



TECHNISCHE
UNIVERSITÄT
WIEN

S E M I N A R A R B E I T

Building portfolios of stocks in the São Paulo Stock Exchange using Random Matrix Theory

ausgeführt am

Institut für
Finanz- und Versicherungsmathematik
TU Wien

unter der Anleitung von

Associate Prof. Dipl.-Ing. Dr.techn. Stefan Gerhold

durch

Markus Grafl

Matrikelnummer: 11778674

Wien, am 28.02.2020

Inhaltsverzeichnis

1	Einleitung	3
2	Zufallsmatrixtheorie	5
2.1	Eigenwerte der Korrelationsmatrizen und deren Verteilung	5
2.2	Eigenvektoren und deren Interpretation	10
2.3	Säubern der Korrelationsmatrix	13
3	Portfoliotheorie nach Markowitz	15
3.1	Effiziente Portfolios: Vorhersage und Realisierung	16
3.2	Reduzierung des Markteffektes durch lineare Regression	20
3.3	Ergebnisse in Zahlen	21
4	Conclusio	26
	Abbildungsverzeichnis	28
	Tabellenverzeichnis	29
	Literaturverzeichnis	30

1 Einleitung

In dieser Arbeit geht es vor allem um eine Studie, die am „Insper“-Institut für Bildung und Forschung (Insper, Instituto de Ensino e Pesquisa) von den Autoren Leonidas Sandoval Junior, Adriana Bruscato und Maria Kelly Venezuela geschrieben wurde. Dies ist eine unabhängige, non-profit Forschungseinrichtung mit Sitz in São Paulo in Brasilien¹. In dieser Studie werden über insgesamt sechs Perioden hinweg (von 2004 bis 2010), wobei jede Periode aus zwei Jahren besteht, die täglichen Kurse einer gewissen Anzahl von 100% liquiden Aktien der brasilianischen Aktien-, Rohstoff- und Terminbörse (Bolsa de Valores, Mercadorias e Futuros de São Paulo, kurz: BM&F-Bovespa) beobachtet. „100% liquide“ bedeutet hier, dass diese Aktien an jedem Tag, an dem die Börse geöffnet war, zum Handel zur Verfügung standen. Die Anzahl der beobachteten Aktien ist für jede Periode unterschiedlich, und zwar waren es 61 Aktien für die Jahre 2004-2005, 72 Aktien für 2005-2006, 86 Aktien für 2006-2007, 105 Aktien für 2007-2008, 148 Aktien für 2008-2009 und 153 Aktien für 2009-2010 (vgl. [1], S. 3).

Ziel ist es, in jeder Periode mit den Kursdaten aus dem ersten Jahr das Risiko eines mit diesen Aktien gebildeten Portfolios im Folgejahr vorherzusagen. Um das Konzept dieser Arbeit zu schildern, betrachten wir nun hauptsächlich die erste Periode von 2004 bis 2005. Für alle anderen Perioden wird analog vorgegangen.

Im Jahr 2004 werden insgesamt 61 verschiedene Aktien über einen Zeitraum von 248 Tagen betrachtet. Dieser Zeitraum entspricht genau jener Anzahl an Tagen, an denen die brasilianische Börse in diesem Jahr geöffnet war. Um zu untersuchen, wie sich die Aktien zueinander verhalten, wird im ersten Abschnitt dieser Arbeit beschrieben, wie man aus den erhaltenen Zeitreihen der 61 Aktienkurse Kreuz-Korrelationsmatrizen bilden kann. Denn neben der Volatilität wird das Risiko vor allem durch Kovarianz- und Korrelationsmatrizen beschrieben, wobei hohe Korrelationen nach dem Diversifikationseffekt auf hohes Risiko und niedrige oder negative Korrelationen umgekehrt auf geringes Risiko schließen lassen (vgl. [1], S. 1). Im nächsten Schritt werden wir mit Hilfe einer Methode der Zufallsmatrixtheorie versuchen, die nützlichen Daten, welche wir zur Risikovorhersage verwenden möchten, von den verrauschten bzw. zufälligen Daten, im englischen als „measurement noise“ oder nur als „noise“ bezeichnet, zu trennen. Dazu werden wir die Eigenwerte der zuerst berechneten Korrelationsmatrix mit einer zufälligen Matrix vergleichen, die wir aus speziellen Zufallsmatrizen bilden

¹vgl. <https://www.insper.edu.br/en/institutional/about-insper>

werden. Daten, die von den zufälligen Ergebnissen abweichen, deuten dann auf die Existenz von wahrer Information hin, Übereinstimmungen hingegen auf verrauschte Daten. Danach werden wir uns Eigenvektoren zu bestimmten Eigenwerten ansehen, die einerseits in der Region der verrauschten Daten und andererseits einmal oberhalb und einmal unterhalb dieser Region liegen. Da die Ergebnisse in der Zufallsregion kaum aussagekräftig sind, werden wir anschließend die Eigenwerte in diesem Bereich abändern und damit eine neue „gesäuberte“ Korrelationsmatrix bilden.

Im zweiten Abschnitt ist das Ziel, auf Basis der Daten aus dem ersten Jahr möglichst risikoarme Portfolios im Sinne der Markowitz'schen Portfoliotheorie zu konstruieren, wobei neben den Grundsätzen dieser Theorie eine Methode erklärt wird, mit der man das Risiko eines Portfolios zu einer vorgegebenen Rendite minimieren kann. Diese Ergebnisse werden danach durch sogenannte „Efficient frontiers“ veranschaulicht und mit den echten Realisierungen des Folgejahres verglichen. Dabei betrachten wir auch die Ergebnisse der gesäuberten Korrelationsmatrix für die erste Periode und vergleichen diese mit den Resultaten der ursprünglichen Korrelationsmatrix.

Im Anschluss daran wird versucht, den Markteffekt, der aufgrund der gemeinsamen Bewegungen aller Aktien entsteht, durch ein Regressionsmodell von den Renditen zu entfernen. Mit den damit erhaltenen Restrenditen werden dann neue Korrelationsmatrizen berechnet, mit welchen wieder Vorhersagen getroffen werden, die mit den Realisierungen verglichen werden.

Zu guter Letzt werden verschiedene Fehlerschätzer betrachtet, welche die Übereinstimmung der Prognosen aus den Daten des ersten Jahres mit den Realisierungen des Folgejahres innerhalb einer Periode messen. Dabei berechnen drei Schätzer die Unterschiede zwischen den vorhergesagten und den realisierten „Efficient frontiers“ und ein weiteres Distanzmaß vergleicht die verschiedenen Korrelationsmatrizen direkt miteinander. Denn die Korrelationsmatrizen werden einmal mit den ursprünglichen Aktienrenditen und einmal mit den Restrenditen des Regressionsmodells berechnet. Zusätzlich werden diese beiden Matrizen einmal von den zufälligen Eigenwerten gesäubert und einmal nicht, wodurch sich insgesamt vier verschiedene Korrelationsmatrizen ergeben, auf Basis derer das Risiko prognostiziert wird. Die Ergebnisse aus diesen vier Fällen sind schließlich nach Perioden unterteilt in zwei Tabellen dargestellt, wobei die erste Leerverkauf verbietet und die zweite Portfolios mit Leerverkauf zulässt.

2 Zufallsmatrixtheorie

Die Zufallsmatrixtheorie beschäftigt sich im Allgemeinen mit den Eigenschaften von Matrizen, deren Einträge Zufallsvariablen sind. Im Allgemeinen kann Zufallsmatrixtheorie immer dann verwendet werden, wenn man Zusammenhänge zwischen großen, aber scheinbar unzusammenhängenden Mengen an Daten untersuchen möchte. Denn durch den Vergleich von scheinbar zufälligen Daten mit tatsächlichen Zufallsdaten, die mit Hilfe von Zufallsvariablen simuliert werden, sollen die nicht zufälligen Informationen zum Vorschein kommen (vgl. [2], S. 1).

Die Ursprünge dieser Theorie gehen zurück auf den ungarischen Physiker Eugene Wigner, der in den 50er-Jahren die Distanzen zwischen Energieniveaus von Atomkernen untersuchte und dabei feststellte, dass diese durch Eigenwerte von Zufallsmatrizen approximiert werden können. Daneben entwickelten sich auch noch weitere Anwendungsgebiete wie die Quantenphysik, Nanotechnologie, Quantengravitation, Finanzmathematik und noch viele andere (vgl. [1], S. 3).

2.1 Eigenwerte der Korrelationsmatrizen und deren Verteilung

In diesem Abschnitt möchten wir uns zunächst die Kreuz-Korrelationsmatrizen von unseren Aktienrenditen berechnen und deren Eigenwerte mit den Eigenwerten einer zufälligen Korrelationsmatrix vergleichen, welche mit Hilfe von Zufallsmatrizen gebildet wird. Durch den Vergleich soll, wie bereits erwähnt, ersichtlich werden, bei welchen Eigenwerten es sich um zufällige bzw. verrauschte Daten handelt und bei welchen es sich um Eigenwerte handelt, aus denen man relevante Informationen für das gemeinsame Verhalten der Aktien ableiten kann.

Wir betrachten nun zuerst die Renditen der $N = 61$ gewählten Aktien aus dem Jahr 2004, mit welchen wir das Risiko für das Jahr 2005 vorhersagen wollen. Aus mathematischer Sicht liefern uns diese Daten 61 Zeitreihen mit $T = 248$ Handelstagen. Sei also

$$r_i(t) = \ln(P_i(t)) - \ln(P_i(t-1)), \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

für jede Aktie i die logarithmische Rendite am Handelstag t . Dabei bezeichnet $P_i(t)$ den Tagesschlusskurs der i -ten Aktie am Tag t (vgl. [1], S. 4). Sobald von nun an

von einer Rendite die Rede ist, soll immer die gerade beschriebene logarithmische Rendite gemeint sein. Um nun die Kreuz-Korrelationsmatrix zwischen den Zeitreihen $(r_i(t))_{t=1, \dots, T}$, $i = 1, \dots, N$, zu berechnen, normieren wir zuerst die Renditen so, dass

$$g_i(t) = \frac{r_i(t) - \hat{r}_i}{\sigma_i}.$$

Dabei sei \hat{r}_i der über die Zeitpunkte $t = 1, \dots, T$ gemittelte Durchschnitt der Renditen (arithmetisches Mittel) und σ_i die empirische Standardabweichung der i-ten Aktie, also

$$\hat{r}_i := \frac{1}{T} \sum_{t=1}^T r_i(t) \tag{2.1}$$

und

$$\sigma_i := \sqrt{\frac{1}{T} \sum_{t=1}^T (r_i(t) - \hat{r}_i)^2}. \tag{2.2}$$

Bildet man nun mit den Einträgen $g_i(t)$ die NxT-Matrix

$$G = (g_i(t))_{i=1..N, t=1..T} = \begin{pmatrix} g_1(1) & \dots & g_1(T) \\ \vdots & & \vdots \\ g_N(1) & \dots & g_N(T) \end{pmatrix},$$

so erhält man die gewünschte Kreuz-Korrelationsmatrix schließlich durch die Formel

$$C = \frac{1}{T} G G^T, \tag{2.3}$$

wobei G^T die transponierte Matrix von G bezeichnet (vgl. [3], S. 4).

Bei der Verlässlichkeit dieser Daten ist allerdings Vorsicht geboten. Denn für die NxN-Kreuz-Korrelationsmatrix müssen insgesamt $\frac{N(N-1)}{2}$ Einträge berechnet werden, welche aus den N Zeitreihen der Länge T bestimmt werden. Vor allem aber, wenn T nicht deutlich größer als N ist, kann es passieren, dass die Daten bzw. Ergebnisse in unserer Korrelationsmatrix C sehr verrauscht und daher zu einem großen Teil vom Zufall bestimmt sind. Deshalb ist es wichtig, in den Daten die sogenannten wahren Signale von den Störsignalen bzw. verrauschten Daten zu unterscheiden (vgl. [2], S. 1).

Um die zufälligen Daten ausfindig zu machen und diese von den wahren Daten zu unterscheiden, vergleichen wir nun unsere Korrelationsmatrix C mit bestimmten Zufallsmatrizen. Dazu betrachten wir zunächst die Zufallsmatrix

$$X = (X_{it})_{i=1..N, t=1..T} = \begin{pmatrix} X_{11} & \dots & X_{1T} \\ \vdots & & \vdots \\ X_{N1} & \dots & X_{NT} \end{pmatrix},$$

bestehend aus unabhängig, identisch verteilten Zufallsvariablen, von denen jede normalverteilt mit Erwartungswert 0 und Varianz σ^2 sein soll, also

$$\forall i = 1..N, t = 1..T : X_{it} \sim \mathcal{N}(0, \sigma^2).$$

Mit dieser Zufallsmatrix X können wir uns nun die sogenannte Wishart-Matrix

$$W = \frac{1}{T} X X^T$$

konstruieren. Diese kann als Kovarianzmatrix von N stochastischen Zeitreihen, bestehend aus je T statistisch unabhängigen Zufallsvariablen, interpretiert werden. Damit würde jedoch auch folgen, dass W keine Kreuz-Korrelationen aufweist, weshalb wir X noch mit einer positiv definiten und symmetrischen $N \times N$ -Matrix ζ multiplizieren, welche die Korrelationen bestimmen soll. Definiert man also eine Matrix M als $M := \zeta^{\frac{1}{2}} X$, dann erhält man die korrelierte Wishart-Matrix durch (vgl. [4], S. 5)

$$W_{corr} = \frac{1}{T} M M^T.$$

Lässt man die Dimensionen N und T immer größer werden, also $N \rightarrow \infty$ und $T \rightarrow \infty$, mit einem fixierten $Q = \frac{T}{N} \in (0, \infty)$, so genügen die Eigenwerte der korrelierten Wishart Matrix einer speziellen Wahrscheinlichkeitsverteilung, nämlich der Marčenko-Pastur Verteilung, welche nach den beiden Mathematikern Vladimir Marčenko und Leonid Pastur benannt wurde. Die Dichtefunktion dieser Verteilung lautet (vgl. [1], S. 4 f.)

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \quad \lambda \in [\lambda_-, \lambda_+],$$

wobei

$$\lambda_- = \sigma^2 \left(1 + \frac{1}{Q} - 2\sqrt{\frac{1}{Q}} \right) \quad \text{und} \quad \lambda_+ = \sigma^2 \left(1 + \frac{1}{Q} + 2\sqrt{\frac{1}{Q}} \right).$$

Für die Periode 2004-2005 beträgt $Q = \frac{T}{N} = \frac{248}{61} \approx 4,06 \in (0, \infty)$ und wir erhalten die Dichtefunktion $p(\lambda)$, die in der linken Graphik von Abbildung 2.1 dargestellt wird. Setzt man für λ_- und λ_+ in die obige Formel ein, so ergibt sich der Definitionsbereich

$[\lambda_-, \lambda_+] = [0.254, 2.238]$. Dieser Bereich wird auch „Wishart-Region“ genannt (vgl. [1], S. 4).

Neben der Dichtefunktion erkennt man in dieser Graphik auch noch ein Histogramm, welches die Verteilung der Eigenwerte der Kreuz-Korrelationsmatrix unserer 61 Aktien aus dem Jahr 2004 beschreibt. Die einzelnen Eigenwerte werden auch noch einmal deutlicher in der rechten Graphik dargestellt. Dort sind sie einzeln als schwarze Striche eingezeichnet. Gut erkennt man auch jene Eigenwerte, die in die „Wishart-Region“ fallen, welche grau eingefärbt ist.

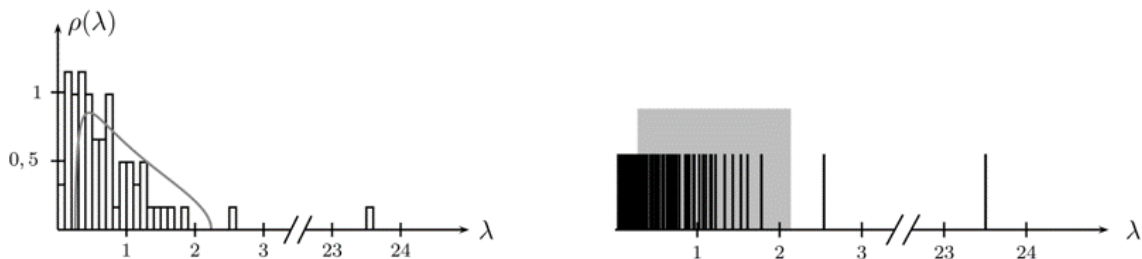


Abbildung 2.1: Verteilung der Eigenwerte, entnommen aus [1], S. 5

Aus diesen beiden Graphiken kann man deutlich erkennen, dass der Großteil der Eigenwerte, nämlich in etwa 72%, im Definitionsbereich unserer Dichtefunktion liegt. In genau diesem Bereich stimmen also die zufälligen Daten mit unseren Eigenwerten überein, weshalb wir darauf schließen, dass die zufälligen bzw. verrauschten Eigenwerte genau diejenigen sind, die innerhalb der „Wishart-Region“ liegen. Geht man aus diesem Bereich heraus, so findet man auch zwei Eigenwerte, die darüber liegen, und insgesamt 15 Eigenwerte, die darunter liegen. Insbesondere sticht der größte aus unseren Daten erhaltene Eigenwert heraus, der in etwa bei 23.5 liegt. Dieser ist somit gleich zehn Mal größer als der größte Eigenwert, der von der Marčenko-Pastur Verteilung vorausgesagt wurde (vgl. [1], S. 5).

Bevor wir allerdings aus diesem Vergleich weitere Schlüsse ziehen, was wir im nächsten Abschnitt im Zusammenhang mit den zu diesen Eigenwerten gehörigen Eigenvektoren machen wollen, möchte ich noch auf zwei Methoden eingehen, die die Verwendung der gewählten Wahrscheinlichkeitsverteilung rechtfertigen. Als erstes wollen wir sogenannte Quantil-Quantil-Diagramme betrachten, welche oft auch nur als Q-Q-Diagramme bezeichnet werden und in Abbildung 2.2 dargestellt sind. In einem solchen werden die Quantile zweier verschiedener Verteilungen einander gegenübergestellt und punktweise abgebildet. In unserem Fall kann man auf der x-Achse die Quantile der Marčenko-Pastur Verteilung und auf der y-Achse jene Quantile, die wir aus der Verteilung der Eigenwerte unserer Daten aus dem Jahr 2004 erhalten, ablesen. Je mehr die Punkte auf der Identitätsgeraden ($y=x$) liegen, desto ähnlicher sind sich die Verteilungen.

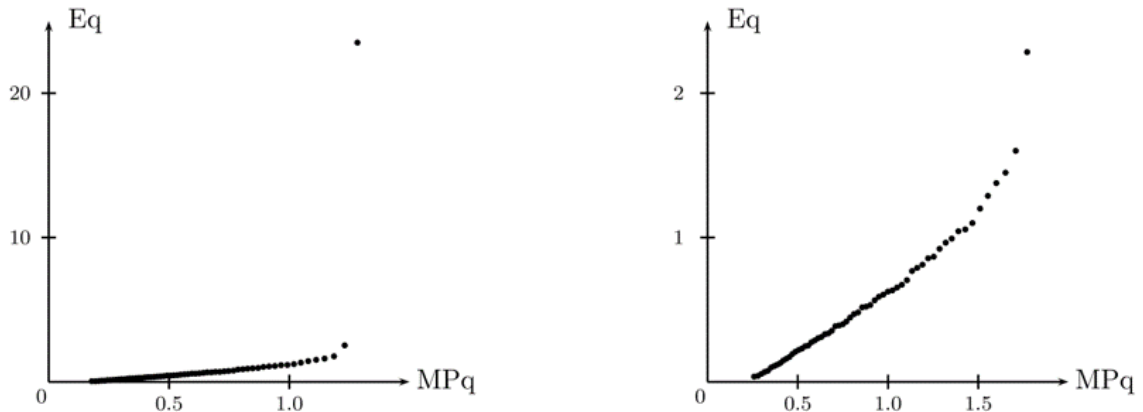


Abbildung 2.2: Quantil-Quantil-Diagramme, entnommen aus [1], S. 7

Es fällt in der linken Graphik sofort auf, dass der größte Eigenwert stark von den anderen abweicht. Diese hingegen liegen nahe an einer Geraden, was auf einen starken linearen Zusammenhang hindeutet. Besonders wenn man die rechte Graphik betrachtet, in der der größte Eigenwert entfernt wurde, sieht man deutlich, dass vor allem die Eigenwerte innerhalb der „Wishart-Region“ ähnlich wie die Marčenko-Pastur Verteilung verteilt sind, weshalb die Dichtefunktion speziell in diesem Bereich eine gute Approximation darstellt (vgl. [1], S. 6 f.).

In einer zweiten Methode wurde die Reihenfolge der Zeitreihen unserer 61 Aktienrenditen beliebig vertauscht. Damit wurde anschließend eine neue Kreuz-Korrelationsmatrix erstellt und deren Eigenwerte berechnet. Diese Simulation wurde insgesamt 10 000 Mal wiederholt. Die Verteilung der daraus resultierenden Eigenwerte wurde in Abbildung 2.3 zusammen mit dem Histogramm der Eigenwerte der ursprünglichen Kreuz-Korrelationsmatrix und der Dichtefunktion der Marčenko-Pastur Verteilung veranschaulicht.

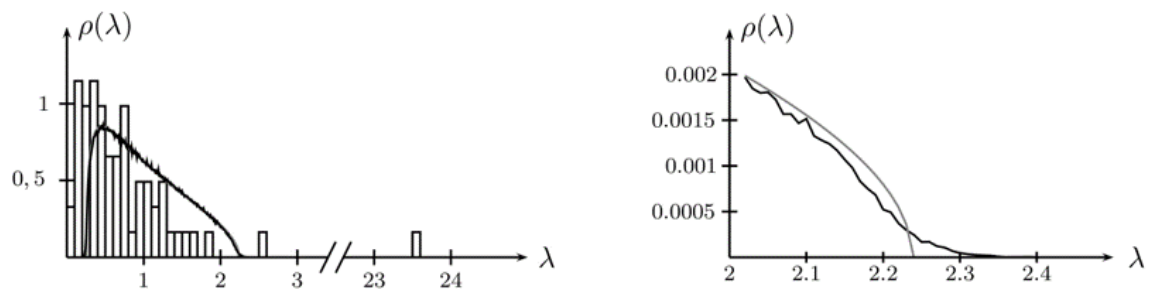


Abbildung 2.3: Verteilung der Eigenwerte mit Simulationen, entnommen aus [1], S.

8

Die Dichtefunktion der Marčenko-Pastur Verteilung ist dabei in grau und das Ergebnis der Simulationen in schwarz eingezeichnet. In der linken Graphik in Abbildung 2.3

erkennt man die Unterschiede zwischen den Simulationen und der Marčenko-Pastur Verteilung kaum, da die Eigenwertverteilungen in beiden Fällen mit freiem Auge fast nicht unterscheidbar sind. Vergrößert man jedoch zum Beispiel, wie in der linken Graphik in Abbildung 2.3, den Bereich um den Wert $\lambda_+ = 2.238$, so erkennt man zwar leichte Unterschiede, doch sind sich die beiden Verteilungen immer noch sehr ähnlich. Etwas größere Differenzen finden sich lediglich an den beiden Enden der Verteilungen (vgl. [1], S. 7 f.).

Beide Methoden lassen letztlich darauf schließen, dass sich die Marčenko-Pastur Verteilung gut für den Vergleich mit unseren Daten eignet. Nun wollen wir erklären, was uns die bisher geschilderten Ergebnisse im Zusammenhang mit Portfolios zeigen. Dafür werden wir uns im nächsten Abschnitt die zu den Eigenwerten gehörigen Eigenvektoren genauer ansehen.

2.2 Eigenvektoren und deren Interpretation

In diesem Abschnitt werden wir uns, wie gerade erwähnt, einige Eigenvektoren ansehen, die zu den in Abbildung 2.1 dargestellten Eigenwerten gehören. Aus unserer 61×61 Kreuz-Korrelationsmatrix C ergeben sich insgesamt 61 Eigenwerte, welche wir nun zuerst aufsteigend vom kleinsten bis zum größten Eigenwert sortieren. Dabei bezeichnet λ_{61} den kleinsten und λ_1 den größten Eigenwert. Der zum i -ten Eigenwert λ_i gehörende Eigenvektor werde für alle $i = 1, \dots, N$ mit e_i bezeichnet.

Wie wir in Abbildung 2.1 gesehen haben, liegen die meisten Eigenwerte innerhalb der „Wishart-Region“, zwei befinden sich darüber und einige andere darunter. Um alle Fälle abzudecken, betrachten wir aus jedem dieser drei Bereiche zwei Eigenvektoren. Die Einträge der Eigenvektoren sind in Abbildung 2.4 abgebildet, wobei weiße Balken positive und graue Balken negative Werte anzeigen. Jeder Eigenvektor e_i , $i = 1, \dots, N$ kann dabei als ein Portfolio $P_i \in \mathbb{R}^{61}$ interpretiert werden, bei dem der j -te Eintrag, $j = 1 \dots 61$, jene Menge angibt, die in die j -te Aktie investiert wird.

Ist dabei der j -te Eintrag negativ, so spricht man auch von „Leerverkauf“ oder im Englischen von „short selling“. Leerverkauf bezeichnet hierbei das Leihen und Verkaufen von Aktien, die sich zum Zeitpunkt des Verkaufs nicht im Eigentum der Verkäuferin oder des Verkäufers befinden. Später muss die Verkäuferin oder der Verkäufer dann die Aktien zum aktuellen Marktpreis wieder kaufen, um sie der ursprünglichen Eigentümerin oder dem ursprünglichen Eigentümer, von der oder dem die Aktien geliehen wurden, wieder zurückgeben zu können. Im Gegensatz zum normalen Kauf von Aktien, bei dem man natürlicherweise auf einen steigenden Kurs hofft, um die Aktien zu einem späteren Zeitpunkt zu einem höheren Preis verkaufen zu können, spekuliert man beim Leerverkauf auf fallende Kurse, da man das Wertpapier zu ei-

nem späteren Zeitpunkt wieder kaufen muss und man dabei einen niedrigeren Preis bezahlen möchte, als man beim früheren Verkauf erhalten hat (vgl. [1], S. 9).

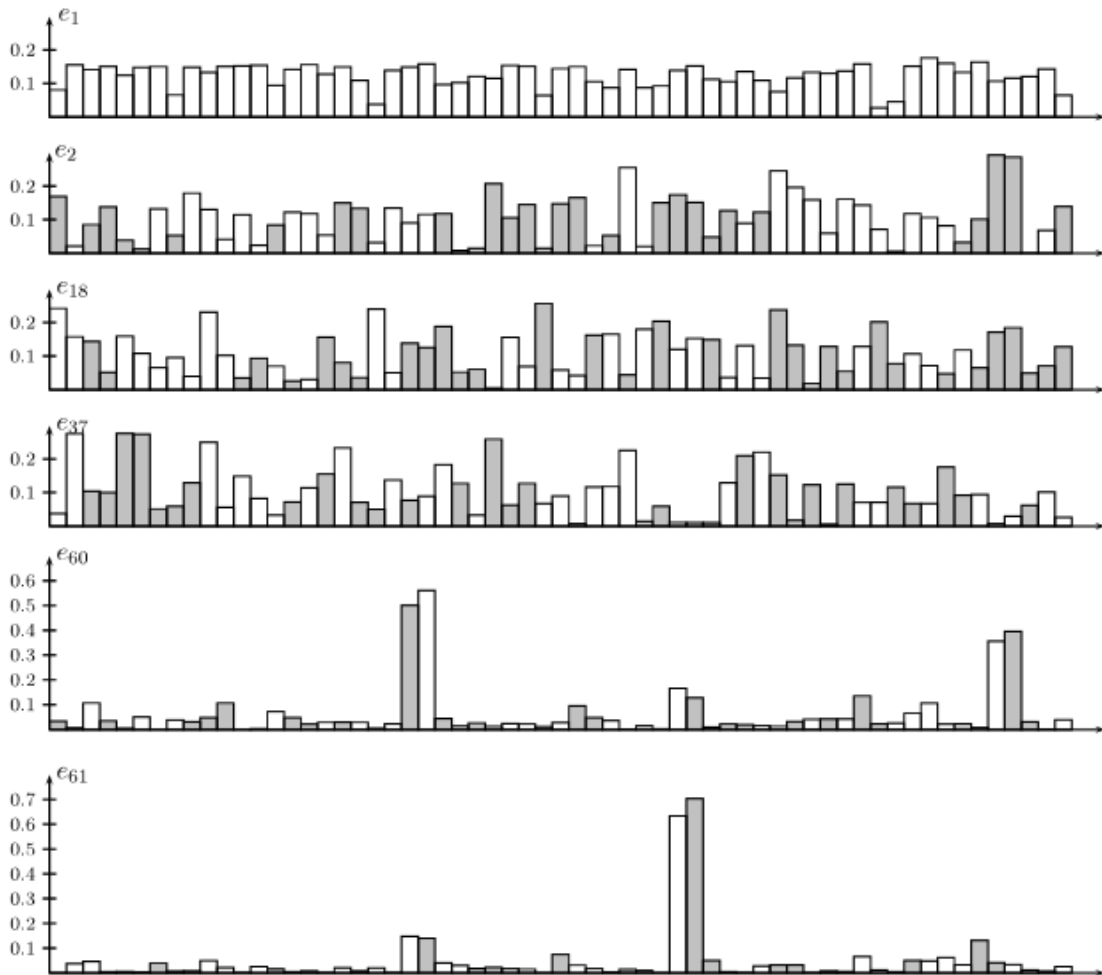


Abbildung 2.4: Eigenvektoren, entnommen aus [1], S. 5

Zuerst werden die beiden Eigenvektoren e_1 und e_2 zu den größten Eigenwerten λ_1 und λ_2 betrachtet, wobei λ_1 mit Abstand der größte ist. Fasst man den Eigenvektor e_1 als Portfolio P_1 auf so fällt auf, dass hier alle Werte positiv sind, also eine Anlegerin oder ein Anleger in alle Aktien investieren würde und hierbei sogar überall ähnlich viel. Man sagt in diesem Zusammenhang auch, dass sich alle Aktien dieses Portfolios ähnlich wie der Marktindex der brasilianischen Börse verhalten. Auch die Korrelation zwischen den logarithmischen Renditen von P_1 und jenen des Ibovespa, der als führender brasilianischer Aktienindex das allgemeine Verhalten des brasilianischen Aktienmarktes beschreibt, bestätigt diesen Sachverhalt. Denn für sie ergibt

sich ein Wert von 0.9865, was darauf hindeutet, dass unser Portfolio P_1 aus dem Jahr 2004 aus einer Kombination von Aktien besteht, die sich alle in gewisser Weise wie der Markt verhalten. Dies tun sie jedoch mit einer viel höheren Volatilität, die sich in der Form der Standardabweichungen ausdrückt. Die Standardabweichung der logarithmischen Renditen von P_1 beträgt 12.51%, jene des Ibovespa nur 1.8%. Diese große Differenz wird durch die unterschiedliche Gewichtung der einzelnen Aktien in unserem Portfolio P_1 und im Ibovespa erklärt. Aufgrund des ähnlichen Verhaltens der Aktien und der hohen Volatilität ist P_1 demnach sehr risikoreich. Betrachtet man den Eigenvektor e_2 zu dem zweitgrößten unserer Eigenwerte $\lambda_2 = 2.540$, welcher viel näher an der „Wishart-Region“ liegt als λ_1 , so wird hier zwar in gewisse Aktien bevorzugt investiert, doch insgesamt erkennt man nur noch geringere Tendenzen (vgl. [1], S. 6).

Dieses Verhalten zeigt sich noch deutlicher in der dritten und vierten Graphik in Abbildung 2.4. Die darin abgebildeten Eigenvektoren e_{18} und e_{37} , deren zugehörige Eigenwerte $\lambda_{18} = 0.853$ und $\lambda_{37} = 0.393$ innerhalb der „Wishart-Region“ liegen, zeigen keinerlei Strukturen mehr. Es gibt größere und kleinere sowie positive und negative Einträge, die in keinem erkennbaren Zusammenhang stehen. Dies bestätigt gewissermaßen die Idee, über die Marčenko-Pastur Verteilung den zufälligen bzw. verrauschten Bereich unsere Daten zu bestimmen. Auch die Korrelation zwischen P_{37} und dem Ibovespa von 0.1824 ist sehr gering. Es besteht also kaum ein linearer Zusammenhang. Die Standardabweichung von P_{37} ist allerdings mit 1.72 jener des Ibovespa sehr ähnlich (vgl. [1], S. 6).

Wie oben schon erwähnt, gibt es auch einige Eigenwerte unterhalb der „Wishart-Region“. Die Eigenvektoren zu den beiden kleinsten Eigenwerten $\lambda_{60} = 0.046$ und $\lambda_{61} = 0.039$ sind unten in Abbildung 2.4 ersichtlich. Aus diesen kann man im Unterschied zu den vorigen beiden doch deutliche Informationen herauslesen. Zum einen sind alle Einträge fast 0. Betrachtet man die Vektoren also wieder als Portfolios, so wird nur in wenige Aktien viel investiert. Außerdem fällt auf, dass bei je zwei hohen Balken der eine einen positiven und der andere einen negativen Wert anzeigt. Analysiert man dabei auch noch genauer, welche Balken welchen Aktien entsprechen, so zeigt sich eine wichtige Erkenntnis. Das Portfolio P_{61} hat zwei solcher Paare von höheren Balken. Untersucht man die zwei größten Balken, so stellt man fest, dass es sich dabei um die Aktien „PETR4“ und „PETR3“ handelt, welche beide zum selben Unternehmen Petrobas gehören. Verfolgt man also diese Strategie, so kauft man die Aktie „PETR4“ und leerverkauft die Aktie „PETR3“, die dem grauen Balken entspricht. Das Risiko ist dabei natürlich minimal. Denn steigen die Aktienwerte dieses Unternehmens, so gewinnt man durch den Kauf der Aktie „PETR4“ und verliert bei der Aktie „PETR3“. Fallen die Kurse, so passiert das gleiche, nur mit umgekehrten Rollen der Aktien. Man kann also weder viel gewinnen noch viel verlieren. Beides hält sich immer ungefähr in Waage. Dasselbe gilt für die anderen beiden ho-

hen Balken. Auch hier handelt es sich um den Kauf (ELET3-Aktie) und Leerverkauf (ELET6-Aktie) von Aktien des Unternehmens Eletrobas. Genauso ist es auch beim Eigenvektor e_{60} . Auch hier stecken hinter jedem Paar, bestehend aus einem positiven und einem negativen Eintrag, der Kauf und Leerverkauf von Aktien desselben Unternehmens und in andere Aktien wird kaum oder nur wenig investiert. Die Eigenvektoren der kleinsten Eigenwerte beschreiben demnach, genau im Gegensatz zu jenen der größten Eigenwerte, also sehr risikolose Aktien. Auch der sehr niedrige Korrelationswert von 0.0932 zwischen P_{61} und dem Ibovespa und eine Standardabweichung dieses Portfolios von 0.44% unterstreichen dieses Resultat (vgl. [1], S. 6).

2.3 Säubern der Korrelationsmatrix

Im letzten Abschnitt des zweiten Kapitels werden wir uns noch etwas genauer mit jenen Eigenwerten beschäftigen, welche unsere verrauschten Daten darstellen. Sie haben wir bisher schon mit Hilfe von Zufallsmatrizen und der Marčenku-Pastur Verteilung in der „Wishart-Region“ ausfindig gemacht, doch haben wir uns noch nicht weiter mit ihnen auseinandergesetzt. Dies wollen wir nun in diesem Abschnitt tun.

Betrachten wir zuerst noch einmal die Kreuz-Korrelationsmatrix $C \in \mathbb{R}^{N \times N}$ aus Gleichung 2.3. Mit Hilfe der gerade genannten Informationen wollen wir die zufälligen bzw. verrauschten Daten so gut es geht von dieser Matrix entfernen. Als Korrelationsmatrix ist diese Matrix symmetrisch und damit auch diagonalisierbar. Es existieren also eine reguläre Matrix $P \in \mathbb{R}^{N \times N}$ und eine Diagonalmatrix $D \in \mathbb{R}^{N \times N}$, sodass C dargestellt werden kann als

$$C = PDP^{-1},$$

wobei

$$P = (e_N, \dots, e_1)$$

und

$$D = \text{diag}(\lambda_N, \dots, \lambda_{W_1}, \dots, \lambda_{W_2}, \dots, \lambda_1).$$

Der Eigenwert λ_{W_1} bezeichnet dabei den kleinsten Eigenwert, der größer als λ_- ist und λ_{W_2} den größten Eigenwert, der kleiner als λ_+ ist. Diese beide Eigenwerte stellen also den kleinsten und größten Eigenwert der „Wishart-Region“ dar, wobei klarerweise auch alle Eigenwerte zwischen λ_{W_1} und λ_{W_2} in diesem verrauschten Bereich liegen. Um die verrauschten Daten einigermaßen zu entfernen, werden nun alle Eigenwerte von λ_{W_1} bis λ_{W_2} durch ihr arithmetisches Mittel ersetzt. Sei deshalb $\bar{\lambda}$ dieser Durchschnitt und n die Anzahl der Eigenwerte innerhalb der Wishart-Region, dann ist

$$\bar{\lambda} = \frac{\lambda_{W1} + \dots + \lambda_{W2}}{n}.$$

Im Jahr 2004 beträgt dieser Durchschnitt 0.748 und im Jahr 2005 hat er den Wert 0.790. Bezeichnen wir mit D_{clean} die neue Diagonalmatrix

$$D_{clean} = \text{diag}(\lambda_N, \dots, \bar{\lambda}, \dots, \bar{\lambda}, \dots, \lambda_1),$$

so erhalten wir damit auch eine neue „gesäuberte“ Korrelationsmatrix C_{clean} , welche sich durch

$$C_{clean} = PD_{clean}P^{-1} \tag{2.4}$$

berechnet (vgl. [1], S. 11).

Von den beiden Korrelationsmatrizen C und C_{clean} ausgehend, werden wir nun im nächsten Kapitel Portfolios mit minimalem Risiko konstruieren bzw. berechnen.

3 Portfoliotheorie nach Markowitz

In diesem Abschnitt möchten wir uns mit der Auswahl möglichst effizienter Portfolios im Sinne der Portfoliotheorie des Mathematikers Harry Markowitz beschäftigen. Dieser schrieb im Jahr 1952 ein Paper mit dem Titel „Portfolio Selection“, welches im „Journal of Finance“ publiziert wurde. Es befasst sich dabei mit einer optimalen Auswahl von Portfolios. Markowitz unterteilt die Auswahl von Portfolios in zwei Phasen. Die erste beginnt dabei mit dem Beobachten und Sammeln von Erfahrungen und endet mit einem festen Glauben über das zukünftige Verhalten der Aktien. Dort beginnt die zweite Phase, die schließlich mit der Wahl eines konkreten Portfolios endet. Sein Paper setzt sich mit der zweiten Phase auseinander (vgl. [5] S. 77).

In seinem Paper hält er folgende wichtige Regel fest, die er selbst als „Expected returns - Variance of returns“ - Regel bezeichnet. Sie beschreibt eine Annahme über das Verhalten einer Investorin oder eines Investors. Nach dieser versucht diese oder dieser stets die zukünftig erwarteten Renditen zu maximieren und gleichzeitig die Varianz der Renditen zu minimieren (vgl. [5], S. 77). Daraus folgt nach Markowitz (1952, S.89) auch ein gewisses Maß an Diversifikation in den resultierenden Portfolios. Diversifikation bedeutet hier, dass man Anlagen so auswählen sollte, dass diese möglichst niedrig oder negativ miteinander korreliert sind. Denn derartige Korrelationen bzw. Varianzen signalisieren ein niedriges Risiko. Auch intuitiv betrachtet ist dieser Sachverhalt einleuchtend. Denn betrachtet man ein Portfolio mit niedrig korrelierten Aktien, so werden sich Änderungen einer Aktie mit jenen der anderen immer etwas ausgleichen. Im Gegensatz dazu werden bei einem Portfolio, welches hohe Korrelationen aufweist, beim Fall oder Anstieg einer Aktie tendenziell auch alle anderen fallen oder steigen, wodurch das Risiko eines solchen Portfolios sehr hoch ist. Nach Markowitz sind wenig diversifizierte Portfolios damit sehr risikoreich und diversifizierte Portfolios eher risikoarm. Man spricht in diesem Kontext auch vom sogenannten „Diversifikationseffekt“.

Klar ist natürlich auch, dass es, bezogen auf die „Expected returns - Variance of returns“ - Regel, kaum ein Portfolio gibt, welches die maximal erwartete Rendite erzielt und gleichzeitig minimale Varianz besitzt. Fixiert man jedoch eines der beiden, so kann man auf zwei Arten effiziente Portfolios berechnen. Entweder fixiert man die erwartete Rendite und minimiert die dazugehörige Varianz, oder man hält die Varianz fest und maximiert die dazugehörige Rendite. Beides liefert Optimierungsprobleme, wobei wir in Hinblick auf den nächsten Abschnitt den ersten Fall betrachten möchten.

Wir fixieren also die Rendite und versuchen dazu ein Portfolio mit minimaler Varianz, also minimalem Risiko, zu finden.

3.1 Effiziente Portfolios: Vorhersage und Realisierung

In diesem Abschnitt werden wir immer zwei Jahre betrachten und versuchen, mit Hilfe der Daten aus dem ersten dieser beiden Jahre, ein Portfolio mit minimalem Risiko zu einer von uns vorgegebenen Rendite zu berechnen. Um dies noch einmal zu verdeutlichen, möchte ich es am Beispiel der Jahre 2004 und 2005 erklären. Wir verwenden demnach die Kovarianzmatrix, die wir aus den Daten von 2004 erhalten, geben uns anschließend verschiedene Erträge des Portfolios vor und versuchen uns dann ein Portfolio zu berechnen, welches zu diesem vorgegebenen Ertrag das minimale Risiko besitzt. Dieses Ergebnis vergleichen wir anschließend mit den tatsächlich eingetretenen Daten aus dem Folgejahr 2005. Analog funktioniert der Vorgang für die anderen Perioden.

Sei $\omega \in \mathbb{R}^N$ nun unser gesuchtes Portfolio, bestehend aus N Aktien, und $R \in \mathbb{R}^N$ der Vektor der Durchschnittsrenditen jeder Aktie aus dem ersten Jahr, also

$$R = \begin{pmatrix} \widehat{r}_1 \\ \vdots \\ \widehat{r}_N \end{pmatrix},$$

mit \widehat{r}_i wie in Gleichung 2.1. Der Ertrag RE eines bestimmten Portfolios ergibt sich dann durch die Formel

$$RE = \omega^T R.$$

Um das Risiko zu einem vorgegebenen Ertrag zu minimieren, stellt sich natürlich zuerst die Frage, welche Renditen wir für die einzelnen Aktien wählen sollten. Da wir in dieser Studie allerdings den Fokus auf die Risikovorhersage und nicht auf eine möglichst akkurate Prognose der Renditen legen wollen, bedienen wir uns in diesem Schritt einfach an den tatsächlich eingetretenen Daten des Folgejahres, welches in der ersten Periode das Jahr 2005 wäre. Wir gehen damit also von perfekten Vorhersagen unserer Renditen aus (vgl. [1], S. 9).

Über die schon bekannte Korrelationsmatrix C (vgl. Gleichung 2.3) und den Standardabweichungen der logarithmischen Renditen σ_i , $i = 1, \dots, N$ (vgl. Gleichung 2.2) können wir nun auch für jedes Jahr die Kovarianzmatrix Σ_R berechnen. Dies geschieht mit Hilfe der Formel

$$\Sigma_R = \text{diag}(\sigma_1, \dots, \sigma_N) C \text{diag}(\sigma_1, \dots, \sigma_N). \quad (3.1)$$

Mit Hilfe dieser Matrix können wir nun das Risiko RI anschreiben als

$$RI = \omega^T \Sigma_R \omega.$$

Gesucht ist nun ein Portfolio $\omega \in \mathbb{R}^N$, sodass,

$$RE = \omega^T R \quad \text{und} \quad RI = \omega^T \Sigma_R \omega \rightarrow \min \quad (3.2)$$

wobei die Einträge von ω durch

$$\sum_{i=1}^N \omega_i = 1$$

begrenzt sein sollen. Zusätzlich nehmen wir vorerst an, dass

$$\omega_i \geq 0, \quad \forall i = 1, \dots, N$$

gilt, also kein Leerverkauf erlaubt ist. Durch diese Methodik aus der Portfoliotheorie erhält man für einen fixen Ertrag RE ein Portfolio mit minimalem Risiko RI. Die möglichen Erträge sollen dabei einerseits positiv sein, da wohl niemand bei einer Investition an einem negativen Ertrag, also einem Verlust, interessiert ist, und andererseits zwischen den minimal und maximal möglichen Erträgen liegen, welche aus den Durchschnittsrenditen des Folgejahres hergeleitet werden. Fasst man all diese erhaltenen Ergebnisse als Punkte (RI, RE) in der Ebene auf, so erhält man die in Abbildung 3.1 dargestellten Graphen, die im englischen als „Efficient frontiers“ bezeichnet werden. „Efficient frontiers“ zeigen dabei alle Portfolios mit minimalem Risiko zu einem fixierten Ertrag (vgl. [1], S.8f). Alle Portfolios, die rechts von der „Efficient frontier“ liegen, können entweder bei gleichem Risiko einen höheren Ertrag erzielen, oder bei gleichem Ertrag mit geringerem Risiko bzw. geringerer Volatilität auskommen. Diese sind also nicht effizient. Auf der linken Seite des Graphen kann es klarerweise auch keine Portfolios geben. Denn würde es ein solches geben, so wäre die „Efficient frontier“ falsch berechnet worden.

In Abbildung 3.1 sieht man nun insgesamt sechs Graphiken, die für jede Periode die vorhergesagten und realisierten „Efficient frontiers“ abbilden. Jeweils auf der horizontalen Achse ist dabei das Risiko RI und auf der vertikalen Achse der Ertrag (Return) RE eines Portfolios dargestellt. Jede Graphik besteht aus einem grauen und einem schwarzen Graphen. Der graue Graph stellt das vorhergesagte Risiko des ersten Jahres und der schwarze Graph das tatsächlich eingetretene Risiko des Folgejahres dar. Um den grauen Graphen für das Jahrespaar 2004-2005 zu erhalten, wurde zuerst die Kovarianzmatrix aus der Korrelationsmatrix C von 2004 und den Standardabweichungen aus dem Jahr 2005 berechnet (vgl. Gleichung 3.1). Für den schwarzen Graphen, der die realisierten Ergebnisse zeigt, werden sowohl Korrelationsmatrix also auch Standardabweichungen aus dem Folgejahr 2005 verwendet. In

beiden Berechnungen werden die Standardabweichungen des Folgejahres benutzt, um in den Ergebnissen den alleinigen Effekt der Korrelationsmatrizen zu sehen. Anschließend wird das minimale Risiko in beiden Fällen mit den Renditen des Folgejahres berechnet (vgl. Gleichung 3.2), um den Unterschied zwischen Vorhersage (2004) und Realisierung (2005) auch unabhängig von einer etwaigen Vorhersage der Renditen darzustellen.

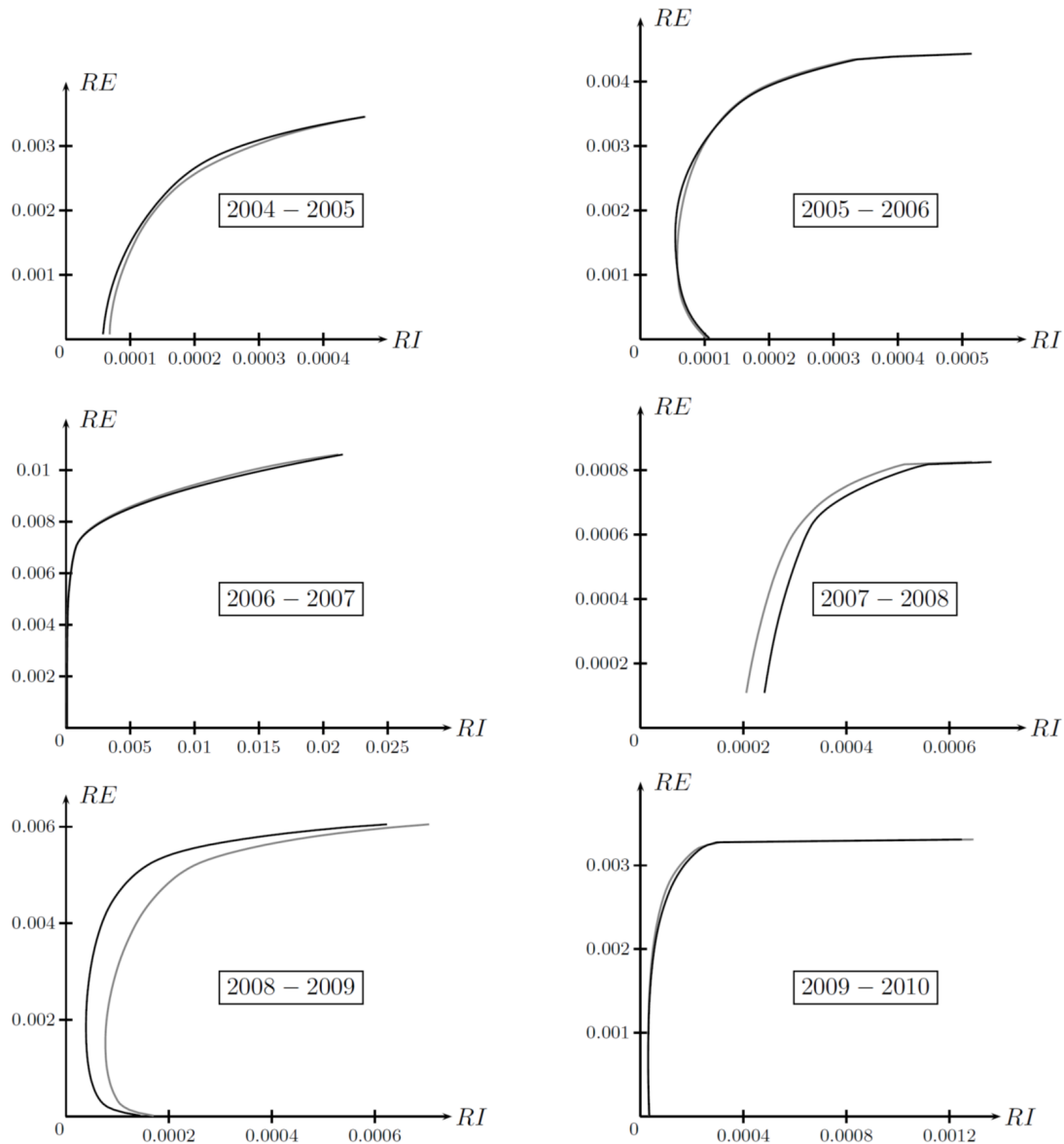


Abbildung 3.1: Efficient frontiers, entnommen aus [1], S. 10

In den sechs Graphiken unterscheiden sich die Kurven einerseits in ihrer allgemeinen Form und andererseits auch in der Höhe der Risiken und Erträge. Beispielsweise lie-

gen die Erträge und Risiken der Jahre 2006-2007 deutlich über jenen aller anderen Jahre. Diese Tatsache liegt hauptsächlich am Telekommunikationsunternehmen Telebras, das in diesen Jahren aufgrund von Gerüchten, wonach die Regierung über dieses Unternehmen Internetdienste für 64% aller Haushalte anbieten würde, stark an Wert zulegte. Da dies allerdings nie verwirklicht wurde, verlor die Aktie dieses Unternehmens wieder stark an Wert. An diesem Beispiel erkennt man auch, welche großen Auswirkungen ein einzelnes Unternehmen auf ein ganzes Portfolio haben kann. Außerdem fällt auf, dass der graue Graph manchmal links und manchmal rechts vom schwarzen Graphen liegt, das Risiko also in einigen Jahren unter- und in anderen Jahren überschätzt wird (vgl. [1], S. 9 ff.). Dass eine Unterschätzung des Risikos für Anlegerinnen oder Anleger sehr gefährlich sein kann, liegt auf der Hand.

Interessant sind insbesondere auch die Jahre von 2007 bis 2009 im Zusammenhang mit der „Subprime Mortgage crisis“, einer Krise auf dem amerikanischen Immobilienmarkt. „Subprime Mortgages“ bezeichnet zweitklassige Hypotheken. Diese Darlehen wurden auch an Kreditnehmer mit geringer Kreditwürdigkeit ausgegeben. Das Ausfallrisiko wurde vom Kreditgeber auf Grund der positiven wirtschaftlichen Lage in den Jahren bis 2006 als sehr gering eingestuft. Außerdem führten niedrige Zinssätze und steigende Häuserpreise zu einer erhöhten Nachfrage nach solchen Darlehen, dessen Zinssätze allerdings variabel waren und je nach wirtschaftlicher Lage angepasst werden konnten. Wider Erwarten kam es in den folgenden Jahren zu steigenden Zinssätzen und fallenden Häuserpreisen, wodurch viele Kreditnehmer ihren Verpflichtungen nicht mehr nachkommen konnten und dazu veranlasst wurden, aus dem Kredit auszutreten. Weltweit verloren unzählige Anlegerinnen und Anleger im späteren Verlauf große Beträge und auch Banken mussten in Konkurs gehen (vgl. [6]).

Im Jahr 2008 erreichten die Auswirkungen dieser Hypothekenkrise auch den brasilianischen Aktienmarkt. Betrachtet man die Jahre 2007-2008 in Abbildung 3.1, so sieht man, dass mit Hilfe der Daten aus dem Jahr 2007, welches sehr geringe Volatilität aufwies, das Risiko für das Jahr 2008 grob unterschätzt wurde. Andererseits wurde das Risiko für das Jahr 2009 durch die Daten des Jahres 2008, welches aufgrund der „Subprime Mortgage Crisis“ sehr volatil war, stark überschätzt.

Verwendet man statt der bisher verwendeten Korrelationsmatrix C die gesäuberte Korrelationsmatrix C_{clean} aus Gleichung 2.4 für die Risikovorhersage, so erhalten wir für die Vergleichsjahre 2004-2005 die „Efficient frontier“ in Abbildung 3.2. Obwohl wir uns vom Säuberungsverfahren in Abschnitt 2.3 eigentlich bessere Ergebnisse erhofft hätten, so zeigt dieser Graph bei genauer Betrachtung sogar teilweise eine etwas schlechtere Vorhersage für das Jahr 2005 als in Abbildung 3.1. In welchen Jahren die gesäuberte und in welchen die ursprüngliche Korrelationsmatrix zu besseren Vorhersagen führte, werden wir im letzten Abschnitt dieses Kapitels noch an Zahlen festmachen.

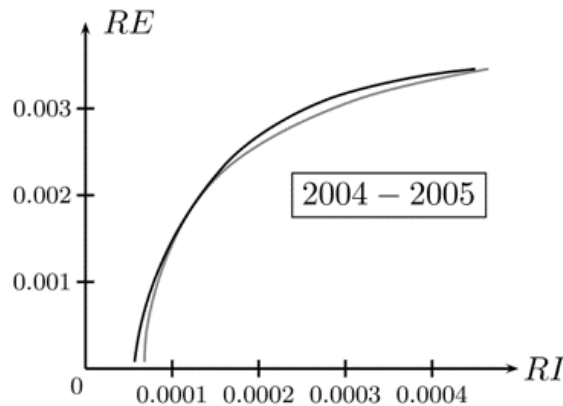


Abbildung 3.2: Efficient frontiers von 2004-2005 mit C_{clean} , entnommen aus [1], S. 11

3.2 Reduzierung des Markteffektes durch lineare Regression

Aufgrund der Volatilität des Marktes stellt die Vorhersage des zukünftig erwarteten Risikos eine schwierige Aufgabe dar. Denn neben der Abhängigkeit der Vermögenswerte untereinander hängen diese auch noch vom Markt ab. Man denke dabei an das bereits beschriebene Beispiel der „Subprime Mortgage Crisis“, in der die unterschiedlichen Volatilitäten in den Jahren 2007 und 2008 deutlich die Risikovorhersagen beeinflusst haben (vgl. [1], S. 12).

Um dieses sogenannte „systemische Risiko“ von unseren Daten zu entfernen, wird ein „Single Index Model“ verwendet. Bei einem solchen geht man davon aus, dass der Großteil des systemischen Risikos durch einen einzigen Index abgebildet werden kann (vgl. [1], S. 2). Dabei wird die logarithmische Rendite einer Aktie R_t zu einem Zeitpunkt t geschrieben als

$$R_t = a + bI_t + E_t,$$

wobei I_t den Marktindex zum Zeitpunkt t beschreibt, die Koeffizienten a und b durch lineare Regression geschätzt werden und E_t den dabei gemachten Fehler bezeichnet. Für den Marktindex liegt es nahe den Ibovespa zu verwenden. Als Alternative zum Ibovespa wird in unserer Studie aber der Index verwendet, der durch die logarithmischen Renditen jenes Aktienportfolios erhalten wird, welches dem Eigenvektor des größten Eigenwertes der jeweiligen Korrelationsmatrix entspricht (vgl. [1], S. 12). Denn wie wir in Kapitel 2 gesehen haben, korreliert dieser stark mit dem Ibovespa und führt daher zu ähnlichen Ergebnissen. Bildet man nun Korrelationsmatrizen zwischen den Restwerten der einzelnen Aktienrenditen, so erkennt man an den daraus

gebildeten „Efficient frontiers“ in den Jahren 2004-2005 deutlich bessere Vorhersagen (vgl. Abbildung 3.3):

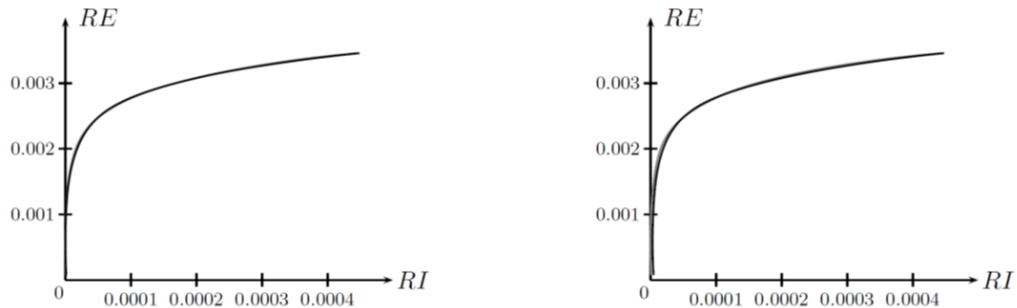


Abbildung 3.3: Efficient frontiers von 2004-2005 mit Regression, entnommen aus [1], S. 12

In beiden Graphiken dieser Abbildung wird das vorhergesagte Risiko von den Korrelationsmatrizen der Restrenditen berechnet. Auf der rechten Seite wird diese Korrelationsmatrix zusätzlich noch gesäubert, also die Eigenwerte innerhalb der „Wishart-Region“ werden durch ihren Durchschnitt ersetzt (vgl. Abschnitt 2.3). Die so mit den Daten von 2004 vorhergesagten Risiken und damit erhaltenen Kurven sind von den Kurven der tatsächlichen eingetretenen Daten kaum noch zu unterscheiden. Wir sehen also, dass unser Regressionsmodell die Risikovorhersage deutlich verbessert.

3.3 Ergebnisse in Zahlen

Um die Qualität der Vorhersagen der verschiedenen Jahre zu vergleichen, werden wir uns in diesem Abschnitt mit verschiedenen Fehlerschätzern auseinandersetzen, welche die Abweichungen zwischen den prognostizierten und den tatsächlich realisierten Daten bzw. den Abstand zwischen den grauen und schwarzen Kurven in Abbildung 3.1 quantifizieren sollen. Die Ergebnisse der einzelnen Fehlermaße AG (englisch für „Agreement“), MSE (Mean Squared Error) und dem Winkel zwischen Vektoren, welche im Folgenden genauer dargelegt werden, sind in den Tabellen 3.1 und 3.2 dargestellt.

In beiden werden die Ergebnisse nach Perioden (von 2004-2005 bis 2009-2010) unterteilt. Man betrachtet einmal jene Resultate, die mit der ungesäuberten Korrelationsmatrix („Without Cleaning“) berechnet wurden und einmal jene, bei denen die Korrelationsmatrix zuvor gesäubert wurde („With Cleaning“). Auch hier wird noch einmal unterteilt, ob die entsprechende Korrelationsmatrix von den ursprünglichen Daten (No Regression) oder von den Restrenditen des Regressionsmodells berechnet wurde („Regression“). Das beste Ergebnis, also der betragsmäßig kleinste Fehler, ist

in jeder Zeile fett markiert. Tabelle 3.1 zeigt die Abweichungen für Portfolios ohne Leerverkauf, also

$$\omega_i \geq 0, \forall i = 1, \dots, N, \quad (3.3)$$

und Tabelle 3.2 enthält die Resultate für Portfolios mit Leerverkauf, wobei die Einträge des Portfolio-Vektors ω in diesem Fall durch die Vorgabe

$$-1 \leq \omega_i \leq 2, \forall i = 1, \dots, N \quad (3.4)$$

beschränkt sind. Leerverkauf ist an sich nicht bei allen finanziellen Transaktionen frei erlaubt, da er das Risiko eines Portfolios deutlich erhöhen kann. Dennoch stellt er für viele Anlegerinnen und Anleger ein wichtiges Werkzeug dar, weshalb auch auf diesen Fall eingegangen wird (vgl. [1], S. 14).

Um die Übereinstimmung der Kurven zu messen, betrachten wir als erstes den Abweichungsschätzer AG, der über die Formel

$$AG = \frac{1}{n} \sum_{i=1}^n \frac{RI_i^{real} - RI_i^{pred}}{RI_i^{pred}}$$

definiert ist. Dabei ist RI_i^{real} das realisierte Risiko und RI_i^{pred} das vorhergesagte Risiko, jeweils für $i = 1, \dots, n$ Werte von fixierten Renditen. Ist AG positiv, so ist das realisierte Risiko im Schnitt größer als das vorhergesagte. Wenn AG negativ ist, dann ist das vorhergesagte Risiko größer als das realisierte Risiko. Dieses „Agreement“-Maß hat leider auch zwei Schwächen. Einerseits wird es aus einer Summe von teilweise positiven und teilweise negativen Werten gebildet, welche sich gegenseitig auslöschen können. Andererseits steht im Nenner die Rendite RI_i^{pred} , die, falls sie nahe bei Null liegt, zu betragsmäßig großen Summanden führen kann. Damit können kleine Renditen bei der Berechnung des AG zu stark gewichtet werden (vgl. [1], S. 11).

Deshalb betrachten wir im nächsten Schritt einen Fehlerschätzer, bei dem dies nicht der Fall ist, nämlich den Mean Squared Error (MSE), welcher definiert ist als

$$MSE = \frac{1}{n} \sum_{i=1}^n \left(RI_i^{real} - RI_i^{pred} \right)^2.$$

Wie der „Agreement“-Schätzer, hängt natürlich auch dieser Wert von der Anzahl der eingetragenen Punkte in der jeweiligen Graphik in Abbildung 3.1 ab, die in den Formeln der Fehlerschätzer mit n bezeichnet wird. In jeder Graphik in Abbildung 3.1 wurden insgesamt 100 Punkte eingezeichnet (vgl. [1], S. 11).

Zusätzlich zu diesen beiden Fehlermaßen wird weiters der Winkel (engl. Angle) zwischen den beiden Vektoren $RI^{pred} (\in \mathbb{R}^n)$ und $RI^{real} (\in \mathbb{R}^n)$ berechnet. Aus der linearen Algebra ist bekannt, dass sich der Winkel θ aus der Formel

$$\cos \theta = \frac{\langle RI^{pred}, RI^{real} \rangle}{\|RI^{pred}\| \cdot \|RI^{real}\|} = \frac{\sum_{i=1}^n RI_i^{pred} RI_i^{real}}{\sqrt{\sum_{i=1}^n (RI_i^{pred})^2} \sqrt{\sum_{i=1}^n (RI_i^{real})^2}}$$

ergibt, wobei $\langle \cdot, \cdot \rangle$ das euklidische Skalarprodukt und $\| \cdot \|$ die daraus abgeleitete euklidische Norm darstellt. Dieses Maß ist im Vergleich zum „Agreement-Maß“ nicht abhängig von der Volatilität der Risiken und Renditen (vgl. [1], S. 13).

Agreement (AG)		Without Cleaning		With Cleaning	
Previous-Predicted		No Regression	Regression	No Regression	Regression
2004 – 2005		-0.073	0.547	-0.066	-3.394
2005 – 2006		-0.017	0.177	-0.039	-0.476
2006 – 2007		0.186	0.064	0.225	-0.486
2007 – 2008		0.119	0.314	0.143	0.330
2008 – 2009		-0.429	0.082	-0.439	-0.872
2009 – 2010		0.073	0.737	0.033	-3.876

Mean Squared Error (MSE)		Without Cleaning		With Cleaning	
Previous-Predicted		No Regression	Regression	No Regression	Regression
2004 – 2005		10.50×10^{-11}	0.46×10^{-11}	24.99×10^{-11}	2.13×10^{-11}
2005 – 2006		2.01×10^{-11}	0.80×10^{-11}	2.60×10^{-11}	2.85×10^{-11}
2006 – 2007		10.87×10^{-8}	11.03×10^{-8}	0.12×10^{-8}	1.53×10^{-8}
2007 – 2008		10.56×10^{-10}	5.13×10^{-10}	15.19×10^{-10}	7.66×10^{-10}
2008 – 2009		39.79×10^{-10}	2.29×10^{-10}	35.94×10^{-10}	0.59×10^{-10}
2009 – 2010		8.53×10^{-11}	11.18×10^{-11}	6.79×10^{-11}	4.77×10^{-11}

Angle (θ)		Without Cleaning		With Cleaning	
Previous-Predicted		No Regression	Regression	No Regression	Regression
2004 – 2005		1.74	0.71	2.10	1.49
2005 – 2006		1.82	1.38	2.10	1.64
2006 – 2007		1.08	0.29	1.12	0.66
2007 – 2008		1.55	6.83	1.31	6.28
2008 – 2009		10.22	2.37	11.52	2.41
2009 – 2010		3.42	3.91	2.05	2.55

Tabelle 3.1: AG, MSE, Winkel (θ) (in Grad) - ohne Leerverkauf, entnommen aus [1], S. 13

Verzichtet man auf Leerverkauf (vgl. Bedingung 3.3) wie in Tabelle 3.1, so erhält man folgende Ergebnisse:

Für den „Agreement“ - Schätzer sind die besten Ergebnisse sehr verstreut. Sie liegen einerseits bei den Vorhersagen durch nicht gesäuberte Korrelationsmatrizen, unabhängig von der Verwendung des Regressionsmodells, und andererseits bei den Prognosen durch gesäuberte Korrelationsmatrizen, aber in diesem Fall ohne die Verwendung der Restrenditen des Regressionsmodells. Die schlechten Resultate für das Regressionsmodell lassen sich dadurch erklären, dass die Renditen und Risiken der Restdaten kleiner sind als die der originalen Daten. (Original meint hierbei, dass nicht die Restrenditen, sondern die originalen Renditen ohne Verwendung des Regressionsmodells zur Erstellung der Korrelationsmatrizen verwendet werden.) Dadurch dividiert

man durch kleinere Werte, wodurch die Summanden des „Agreement“-Schätzers, wie oben erwähnt, größer werden. Im Fall des Mean Squared Errors zeigt das Regressionsmodell schon mehr Wirkung. Die geringsten Fehler erhielt man dabei bei der Anwendung des Regressionsmodells und keiner Säuberung der Korrelationsmatrizen der Restrenditen, manchmal auch für die gesäuberte Korrelationsmatrix, allerdings hierbei ohne dem Regressionsmodell. Noch deutlicher spiegelt sich dieses Ergebnis beim Winkel zwischen den beiden Vektoren RI^{pred} und RI^{real} ab. Die kleinsten Unterschiede bekommt man hier deutlich für die ungesäuberten Korrelationsmatrizen der Restrenditen. Nur in wenigen Perioden führten die gesäuberten Korrelationsmatrizen der originalen Daten beim Winkelmaß zu den besten Vorhersagen (vgl. [1], S. 13 f.).

Wird nun auch Leerverkauf unter der Beschränkung 3.4 zugelassen, so ergeben sich die Ergebnisse aus Tabelle 3.2, die etwas genauere Vorzüge im Hinblick auf die unterschiedlichen Methoden erkennen lassen:

Agreement (AG) Previous-Predicted	Without Cleaning		With Cleaning	
	No Regression	Regression	No Regression	Regression
2004 – 2005	0.588	0.296	0.528	0.162
2005 – 2006	0.302	0.291	0.608	0.185
2006 – 2007	-0.004	0.048	0.494	0.273
2007 – 2008	0.537	0.568	0.690	0.486
2008 – 2009	0.125	0.024	0.655	-0.128
2009 – 2010	0.398	0.433	0.529	0.239

Mean Squared Error (MSE) Previous-Predicted	Without Cleaning		With Cleaning	
	No Regression	Regression	No Regression	Regression
2004 – 2005	31.82×10^{-6}	30.43×10^{-6}	6.24×10^{-6}	4.68×10^{-6}
2005 – 2006	16.10×10^{-6}	31.55×10^{-6}	11.52×10^{-6}	2.62×10^{-6}
2006 – 2007	2.11×10^{-5}	7.30×10^{-5}	1.57×10^{-5}	4.05×10^{-5}
2007 – 2008	6.98×10^{-3}	10.36×10^{-3}	5.16×10^{-3}	11.21×10^{-3}
2008 – 2009	0.60×10^{-4}	5.01×10^{-4}	11.80×10^{-4}	1.00×10^{-4}
2009 – 2010	3.50×10^{-4}	4.42×10^{-4}	4.68×10^{-4}	2.93×10^{-4}

Angle (θ) Previous-Predicted	Without Cleaning		With Cleaning	
	No Regression	Regression	No Regression	Regression
2004 – 2005	4.70	1.96	4.71	1.29
2005 – 2006	2.37	2.16	6.91	3.36
2006 – 2007	0.69	0.46	2.15	2.92
2007 – 2008	4.50	3.54	11.28	8.09
2008 – 2009	2.69	2.33	2.51	1.65
2009 – 2010	1.75	2.36	4.37	1.66

Tabelle 3.2: AG, MSE, Winkel (θ) - mit Leerverkauf, entnommen aus [1], S. 14

Beim „Agreement“-Schätzer erhält man die beste Übereinstimmung zwischen Vorhersage und Realisierung durch Verwendung der gesäuberten Korrelationsmatrix der

Restrenditen. Auch beim Mean Squared Error liefern die gesäuberten Korrelationsmatrizen die besten Ergebnisse. Hierbei ist es jedoch unerheblich, ob die Korrelationsmatrix von den originalen Renditen oder den Restrenditen gebildet wird. Für den Winkel ergeben sich die besten Resultate für die Korrelationsmatrix der Restrenditen. Dabei macht die Säuberung der Matrix keine Unterschiede (vgl. [1], S. 14).

Neben diesen drei Abweichungs-Schätzern gibt es auch andere, bei denen direkt der Unterschied zwischen zwei Matrizen gemessen wird. Für diese Methoden werden die Renditen und Risiken der „Efficient frontiers“ nicht benötigt. Auch das Verwenden oder Verzichten von Leerverkäufen macht dabei keinen Unterschied. Betrachtet man also zwei quadratische Matrizen $A, B \in R^{N \times N}$ so sei der Distanz-Schätzer (Dist) zwischen diesen Matrizen definiert als

$$\text{Dist} = \text{Tr} \left((A - B)^T (A - B) \right), \quad (3.5)$$

wobei $\text{Tr}(\cdot)$ hier die Spur (engl. trace) dieses Matrixproduktes bezeichnet. Diese entspricht der Summe aller Diagonaleinträge einer Matrix. Vergleicht man nun die Korrelationsmatrizen, die aus den vergangenen Daten gebildet werden, mit den Korrelationsmatrizen der tatsächlich realisierten Daten des Folgejahres, und berechnet man durch Gleichung 3.5 deren Distanz, so ergeben sich die Ergebnisse, welche in Tabelle 3.3 ersichtlich sind. Auch hierbei wurden die Ergebnisse für die jeweiligen Perioden wieder in alle vier möglichen Kombinationen (ungesäuberte Korrelationsmatrix ohne Regression, ungesäuberte Korrelationsmatrix mit Regression, gesäuberte Korrelationsmatrix ohne Regression, gesäuberte Korrelationsmatrix mit Regression) unterteilt. Zu beachten ist auch, dass die Werte dieser Matrix-Distanz sehr von den Größen bzw. Dimensionen der jeweiligen Korrelationsmatrizen abhängen, weshalb die Ergebnisse einer Periode nicht mit einer anderen verglichen werden können. Dennoch ist für jede einzelne Periode deutlich zu erkennen, dass die kleinsten Unterschiede zwischen den vorhergesagten und realisierten Korrelationen durch die gesäuberte Korrelationsmatrix und der Verwendung der Restrenditen aus dem Regressionsmodell erreicht werden (vgl. [1], S. 14 f.).

Distance (Dist) Previous-Predicted	Without Cleaning		With Cleaning	
	No Regression	Regression	No Regression	Regression
2004 – 2005	50	37	38	25
2005 – 2006	68	52	49	36
2006 – 2007	112	78	92	52
2007 – 2008	222	149	202	112
2008 – 2009	794	277	695	172
2009 – 2010	268	226	182	133

Tabelle 3.3: Distanz zwischen den Korrelationsmatrizen, entnommen aus [1], S. 15

4 Conclusio

In dieser Arbeit wurde eine Studie genauer beleuchtet, bei der über insgesamt sechs Zwei-Jahres-Perioden hinweg, von 2004 bis 2010, eine bestimmte Anzahl an Kursen von 100% liquiden Aktien an der brasilianischen Börse beobachtet wurden. Die Daten der Aktienrenditen, die aus diesen erfassten Daten erhalten wurden, wurden anschließend dazu verwendet, um jeweils das zukünftig erwartete Risiko eines aus diesen Aktien bestehenden Portfolios im Folgejahr zu prognostizieren. Dafür wurden aus den Aktienrenditen des jeweiligen ersten Jahres einer Periode Korrelationsmatrizen gebildet, welche die Abhängigkeit der einzelnen Aktien untereinander beschreiben.

Danach wurden die Eigenwerte dieser Matrix berechnet und mit den Eigenwerten einer Zufallsmatrix verglichen. Durch diese übliche Methodik aus der Zufallsmatrixtheorie konnte man jene Eigenwerte ausfindig machen, die, im Vergleich mit echten Daten, als verrauschte Daten oder Störsignale bezeichnet werden. Auch die Eigenvektoren zu diesen Eigenwerten brachten keine klaren Ergebnisse. Neben den verrauschten Eigenwerten innerhalb der „Wishart-Region“ zeigten sich allerdings ebenso Eigenwerte, die nicht mit den zufälligen Ergebnissen übereinstimmten. Bei der Analyse der Eigenvektoren dieser Eigenwerte stellte man fest, dass sich der Eigenvektor des größten Eigenwertes (als Portfolio betrachtet) ähnlich wie der Marktindex verhält und die Eigenvektoren der kleinsten Eigenwerte kaum mit dem Markt korrelieren und auch wenig Risiko beinhalten.

Das Ziel war, auf Basis der Daten aus dem ersten Jahr einer Periode, möglichst risikoarme Portfolios im Sinne der Markowitz'schen Portfoliotheorie zu konstruieren bzw. dieses Risiko bestmöglich zu prognostizieren. Um das Risiko der Portfolios besser vorherzusagen, wurden zwei Methoden eingesetzt. Einerseits wurde eine neue gesäuberte Korrelationsmatrix gebildet, indem man die Eigenwerte innerhalb der „Wishart-Region“ durch ihren Durchschnitt ersetzt hat und andererseits wurde durch ein Regressionsmodell der Markteffekt von den Renditen entfernt. Mit den daraus entstehenden Restrenditen wurden wieder neue Korrelationsmatrizen gebildet, welche erneut einmal von den verrauschten Daten gesäubert wurden und einmal nicht.

Schließlich erhielt man vier verschiedene Korrelationsmatrizen, auf deren Basis Risikovorhersagen getroffen wurden. Um die Vorhersagen mit den realisierten Daten des Folgejahres zu vergleichen und jene Methode zu finden, welche die präziseste Ri-

sikovorhersage ermöglicht, wurden die Ergebnisse zuerst durch „Efficient frontiers“ veranschaulicht. Es zeigte sich dabei, dass die Vorhersage sehr von der Volatilität jenes Jahres abhängt, dessen Daten der Vorhersage zu Grunde gelegt wurden. War die Volatilität im ersten Jahr geringer als im Folgejahr, so wurde das Risiko eher unterschätzt, war sie jedoch höher, so wurde das Risiko überschätzt.

Anschließend wurden die Abweichungen der vorhergesagten und der eingetretenen Daten mit Hilfe von verschiedenen Fehlerschätzern, wie dem „Agreement“ - Schätzer (AG), dem Mean Squared Error (MSE) und dem Winkel zwischen vorhergesagtem und realisiertem Risikovektor, ausgewertet. Hierbei wurde auch zwischen Portfolios mit und ohne Leerverkauf unterschieden. Zusätzlich wurde noch eine Matrix-Distanz zwischen den verschiedenen Korrelationsmatrizen des ersten Jahres und der tatsächlichen Korrelationsmatrix des Folgejahres berechnet.

Bei diesen Auswertungen führte die Säuberung der Korrelationsmatrizen allein leider zu keinen Verbesserungen. Im Gegensatz dazu lieferte das Regressionsmodell in zumindest 65% der Fälle des Mean Squared Errors und des Winkels genauere Vorhersagen. Kombinierte man das Regressionsmodell mit dem Säubern der Korrelationsmatrizen, so ergaben sich für Portfolios ohne Leerverkauf in 65% der Fälle des „Agreement“ - Schätzers und in 83% der Fälle des Mean Squared Errors und des Winkels präzisere Ergebnisse. Wurde Leerverkauf zugelassen, so führte dies beim „Agreement“ - Schätzer in 65%, beim Mean Squared Error in 83% und beim Winkel in 50% der Fälle zu genaueren Vorhersagen (vgl. [1], S.20). Bei der Matrix-Distanz erwies sich die Kombination beider Methoden in jeder Periode als beste Vorhersage. Insgesamt kann man sagen, dass die Verwendung des Regressionsmodells tendenziell zu besseren Risikovorhersagen führt, das Säubern der Korrelationsmatrix allerdings in dieser Studie eher mit Vorsicht zu genießen ist.

Abbildungsverzeichnis

2.1	Verteilung der Eigenwerte	8
2.2	Quantil-Quantil-Diagramme	9
2.3	Verteilung der Eigenwerte mit Simulationen	9
2.4	Eigenvektoren	11
3.1	Efficient frontiers	18
3.2	Efficient frontiers von 2004-2005 mit C_{clean}	20
3.3	Efficient frontiers von 2004-2005 mit Regression	21

Tabellenverzeichnis

3.1	AG, MSE, Winkel (θ) - ohne Leerverkauf	23
3.2	AG, MSE, Winkel (θ) - mit Leerverkauf	24
3.3	Distanz zwischen den Korrelationsmatrizen	25

Literaturverzeichnis

- [1] Leonidas Sandoval Junior, Adriana Bruscato und Maria Kelly Venezuela. (2013), *Building portfolios of stocks in the São Paulo Stock Exchange using Random Matrix Theory*, <https://arxiv.org/pdf/1201.0625.pdf> (14.10.2019).
- [2] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud und Marc Potters. (1998), *Noise Dressing of Financial Correlation Matrices*, <https://arxiv.org/pdf/cond-mat/9810255v1.pdf> (05.12.2019).
- [3] T. Conlon, H.J. Ruskin und M. Crane. (2010) *Cross-Correlation Dynamics in Financial Time Series*, <https://arxiv.org/pdf/1002.0321v1.pdf> (12.12.2019).
- [4] Hirdesh K. Pharasi, Kiran Sharma, Anirban Chakraborti und Thomas H. Seligman. (2018) *Complex market dynamics in the light of random matrix theory*, <https://arxiv.org/pdf/1809.07100.pdf> (06.12.2019).
- [5] Harry M. Markowitz. (1952), *Portfolio Selection*, The Journal of Finance, S. 77-91.
- [6] Institut der deutschen Wirtschaft Köln Medien GmbH. (2010), *Subprime Mortgage Crisis*, <https://www.wirtschaftundschule.de/wirtschaftslexikon/s/subprime-mortgage-crisis/> (21.02.2020).