen Dimensionen der Matrizen verwendet.

```
#####
# Vergleiche natuerliche Reihenfolge mit outsample #
#####

for(k in 1:1){
  #Daten von Optimierungen aus POEM einlesen
  files <- list.files(pattern = "\\\\.xlsx$")

  #restult matrix und outsample matrix definieren
  res_out <- matrix(rep(0,420), byrow = FALSE, nrow = 20)
  place_out <- matrix(rep(0,420), byrow = FALSE, nrow = 20)

  data_in <- read_excel("Opt_Q5_02_05211.xlsx" , na="NA",
                        sheet="Results_InSample", col_names = TRUE)
  res_mtrx_out <- data_in[,1]
  res_mtrx_out <- res_mtrx_out[-21,]

  num <- c(0:20)
}
for (m in 1:length(files)){

  #daten des outsamples einer Optimization einlesen
  data_out<- read_excel(files[m] , na="NA",
                        sheet="Results_OutSample", col_names = TRUE)

  #Spalten mit Name und Volatility herausfiltern
  vola_out <- select(data_out, WeightId, Volatility)
```

```

#Zeilen des Outsample sheets nach
# aufsteigender Volatilitaet sortieren
vola_out_sort <- vola_out[with(vola_out, order(Volatility)),]
vola_out_sorted <- bind_cols(vola_out_sort, "place_out" = num)
vola_out_sorted <- vola_out_sorted[-1,]

for(t in 1:20){
  for(i in 1:20){
    if(res_mtrx_out[t,1] == vola_out_sorted[i,1]){
      {res_out[t,m] <- t-i
        # Differenz zw. natuerliche Reihenfolge & outsample
        place_out[t,m] <- i }
    }
  }
}

all_out <- bind_cols( res_mtrx_out[,1], "place_in" = place_out[,m],
                     "difference" = res_out[,m])

# res_mtrx_out ist die Vereinigung aller "all" Matrizen
# und stellt das Resultat am Ende dar
res_mtrx_out <- full_join(res_mtrx_out, all_out , by = "WeightId")

#Matrix mit allen Informationen fuer jede Optimierung ausgeben
print(files[m])
print(all_out)

# Balkendiagramm ueber Differenzen aller Quantile pro Optimierung
print(ggplot(all_out, aes(x = difference, y = fct_inorder(WeightId))) +
      geom_bar( col = ifelse(res_out[,m] < 0 , "red", "green"),
               stat = "identity", width=0.3) +
      ggtitle(label = files[m]))
}

#Boxplot ber alle Optimierungen
rownames(res_out) <- t(res_mtrx_out[,1])
boxplot.matrix(res_out , use.cols = FALSE,
               col = "lightblue", horizontal = TRUE ,
               las=1, cex.axis=0.45,
               xlab = "difference_of_id_and_outsample")

#Durchschnitt aller Differenzen
row_sum_out <-matrix( rowSums(abs(res_out[,c(1:21)])) , na.rm=TRUE) ,
                    20 , byrow = TRUE)
av_row_sum_out <- (1/21)*row_sum_out
mean(abs(av_row_sum_out))

```

```
#####
#           Distribution functions           #
#####

#  $P(X \leq k)$  mit  $X: S_n \rightarrow \{3, \dots, n+(n-1)\}$ ,  $pi \rightarrow pi(i) + pi(j)$  ,
# Summe von 2 Stellen
# In der Arbeit werden nur aufeinanderfolgende Stellen am
# Anfang/Ende des Vektors betrachtet
distr_1_2 <- function(k,n){
  l <- 0
  if(k < 3 | k > 2*n-1){ print("k_muss_zwischen_3_und_2n-1_liegen")}
  else{
    k <- floor(k)
    for (i in k:k){
      for(j in 1:n){
        for(h in 1:n){
          if(j != h){
            if( j + h == i){l <- l + 1 }
          }
        }
      }
    }
    value <- l*(factorial(n-2))/(factorial(n))
    print(l)
    print(value)
  }
}

#  $P(X = k)$ 
distr_12 <- function(k,n){
  l <- 0
  for(h in 1:n){
    for(t in 1:n){
      if( t!= h){
        if( h + t == k){l <- l + 1 }
      }
    }
  }

  value2 <- l*(factorial(n-2))/(factorial(n))
  # print(l)
  # print(value2)
  return(value2*k)
}

#EZ = sum k * P(X = k) f r n=20
E2 = 0
```

```

for (k in c(3:57)) {
  E2 = E2 + distr_12(k,20)
}
print(E2)

```

```

# Standardabweichung von X f r n=20

```

```

e.2 = 0

```

```

for (k in c(3:57)) {
  e.2 = e.2 + k*distr_12(k,20)
}

```

```

sd.x = sqrt(e.2 - E2^2)

```

```

sd.x

```

```

#####

```

```

#           3 Stellen betrachten           #

```

```

#####

```

```

# P(Z <= k ) mit Z: S_n -> {6,...,n+ (n-1)+(n-2)}, pi ->

```

```

      pi(h) + pi(i) + pi(j)

```

```

# Summe von 3 Stellen

```

```

distr_1_2_3 <- function(k,n){

```

```

  l <- 0

```

```

  if(k < 6 | k > 3*n-3){ print("k_muss_zwischen_6_und_3n-3_liegen")}

```

```

  else{

```

```

    k <- floor(k)

```

```

    for (i in 1:k){

```

```

      for(j in 1:n){

```

```

        for(h in 1:n){

```

```

          for(t in 1:n){

```

```

            if(t != j){

```

```

              if( t!= h){

```

```

                if(j != h){

```

```

                  if( j + h + t == i){l <- l + 1 }

```

```

                }

```

```

              }

```

```

            }

```

```

          }

```

```

        }

```

```

      }

```

```

    }

```

```

    value <- l*(factorial(n-3))/(factorial(n))

```

```

    print(l)

```

```

    print(value)

```

```

  }

```

```

}

```

```

# P(Z = k)

```

```

distr_123 <- function(k,n){

```

```

l <- 0
for(j in 1:n){
  for(h in 1:n){
    for(t in 1:n){
      if(t != j){
        if(t != h){
          if(j != h){
            if(j + h + t == k){l <- l + 1 }
          }
        }
      }
    }
  }
}

value <- l*(factorial(n-3))/(factorial(n))
#   print(l)
#   print(value)
return(value*k)
}

#EZ = sum k*P(Z = k) f r n=20
E = 0
for (k in c(6:57)) {
  E = E + distr_123(k,20)
}
print(E)

# Standardabweichung von Z f r n=20
e.3 = 0
for (k in c(6:57)) {
  e.3 = e.3 + k*distr_123(k,20)
}
sd.z = sqrt(e.3 - E^2)
sd.z

#####
#                               p-Wert bestimmen                               #
#####

(1/20)*rowSums(place_out)

# Mittelwert des Quantils auf Platz 1 bestimmen
sum(c(1:20))*(1/20) #oder
mul <- mean(c(1:20))

```



```

mu2 <- mean(c(1:20) + c(1:20))
mu3 <- mean(3*c(1:20))

# Empirisches Mittel ber das Quantil auf Platz 1 ( f r sigma)
i_emp <- mean(place_out[1,])
# Empirisches Mittel ber das Quantil auf Platz 1 und 2
i_emp2 <- mean(place_out[1,] + place_out[2,])
# Empirisches Mittel ber das Quantil auf Platz 1 und 2 und 3
i_emp3 <- mean(place_out[1,] + place_out[2,] + place_out[3,])

# Standardabweichung
sigma_emp <- sqrt( 1/20*(sum((place_out[1,] - i_emp)^2)))
sigma1 <- sd(place_out[1,])
sigma_emp2 <- sd(place_out[1,] + place_out[2,])
sigma2 <- sd.x
sigma_emp3 <- sd(place_out[1,] + place_out[2,] + place_out[3,])
sigma3 <- sd.z

# Z - Test (nur Platz 1)
x1 <- place_out[1,]
z.test(x1,y = NULL, alternative = "less",
       mu = mu1, sigma1 , conf.level = 0.95 )

# Z - Test (Platz 1 & 2)
x2 <- place_out[1,] + place_out[2,]
z.test(x2,y = NULL, alternative = "less",
       mu = mu2, sigma2 , conf.level = 0.95 )

# Z - Test (Platz 1 & 2 & 3)
x <- place_out[1,] + place_out[2,] +place_out[3,]
z.test(x ,y = NULL, alternative = "less",
       mu = mu3, sigma3 , conf.level = 0.95 )

# Quantiles on last 3 places in outsample
i_emp_last3 = mean (place_out[18,] + place_out[19,] +place_out[20,])
sigma_emp_last3 = sd.z

# z test(platz (18,19,20))
x3.1 <- place_out[18,] + place_out[19,] +place_out[20,]
z.test(x3.1 , alternative = "greater",
       mu = mu3, sigma3 , conf.level = 0.95 )

# Quantiles in the middle of outsample
i_emp_middle3 = mean (place_out[9,] + place_out[10,] +place_out[11,])
sigma_emp_middle3 = sd.z

# z test(platz (9,10,11))

```

```
x3.2 <- place_out[9,] + place_out[10,] +place_out[11,]  
z.test(x3.2,y = NULL , alternative = "two.sided",  
       mu = mu3, sigma3 , conf.level = 0.95 )
```