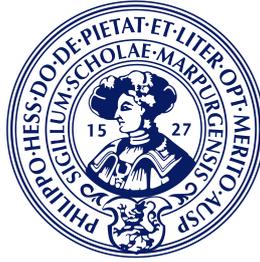


Philipps



Universität
Marburg

Masterarbeit

im Studiengang Wirtschaftsmathematik
am Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg

Berücksichtigung von Schätzunsicherheit bei der Portfolio-Optimierung nach Markowitz

eingereicht von
Sabrina Thorn

25. Januar 2013

Betreuer: Prof. Dr. H. Holzmann

Inhaltsverzeichnis

1	Einleitung	1
2	Portfoliotheorie nach Markowitz	3
2.1	Grundlagen	3
2.1.1	Risiko	3
2.1.2	Risikodiversifikation	3
2.1.3	Statistisches Modell für Assetpreise und Renditen	5
2.2	Mean-Variance-Optimierung	7
2.2.1	Modell	7
2.2.2	Minimum-Varianz-Kurve und Effizienzkurve	7
2.2.3	Formulierung und Lösung des Mean-Variance-Optimierungsproblems	10
2.3	Darstellung von Minimum-Varianz-Portfolios und die Geometrie der Minimum-Varianz-Kurve	20
2.4	Das Markowitz'sche Optimierungsenigma	21
3	Verschiedene Ansätze zum Schätzen von μ, Σ und w^*	23
3.1	Empirische Schätzer für μ und Σ	23
3.2	Multifaktor-Pricing-Modelle	25
3.3	Bootstrapping und die <i>resampled frontier</i>	27
3.4	Bayes- und Shrinkage-Schätzer für μ und Σ	27
3.4.1	Bayes-Schätzer	28
3.4.2	Empirische Bayes-Schätzer	32
3.4.3	Shrinkage-Schätzer	32
4	Verteilung der empirischen Minimum-Varianz-Kurve	34
4.1	Analyse der empirischen Effizienzmengenkonstanten	34
4.2	Die vier Minimum-Varianz-Kurven	37
4.3	<i>In-sample</i> Performance von empirischen Minimum-Varianz-Portfolios	39
4.3.1	Verteilung der empirischen Minimum-Varianz-Kurve	39
4.3.2	Verteilung der Gewichte eines empirischen Minimum-Varianz-Portfolios	43
4.4	<i>Out-of-sample</i> Performance von empirischen Minimum-Varianz-Portfolios	44
4.5	Beweise	47
5	Simulationsstudie	83

5.1	Die Effizienzkurve mit und ohne Leerverkaufsrestriktion	83
5.2	Vergleich der Schätzmethoden zur Approximation der wahren Effizienzkurve	86
5.3	Die empirische Minimum-Varianz-Kurve	88
6	Fazit und Ausblick	101
	Literaturverzeichnis	I
A	Capital Asset Pricing Model	V
B	Bayes'sche Entscheidungstheorie	VII
C	Verteilungen	IX
C.1	Multivariate Normalverteilung	IX
C.2	χ^2 -Verteilung	XII
C.3	Nichtzentrale χ^2 -Verteilung	XIII
C.4	Wishart-Verteilung	XIV
C.5	F -Verteilung	XV
C.6	Nichtzentrale F -Verteilung	XV
C.7	Multivariate t -Verteilung	XV
C.8	Matrix- t -Verteilung	XV
D	Sätze und Lemmata anderer Autoren	XVII
D.1	Theorem 3.2 von Haff (1979)	XVII
D.2	Theoreme und Korollar von Muirhead (1982)	XVII
D.3	Lemma 1 von Sawa (1972)	XVIII
E	Sonstiges	XIX
E.1	Suffizienz	XIX
E.2	Optimierung	XX
E.3	Konfluente hypergeometrische Funktion und Gammafunktion	XXII
E.4	Zerlegung der Kovarianz	XXII
F	Notationen	XXIII

1. Einleitung

Investoren können auf einen breiten Markt an Anlagemöglichkeiten zurückgreifen. Jeder rational denkende Investor möchte dabei diejenige Anlagestrategie wählen, die ihm eine möglichst hohe Rendite unter möglichst niedrigem Risiko verspricht. Die moderne Portfoliotheorie beschäftigt sich mit der Frage, wie ein Investor sein Vermögen auf die am Markt vorhandenen Assets verteilen muss, um das für ihn bestmögliche Portfolio zu erhalten. Den Grundstein hierfür legte Harry Markowitz 1952 mit seinem Artikel „Portfolio Selection“. Darin quantifizierte er das Risiko eines Portfolios durch die Varianzen und Kovarianzen der Renditen der im Portfolio enthaltenen Assets und formulierte das „Mean-Variance-Optimierungsproblem“ mit den Prinzipien „Maximiere zu gegebenem Risiko die erwartete Rendite“ und „Minimiere zu gegebener erwarteter Rendite das Risiko“. Für seine Portfoliotheorie wurde Markowitz 1990 mit dem Nobelpreis für Wirtschaftswissenschaften ausgezeichnet. Auf diese Theorie aufbauend gibt es inzwischen unzählige weiterführende Arbeiten zur Weiterentwicklung der modernen Portfoliotheorie. Einen Überblick über die Entwicklung der modernen Portfoliotheorie liefern Elton und Gruber (1997).

Das Mean-Variance-Optimierungsproblem lässt sich folgendermaßen formulieren:

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \quad \mathbf{w}^T \mathbf{1} = 1.$$

Hierbei seien \mathbf{w} die Portfoliogewichte, mit denen die Assets in das Portfolio eingehen, und $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ der Erwartungswertvektor und die Kovarianzmatrix der Assetrenditen. Es soll also das Risiko eines Portfolios minimiert und dabei die Nebenbedingung eingehalten werden, dass eine bestimmte erwartete Zielrendite μ_* erreicht werden soll. Die Gewichte müssen sich dabei zu eins summieren. Dieses Optimierungsproblem lässt sich unter bestimmten Bedingungen in äquivalente Probleme umformen, was unter anderem Gegenstand dieser Arbeit ist. Durch Lösen des Problems für verschiedene Werte für μ_* erhält man Portfolios, die für eine gegebene erwartete Zielrendite jeweils mit dem kleinstmöglichen Risiko verbunden sind. Mit diesen Portfolios erhält man deshalb die sogenannte *Minimum-Varianz-Kurve*. Deren obere Hälfte wird von denjenigen Portfolios gebildet, welche die beiden Prinzipien der Mean-Variance-Optimierung erfüllen. Solche Portfolios werden als effizient bezeichnet, die obere Hälfte der Minimum-Varianz-Kurve heißt daher auch *Effizienzkurve*.

Markowitz geht in seiner Theorie davon aus, dass $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ bekannt sind. Diese Parameter sind in der Praxis jedoch unbekannt. Damit Investoren annähernd effiziente Portfolios konstruieren können, muss das Mean-Variance-Optimierungsproblem mit geeigneten Schätzern für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ gelöst werden. In dieser Arbeit werden mehrere Möglichkeiten vorgestellt, um $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ zu schätzen. Aufgrund ihrer häufigen Anwendung in der Praxis werden zunächst die empirischen Schätzer von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ betrachtet. Da die geschätzte Effizienzkurve aufgrund von Schätzfehlern von der wahren Kurve stark abweichen kann, werden

ferner Ansätze vorgestellt, die das Schätzrisiko reduzieren sollen. Dazu zählen Multifaktormodelle, mit denen durch die Dimensionsreduktion der zu schätzenden Parameter die Gefahr für Schätzfehler gesenkt wird. Des Weiteren soll mit der von Michaud (1989) vorgeschlagenen Methode mittels Bootstrap die Verzerrung der Portfoliogewichte korrigiert werden, die aus der Verwendung von Schätzern anstelle der wahren Parameter im Mean-Variance-Optimierungsproblem resultiert. Die Kurve, die man auf diese Weise erhält, ist als *resampled frontier* bekannt. Außerdem werden in dieser Arbeit Bayes-Schätzer und andere Shrinkage-Schätzer hergeleitet. Dabei bilden die Bayes-Schätzer für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ eine konvexe Kombination aus den jeweiligen empirischen Schätzern und dem entsprechenden A-priori-Erwartungswert. Der von Ledoit und Wolf (2003) entwickelte Shrinkage-Schätzer für $\boldsymbol{\Sigma}$ ist eine konvexe Kombination aus der empirischen Kovarianzmatrix und einer Single-Index-Kovarianzmatrix, die man aus einem Einfaktor-Modell der Assetrenditen erhält.

Da es sich bei den Schätzern von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ um Zufallsvariablen handelt, besitzt die geschätzte Minimum-Varianz-Kurve eine Wahrscheinlichkeitsverteilung, mit deren Hilfe sich die Lage der geschätzten zur wahren Kurve untersuchen lässt. Aufgrund der häufigen Verwendung der empirischen Schätzer in der Praxis wird in dieser Arbeit die Verteilung der empirischen Minimum-Varianz-Kurve analysiert, die man durch Lösen des Mean-Variance-Optimierungsproblems mit den empirischen Schätzern von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ erhält. Dabei wird unterschieden zwischen der *in-sample* Kurve, die man allein mit den empirischen Schätzern bestimmt, und der *out-of-sample* Kurve, die man erhält, wenn man die erwartete Rendite und das Risiko der mithilfe der geschätzten Parameter ermittelten Portfolios mit den wahren Parametern berechnet.

Aufgrund der eingeschränkten Zulässigkeit von Leerverkäufen in der Praxis wird ferner untersucht, wie sich ein Leerverkaufsverbot auf die jeweiligen Ergebnisse auswirkt. Um Leerverkäufe auszuschließen, wird das Mean-Variance-Optimierungsproblem um die Nebenbedingung $\mathbf{w} \geq 0$ ergänzt.

Nach der Vorstellung der Grundlagen des Mean-Variance-Optimierungsproblems und der Einführung von Effizienz- und Minimum-Varianz-Kurven werden in Kapitel 2 verschiedene Formulierungen des Mean-Variance-Optimierungsproblems vorgestellt und Bedingungen, unter denen diese Formulierungen äquivalent sind, hergeleitet. In Kapitel 3 werden die verschiedenen Methoden zum Schätzen von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ vorgestellt. Kapitel 4 beschäftigt sich mit der Untersuchung der Verteilung der empirischen Minimum-Varianz-Kurve für den Fall, dass Leerverkäufe zulässig sind. Während den Kapiteln 2 und 3 die Arbeit von Lai, Xing und Chen (2011) als Grundlage dient, basiert Kapitel 4 auf der Arbeit von Kan und Smith (2008). In der Simulationsstudie in Kapitel 5 werden die hergeleiteten Ergebnisse mit zum Teil historischen Daten veranschaulicht. Dabei wird zunächst die Auswirkung einer Leerverkaufsbeschränkung auf die Effizienzkurve betrachtet. Des Weiteren werden die verschiedenen Schätzmethode aus Kapitel 3 miteinander verglichen und die Ergebnisse aus Kapitel 4 empirisch überprüft. Die Schlussfolgerungen, die sich aus den Untersuchungen ergeben, sowie weiterführende Fragestellungen werden in Kapitel 6 vorgestellt. Im Anhang sind Definitionen, Eigenschaften und Notationen zusammengefasst, die in der Arbeit verwendet werden.

2. Portfoliotheorie nach Markowitz

2.1. Grundlagen

2.1.1. Risiko

Eine der ersten Definitionen von Risiko liefert Knight (1921), der Risiko als messbare Unsicherheit charakterisiert. Bezogen auf die Finanzwelt definiert Cool (1993) Risiko als den absoluten Wert des erwarteten Verlusts, wohingegen Jorion (2000) mit Risiko die Volatilität der erwarteten Werte von Assets bezeichnet. Sharpe (1964) fand heraus, dass sich das Gesamtrisiko eines Assets bzw. eines Portfolios in ein systematisches und ein unsystematisches Risiko aufteilen lässt. Das systematische Risiko, auch Marktrisiko genannt, umfasst alle Risiken, die von allgemeinen Marktbewegungen, wie zum Beispiel Aktienkurs-, Zins- oder Wechselkursänderungen, ausgehen. Das unsystematische Risiko stammt von spezifischen Einflüssen, die unabhängig vom Markt sind.

2.1.2. Risikodiversifikation

Durch die geeignete Verteilung des Vermögens auf verschiedene riskante Assets lassen sich Portfolios konstruieren, deren Risiko geringer ist als die gewichtete Summe der Risiken der einzelnen Assets. Der Grund dafür ist, dass man mit einer solchen Verteilungsstrategie das Risiko breiter streuen kann. Wie stark sich das Risiko eines Portfolios reduzieren lässt, hängt von der Korrelation der in dem Portfolio enthaltenen Assets ab. Da alle Assets dem systematischen Risiko gleichermaßen ausgesetzt sind, lässt sich allerdings nur das unsystematische Risiko durch Diversifikation eliminieren. Die folgenden Überlegungen zur Risikodiversifikation orientieren sich an Kremer (2006-a).

Ein gebräuchliches Maß für das Risiko eines Portfolios ist die Varianz bzw. die Standardabweichung seiner Rendite. Betrachtet man ein Portfolio aus m riskanten Assets, lässt sich die Varianz der Portfoliorendite r_{Pf} schreiben als

$$\text{Var}(r_{Pf}) = \sum_{i=1}^m \alpha_i^2 \text{Var}(r_i) + \sum_{\substack{i,j=1, \\ i \neq j}}^m \alpha_i \alpha_j \text{Cov}(r_i, r_j),$$

wobei r_i die Rendite von Asset i bezeichne und $\alpha_i \in [0, 1]$ der Anteil von Asset i am Portfolio sei. Der Risikodiversifikationseffekt lässt sich leicht an einem Portfolio aus zwei Assets zeigen, wobei sich die Ergebnisse induktiv auf Portfolios mit mehr als zwei Assets erweitern lassen. Für ein Portfolio aus zwei Assets, A_1 und A_2 , gilt

$$\begin{aligned} \text{Var}(r_{Pf}) &= \alpha^2 \text{Var}(r_1) + (1 - \alpha)^2 \text{Var}(r_2) + 2\alpha(1 - \alpha) \text{Cov}(r_1, r_2) \\ &= (\alpha\sigma_1)^2 + ((1 - \alpha)\sigma_2)^2 + 2\alpha(1 - \alpha)\sigma_1\sigma_2\rho, \end{aligned}$$

wobei $\alpha \in [0, 1]$ der Anteil von A_1 am Portfolio sei, σ_1 und σ_2 die jeweilige Standardabweichung der Renditen von A_1 und A_2 und ρ ihr Korrelationskoeffizient.

Falls $\rho = 1$ ist, also die Renditen der beiden Assets vollständig positiv miteinander korreliert sind, gilt für $\sigma_{Pf} = \sqrt{\text{Var}(r_{Pf})}$

$$\sigma_{Pf} = \alpha\sigma_1 + (1 - \alpha)\sigma_2.$$

Das Risiko der Portfoliorendite ist dann also die Summe der Risiken der einzelnen Assets, gewichtet mit ihrem jeweiligen Anteil am Portfolio. Das bedeutet, dass sich das Risiko bei vollständig positiver Korrelation der Renditen nicht reduzieren lässt.

Für den anderen Extremfall, nämlich dass $\rho = -1$ ist, gilt

$$\begin{aligned} \sigma_{Pf} &= |\alpha\sigma_1 - (1 - \alpha)\sigma_2| \\ &= |-\sigma_2 + \alpha(\sigma_1 + \sigma_2)| \\ &= \begin{cases} \sigma_2 - \alpha(\sigma_1 + \sigma_2), & \text{wenn } 0 \leq \alpha \leq \frac{\sigma_2}{\sigma_1 + \sigma_2} \\ -\sigma_2 + \alpha(\sigma_1 + \sigma_2), & \text{wenn } \frac{\sigma_2}{\sigma_1 + \sigma_2} \leq \alpha \leq 1. \end{cases} \end{aligned}$$

Mit $\alpha = \sigma_2/(\sigma_1 + \sigma_2)$ lässt sich bei perfekter negativer Korrelation der Assetrenditen die Varianz der Portfoliorendite vollständig eliminieren. Dieser Fall wird in der Praxis allerdings niemals eintreten, da alle Assets dem von allgemeinen Marktbewegungen ausgehenden systematischen Risiko gleichermaßen ausgesetzt sind, sodass sich dieses Risiko durch Kombination verschiedener Assets nicht reduzieren lässt.

Falls $-1 < \rho < 1$ ist, gilt für die Varianz der Portfoliorendite

$$\begin{aligned} \sigma_{Pf}^2 &= (\alpha\sigma_1)^2 + ((1 - \alpha)\sigma_2)^2 + 2\alpha(1 - \alpha)\sigma_1\sigma_2\rho \\ &= (\alpha\sigma_1 + (1 - \alpha)\sigma_2)^2 - 2\alpha(1 - \alpha)\sigma_1\sigma_2(1 - \rho). \end{aligned}$$

Da $2\alpha(1 - \alpha)\sigma_1\sigma_2(1 - \rho)$ für jeden Wert von ρ zwischen -1 und 1 positiv ist, folgt

$$\sigma_{Pf} < \alpha\sigma_1 + (1 - \alpha)\sigma_2.$$

Das Risiko des Portfolios ist in diesem Fall also stets geringer als die gewichtete Summe der Risiken der einzelnen Assets. Je stärker die Renditen negativ korreliert sind, umso stärker ist der Diversifikationseffekt.

Ist μ_1 die erwartete Rendite von A_1 , also $\mu_1 = E[r_1]$, und μ_2 die erwartete Rendite von A_2 , so gilt für die erwartete Portfoliorendite

$$\mu_{Pf} = \alpha\mu_1 + (1 - \alpha)\mu_2.$$

Die Risiko-Rendite-Kombinationen der Assets bzw. Portfolios lassen sich als Punkte in einem (μ, σ) -Diagramm darstellen. In einem solchen Diagramm wird die Standardabweichung der Renditen auf der x -Achse und ihr Erwartungswert auf der y -Achse abgetragen. Die Risiko-Rendite-Kombination eines Portfolios aus zwei Assets liegt auf einer Linie, die die Risiko-Rendite-Kombinationen der beiden Assets miteinander verbindet. Die verschiedenen Kombinationen auf dieser Verbindungslinie erhält man durch die stetige Variation

von α zwischen 0 und 1. Der Verlauf der Verbindungslinie hängt von der Korrelation der beiden Assets ab. In Abbildung 2.1 ist dies für

$$\begin{aligned}\mu_1 &= 0.115, & \sigma_1 &= 0.074, \\ \mu_2 &= 0.322, & \sigma_2 &= 0.158\end{aligned}$$

und verschiedene Werte für ρ exemplarisch dargestellt.

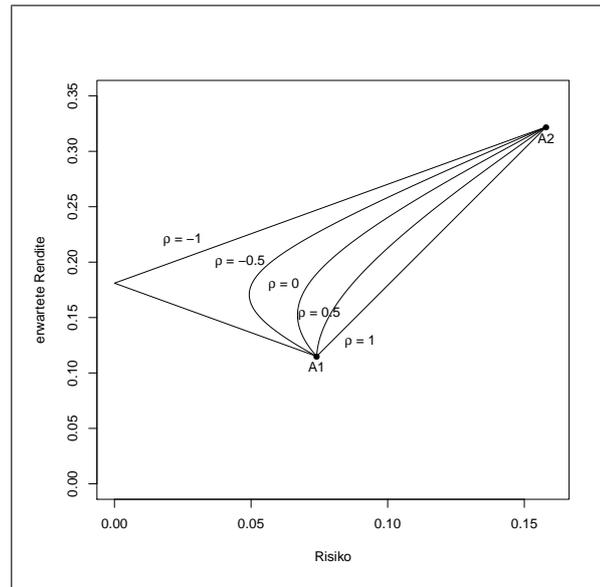


Abbildung 2.1.: Risikodiversifikation in Abhängigkeit der Korrelation zwischen zwei Assets

2.1.3. Statistisches Modell für Assetpreise und Renditen

Bezeichnet man mit P_{it} den Wert eines Assets i im Zeitpunkt t und hält man eine Einheit des Assets von $t - 1$ bis t , so erhält man in t die *einfache Rendite*

$$R_{it} = \frac{P_{it} - P_{i,t-1}}{P_{i,t-1}},$$

also die relative Wertänderung des Assets von $t - 1$ bis t . Häufig betrachtet man aber auch die logarithmierten Renditen (*log>Returns*)

$$r_{it} = \ln \left(\frac{P_{it}}{P_{i,t-1}} \right) = \ln(1 + R_{it}). \quad (2.1)$$

Je kürzer eine Periode bzw. je kleiner die Wertänderung des Assets innerhalb einer Periode, umso genauer ist die Approximation

$$r_{it} \approx R_{it}.$$

Deswegen genügt es, wenn im weiteren Verlauf von Renditen die Rede ist, nur die log>Returns zu betrachten.

Ein gebräuchliches Modell für die stetige Wertentwicklung eines riskanten Assets i liefert die geometrische Brownsche Bewegung. Sei

$$\frac{dP_{it}}{P_{it}} = \theta_i dt + \sigma_i dB_t^{(i)}, \quad (2.2)$$

wobei $\{B_t^{(i)}, t \geq 0\}$ die Standard-Brownsche-Bewegung ist und θ_i die Drift und σ_i die Volatilität des Preisprozesses sind; siehe dazu z.B. Kremer (2006-b). Die eindeutige Lösung der stochastischen Differentialgleichung

$$dP_{it} = P_{it} \left(\theta_i dt + \sigma_i dB_t^{(i)} \right)$$

ist die geometrische Brownsche Bewegung:

$$P_{it} = P_{i0} e^{(\theta_i - \frac{1}{2}\sigma_i^2)t + \sigma_i B_t^{(i)}}.$$

Setzt man diese in (2.1) ein, erhält man

$$\begin{aligned} r_{it} &= \log \left(\frac{P_{i0} e^{(\theta_i - \frac{1}{2}\sigma_i^2)t + \sigma_i B_t^{(i)}}}{P_{i0} e^{(\theta_i - \frac{1}{2}\sigma_i^2)(t-1) + \sigma_i B_{t-1}^{(i)}}} \right) \\ &= \log \left(e^{\theta_i - \frac{1}{2}\sigma_i^2 + \sigma_i (B_t^{(i)} - B_{t-1}^{(i)})} \right) \\ &= \theta_i - \frac{1}{2}\sigma_i^2 + \sigma_i (B_t^{(i)} - B_{t-1}^{(i)}). \end{aligned}$$

Da für Inkremente B_s und B_t einer Brownschen Bewegung mit $s < t$ gilt, dass sie normalverteilt mit Erwartungswert 0 und Varianz $t - s$ sind, gilt $B_t^{(i)} - B_{t-1}^{(i)} \sim \mathcal{N}(0, 1)$. Somit ist $r_{it} = \theta_i - \frac{1}{2}\sigma_i^2 + \sigma_i (B_t^{(i)} - B_{t-1}^{(i)})$ normalverteilt mit Erwartungswert

$$E[r_{it}] = E \left[\theta_i - \frac{1}{2}\sigma_i^2 + \sigma_i (B_t^{(i)} - B_{t-1}^{(i)}) \right] = \theta_i - \frac{1}{2}\sigma_i^2$$

und Varianz

$$Var(r_{it}) = Var \left(\theta_i - \frac{1}{2}\sigma_i^2 + \sigma_i (B_t^{(i)} - B_{t-1}^{(i)}) \right) = \sigma_i^2 Var(B_t^{(i)} - B_{t-1}^{(i)}) = \sigma_i^2.$$

Nach Karlin und Taylor (1981) sind $P_{i0}, P_{i1}/P_{i0}, P_{i2}/P_{i1}, \dots$ unabhängig. Es gilt also unter der Annahme (2.2) insgesamt, dass die log>Returns eines Assets unabhängig und identisch $\mathcal{N}(\theta_i - \sigma_i^2/2, \sigma_i^2)$ -verteilt sind.

2.2. Mean-Variance-Optimierung

2.2.1. Modell

Mithilfe der Theorie der Mean-Variance-Optimierung von Markowitz lassen sich Portfolios konstruieren, die aufgrund der Diversifikationseffekte ihrer Assets eine bestimmte erwartete Rendite unter minimalem Risiko bzw. zu einem gegebenen Risiko die maximale erwartete Rendite liefern. Dieser Theorie liegt ein Marktmodell zugrunde, in dem ein Investor die Möglichkeit hat, in m riskante Assets zu investieren. Dabei investiert er sein vorhandenes Vermögen im Zeitpunkt t und hält seine Position bis $t+1$. Es handelt sich also um ein Ein-Perioden-Modell. Zudem wird vereinfachend angenommen, dass keine Transaktionskosten erhoben werden und dass die Assets beliebig teilbar sind. Weiterhin seien

- r_{it} die Rendite von Asset i zum Zeitpunkt t ,
- $\mu_i = E[r_{i,t+1}]$ die erwartete Rendite von Asset i in $t+1$,
- $\Sigma = (\sigma_{i,j})_{i,j=1,\dots,m}$ die Kovarianzmatrix der Renditen mit $\sigma_{i,j} = Cov(r_{i,t+1}, r_{j,t+1})$ für $i \neq j$ und $\sigma_{ii} = \sigma_i^2 = Var(r_{i,t+1})$ und
- w_i das Portfoliogewicht von Asset i .

Es sei $\mathbf{r}_t = (r_{1t}, \dots, r_{mt})^T$ der Vektor der Renditen aller m Assets und $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)^T$ der Vektor ihrer erwarteten Renditen. Die Portfoliogewichte werden in dem Vektor $\mathbf{w} = (w_1, \dots, w_m)^T$ zusammengefasst. Da w_i der relative Anteil von Asset i am Portfolio ist, gilt $\sum_{i=1}^m w_i = 1$. Weiterhin sei $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^m$ ein m -dimensionaler Vektor, dessen Komponenten alle gleich 1 sind. Außerdem werden mit

- $\mu_{Pf} = \mathbf{w}^T \boldsymbol{\mu}$ die erwartete Portfoliorendite und mit
- $\sigma_{Pf}^2 = \mathbf{w}^T \Sigma \mathbf{w}$ die Varianz der Portfoliorendite

bezeichnet.

2.2.2. Minimum-Varianz-Kurve und Effizienzkurve

Portfolios, die zu gegebener erwarteter Rendite das geringste Risiko aufweisen bzw. zu gegebenem Risiko die höchste erwartete Rendite liefern, heißen (μ, σ) -effizient, im Folgenden auch einfach als *effizient* bezeichnet. Trägt man die Risiko-Rendite-Kombinationen jedes möglichen Portfolios in ein (μ, σ) -Diagramm ein, erhält man den sogenannten *Opportunitätsbereich*; siehe Kremer (2006-a). Dieser Bereich wird durch die *Minimum-Varianz-Kurve* begrenzt, die die Form des rechten Asts einer Hyperbel annimmt (im Fall ohne Leerverkäufe nur annähernd); siehe Kan und Smith (2008). Abbildung 2.2 skizziert den Opportunitätsbereich und die Minimum-Varianz-Kurve für sieben riskante Assets unter der Annahme, dass Leerverkäufe unzulässig sind. Auf die Ausprägung des rechten Randes des Opportunitätsbereichs ist später noch genauer einzugehen.

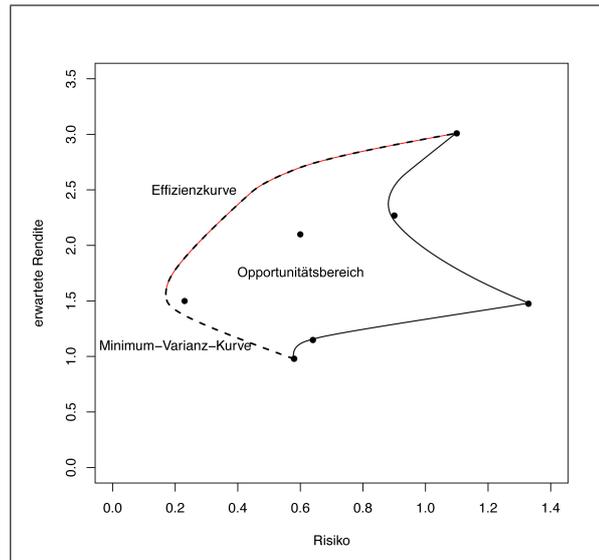


Abbildung 2.2.: Opportunitätsbereich, Minimum-Varianz-Kurve (gestrichelte Linie) und Effizienzkurve (farbige Linie) für sieben riskante Assets unter Leerverkaufsverbot

Die Minimum-Varianz-Kurve stellt die Risiko-Rendite-Kombinationen solcher Portfolios dar, die eine bestimmte erwartete Rendite unter dem jeweils kleinstmöglichen Risiko liefern. Allerdings sind nicht alle dieser Portfolios auch effizient, da zu jedem Portfolio auf der unteren Hälfte der Hyperbel ein Portfolio auf der oberen Hälfte existiert, das zum gleichen Risiko eine höhere Rendite erwarten lässt. Diese obere Hälfte der Minimum-Varianz-Kurve wird von den effizienten Portfolios gebildet und somit auch als *Effizienzkurve* oder *effizienter Rand* bezeichnet. In Abbildung 2.2 ist die Effizienzkurve durch die farbige Linie gekennzeichnet. In Kapitel 2.1.2 wird gezeigt, dass für zwei nicht vollständig positiv oder negativ korrelierte Assets mit erwarteter Rendite μ_1 bzw. μ_2 und Standardabweichung σ_1 bzw. σ_2

$$\sigma_{Pf} < \alpha\sigma_1 + (1 - \alpha)\sigma_2 \quad (2.3)$$

gilt. Alle möglichen (σ_{Pf}, μ_{Pf}) -Kombinationen, die man mit der stetigen Variation von α von 0 bis 1 erhält, liegen auf einer Kurve, die (σ_1, μ_1) und (σ_2, μ_2) miteinander verbindet und die wegen (2.3) stets links von der direkten Verbindungsgeraden der beiden Punkte liegt. Dies hat zur Folge, dass die Minimum-Varianz-Kurve nach links hin gewölbt bzw. der effiziente Rand konkav ist. Auch wenn Leerverkäufe zulässig sind, also die Portfoliogewichte jeden beliebigen – also insbesondere auch negativen – Wert annehmen können, wobei die Summe der Gewichte aber wieder 1 ergeben muss, nimmt die Effizienzkurve diese Form an, denn sonst könnte man ein Portfolio finden, dessen Risiko-Rendite-Kombination besser ist als die Kombinationen auf der Effizienzkurve; siehe Hausmann, Diener und Käsler (2002). Ohne Leerverkäufe lässt sich der Opportunitätsbereich aufgrund (2.3) wie in Abbildung 2.2 skizzieren. Falls Leerverkäufe zugelassen sind, ist der Opportunitätsbereich dagegen nach

rechts hin unbeschränkt. Das Portfolio, dessen Risiko-Rendite-Kombination ganz links auf der Effizienzkurve liegt, wird als *globales varianzminimales Portfolio* bezeichnet. Dieses Portfolio ist mit dem kleinstmöglichen Risiko verbunden, das mit den verfügbaren Assets erzielt werden kann.

Besteht nun neben den riskanten Assets zusätzlich die Möglichkeit in eine risikolose Anlage mit sicherer Rendite r_f zu investieren, kommen zum Opportunitätsbereich neue Risiko-Rendite-Kombinationen hinzu. Die Effizienzkurve nimmt dann eine etwas andere Form an. Fasst man ein beliebiges riskantes Portfolio P und die sichere Anlage S zu einem neuen Portfolio Q zusammen, so erhält man Risiko-Rendite-Kombinationen, die auf der Geraden liegen, die durch $(0, r_f)$ und (σ_P, μ_P) verläuft. Diese Gerade besitzt die Geradengleichung $\mu_Q = r_f + \frac{\mu_P - r_f}{\sigma_P} \sigma_Q$ (vgl. Kremer (2006-a)). Anhand dieser Gleichung lässt sich das Verhältnis zwischen Überschussrendite bzw. Risikoprämie $\mu_Q - r_f$ und dem dafür einzugehenden Risiko σ_Q erkennen. Es gilt

$$\mu_Q - r_f = \frac{\mu_P - r_f}{\sigma_P} \sigma_Q,$$

wobei $\frac{\mu_P - r_f}{\sigma_P}$ die Steigung der Geraden ist. Diese Steigung wird auch als *Sharpe Ratio* bezeichnet. Je höher die Sharpe Ratio, umso höher ist die Risikoprämie im Verhältnis zum aufgenommenen Risiko. Die höchste Steigung solcher Geraden, die $(0, r_f)$ mit den Risiko-Rendite-Kombinationen riskanter Portfolios verbinden, besitzt die Tangente der Effizienzkurve. Das Portfolio, das von dieser Geraden tangiert wird, heißt *Tangentialportfolio* oder *Marktportfolio*. Es enthält alle am Markt gehandelten riskanten Assets, wobei ihr Anteil am Marktportfolio ihrem jeweiligen Anteil am Gesamtwert aller Assets entspricht; siehe z.B. Berk und DeMarzo (2011). Die Tangente bezeichnet man auch als *Kapitalmarktlinie*. Sie ist in Abbildung 2.3 dargestellt, wobei der Punkt M die Risiko-Rendite-Kombination des Marktportfolios und S die sichere Anlage kennzeichnet. Sind keine Leerverkäufe zugelassen, setzt sich die Effizienzkurve aus den Risiko-Rendite-Kombinationen auf der Kapitalmarktlinie links vom Marktportfolio und den Kombinationen auf dem effizienten Rand des ursprünglichen Opportunitätsbereichs (also ohne sichere Anlage) rechts vom Marktportfolio zusammen. Sind dagegen Leerverkäufe erlaubt, liegen alle effizienten Risiko-Rendite-Kombinationen auf der Kapitalmarktlinie. Ein Investor erhält dann also durch die Kombination des Marktportfolios und der sicheren Anlage stets den bestmöglichen Trade-off zwischen Risiko und erwarteter Rendite. Die Steigung der Kapitalmarktlinie $\frac{\mu_M - r_f}{\sigma_M}$ wird als *Marktpreis des Risikos* bezeichnet, wobei μ_M und σ_M die erwartete Rendite und das Risiko des Marktportfolios sind. Jeder Investor wird nun ein Portfolio wählen, das eine Kombination aus sicherer Anlage und Marktportfolio darstellt. Dabei investieren risikoaverse Investoren einen höheren Anteil in die sichere Anlage (und erhalten somit Risiko-Rendite-Kombinationen auf der Kapitalmarktlinie links vom Marktportfolio), während risikofreudige Investoren sogar einen Kredit zum risikolosen Zins aufnehmen, um noch mehr als ihr vorhandenes Vermögen in das Marktportfolio zu investieren (und Kombinationen rechts vom Marktportfolio zu erhalten); vgl. Kremer (2006-a).

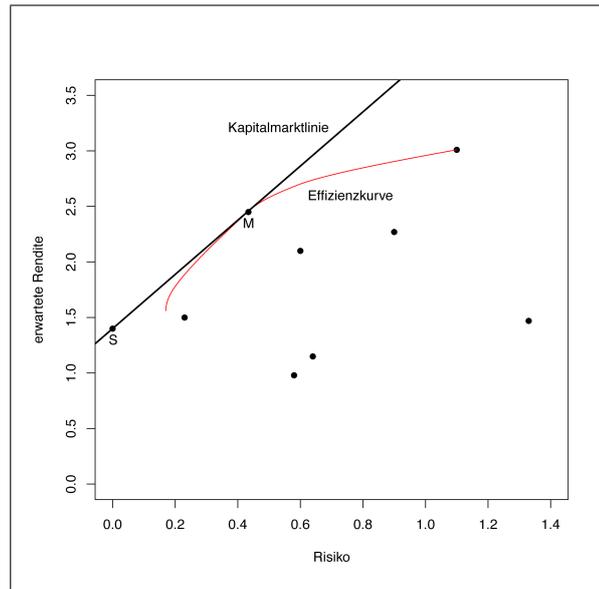


Abbildung 2.3.: Kapitalmarktklinie

2.2.3. Formulierung und Lösung des Mean-Variance-Optimierungsproblems

In diesem Abschnitt werden verschiedene Formulierungen des Markowitz'schen Mean-Variance-Optimierungsproblems vorgestellt, die sich in ihrer Parametrisierung unterscheiden. Durch das Lösen dieser Probleme erhält man Portfoliogewichte, mit denen effiziente Portfolios konstruiert werden können. Die Probleme sind unter bestimmten Bedingungen an ihre Parameter äquivalent, d.h. sie besitzen dann dieselben Lösungen.

Durch Lösen des Problems

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \Sigma \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0 \quad (2.4)$$

erhält man ein Portfolio, das eine bestimmte erwartete Rendite μ_* unter kleinstmöglichem Risiko liefert. Die Lösungen für verschiedene Werte für μ_* bilden folglich die Minimum-Varianz-Kurve. Die Einschränkung $\mathbf{w} \geq 0$ (also $w_i \geq 0$ für alle i) bedeutet, dass Leerverkäufe nicht zulässig sind. Es handelt sich in (2.4) um eine quadratische Zielfunktion unter linearen Restriktionen und Nichtnegativitätsbedingungen, die somit mittels numerischer Methoden der quadratischen Optimierung minimiert werden muss. Das Problem ist eindeutig lösbar, wenn Σ positiv definit, also invertierbar ist; siehe Fletcher (1991). Sind Leerverkäufe erlaubt, entfällt die Einschränkung $\mathbf{w} \geq 0$ und man erhält das Optimierungsproblem

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \Sigma \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \mathbf{w}^T \mathbf{1} = 1, \quad (2.5)$$

das genau dann eine explizit darstellbare Lösung besitzt, wenn Σ invertierbar ist. Die Lösung von (2.4) bzw. (2.5) hängt vom Parameter μ_* ab. Unter bestimmten Bedingungen,

die in diesem Abschnitt noch genauer untersucht werden, lässt sich das Problem (2.4) bzw. (2.5), in das Problem

$$\max_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\mu} - \lambda \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0 \quad (2.6)$$

bzw.

$$\max_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\mu} - \lambda \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1 \quad (2.7)$$

überführen. Mit diesem Problem soll der Trade-off zwischen erwarteter Rendite und Varianz maximiert werden, wobei das Eingehen eines höheren Risikos durch den Risikoaversionsparameter λ bestraft wird. Da Markowitz in seinem Modell davon ausgeht, dass jeder Investor risikoavers ist, gilt im Folgenden stets $\lambda > 0$. Je risikoaverser ein Investor ist, umso höher ist dieser Parameter zu wählen. Die Lösung von (2.6) bzw. (2.7) hängt also schließlich von λ ab. Dieses Maximierungsproblem lässt sich durch entsprechende Anpassung des Parameters auch als Minimierungsproblem formulieren. Dieses lautet dann

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} - \gamma \mathbf{w}^T \boldsymbol{\mu} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0 \quad (2.8)$$

bzw.

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} - \gamma \mathbf{w}^T \boldsymbol{\mu} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1 \quad (2.9)$$

mit $\gamma = \frac{1}{\lambda}$. Bezeichnet nämlich $f_\lambda(\mathbf{w})$ die in λ parametrisierte Zielfunktion von (2.6) bzw. (2.7) und $g_\gamma(\mathbf{w})$ die Zielfunktion von (2.8) bzw. (2.9), parametrisiert in γ , so ergibt sich die Beziehung

$$-\frac{1}{\lambda} f_\lambda(\mathbf{w}) = g_{\frac{1}{\lambda}}(\mathbf{w}). \quad (2.10)$$

Falls nun \mathbf{w}^* für ein festes λ die Funktion $f_\lambda(\mathbf{w})$ maximiert, so wird $-f_\lambda(\mathbf{w})$ durch \mathbf{w}^* minimiert. Da $\frac{1}{\lambda}$ konstant ist, ändert sich die Lösung des Optimierungsproblems nicht, wenn man die Zielfunktion mit diesem Faktor skaliert. Wegen (2.10) minimiert \mathbf{w}^* somit $g_\gamma(\mathbf{w})$, wenn $\gamma = \frac{1}{\lambda}$ ist. Analog lässt sich zeigen, dass die Lösung des Minimierungsproblems (2.8) bzw. (2.9) auch das Maximierungsproblem (2.6) bzw. (2.7) löst. Somit sind die beiden Probleme äquivalent.

Steinbach (2001) zeigt, dass unter bestimmten Annahmen in Bezug auf γ und μ_* die Probleme (2.5) und (2.9) äquivalent sind, indem er untersucht, unter welchen Bedingungen sie die gleiche Lösung besitzen. Dazu werden zunächst die Lösungen der Probleme bestimmt. Zur einfacheren Formulierung der Lösungen gelte im Folgenden

$$\begin{aligned} A &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \\ B &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \\ C &= \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}, \\ D &= BC - A^2. \end{aligned} \quad (2.11)$$

Lemma 2.2.1 *Das Problem (2.5)*

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \mathbf{w}^T \mathbf{1} = 1$$

besitzt die eindeutige Lösung

$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1}(\tilde{\gamma}^* \boldsymbol{\mu} + \tilde{\lambda}^* \mathbf{1})}{2} =: \mathbf{w}_{\mu_*}^* \quad (2.12)$$

$$\tilde{\gamma}^* = \frac{2(C\mu_* - A)}{D} \quad (2.13)$$

$$\tilde{\lambda}^* = \frac{2(B - A\mu_*)}{D}. \quad (2.14)$$

Beweis. Die zum Optimierungsproblem (2.5) gehörige Lagrangefunktion lautet

$$L(\mathbf{w}, \tilde{\gamma}, \tilde{\lambda}) = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} - \tilde{\gamma}(\mathbf{w}^T \boldsymbol{\mu} - \mu_*) - \tilde{\lambda}(\mathbf{w}^T \mathbf{1} - 1).$$

Der Parameter $\tilde{\gamma}$ wird auch als *Renditemultiplikator* und $\tilde{\lambda}$ als *Budgetmultiplikator* bezeichnet. Wenn man die Lagrangefunktion jeweils bezüglich \mathbf{w} , $\tilde{\gamma}$ und $\tilde{\lambda}$ differenziert, die Ableitungen gleich null setzt und diese geeignet umstellt, erhält man das Gleichungssystem

$$\mathbf{w} = \frac{\boldsymbol{\Sigma}^{-1}(\tilde{\gamma} \boldsymbol{\mu} + \tilde{\lambda} \mathbf{1})}{2} \quad (2.15)$$

$$\boldsymbol{\mu}^T \mathbf{w} = \mu_* \quad (2.16)$$

$$\mathbf{1}^T \mathbf{w} = 1. \quad (2.17)$$

Einsetzen von (2.15) in (2.16) und (2.17) liefert das Gleichungssystem

$$\begin{aligned} & \begin{cases} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}(\tilde{\gamma} \boldsymbol{\mu} + \tilde{\lambda} \mathbf{1}) = 2\mu_* \\ \mathbf{1}^T \boldsymbol{\Sigma}^{-1}(\tilde{\gamma} \boldsymbol{\mu} + \tilde{\lambda} \mathbf{1}) = 2 \end{cases} \\ \Leftrightarrow & \begin{cases} \tilde{\gamma} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \tilde{\lambda} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = 2\mu_* \\ \tilde{\gamma} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \tilde{\lambda} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = 2. \end{cases} \end{aligned}$$

In Matrixschreibweise und mit den Definitionen in (2.11) gilt dann

$$\begin{aligned} & \begin{pmatrix} B & A \\ A & C \end{pmatrix} \begin{pmatrix} \tilde{\gamma} \\ \tilde{\lambda} \end{pmatrix} = \begin{pmatrix} 2\mu_* \\ 2 \end{pmatrix} \\ \Leftrightarrow & \begin{pmatrix} \tilde{\gamma} \\ \tilde{\lambda} \end{pmatrix} = \begin{pmatrix} B & A \\ A & C \end{pmatrix}^{-1} \begin{pmatrix} 2\mu_* \\ 2 \end{pmatrix} \\ \Leftrightarrow & \begin{pmatrix} \tilde{\gamma} \\ \tilde{\lambda} \end{pmatrix} = \frac{1}{\det \begin{pmatrix} B & A \\ A & C \end{pmatrix}} \begin{pmatrix} C & -A \\ -A & B \end{pmatrix} \begin{pmatrix} 2\mu_* \\ 2 \end{pmatrix} \\ \Leftrightarrow & \begin{pmatrix} \tilde{\gamma} \\ \tilde{\lambda} \end{pmatrix} = \frac{1}{BC - A^2} \begin{pmatrix} 2C\mu_* - 2A \\ 2B - 2A\mu_* \end{pmatrix} \\ \Leftrightarrow & \begin{pmatrix} \tilde{\gamma} \\ \tilde{\lambda} \end{pmatrix} = \frac{2}{D} \begin{pmatrix} C\mu_* - A \\ B - A\mu_* \end{pmatrix} =: \begin{pmatrix} \tilde{\gamma}^* \\ \tilde{\lambda}^* \end{pmatrix}. \end{aligned}$$

Durch Einsetzen der Lagrangemultiplikatoren $\tilde{\gamma}^*$ und $\tilde{\lambda}^*$ in (2.15) erhält man die effizienten Portfoliogewichte $\mathbf{w}_{\mu^*}^*$. \square

Lemma 2.2.2 *Das Problem (2.9)*

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \Sigma \mathbf{w} - \gamma \mathbf{w}^T \boldsymbol{\mu} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1$$

besitzt die eindeutige Lösung

$$\mathbf{w}^* = \frac{\Sigma^{-1} (\gamma \boldsymbol{\mu} + \tilde{\lambda}^* \mathbf{1})}{2} =: \mathbf{w}_{\gamma}^* \quad (2.18)$$

$$\tilde{\lambda}^* = \frac{2 - \gamma A}{C} \quad (2.19)$$

mit zugehöriger erwarteter Portfoliorendite

$$\mu_{Pf,\gamma} = \frac{2A + \gamma D}{2C}. \quad (2.20)$$

Beweis. Die zum Optimierungsproblem gehörige Lagrangefunktion lautet

$$L(\mathbf{w}, \tilde{\lambda}) = \mathbf{w}^T \Sigma \mathbf{w} - \gamma \mathbf{w}^T \boldsymbol{\mu} - \tilde{\lambda} (\mathbf{w}^T \mathbf{1} - 1).$$

Setzt man ihre partiellen Ableitungen gleich null, erhält man durch entsprechende Umstellung das Gleichungssystem

$$\mathbf{w} = \frac{\Sigma^{-1} (\gamma \boldsymbol{\mu} + \tilde{\lambda} \mathbf{1})}{2} \quad (2.21)$$

$$\mathbf{1}^T \mathbf{w} = 1. \quad (2.22)$$

Einsetzen von (2.21) in (2.22) liefert

$$\begin{aligned} \mathbf{1}^T \Sigma^{-1} (\gamma \boldsymbol{\mu} + \tilde{\lambda} \mathbf{1}) &= 2 \\ \Leftrightarrow \gamma \mathbf{1}^T \Sigma^{-1} \boldsymbol{\mu} + \tilde{\lambda} \mathbf{1}^T \Sigma^{-1} \mathbf{1} &= 2 \\ \Leftrightarrow \gamma A + \tilde{\lambda} C &= 2 \\ \Leftrightarrow \tilde{\lambda} = \frac{2 - \gamma A}{C} &=: \tilde{\lambda}^*. \end{aligned}$$

Setzt man den Lagrangemultiplikator $\tilde{\lambda}^*$ in (2.21) ein, erhält man die effizienten Portfoliogewichte

$$\mathbf{w}_{\gamma}^* = \frac{\gamma \Sigma^{-1} \boldsymbol{\mu} + \frac{2 - \gamma A}{C} \Sigma^{-1} \mathbf{1}}{2},$$

mit denen sich die zugehörige erwartete Portfoliorendite bestimmen lässt:

$$\begin{aligned}
 \mu_{Pf,\gamma} &= \mathbf{w}_\gamma^{*T} \boldsymbol{\mu} \\
 &= \frac{\left(\gamma \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} + \frac{2-\gamma A}{C} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \right) \boldsymbol{\mu}}{2} \\
 &= \frac{\gamma B + \frac{2-\gamma A}{C} A}{2} \\
 &= \frac{\gamma \frac{BC}{C} + \frac{2A-\gamma A^2}{C}}{2} \\
 &= \frac{2A + \gamma D}{2C}.
 \end{aligned}$$

□

Bezeichnen $\tilde{\gamma}_{\mu_*}^*$, $\tilde{\lambda}_{\mu_*}^*$ und $\tilde{\lambda}_\gamma^*$ die optimalen Rendite- bzw. Budgetmultiplikatoren des jeweils entsprechend parametrisierten Problems, so sind die Lösungen

$$\mathbf{w}_{\mu_*}^* = \frac{\boldsymbol{\Sigma}^{-1} \left(\tilde{\gamma}_{\mu_*}^* \boldsymbol{\mu} + \tilde{\lambda}_{\mu_*}^* \mathbf{1} \right)}{2} \quad \text{und} \quad \mathbf{w}_\gamma^* = \frac{\boldsymbol{\Sigma}^{-1} \left(\gamma \boldsymbol{\mu} + \tilde{\lambda}_\gamma^* \mathbf{1} \right)}{2}$$

der beiden Probleme genau dann gleich, wenn $\tilde{\gamma}_{\mu_*}^* = \gamma$ und $\tilde{\lambda}_{\mu_*}^* = \tilde{\lambda}_\gamma^*$ ist.

Satz 2.2.3 Das Problem (2.5)

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \quad \mathbf{w}^T \mathbf{1} = 1$$

mit Parameter μ_* und das Problem (2.9)

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} - \gamma \mathbf{w}^T \boldsymbol{\mu} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1$$

mit Parameter γ sind genau dann äquivalent, wenn $\gamma = \tilde{\gamma}_{\mu_*}^*$ gilt, wobei $\tilde{\gamma}_{\mu_*}^* = \frac{2(C\mu_* - A)}{D}$ der optimale Renditemultiplikator von Problem (2.5) ist, oder wenn äquivalent dazu $\mu_* = \mu_{Pf,\gamma}$ gilt, wobei $\mu_{Pf,\gamma} = \frac{2A + \gamma D}{2C}$ die Portfoliorendite ist, die man mit der Lösung von Problem (2.9) erhält.

Beweis. Die Äquivalenz $\gamma = \tilde{\gamma}_{\mu_*}^* \Leftrightarrow \mu_* = \mu_{Pf,\gamma}$ gilt, da für $\gamma = \tilde{\gamma}_{\mu_*}^*$

$$\mu_{Pf,\gamma} = \frac{2A + \gamma D}{2C} = \frac{2A + 2 \cdot \frac{C\mu_* - A}{D} D}{2C} = \mu_*$$

ist und für $\mu_* = \mu_{Pf,\gamma}$

$$\tilde{\gamma}_{\mu_*}^* = \frac{2(C\mu_* - A)}{D} = \frac{2 \left(C \frac{2A + \gamma D}{2C} - A \right)}{D} = \gamma$$

gilt.

Mit $\gamma = \tilde{\gamma}_{\mu_*}^*$ bzw. $\mu_* = \mu_{Pf,\gamma}$ folgt für die optimalen Budgetmultiplikatoren der beiden Probleme

$$\begin{aligned}\tilde{\lambda}_\gamma^* &= \frac{2 - \gamma A}{C} = \frac{2D - \gamma AD}{CD} = \frac{2D - \frac{2(C\mu_* - A)}{D}AD}{CD} = \frac{2(D - AC\mu_* + A^2)}{CD} \\ &= \frac{2(BC - A^2 - AC\mu_* + A^2)}{CD} = \frac{2(B - A\mu_*)}{D} = \tilde{\lambda}_{\mu_*}^*.\end{aligned}$$

Damit sind die Lösungen \mathbf{w}_γ^* und $\mathbf{w}_{\mu_*}^*$ der beiden Probleme identisch.

Sind andererseits die beiden Probleme äquivalent, so folgt aus der Gleichheit ihrer Lösungen direkt $\tilde{\lambda}_\gamma^* = \tilde{\lambda}_{\mu_*}^*$ und $\gamma = \tilde{\gamma}_{\mu_*}^*$. \square

Die Probleme (2.5) und (2.9) lassen sich also ineinander überführen, wenn man den Parameter des einen Problems entsprechend in Abhängigkeit des Parameters des anderen Problems setzt, genauer gesagt, wenn man für eine gegebene Zielrendite μ_*

$$\gamma = \gamma(\mu_*) = \frac{2(C\mu_* - A)}{D}$$

oder für gegebenen inversen Risikoaversionsparameter γ

$$\mu_* = \mu_*(\gamma) = \frac{2A + \gamma D}{2C}$$

setzt. Wegen $\lambda = \frac{1}{\gamma}$ lassen sich dementsprechend das in μ_* parametrisierte Problem und das in λ parametrisierte Maximierungsproblem (2.7)

$$\max_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\mu} - \lambda \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1$$

durch

$$\lambda = \lambda(\mu_*) = \frac{D}{2(C\mu_* - A)}$$

bzw.

$$\mu_* = \mu_*(\lambda) = \frac{2\lambda A + D}{2\lambda C}$$

ineinander überführen. Es ist jedoch zu berücksichtigen, dass sich die Probleme (2.5) und (2.7) (bzw. (2.9)) nur für solche Werte für μ_* ineinander überführen lassen, für die die Bedingung $\lambda > 0$ bestehen bleibt. Dies ist dann der Fall, wenn $D/(2(C\mu_* - A)) > 0$, also $\mu_* > A/C$ gilt. Bei A/C handelt es sich um die erwartete Rendite des globalen varianzminimalen Portfolios. Dies lässt sich zeigen, indem man Problem (2.5) ohne die Nebenbedingung an die erwartete Zielrendite löst. Die Lagrangefunktion ist dann

$$L(\mathbf{w}, \tilde{\lambda}) = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} - \tilde{\lambda} (\mathbf{w}^T \mathbf{1} - 1).$$

Ihre partiellen Ableitungen liefern durch Nullsetzen und Umstellen das Gleichungssystem

$$\mathbf{w} = \frac{\tilde{\lambda} \boldsymbol{\Sigma}^{-1} \mathbf{1}}{2} \quad (2.23)$$

$$1 = \mathbf{1}^T \mathbf{w}. \quad (2.24)$$

Setzt man (2.23) in (2.24) ein, erhält man den Lagrangemultiplikator

$$\tilde{\lambda} = \frac{2}{C} =: \tilde{\lambda}^*.$$

Mit diesem lassen sich die Portfoliogewichte

$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}}{C} =: \mathbf{w}_g^*$$

des globalen varianzminimalen Portfolios bestimmen. Die mit diesen Gewichten zu erwartende Rendite ist dann

$$\mu_g := \mathbf{w}_g^{*T} \boldsymbol{\mu} = \frac{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{C} = \frac{A}{C}.$$

Damit gilt also, dass man durch Lösen von Problem (2.7) für alle $\lambda > 0$ bzw. durch Lösen von Problem (2.5) für $\mu_* > \mu_g$ die Effizienzkurve, also die obere Hälfte der Minimum-Varianz-Kurve, erhält. Die untere Hälfte der Minimum-Varianz-Kurve dagegen erhält man nur durch Lösen von Problem (2.5) für Werte von μ_* , die kleiner als μ_g sind.

Satz 2.2.4 *Das mit der Lösung des Portfolio-Optimierungsproblems (2.5), (2.7) bzw. (2.9) verbundene Risiko ist*

$$\sigma^2(\mu_*) = \frac{C\mu_*^2 - 2A\mu_* + B}{D}$$

bzw.

$$\sigma^2(\gamma) = \frac{1}{C} + \frac{D}{4C} \cdot \gamma^2$$

bzw.

$$\sigma^2(\lambda) = \frac{1}{C} + \frac{D}{4C} \cdot \frac{1}{\lambda^2}.$$

Beweis. Es gilt mit den effizienten Portfoliogewichten $\mathbf{w}_{\mu_*}^*$

$$\begin{aligned} \sigma^2(\mu_*) &= \mathbf{w}_{\mu_*}^{*T} \boldsymbol{\Sigma} \mathbf{w}_{\mu_*}^* \\ &= \frac{1}{4} \left(\boldsymbol{\Sigma}^{-1} (\tilde{\lambda}^* \mathbf{1} + \tilde{\gamma}^* \boldsymbol{\mu}) \right)^T \boldsymbol{\Sigma} \left(\boldsymbol{\Sigma}^{-1} (\tilde{\lambda}^* \mathbf{1} + \tilde{\gamma}^* \boldsymbol{\mu}) \right) \\ &= \frac{1}{4} \left(\tilde{\lambda}^* \mathbf{1} + \tilde{\gamma}^* \boldsymbol{\mu} \right)^T \boldsymbol{\Sigma}^{-1} \left(\tilde{\lambda}^* \mathbf{1} + \tilde{\gamma}^* \boldsymbol{\mu} \right) \\ &= \frac{1}{4} \left(\tilde{\lambda}^{*2} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} + \tilde{\lambda}^* \tilde{\gamma}^* \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \tilde{\lambda}^* \tilde{\gamma}^* \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} + \tilde{\gamma}^{*2} \boldsymbol{\mu}^T \boldsymbol{\Sigma} \boldsymbol{\mu} \right) \\ &= \frac{1}{4} \left(\tilde{\lambda}^{*2} C + 2\tilde{\lambda}^* \tilde{\gamma}^* A + \tilde{\gamma}^{*2} B \right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{4} \left(\tilde{\lambda}^* \left(\tilde{\lambda}^* C + \tilde{\gamma}^* A \right) + \tilde{\gamma}^* \left(\tilde{\lambda}^* A + \tilde{\gamma}^* B \right) \right) \\
 &= \frac{1}{4} \left(\tilde{\lambda}^* \left(\frac{2BC - 2AC\mu_*}{D} + \frac{2AC\mu_* - 2A^2}{D} \right) \right. \\
 &\quad \left. + \tilde{\gamma}^* \left(\frac{2AB - 2A^2\mu_*}{D} + \frac{2BC\mu_* - 2AB}{D} \right) \right) \\
 &= \frac{1}{2} \left(\tilde{\lambda}^* + \tilde{\gamma}^* \mu_* \right) \\
 &= \frac{1}{2} \left(\frac{2B - 2A\mu_*}{D} + \frac{2C\mu_*^2 - 2A\mu_*}{D} \right) \\
 &= \frac{C\mu_*^2 - 2A\mu_* + B}{D}.
 \end{aligned}$$

In Abhängigkeit von γ gilt dann

$$\begin{aligned}
 \sigma^2(\gamma) &= \frac{C \frac{2A+\gamma D}{2C} \cdot \frac{2A+\gamma D}{2C} - 2A \frac{2A+\gamma D}{2C} + B}{D} \\
 &= \frac{4A^2 + 4\gamma AD + \gamma^2 D^2 - 8A^2 + 4\gamma AD + 4BC}{4CD} \\
 &= \frac{-4A^2 + \gamma^2 D^2 + 4BC}{4CD} \\
 &= \frac{4 + \gamma^2 D}{4C} \\
 &= \frac{1}{C} + \frac{D}{4C} \gamma^2.
 \end{aligned}$$

Als Funktion von λ ist das Risiko somit

$$\sigma^2(\lambda) = \frac{1}{C} + \frac{D}{4C} \cdot \frac{1}{\lambda^2}.$$

□

Zeichnet man den Graph von $\sigma(\cdot) = \sqrt{\sigma^2(\cdot)}$ in Abhängigkeit von $\mu_* > A/C$, $\gamma > 0$ oder $\lambda > 0$, erhält man die Effizienzkurve.

Falls es ein Leerverkaufsverbot gibt, müssen die Nebenbedingungen der verschiedenen Problemformulierungen um die Einschränkung $\mathbf{w} \geq 0$ ergänzt werden. In diesem Fall lassen sich die Lösungen allerdings nicht mehr explizit angeben, da aufgrund der Ungleichungsnebenbedingung die Lösungen numerisch angenähert werden müssen. Man kann dennoch zeigen, dass die unterschiedlich parametrisierten Probleme auch mit der Nichtnegativitätsbedingung an die Portfoliogewichte unter bestimmten Bedingungen äquivalent sind. Da es sich bei Problem (2.4)

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \Sigma \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0$$

und Problem (2.8)

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \Sigma \mathbf{w} - \gamma \mathbf{w}^T \boldsymbol{\mu} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0$$

aufgrund der konvexen Zielfunktionen und Ungleichungsrestriktionen sowie der linearen Gleichungsnebenbedingungen um konvexe Optimierungsprobleme handelt, kann man mit Satz E.2.4 (im Anhang) nachweisen, dass sich die beiden Probleme ineinander überführen lassen. Dazu zeigt man für einen KKT-Punkt des einen Problems, dass dieser auch ein KKT-Punkt des anderen Problems ist. Damit folgt, dass es Lagrangemultiplikatoren gibt, mit denen man einen KKT-Punkt des anderen Problems erhält, sodass nach Satz E.2.4 dieser Punkt das Minimum des Problems ist.

Lemma 2.2.5 *Die KKT-Bedingungen von Problem (2.4)*

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \Sigma \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0$$

lauten

$$\begin{aligned} 0 &= 2\Sigma\mathbf{w} - \tilde{\nu} - \tilde{\gamma}\boldsymbol{\mu} - \tilde{\lambda}\mathbf{1} \\ \mu_* &= \mathbf{w}^T \boldsymbol{\mu} \\ 1 &= \mathbf{w}^T \mathbf{1} \\ 0 &\leq \tilde{\nu} \\ 0 &\geq -\mathbf{w} \\ 0 &= -\tilde{\nu}^T \mathbf{w}. \end{aligned}$$

Beweis. Die Lagrangefunktion zu Problem (2.4) ist

$$L(\mathbf{w}, \tilde{\nu}, \tilde{\gamma}, \tilde{\lambda}) = \mathbf{w}^T \Sigma \mathbf{w} - \tilde{\nu}^T \mathbf{w} - \tilde{\gamma} (\mathbf{w}^T \boldsymbol{\mu} - \mu_*) - \tilde{\lambda} (\mathbf{w}^T \mathbf{1} - 1).$$

Differenziert man die Lagrangefunktion bezüglich \mathbf{w} und setzt man die Ableitung gleich 0, erhält man die erste KKT-Bedingung

$$2\Sigma\mathbf{w} - \tilde{\nu} - \tilde{\gamma}\boldsymbol{\mu} - \tilde{\lambda}\mathbf{1} = 0.$$

Die übrigen KKT-Bedingungen ergeben sich direkt aus Definition E.2.3 im Anhang. \square

Lemma 2.2.6 *Die KKT-Bedingungen von Problem (2.8)*

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \Sigma \mathbf{w} - \gamma \mathbf{w}^T \boldsymbol{\mu} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0$$

lauten

$$\begin{aligned} 0 &= 2\Sigma\mathbf{w} - \tilde{\nu} - \gamma\boldsymbol{\mu} - \tilde{\lambda}\mathbf{1} \\ 1 &= \mathbf{w}^T \mathbf{1} \\ 0 &\leq \tilde{\nu} \\ 0 &\geq -\mathbf{w} \\ 0 &= -\tilde{\nu}^T \mathbf{w}. \end{aligned}$$

Beweis. Die Lagrangefunktion zu Problem (2.4) ist

$$L(\mathbf{w}, \tilde{\nu}, \tilde{\lambda}) = \mathbf{w}^T \Sigma \mathbf{w} - \gamma \mathbf{w}^T \boldsymbol{\mu} - \tilde{\nu}^T \mathbf{w} - \tilde{\lambda} (\mathbf{w}^T \mathbf{1} - 1).$$

Differenziert man die Lagrangefunktion bezüglich \mathbf{w} und setzt die Ableitung gleich 0, erhält man die erste KKT-Bedingung

$$2\Sigma \mathbf{w} - \tilde{\nu} - \gamma \boldsymbol{\mu} - \tilde{\lambda} \mathbf{1} = 0.$$

Die übrigen KKT-Bedingungen ergeben sich direkt aus Definition E.2.3 im Anhang. \square

Stellt man die jeweils erste KKT-Bedingung der beiden Probleme nach \mathbf{w} um, erhält man die Portfoliogewichte als Funktion der jeweiligen Lagrangemultiplikatoren. Ist $\mathbf{w}_{\mu_*}^*$ die Lösung von Problem (2.4) und \mathbf{w}_γ^* die Lösung von Problem (2.8) und bezeichnet man mit $\tilde{\nu}_{\mu_*}^*$, $\tilde{\gamma}_{\mu_*}^*$, $\tilde{\lambda}_{\mu_*}^*$, $\tilde{\nu}_\gamma^*$ und $\tilde{\lambda}_\gamma^*$ die optimalen Lagrangemultiplikatoren des jeweils entsprechend parametrisierten Problems, so gilt

$$\mathbf{w}_{\mu_*}^* = \frac{\Sigma^{-1} (\tilde{\nu}_{\mu_*}^* + \tilde{\gamma}_{\mu_*}^* \boldsymbol{\mu} + \tilde{\lambda}_{\mu_*}^* \mathbf{1})}{2} \quad \text{und} \quad \mathbf{w}_\gamma^* = \frac{\Sigma^{-1} (\tilde{\nu}_\gamma^* + \gamma \boldsymbol{\mu} + \tilde{\lambda}_\gamma^* \mathbf{1})}{2}.$$

Die Lösungen sind also dann gleich, wenn $\tilde{\nu}_{\mu_*}^* = \tilde{\nu}_\gamma^*$, $\tilde{\gamma}_{\mu_*}^* = \gamma$ und $\tilde{\lambda}_{\mu_*}^* = \tilde{\lambda}_\gamma^*$ ist.

Satz 2.2.7 *Das Problem (2.4)*

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \Sigma \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0$$

mit Parameter μ_* und das in γ parametrisierte Problem (2.8)

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \Sigma \mathbf{w} - \gamma \mathbf{w}^T \boldsymbol{\mu} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0$$

sind genau dann äquivalent, wenn Problem (2.4) einen KKT-Punkt $(\mathbf{w}_{\mu_*}^*, \tilde{\nu}_{\mu_*}^*, \tilde{\gamma}_{\mu_*}^*, \tilde{\lambda}_{\mu_*}^*)$ besitzt und man für Problem (2.8) $\gamma = \tilde{\gamma}_{\mu_*}^*$ wählt bzw. wenn Problem (2.8) einen KKT-Punkt $(\mathbf{w}_\gamma^*, \tilde{\nu}_\gamma^*, \tilde{\lambda}_\gamma^*)$ besitzt und man für Problem (2.4) $\mu_* = \mu_{Pf,\gamma}$ wählt, wobei $\mu_{Pf,\gamma}$ die Portfoliorendite ist, die man mit der Lösung von Problem (2.8) erhält.

Beweis. Sei $(\mathbf{w}_{\mu_*}^*, \tilde{\nu}_{\mu_*}^*, \tilde{\gamma}_{\mu_*}^*, \tilde{\lambda}_{\mu_*}^*)$ für eine bestimmte erwartete Zielrendite μ_* ein KKT-Punkt von Problem (2.4). Wählt man für den Parameter γ von (2.8) $\gamma = \tilde{\gamma}_{\mu_*}^*$, dann werden mit

$$\mathbf{w}_\gamma = \mathbf{w}_{\mu_*}^*, \quad \tilde{\nu}_\gamma = \tilde{\nu}_{\mu_*}^*, \quad \tilde{\lambda}_\gamma = \tilde{\lambda}_{\mu_*}^*$$

die KKT-Bedingungen von Problem (2.8) erfüllt.

Sei nun $(\mathbf{w}_\gamma^*, \tilde{\nu}_\gamma^*, \tilde{\lambda}_\gamma^*)$ ein KKT-Punkt von Problem (2.8) für einen bestimmten Wert für γ . Wählt man für den Parameter μ_* von (2.4) $\mu_* = \mu_{Pf,\gamma}$, dann werden mit

$$\mathbf{w}_{\mu_*} = \mathbf{w}_\gamma^*, \quad \tilde{\nu}_{\mu_*} = \tilde{\nu}_\gamma^*, \quad \tilde{\gamma}_{\mu_*} = \gamma, \quad \tilde{\lambda}_{\mu_*} = \tilde{\lambda}_\gamma^*$$

die KKT-Bedingungen von Problem (2.4) erfüllt.

Sind andererseits die beiden Probleme äquivalent, so folgt $\tilde{\nu}_{\mu_*}^* = \tilde{\nu}_\gamma^*$, $\tilde{\gamma}_{\mu_*}^* = \gamma$ und $\tilde{\lambda}_{\mu_*}^* = \tilde{\lambda}_\gamma^*$. \square

Ersetzt man in Satz 2.2.7 Problem (2.8) durch das äquivalente Problem (2.6)

$$\max_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\mu} - \lambda \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq 0,$$

liefert der Satz mit $\gamma = \frac{1}{\lambda}$ auch die Äquivalenz von Problem (2.4) und (2.6)

2.3. Darstellung von Minimum-Varianz-Portfolios und die Geometrie der Minimum-Varianz-Kurve

Durch Lösen von Problem (2.5)

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \mathbf{w}^T \mathbf{1} = 1$$

erhält man Portfoliogewichte, mit denen man ein Portfolio konstruieren kann, das eine vorgegebene erwartete Zielrendite μ_* mit dem kleinstmöglichen Risiko liefert. Solche Portfolios werden auch als *Minimum-Varianz-Portfolios* bezeichnet; siehe Kan und Smith (2008). Trägt man für viele nahe beieinanderliegende Werte für μ_* die Risiko-Rendite-Kombinationen solcher Portfolios in ein (μ, σ) -Diagramm ein, so erhält man die Minimum-Varianz-Kurve. In Lemma 2.2.1 wird gezeigt, dass Problem (2.5) die Lösung

$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1} (\tilde{\gamma}^* \boldsymbol{\mu} + \tilde{\lambda}^* \mathbf{1})}{2}$$

mit

$$\begin{aligned} \tilde{\gamma}^* &= \frac{2(C\mu_* - A)}{D}, \\ \tilde{\lambda}^* &= \frac{2(B - A\mu_*)}{D} \end{aligned}$$

und

$$\begin{aligned} A &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \\ B &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \\ C &= \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}, \\ D &= BC - A^2 \end{aligned}$$

besitzt. Man bezeichnet (A, B, C) auch als *Effizienzmengenkostanten*, da sie die Lösung von Problem (2.5) und somit insbesondere die effizienten Portfolios charakterisieren; siehe Merton (1972). Durch Einsetzen der Rendite- und Budgetmultiplikatoren $\tilde{\gamma}^*$ und $\tilde{\lambda}^*$ in den Ausdruck für \mathbf{w}^* erhält man für die Lösung

$$\mathbf{w}^* = (B\boldsymbol{\Sigma}^{-1}\mathbf{1} - A\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mu_* (C\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - A\boldsymbol{\Sigma}^{-1}\mathbf{1})) / D.$$

Fasst man die Effizienzmengenkostanten in einer Matrix zusammen, etwa

$$\mathbf{M} = (\boldsymbol{\mu}, \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}, \mathbf{1}) = \begin{pmatrix} B & A \\ A & C \end{pmatrix},$$

lässt sich die Lösung von Problem (2.5) auch kompakter schreiben als

$$\mathbf{w}^* = \Sigma^{-1} (\boldsymbol{\mu}, \mathbf{1}) \mathbf{M}^{-1} (\mu_*, 1)^T.$$

Das zugehörige Risiko eines Minimum-Varianz-Portfolios, das eine bestimmte erwartete Zielrendite μ_* liefert, ist nach Satz 2.2.4 durch

$$\sigma^2 = \sigma^2(\mu_*) = \frac{C\mu_*^2 - 2A\mu_* + B}{D} \quad (2.25)$$

gegeben. Man kann (2.25) folgendermaßen umformen:

$$\begin{aligned} \sigma^2 (BC - A^2) - C\mu_*^2 + 2A\mu_* - B &= 0 \\ \Leftrightarrow \sigma^2 \left(B - \frac{A^2}{C} \right) - \left(\mu_* - \frac{A}{C} \right)^2 + \frac{A^2}{C^2} - \frac{B}{C} &= 0 \\ \Leftrightarrow \left(\sigma^2 - \frac{1}{C} \right) \left(B - \frac{A^2}{C} \right) - \left(\mu_* - \frac{A}{C} \right)^2 &= 0 \\ \Leftrightarrow \sigma^2 - \frac{\left(\mu_* - \frac{A}{C} \right)^2}{\left(B - \frac{A^2}{C} \right)} &= \frac{1}{C} \\ \Leftrightarrow \frac{\sigma^2}{\frac{1}{C}} - \frac{\left(\mu_* - \frac{A}{C} \right)^2}{\left(\frac{B}{C} - \frac{A^2}{C^2} \right)} &= 1. \end{aligned} \quad (2.26)$$

Dabei handelt es sich um eine allgemeine Hyperbelgleichung (siehe z.B. Tröls (2008)), mit der sich die Minimum-Varianz-Kurve beschreiben lässt. Da das Risiko σ^2 nicht negativ sein kann, handelt es sich hierbei genauer gesagt um den rechten Ast der Hyperbel mit dem Mittelpunkt $(0; A/C)$ und den Asymptoten $\mu_* = A/C \pm \sqrt{B - (A^2/C)}\sigma$. Wie bereits gezeigt wurde, ist A/C die erwartete Rendite des globalen varianzminimalen Portfolios.

2.4. Das Markowitz'sche Optimierungsenigma

Markowitz geht in seiner Theorie davon aus, dass die erwarteten Renditen $\boldsymbol{\mu}$ und ihre Kovarianzmatrix Σ bekannt sind. In der Praxis ist dies jedoch nicht der Fall, weil es sich dabei um unbekannte Populationsgrößen handelt. Die Parameter müssen also geeignet geschätzt werden. Dafür werden in der Regel historische Daten herangezogen, wobei üblicherweise angenommen wird, dass die Renditen unabhängig und identisch verteilt sind. Ersetzt man in einem der Optimierungsprobleme (2.4) - (2.9) die Parameter $\boldsymbol{\mu}$ und Σ durch die Schätzer $\hat{\boldsymbol{\mu}}$ und $\hat{\Sigma}$, erhält man durch das Lösen dieser Probleme Portfolios, deren Risiko-Rendite-Kombinationen weit unterhalb der tatsächlichen Effizienzkurve liegen können. Das liegt vor allem daran, dass Assets, deren erwartete Rendite zu hoch bzw. deren Risiko zu gering geschätzt werden, ein zu hohes Portfoliogewicht erhalten, während Assets, deren erwartete Rendite zu niedrig bzw. deren Risiko zu hoch geschätzt werden, ein zu niedriges Gewicht zugewiesen bekommen. Im schlimmsten Fall erhalten diejenigen Assets

das größte Gewicht, denen der größte Schätzfehler zugrunde liegt. Aus diesem Grund bezeichnet Michaud (1989) das Mean-Variance-Optimierungsproblem auch als „Schätzfehler-Maximierer“. Best und Grauer (1991) fanden in diesem Zusammenhang heraus, dass effiziente Portfolios sehr empfindlich auf Änderungen im Erwartungswert der Renditen reagieren. Dieses Phänomen wird als *Markowitz'sches Optimierungsenigma* bezeichnet.

3. Verschiedene Ansätze zum Schätzen von $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ und \mathbf{w}^*

Die Mean-Variance-Optimierung nach Markowitz setzt die Kenntnis der erwarteten Renditen und ihrer Varianzen und Kovarianzen voraus. In der Praxis ist es allerdings nicht möglich, die wahren Werte für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ zu kennen, da es sich um unbekannte Populationsgrößen handelt. Sie müssen also geeignet geschätzt werden. Es liegt zunächst nahe, die unbekannt Parameter durch ihre empirischen Pendanten zu ersetzen, die sich aus historischen Daten ermitteln lassen. Doch diverse Studien belegen, dass man, wenn man die effizienten Portfoliogewichte schätzt, indem man die empirischen Parameter in das Mean-Variance-Optimierungsproblem einsetzt, eine Annäherung der Effizienzkurve erhält, die schon bei kleinen Schätzfehlern weit unterhalb der wahren Effizienzkurve liegen kann (vgl. z.B. Frankfurter, Phillips und Seagle (1976)). Dabei fanden Best und Grauer (1991) heraus, dass μ - σ -effiziente Portfolios sehr sensibel auf Änderungen in den Erwartungswerten der Assetrenditen reagieren.

Es gibt verschiedene Ansätze, die das Problem der ineffizienten Portfolios durch fehlerhaft geschätzte Parameter angehen. Eine hohe Anzahl an vorhandenen Assets hat zur Folge, dass sehr viele Werte geschätzt werden müssen, was die Gefahr für Schätzfehler wachsen lässt. Multifaktormodelle bieten eine Möglichkeit, die Dimension der zu schätzenden Parameter und somit auch die Schätzfehler erheblich zu reduzieren. Ein zweiter Ansatz versucht die Verzerrung der Portfoliogewichte, die man durch Einsetzen von Schätzern in das Mean-Variance-Optimierungsproblem erhält, mittels Bootstrap zu korrigieren (siehe Michaud (1989, 1999)). In einem dritten Ansatz werden Bayes-Schätzer oder andere Shrinkage-Schätzer anstelle von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ verwendet (siehe Lai, Xing und Chen (2011) und Ledoit und Wolf (2003)).

Diese verschiedenen Methoden zum Schätzen der Parameter des Mean-Variance-Optimierungsproblems und der effizienten Portfoliogewichte werden in den folgenden Abschnitten genauer vorgestellt.

3.1. Empirische Schätzer für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$

Es seien $\mathbf{r}_t = (r_{1t}, \dots, r_{mt})^T$, $t = 1, \dots, n$, die vergangenen Renditen von m Assets aus n Perioden. Eine gebräuchliche Annahme ist, dass die Renditen unabhängig und identisch normalverteilt sind, obwohl Mandelbrot (1963) herausfand, dass die Verteilung von Renditen schwerere *tails* als eine Normalverteilung besitzt, da die Wahrscheinlichkeit für extremere Preisänderungen in Wirklichkeit höher ist als unter der Annahme, dass die Renditen normalverteilt sind. Dennoch unterstellt man Renditen sehr häufig eine Normalverteilung. Ist

$\boldsymbol{\mu}$ der Erwartungswert und $\boldsymbol{\Sigma}$ die Kovarianzmatrix der Renditen, so wird also $\mathbf{r}_t \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ angenommen. Der empirische Mittelwert der Renditen ist definiert als

$$\bar{\mathbf{r}} = \frac{1}{n} \sum_{t=1}^n \mathbf{r}_t$$

und ihre empirische Kovarianzmatrix als

$$\mathbf{S}^* = \frac{1}{n-1} \sum_{t=1}^n (\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^T.$$

Beide haben die Eigenschaft, dass sie unverzerrte Schätzer des wahren Erwartungswerts bzw. der wahren Kovarianzmatrix sind. Unter der Normalverteilungsannahme der Renditen gilt außerdem, dass $\bar{\mathbf{r}}$ und $\mathbf{S} := (n-1)/n \mathbf{S}^*$ die Maximum-Likelihood-Schätzer (ML-Schätzer) von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ sind. Eine wichtige Eigenschaft von ML-Schätzern ist, dass sie asymptotisch effizient sind, d.h. für einen ML-Schätzer $\hat{\theta}_n$ gilt die Verteilungskonvergenz

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1}), \quad (3.1)$$

wobei θ_0 der unbekannte wahre Parameter und $\mathcal{I}(\theta_0)^{-1}$ die inverse Fisher-Information ist; siehe Ferguson (1996). Für große n ist also

$$\hat{\theta}_n \approx \theta_0 + \frac{1}{\sqrt{n}} Z$$

mit $Z \sim \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1})$. Das bedeutet, dass $\hat{\theta}_n$ für große n im Mittel gleich θ_0 ist und mit einer Varianz von $(n\mathcal{I}(\theta_0))^{-1}$ um diesen Wert streut. Mit steigendem n wird diese Varianz aber immer kleiner. Wichtig ist hierbei auch, dass die inverse Fisher-Information die kleinstmögliche Varianz eines Schätzers für den wahren Parameter ist. Aus diesem Grund wird im Folgenden, wenn von der empirischen Kovarianzmatrix als Schätzer von $\boldsymbol{\Sigma}$ die Rede ist, der ML-Schätzer \mathbf{S} verwendet.

Nach Muirhead (1982) sind $\bar{\mathbf{r}} =: \hat{\boldsymbol{\mu}}_{emp}$ und $\mathbf{S} =: \hat{\boldsymbol{\Sigma}}_{emp}$ unabhängig voneinander und besitzen die Verteilungen

$$\hat{\boldsymbol{\mu}}_{emp} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n), \quad (3.2)$$

$$\hat{\boldsymbol{\Sigma}}_{emp} \sim \mathcal{W}_m(n-1, \boldsymbol{\Sigma}/n), \quad (3.3)$$

wobei $\mathcal{W}_m(n-1, \boldsymbol{\Sigma}/n)$ eine Wishart-Verteilung mit $n-1$ Freiheitsgraden ist. Für die Invertierbarkeit von $\hat{\boldsymbol{\Sigma}}_{emp}$ ist es notwendig, dass $n > m$ ist. Dies lässt sich zeigen, indem man die empirische Kovarianzmatrix in der Form $\hat{\boldsymbol{\Sigma}}_{emp} = 1/n \mathbf{D} \mathbf{D}^T$ darstellt, wobei \mathbf{D} eine $(m \times n)$ -Matrix ist, deren Spalten $\mathbf{d}_1, \dots, \mathbf{d}_n$ den Abweichungen der Renditen vom empirischen Mittelwert entsprechen, also $\mathbf{d}_t = (\mathbf{r}_t - \hat{\boldsymbol{\mu}}_{emp})$, $t = 1, \dots, n$. Da $\sum_{t=1}^n \mathbf{d}_t = \sum_{t=1}^n \mathbf{r}_t - n \hat{\boldsymbol{\mu}}_{emp} = \mathbf{0}_m$ ist, die Zeilensummen von \mathbf{D} also gleich null sind, kann der Spaltenrang von \mathbf{D} nicht größer als $n-1$ sein. Es gilt also $rg(\mathbf{D}) \leq \min\{m, n-1\}$. Da der Rang des

Produkts zweier Matrizen höchstens dem kleineren Rang der beiden Matrizen entsprechen kann, gilt somit insgesamt

$$rg(\widehat{\boldsymbol{\Sigma}}_{emp}) = rg(\mathbf{D}\mathbf{D}^T) \leq rg(\mathbf{D}) \leq \min\{n-1, m\}.$$

Damit $\widehat{\boldsymbol{\Sigma}}_{emp}$ vollen Rang besitzt, muss also $n > m$ gelten.

Da in der Praxis üblicherweise die empirischen Schätzer zur Annäherung der erwarteten Renditen, Varianzen und Kovarianzen verwendet werden, befasst sich Kapitel 4 intensiver mit der Auswirkung der empirischen Schätzer auf die Approximation der Minimum-Varianz-Kurve.

3.2. Multifaktor-Pricing-Modelle

Bei einer hohen Anzahl m an Assets wird man beim Schätzen der Parameter für das Mean-Variance-Optimierungsproblem schnell mit einem Dimensionsproblem konfrontiert, da man jeweils m Erwartungswerte und Varianzen und $m(m-1)/2$ Kovarianzen, also insgesamt $2m + m(m-1)/2$ Parameter zu schätzen hat, was für große m mit einem enormen Aufwand verbunden ist. Mithilfe von Faktormodellen lässt sich die Anzahl der zu schätzenden Parameter erheblich reduzieren; siehe z.B. Tsay (2010). In solchen Modellen geht man davon aus, dass die Renditen aller Assets im Wesentlichen von einem oder mehreren gemeinsamen Faktoren beeinflusst werden. Die Rendite r_{it} lässt sich dann schreiben als

$$r_{it} = \alpha_i + \boldsymbol{\beta}_i^T \mathbf{f}_t + \epsilon_{it}, \quad (3.4)$$

wobei $\mathbf{f}_t = (f_{1t}, \dots, f_{kt})^T$ mit $k < m$ die Werte der k Faktoren sind, von denen anzunehmen ist, dass sie die Renditen maßgeblich beeinflussen. Die Koeffizienten $\alpha_i \in \mathbb{R}$ und $\boldsymbol{\beta}_i \in \mathbb{R}^k$ sind konstant. Da die Werte von $\boldsymbol{\beta}_i$ angeben, wie stark der Einfluss der einzelnen Faktoren auf die Renditen ist, nennt man sie auch Faktorladungen. Aufgrund der zufälligen Entwicklung der Renditen, wird der Fehler ϵ_{it} hinzuaddiert. Für jede Rendite ist dieser Fehler im Mittel gleich 0 und mit den Fehlern der anderen Renditen sowie mit den einzelnen Faktoren unkorreliert. Für ϵ_{it} , $i = 1, \dots, m$, gilt also $E\epsilon_{it} = 0$, $E\epsilon_{it}\epsilon_{jt} = 0$ für $i \neq j$ und $Cov(f_{lt}, \epsilon_{it}) = 0$ mit $l = 1, \dots, k$. Die Koeffizienten α_i und $\boldsymbol{\beta}_i$ lassen sich mittels einfacher linearer Regression schätzen. Mit dem Erwartungswertvektor $\boldsymbol{\zeta} = E[\mathbf{f}_{t+1}]$, der Kovarianzmatrix $\boldsymbol{\Sigma}_f = Var(\mathbf{f}_{t+1})$ und der Varianz der zufälligen Fehler $\sigma_{\epsilon_i}^2 = Var(\epsilon_{i,t+1})$, $i = 1, \dots, m$ gilt

$$\mu_i = \alpha_i + \boldsymbol{\beta}_i^T \boldsymbol{\zeta}, \quad \sigma_i^2 = \boldsymbol{\beta}_i^T \boldsymbol{\Sigma}_f \boldsymbol{\beta}_i + \sigma_{\epsilon_i}^2$$

und

$$\begin{aligned} \sigma_{i,j} &= E[(r_{i,t+1} - \mu_i)(r_{j,t+1} - \mu_j)] = E[(\boldsymbol{\beta}_i^T (\mathbf{f}_t - \boldsymbol{\zeta}) + \epsilon_{it})(\boldsymbol{\beta}_j^T (\mathbf{f}_t - \boldsymbol{\zeta}) + \epsilon_{jt})] \\ &= E[\boldsymbol{\beta}_i^T (\mathbf{f}_t - \boldsymbol{\zeta})(\mathbf{f}_t - \boldsymbol{\zeta})^T \boldsymbol{\beta}_j] = \boldsymbol{\beta}_i^T \boldsymbol{\Sigma}_f \boldsymbol{\beta}_j. \end{aligned}$$

In diesem Modell sind mit α_i , $\boldsymbol{\beta}_i$, $\sigma_{\epsilon_i}^2$, $\boldsymbol{\zeta}$ und $\boldsymbol{\Sigma}_f$ also insgesamt $2m + km + 2k + k(k-1)/2$ Parameter zu schätzen. Je weniger Faktoren in das Modell aufgenommen werden, umso

stärker verringert sich die Anzahl der zu schätzenden Parameter. Für $k = 1$ sind zum Beispiel nur noch $3m + 2$ Parameter zu schätzen.

Falls man nur einen Faktor in das Modell aufnimmt, spricht man von einem Einfaktor-Modell. Ein Spezialfall eines solchen Modells ist das *Capital Asset Pricing Model* (CAPM) von Sharpe (1964), da für $k = 1$, $\alpha_i = r_f$ und $\mathbf{f}_t = r_{Mt} - r_f$, wobei r_f die Rendite der sicheren Anlage am Markt und r_{Mt} die Rendite des Marktportfolios sei,

$$r_{it} = r_f + (r_{Mt} - r_f) \beta_i + \epsilon_{it}$$

gilt. Bildet man von diesem Ausdruck den Erwartungswert, so erhält man die CAPM-Gleichung

$$\mu_i = r_f + (\mu_M - r_f) \beta_i.$$

Das Marktportfolio ist allerdings nur ein theoretisches Portfolio und wird daher in der Praxis häufig durch einen breiten Marktindex wie den DAX oder den S&P 500 approximiert.

Die Darstellung der Renditen mithilfe dieses Faktormodells verdeutlicht die Aufteilung des Gesamtrisikos einer Rendite r_{it} in ein systematisches und ein unsystematisches Risiko: Bildet man die Varianz

$$\text{Var}(r_{it}) = \text{Var}(r_f + (r_{Mt} - r_f) \beta_i + \epsilon_{it}) = \beta_i^2 \sigma_{Mt}^2 + \sigma_{\epsilon_i}^2,$$

wobei $\sigma_{\epsilon_i}^2$ die Varianz von ϵ_i ist, erhält man die Aufteilung in das allein vom Markt ausgehende und somit systematische Risiko gewichtet mit dem Beta-Faktor und das allein vom spezifischen zufälligen Fehler ϵ_{it} resultierende und somit unsystematische Risiko.

Eine weitere Kapitalmarkttheorie ist die von Ross (1976) entwickelte *Arbitrage Pricing Theorie* (APT), bei der ebenfalls angenommen wird, dass sich die Rendite eines Assets durch ein Faktormodell beschreiben lässt und somit die Form

$$r_{it} = \alpha_i + \boldsymbol{\beta}_i^T \mathbf{f}_t + \epsilon_{it}$$

mit $\alpha_i \in \mathbb{R}$, $\boldsymbol{\beta}_i \in \mathbb{R}^k$, $\mathbf{f}_t \in \mathbb{R}^k$, $k < m$, und den entsprechenden Annahmen an die zufälligen Fehler ϵ_{it} hat. Die APT besagt, dass unter der Annahme von Arbitragefreiheit und für eine genügend große Anzahl Assets m die erwartete Rendite gleich

$$\mu_i = r_f + \boldsymbol{\beta}_i^T \boldsymbol{\lambda}$$

ist, wobei $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^T$ geeignete Konstanten sind.

Der Vorteil der APT gegenüber dem CAPM besteht darin, dass die APT die Renditen mit mehr als nur einem Faktor erklärt, was für die Praxis realistischer erscheint. Problematisch ist hierbei allerdings, dass die Frage, mit welchen und wie vielen Faktoren die Renditen ausreichend beschrieben werden können, bisher nicht befriedigend beantwortet werden konnte.

3.3. Bootstrapping und die *resampled frontier*

Löst man das Mean-Variance-Optimierungsproblem mit geschätzten Parametern, besteht die Gefahr, dass Assets, deren erwartete Rendite zu hoch und deren Risiko zu niedrig geschätzt werden, mit einem zu hohen Gewicht in das Portfolio eingehen, während Assets, deren Schätzungen für die erwartete Rendite zu niedrig und für das Risiko zu hoch ausfallen, ein zu geringes Gewicht bekommen, sodass man auf diese Weise oft sehr ineffiziente Portfolios erhält. Um dieser Verzerrung der Gewichte entgegenzuwirken, schlägt Michaud (1989, 1999) eine Methode vor, in der man mittels Bootstrap für eine bestimmte erwartete Zielrendite immer neue Portfoliogewichte bestimmt und anschließend den Durchschnitt dieser Gewichte bildet. Dazu bestimmt man zunächst den empirischen Mittelwert $\hat{\mu}_{emp}$ und die empirische Kovarianzmatrix $\hat{\Sigma}_{emp}$ einer Menge historischer Renditen $\{r_1, \dots, r_n\}$ und generiert anschließend aus einer Verteilung mit diesen Parametern B Bootstrap-Stichproben $\{r_{b1}^*, \dots, r_{bn}^*\}$, $b = 1, \dots, B$, der Größe n . Für jede dieser Bootstrap-Stichproben lassen sich die effizienten Portfoliogewichte \hat{w}_b^* ermitteln, indem man das Mean-Variance-Optimierungsproblem mit dem empirischen Mittelwert $\hat{\mu}_b^*$ und der empirischen Kovarianzmatrix $\hat{\Sigma}_b^*$ der jeweiligen Bootstrap-Stichprobe löst. Bestimmt man zu jeder Bootstrap-Stichprobe für viele nahe beieinanderliegende Werte der Zielrendite μ_* die effizienten Gewichte und berechnet anschließend den Mittelwert der zu jeweils demselben Wert $\tilde{\mu}_*$ gehörenden Gewichte, also

$$\bar{w}(\tilde{\mu}_*) = \frac{1}{B} \sum_{b=1}^B \hat{w}_b^*(\tilde{\mu}_*),$$

erhält man die Portfoliogewichte, mit denen die *resampled frontier* gebildet wird. Es sei an dieser Stelle allerdings angemerkt, dass die aus diesem Verfahren resultierenden Gewichte keine μ - σ -effizienten Portfolios bilden, d.h. sie sind keine Lösung des Mean-Variance-Optimierungsproblems von Markowitz.

3.4. Bayes- und Shrinkage-Schätzer für μ und Σ

In der Statistik gibt es zwei grundlegend verschiedene Ansichten in Bezug auf die Definition von Wahrscheinlichkeiten (siehe z.B. Samaniego (2010)). Die sogenannten Frequentisten definieren die Wahrscheinlichkeit für das Eintreten eines Ereignisses als seine relative Häufigkeit bei der ständigen Wiederholung eines Zufallsexperiments. Dieser sehr objektive Ansatz hat jedoch den Nachteil, dass es in der Realität oft nicht möglich ist, dass ein Ereignis sich wiederholt, sodass dessen Eintrittswahrscheinlichkeit auf frequentistische Weise nicht bestimmt werden kann. Dies ist jedoch möglich mit der Bayes'schen Definition des Wahrscheinlichkeitsbegriffs. In der Bayes'schen Statistik wird die Wahrscheinlichkeit als Maß für den Grad der subjektiven Überzeugung für das Eintreten eines Ereignisses interpretiert, wobei verfügbare Informationen zur Bildung von Wahrscheinlichkeiten herangezogen werden. Vor allem bedingte Wahrscheinlichkeiten, die mithilfe des Bayes-Theorems bestimmt werden können, sind hier von Bedeutung.

3.4.1. Bayes-Schätzer

Unter der Annahme, dass die erwartete Rendite $\boldsymbol{\mu}$ und die Kovarianzmatrix $\boldsymbol{\Sigma}$ im Mean-Variance-Optimierungsproblem gegeben sind, werden die Renditen \mathbf{r}_t , $t = 1, \dots, n$, häufig als unabhängig und identisch $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -verteilt angenommen. Sind $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ unbekannt und möchte man sie in einem Bayes'schen Modell schätzen, werden sie als Zufallsvariablen betrachtet. Die Normalverteilungsannahme an die Renditen ist dann nicht mehr gültig. Es sei $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ die A-priori-Verteilung von $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ und $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathcal{R}_n)$ ihre A-posteriori-Verteilung gegeben die aus den letzten n Perioden verfügbare Information über die vergangenen Renditen $\mathcal{R}_n = \{\mathbf{r}_1, \dots, \mathbf{r}_n\}$. Gehören diese beiden Verteilungen derselben parametrischen Verteilungsfamilie an, so spricht man von einer *konjugierten* A-priori-Verteilung. Zur Schätzung von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ genügt es, die verfügbare Information über die vergangenen Renditen als empirischen Mittelwert $\bar{\mathbf{r}} = 1/n \sum_{t=1}^n \mathbf{r}_t$ und empirische Kovarianzmatrix $\mathbf{S} = 1/n \sum_{t=1}^n (\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^T$ zusammenzufassen, da sie suffiziente Statistiken für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ sind. Die gemeinsame Verteilung von $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ lässt sich in die bedingte Verteilung von $\boldsymbol{\mu}$ gegeben $\boldsymbol{\Sigma}$ und die Marginalverteilung von $\boldsymbol{\Sigma}$ aufspalten. Eine für die Parameter von Renditen häufig angenommene konjugierte A-priori-Verteilung ist

$$\boldsymbol{\mu} | \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/\kappa) \quad (3.5)$$

mit $\boldsymbol{\mu}_0 \in \mathbb{R}^m$ und $\kappa \in \mathbb{R}_+$ und

$$\boldsymbol{\Sigma} \sim \mathcal{IW}_m(\nu_0, \nu_0 \boldsymbol{\Sigma}_0), \quad (3.6)$$

wobei $\mathcal{IW}_m(\nu_0, \nu_0 \boldsymbol{\Sigma}_0)$ die inverse Wishart-Verteilung mit ν_0 Freiheitsgraden und symmetrischer und positiv definiten $(m \times m)$ -Matrix $\boldsymbol{\Sigma}_0$ ist; siehe Meucci (2009). Die A-posteriori-Verteilung lässt sich mithilfe der stetigen Version des Bayes-Theorems bestimmen:

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathcal{R}_n) = \frac{f(\mathcal{R}_n | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\int_{\Theta} f(\mathcal{R}_n | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) d(\boldsymbol{\mu}, \boldsymbol{\Sigma})}.$$

Dabei ist $f(\mathcal{R}_n | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ die Likelihoodfunktion $l(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathcal{R}_n)$ und Θ der Parameterraum von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$, also $\Theta = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) | \boldsymbol{\mu} \in \mathbb{R}^m, \boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}, \boldsymbol{\Sigma} \text{ positiv definit}\}$. Der folgende Satz und dessen Beweis sind bei Meucci (2009) zu finden.

Satz 3.4.1 *Ist die A-priori-Verteilung von $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ wie in (3.5) und (3.6), dann ist die A-posteriori-Verteilung von $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ von der Form*

$$\boldsymbol{\mu} | \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}/\lambda) \quad (3.7)$$

mit $\boldsymbol{\mu}_1 = \kappa/(n + \kappa)\boldsymbol{\mu}_0 + n/(n + \kappa)\bar{\mathbf{r}}$ und $\lambda = n + \kappa$ und

$$\boldsymbol{\Sigma} \sim \mathcal{IW}_m(\nu_1, \nu_1 \boldsymbol{\Sigma}_1) \quad (3.8)$$

mit $\nu_1 = n + \nu_0$, $\boldsymbol{\Sigma}_1 = (n\mathbf{S} + \nu_0 \boldsymbol{\Sigma}_0 + \boldsymbol{\Phi})/\nu_1$ und $\boldsymbol{\Phi} = n\kappa/(\kappa + n)(\bar{\mathbf{r}} - \boldsymbol{\mu}_0)(\bar{\mathbf{r}} - \boldsymbol{\mu}_0)^T$.

Beweis. Im Folgenden bezeichnen $\gamma_1, \gamma_2, \dots$ Normierungskonstanten. Statt der inversen Wishart-Verteilung von $\boldsymbol{\Sigma}$ ist es einfacher die Verteilung von $\boldsymbol{\Sigma}^{-1} =: \boldsymbol{\Omega}$ zu betrachten. Es gilt

$$\boldsymbol{\Omega} \sim \mathcal{W}_m \left(\nu_0, (\nu_0 \boldsymbol{\Sigma}_0)^{-1} \right)$$

und

$$\boldsymbol{\mu} | \boldsymbol{\Omega} \sim \mathcal{N} \left(\boldsymbol{\mu}_0, (\kappa \boldsymbol{\Omega})^{-1} \right).$$

Damit folgt für die gemeinsame A-priori-Verteilung von $\boldsymbol{\mu}$ und $\boldsymbol{\Omega}$

$$\begin{aligned} \pi(\boldsymbol{\mu}, \boldsymbol{\Omega}) &= \pi(\boldsymbol{\mu} | \boldsymbol{\Omega}) \pi(\boldsymbol{\Omega}) \\ &= \gamma_1 |\boldsymbol{\Omega}|^{\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T (\kappa \boldsymbol{\Omega}) (\boldsymbol{\mu} - \boldsymbol{\mu}_0)} |\boldsymbol{\Sigma}_0|^{\frac{\nu_0}{2}} |\boldsymbol{\Omega}|^{\frac{\nu_0 - m - 1}{2}} e^{-\frac{1}{2} \text{tr}(\nu_0 \boldsymbol{\Sigma}_0 \boldsymbol{\Omega})}. \end{aligned}$$

Da für den empirischen Mittelwert

$$\bar{\mathbf{r}} \sim \mathcal{N} \left(\boldsymbol{\mu}, (n \boldsymbol{\Omega})^{-1} \right)$$

und für die empirische Kovarianzmatrix

$$n\mathbf{S} \sim \mathcal{W}_m(n-1, \boldsymbol{\Sigma})$$

gilt und die beiden Größen unabhängig sind, ist die Wahrscheinlichkeitsdichte von $(\bar{\mathbf{r}}, n\mathbf{S})$ gegeben $(\boldsymbol{\mu}, \boldsymbol{\Omega})$ gleich

$$\begin{aligned} f(\bar{\mathbf{r}}, n\mathbf{S} | \boldsymbol{\mu}, \boldsymbol{\Omega}) &= f(\bar{\mathbf{r}} | \boldsymbol{\mu}, \boldsymbol{\Omega}) f(n\mathbf{S} | \boldsymbol{\mu}, \boldsymbol{\Omega}) \\ &= \gamma_2 |\boldsymbol{\Omega}|^{\frac{1}{2}} e^{-\frac{1}{2}(\bar{\mathbf{r}} - \boldsymbol{\mu})^T (n \boldsymbol{\Omega}) (\bar{\mathbf{r}} - \boldsymbol{\mu})} |\boldsymbol{\Omega}|^{\frac{n-1}{2}} |\mathbf{S}|^{\frac{n-m-2}{2}} e^{-\frac{1}{2} \text{tr}(n \boldsymbol{\Omega} \mathbf{S})}. \end{aligned}$$

Die gemeinsame Wahrscheinlichkeitsdichte von $\bar{\mathbf{r}}, n\mathbf{S}, \boldsymbol{\mu}$ und $\boldsymbol{\Omega}$ ist also von der Form

$$\begin{aligned} f(\bar{\mathbf{r}}, n\mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Omega}) &= f(\bar{\mathbf{r}}, n\mathbf{S} | \boldsymbol{\mu}, \boldsymbol{\Omega}) \pi(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ &= \gamma_3 e^{-\frac{1}{2} \{ (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T (\kappa \boldsymbol{\Omega}) (\boldsymbol{\mu} - \boldsymbol{\mu}_0) + (\boldsymbol{\mu} - \bar{\mathbf{r}})^T (n \boldsymbol{\Omega}) (\boldsymbol{\mu} - \bar{\mathbf{r}}) \}} \\ &\quad |\mathbf{S}|^{\frac{n-m-2}{2}} |\boldsymbol{\Sigma}_0|^{\frac{\nu_0}{2}} |\boldsymbol{\Omega}|^{\frac{n+\nu_0-m}{2}} e^{-\frac{1}{2} \text{tr}(n \boldsymbol{\Omega} \mathbf{S} + \nu_0 \boldsymbol{\Sigma}_0 \boldsymbol{\Omega})}. \end{aligned}$$

Durch geeignetes Erweitern und Umstellen des Ausdrucks in geschwungenen Klammern kann man diesen umschreiben als

$$\{\dots\} = (\boldsymbol{\mu} - \boldsymbol{\mu}_1)^T \lambda \boldsymbol{\Omega} (\boldsymbol{\mu} - \boldsymbol{\mu}_1) + \text{tr}(\boldsymbol{\Phi} \boldsymbol{\Omega}),$$

wobei

$$\begin{aligned} \lambda &:= \kappa + n \\ \boldsymbol{\mu}_1 &:= \frac{\kappa \boldsymbol{\mu}_0 + n \bar{\mathbf{r}}}{\kappa + n} \\ \boldsymbol{\Phi} &:= \frac{n \kappa}{\kappa + n} (\bar{\mathbf{r}} - \boldsymbol{\mu}_0) (\bar{\mathbf{r}} - \boldsymbol{\mu}_0)^T \end{aligned}$$

gilt. Indem man ferner

$$\boldsymbol{\Sigma}_1 := \frac{n\mathbf{S} + \nu_0\boldsymbol{\Sigma}_0 + \Phi}{\nu_1}$$

setzt, wobei ν_1 noch zu definieren ist, kann die gemeinsame Dichte von $\bar{\mathbf{r}}, n\mathbf{S}, \boldsymbol{\mu}$ und $\boldsymbol{\Omega}$ folgendermaßen umgeschrieben werden:

$$f(\bar{\mathbf{r}}, n\mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Omega}) = \gamma_3 e^{-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_1)^T \lambda \boldsymbol{\Omega} (\boldsymbol{\mu}-\boldsymbol{\mu}_1)} |\mathbf{S}|^{\frac{n-m-2}{2}} |\boldsymbol{\Sigma}_0|^{\frac{\nu_0}{2}} |\boldsymbol{\Omega}|^{\frac{n+\nu_0-m}{2}} e^{-\frac{1}{2}tr(\nu_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Omega})}.$$

Nun lässt sich die Marginalverteilung von $(\bar{\mathbf{r}}, n\mathbf{S})$ mittels Integration über $(\boldsymbol{\mu}, \boldsymbol{\Omega})$ bestimmen:

$$\begin{aligned} f(\bar{\mathbf{r}}, n\mathbf{S}) &= \int f(\bar{\mathbf{r}}, n\mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Omega}) d\boldsymbol{\mu} d\boldsymbol{\Omega} \\ &= \gamma_4 \int \left\{ \int \gamma_5 |\boldsymbol{\Omega}|^{\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_1)^T \lambda \boldsymbol{\Omega} (\boldsymbol{\mu}-\boldsymbol{\mu}_1)} d\boldsymbol{\mu} \right\} \\ &\quad |\mathbf{S}|^{\frac{n-m-2}{2}} |\boldsymbol{\Sigma}_0|^{\frac{\nu_0}{2}} |\boldsymbol{\Omega}|^{\frac{n+\nu_0-m-1}{2}} e^{-\frac{1}{2}tr(\nu_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Omega})} d\boldsymbol{\Omega} \\ &= \gamma_4 \int |\mathbf{S}|^{\frac{n-m-2}{2}} |\boldsymbol{\Sigma}_0|^{\frac{\nu_0}{2}} |\boldsymbol{\Omega}|^{\frac{n+\nu_0-m-1}{2}} e^{-\frac{1}{2}tr(\nu_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Omega})} d\boldsymbol{\Omega}. \end{aligned}$$

Dabei nutzt man im letzten Schritt aus, dass in den geschwungenen Klammern die Dichte einer Normalverteilung über den gesamten Raum integriert wird und somit eins ergibt. Setzt man nun

$$\nu_1 := n + \nu_0$$

und erweitert den Ausdruck für $f(\bar{\mathbf{r}}, n\mathbf{S})$ mit $|\boldsymbol{\Sigma}_1|^{-\nu_1/2} |\boldsymbol{\Sigma}_1|^{\nu_1/2}$, so erhält man schließlich

$$\begin{aligned} f(\bar{\mathbf{r}}, n\mathbf{S}) &= \gamma_6 |\mathbf{S}|^{\frac{n-m-2}{2}} |\boldsymbol{\Sigma}_0|^{\frac{\nu_0}{2}} |\boldsymbol{\Sigma}_1|^{-\frac{\nu_1}{2}} \left\{ \int \gamma_7 |\boldsymbol{\Sigma}_1|^{\frac{\nu_1}{2}} |\boldsymbol{\Omega}|^{\frac{\nu_1-m-1}{2}} e^{-\frac{1}{2}tr(\nu_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Omega})} d\boldsymbol{\Omega} \right\} \\ &= \gamma_6 |\mathbf{S}|^{\frac{n-m-2}{2}} |\boldsymbol{\Sigma}_0|^{\frac{\nu_0}{2}} |\boldsymbol{\Sigma}_1|^{-\frac{\nu_1}{2}}, \end{aligned}$$

wobei im letzten Schritt ausgenutzt wird, dass in den geschwungenen Klammern die Dichte einer Wishart-Verteilung über den gesamten Raum integriert wird und der Ausdruck somit eins ergibt. Durch Dividieren der gemeinsamen Verteilung von $(\bar{\mathbf{r}}, n\mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Omega})$ durch die Randverteilung von $(\bar{\mathbf{r}}, n\mathbf{S})$ erhält man die A-posteriori-Verteilung von $(\boldsymbol{\mu}, \boldsymbol{\Omega})$:

$$\begin{aligned} \pi(\boldsymbol{\mu}, \boldsymbol{\Omega} | \bar{\mathbf{r}}, n\mathbf{S}) &= \frac{f(\bar{\mathbf{r}}, n\mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Omega})}{f(\bar{\mathbf{r}}, n\mathbf{S})} \\ &= \gamma_8 |\boldsymbol{\Omega}|^{\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_1)^T \lambda \boldsymbol{\Omega} (\boldsymbol{\mu}-\boldsymbol{\mu}_1)} |\boldsymbol{\Sigma}_1|^{\frac{\nu_1}{2}} |\boldsymbol{\Omega}|^{\frac{\nu_1-m-1}{2}} e^{-\frac{1}{2}tr(\nu_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Omega})}. \end{aligned}$$

Daraus folgt nun für die A-posteriori-Verteilung von $(\boldsymbol{\mu}, \boldsymbol{\Omega})$ gegeben $(\bar{\mathbf{r}}, n\mathbf{S})$

$$\boldsymbol{\mu} | \boldsymbol{\Omega} \sim \mathcal{N}(\boldsymbol{\mu}_1, (\lambda \boldsymbol{\Omega})^{-1})$$

und

$$\boldsymbol{\Omega} \sim \mathcal{W}_m(\nu_1, (\nu_1 \boldsymbol{\Sigma}_1)^{-1}),$$

und somit für die A-posteriori-Verteilung von $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ gegeben \mathcal{R}_n

$$\boldsymbol{\mu} | \boldsymbol{\Sigma} \sim \mathcal{N}\left(\boldsymbol{\mu}_1, \frac{\boldsymbol{\Sigma}}{\lambda}\right)$$

und

$$\boldsymbol{\Sigma} \sim \mathcal{IW}_m(\nu_1, \nu_1 \boldsymbol{\Sigma}_1).$$

□

Nach Satz B.0.2 (im Anhang) ist der Bayes-Schätzer $\hat{\boldsymbol{\mu}}_{Bayes}$ für $\boldsymbol{\mu}$ unter der A-priori-Verteilung von $\boldsymbol{\mu}$ gegeben $\boldsymbol{\Sigma}$ in (3.5) und der quadratischen Verlustfunktion der bedingte Erwartungswert $E[\boldsymbol{\mu} | \mathcal{R}_n]$:

$$\hat{\boldsymbol{\mu}}_{Bayes} = E[\boldsymbol{\mu} | \mathcal{R}_n] = \boldsymbol{\mu}_1 = \frac{\kappa \boldsymbol{\mu}_0 + n \bar{\mathbf{r}}}{\kappa + n} = \frac{\kappa}{\kappa + n} \boldsymbol{\mu}_0 + \frac{n}{\kappa + n} \bar{\mathbf{r}}. \quad (3.9)$$

Den Bayes-Schätzer $\hat{\boldsymbol{\Sigma}}_{Bayes}$ für $\boldsymbol{\Sigma}$ erhält man analog durch

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{Bayes} &= E[\boldsymbol{\Sigma} | \mathcal{R}_n] = \frac{\nu_1 \boldsymbol{\Sigma}_1}{\nu_1 - m - 1} = \frac{(n + \nu_0) \frac{n \mathbf{S} + \nu_0 \boldsymbol{\Sigma}_0 + \boldsymbol{\Phi}}{n + \nu_0}}{n + \nu_0 - m - 1} \\ &= \frac{\sum_{t=1}^n (\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^T + \nu_0 \boldsymbol{\Sigma}_0 + \frac{n\kappa}{\kappa + n} (\bar{\mathbf{r}} - \boldsymbol{\mu}_0)(\bar{\mathbf{r}} - \boldsymbol{\mu}_0)^T}{n + \nu_0 - m - 1} \\ &= \frac{\nu_0 - m - 1}{n + \nu_0 - m - 1} \frac{\nu_0 \boldsymbol{\Sigma}_0}{\nu_0 - m - 1} + \\ &\quad \frac{n}{n + \nu_0 - m - 1} \left\{ \frac{1}{n} \sum_{t=1}^n (\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^T + \frac{\kappa}{\kappa + n} (\bar{\mathbf{r}} - \boldsymbol{\mu}_0)(\bar{\mathbf{r}} - \boldsymbol{\mu}_0)^T \right\}. \end{aligned} \quad (3.10)$$

Die A-posteriori-Verteilung von $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ lässt sich somit auch schreiben als

$$\boldsymbol{\mu} | \boldsymbol{\Sigma} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_{Bayes}, \boldsymbol{\Sigma}/\lambda)$$

und

$$\boldsymbol{\Sigma} \sim \mathcal{IW}_m(\nu_1, (n + \nu_0 - m - 1) \hat{\boldsymbol{\Sigma}}_{Bayes}).$$

Betrachtet man den Bayes-Schätzer für $\boldsymbol{\mu}$, erkennt man, dass dieser eine konvexe Kombination aus dem Stichprobenmittelwert $\bar{\mathbf{r}}$, also dem ML-Schätzer von $\boldsymbol{\mu}$ unter der Normalverteilungsannahme an die Renditen, und dem A-priori-Erwartungswert $\boldsymbol{\mu}_0$ ist. Damit berücksichtigt der Bayes-Schätzer neben der Schätzung für $\boldsymbol{\mu}$ auch die Annahmen an dessen A-priori-Verteilung. Auch der Bayes-Schätzer der Kovarianzmatrix ist eine konvexe Kombination aus dem A-priori-Erwartungswert $\nu_0 \boldsymbol{\Sigma}_0 / (\nu_0 - m - 1)$ und einer angepassten Form der Stichprobenkovarianzmatrix, also des ML-Schätzers von $\boldsymbol{\Sigma}$ unter der Normalverteilungsannahme der Renditen. Die Anpassung erfolgt durch das Hinzuaddieren von $\kappa / (\kappa + n) (\bar{\mathbf{r}} - \boldsymbol{\mu}_0)(\bar{\mathbf{r}} - \boldsymbol{\mu}_0)^T$, wodurch der Unsicherheit Rechnung getragen wird, die dadurch entsteht, dass für die Berechnung der Stichprobenkovarianzmatrix der Schätzer $\bar{\mathbf{r}}$ anstelle von $\boldsymbol{\mu}$ verwendet wird, da $\boldsymbol{\mu}$ unbekannt ist; siehe Lai, Xing und Chen (2011). In beiden Fällen wird also der ML-Schätzer in Richtung des A-priori-Erwartungswertes gelenkt bzw. „geschrumpft“, weshalb diese Bayes-Schätzer auch *Shrinkage*-Schätzer sind.

3.4.2. Empirische Bayes-Schätzer

Anstatt die Hyperparameter der A-priori-Verteilung von $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ selbst zu spezifizieren, können sie auch mittels Methoden wie der Maximum-Likelihood- oder der Momentenmethode mithilfe der beobachteten Werte approximiert werden. Schätzt man $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ in einem Bayes'schen Modell, in dem die A-priori-Verteilung empirisch geschätzt wurde, nennt man $\hat{\boldsymbol{\mu}}_{Bayes}$ und $\hat{\boldsymbol{\Sigma}}_{Bayes}$ auch *empirische Bayes-Schätzer* (siehe Robert (2007)).

3.4.3. Shrinkage-Schätzer

Das Bayes'sche Schätzverfahren aus Abschnitt 3.4.1 ist mit einigen Nachteilen verbunden. Zu der Notwendigkeit zusätzliche Hyperparameter $\boldsymbol{\mu}_0, \kappa, \nu_0$ und $\boldsymbol{\Sigma}_0$ zu spezifizieren kommt hinzu, dass die empirische Kovarianzmatrix \mathbf{S} wenig Struktur aufweist, sie je nach Anzahl vorhandener Assets m und historischer Daten n pro Asset sehr große Dimensionen annehmen kann und insbesondere für $m > n$ stets singular ist. Zwar ist die Matrix \mathbf{S} unter der Normalverteilungsannahme der ML-Schätzer für $\boldsymbol{\Sigma}$, doch für kleine Stichprobengrößen n stellt sie keinen guten Schätzer dar. Um diese Probleme zu umgehen, schlagen Ledoit und Wolf (2003) vor, dem Schätzer der Kovarianzmatrix eine Faktor-Struktur aufzuerlegen, um damit eine Dimensionsreduktion und somit bessere Schätzergebnisse zu erzielen. Für die Struktur verwenden sie das *Single-Index-Modell* von Sharpe (1963), das kurz nach seiner Entwicklung durch das CAPM erweitert wurde. Im Single-Index-Modell wird den Assetrenditen unterstellt, dass sie sich durch einen einzigen gemeinsamen Faktor r_{0t} erklären lassen und somit in der Form

$$r_{it} = \alpha_i + \beta_i r_{0t} + \epsilon_{it} \quad \text{für } i = 1, \dots, m, \quad t = 1, \dots, n \quad (3.11)$$

darstellbar sind. Dabei soll nach Sharpe (1963) für r_{0t} derjenige Faktor gewählt werden, bei dem davon ausgegangen wird, dass er allein den größten Einfluss auf die Renditen besitzt. Ledoit und Wolf (2003) definieren r_{0t} als Rendite des Marktportfolios. In der Praxis wird zur Approximation für r_{0t} häufig ein Marktindex verwendet. In (3.11) sind α_i und β_i Regressionsparameter. Für die zufälligen Fehler ϵ_{it} gilt, dass sie untereinander und mit r_{0t} unkorreliert sind, Erwartungswert 0 und Varianz δ_{ii} haben. Für die Kovarianz der Renditen zweier Assets gilt

$$\begin{aligned} Cov(r_{it}, r_{jt}) &= E[(r_{it} - Er_{it})(r_{jt} - Er_{jt})] \\ &= E[(\alpha_i + \beta_i r_{0t} + \epsilon_{it} - (\alpha_i + \beta_i Er_{0t}))(\alpha_j + \beta_j r_{0t} + \epsilon_{jt} - (\alpha_j + \beta_j Er_{0t}))] \\ &= E[(\beta_i (r_{0t} - Er_{0t}) + \epsilon_{it})(\beta_j (r_{0t} - Er_{0t}) + \epsilon_{jt})] \\ &= E[\beta_i \beta_j (r_{0t} - Er_{0t})^2] + E[\epsilon_{it} \epsilon_{jt}] \\ &= \beta_i \beta_j Var(r_{0t}) + \begin{cases} \delta_{ii}, & \text{falls } i = j \\ 0, & \text{sonst.} \end{cases} \end{aligned}$$

Somit ist die Kovarianzmatrix der Assetrenditen von der Form

$$\boldsymbol{\Phi} = \sigma_{00}^2 \boldsymbol{\beta} \boldsymbol{\beta}^T + \boldsymbol{\Delta},$$

wobei σ_{00}^2 die Varianz von r_{0t} , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^T$ der Vektor der Regressionsparameter und $\boldsymbol{\Delta}$ die Diagonalmatrix mit den Varianzen δ_{ii} der Residuen sei. Schätzt man nun $\boldsymbol{\beta}$ mittels einfacher Regression durch \mathbf{b} und $\boldsymbol{\Delta}$ durch \mathbf{D} , so erhält man

$$\mathbf{F} = s_{00}^2 \mathbf{b} \mathbf{b}^T + \mathbf{D}$$

als Schätzer für $\boldsymbol{\Phi}$, wobei s_{00}^2 die empirische Varianz von r_{0t} sei. Diese Single-Index-Kovarianzmatrix stammt nun von einem Einfaktor-Modell. Fasst man dagegen jedes Asset als eigenen Faktor auf und geht man davon aus, dass es keine Residuen gibt, so stammt die empirische Kovarianzmatrix \mathbf{S} von einem m -Faktoren-Modell. Ledoit und Wolf (2003) nehmen an, dass das optimale Modell zwischen diesen beiden Extrema liegt, also ein k -Faktorenmodell mit $1 < k < m$ ist. Aufgrund der unpräzisen Annahme, dass die Renditen von nur einem Faktor abhängen, ist der Schätzer \mathbf{F} der Single-Index-Kovarianzmatrix zum einen stark verzerrt, führt zum anderen aber nur zu geringen Schätzfehlern. Die empirische Kovarianzmatrix \mathbf{S} hingegen ist (asymptotisch) unverzerrt, bringt aber dafür große Schätzfehler mit sich. Für den optimalen Trade-off zwischen Bias und Schätzfehler muss man \mathbf{S} geeignet gegen \mathbf{F} „schrumpfen“. Dazu bildet man eine konvexe Kombination von \mathbf{F} und \mathbf{S} und erhält damit den Shrinkage-Schätzer

$$\widehat{\boldsymbol{\Sigma}}_{Shrink} = \widehat{\alpha} \mathbf{F} + (1 - \widehat{\alpha}) \mathbf{S}$$

für $\boldsymbol{\Sigma}$, wobei $\widehat{\alpha}$ ein Schätzer für den optimalen Shrinkage-Faktor ist (siehe Ledoit und Wolf (2003)). In ihrer Arbeit diskutieren Ledoit und Wolf (2003) weitere Modelle, aus denen man strukturierte Kovarianzmatrizen erhalten kann. Dabei stellen sie fest, dass eine Kovarianzmatrix, die aus dem von Elton und Gruber (1973) vorgeschlagenen Modell hergeleitet wird, vergleichbar gute Resultate liefert. Bei diesem Modell wird angenommen, dass alle Assetpaare denselben Korrelationskoeffizienten aufweisen. Ledoit und Wolf (2003) empfehlen diesen Shrinkage-Schätzer anstelle von \mathbf{S} für die Approximation der Effizienzkurve.

4. Verteilung der empirischen Minimum-Varianz-Kurve

Trotz der Gefahr enormer Schätzfehler werden in der Praxis – mit dem Ziel annähernd effiziente Portfolios zu bestimmen – häufig die empirischen Schätzer für die erwartete Rendite $\boldsymbol{\mu}$ und das Risiko $\boldsymbol{\Sigma}$ verwendet. Aus diesem Grund wird in diesem Kapitel die Verteilung der empirischen Minimum-Varianz-Kurve untersucht. Die Beweise für dieses Kapitel sind in Abschnitt 4.5 zusammengestellt.

4.1. Analyse der empirischen Effizienzmengenkonstanten

Durch Lösen von Problem (2.5)

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \} \quad \text{unter} \quad \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \quad \mathbf{w}^T \mathbf{1} = 1$$

erhält man die Gewichte des Minimum-Varianz-Portfolios, das eine bestimmte erwartete Zielrendite μ_* liefert. In Abschnitt 2.3 wird gezeigt, dass sich ein Minimum-Varianz-Portfolio mithilfe der Effizienzmengenkonstanten (A, B, C) mit

$$A = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \quad B = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \quad C = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$$

charakterisieren lässt. Mit der Matrix

$$\mathbf{M} = (\boldsymbol{\mu}, \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}, \mathbf{1}) = \begin{pmatrix} B & A \\ A & C \end{pmatrix} \quad (4.1)$$

lässt sich dieses Portfolio folgendermaßen darstellen:

$$\mathbf{w}^* = \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}, \mathbf{1}) \mathbf{M}^{-1} (\mu_*, 1)^T. \quad (4.2)$$

Kan und Smith (2008) definieren allerdings, wie u.a. auch schon Jobson (1991) bei seiner Analyse der Effizienzkurve, drei andere Konstanten, welche die Lösung des Minimum-Varianz-Problems (2.5) ebenfalls eindeutig beschreiben, für die Analyse gegenüber (A, B, C) jedoch besser geeignet sind. Betrachtet man die Herleitung der Hyperbelgleichung der Minimum-Varianz-Kurve in (2.26), lassen sich

$$\psi^2 = B - \frac{A^2}{C}, \quad \mu_g = \frac{A}{C}, \quad \sigma_g^2 = \frac{1}{C}$$

als solche Konstanten identifizieren. Wie in Abschnitt 2.2.3 bereits gezeigt, handelt es sich bei μ_g um die erwartete Rendite des globalen varianzminimalen Portfolios. σ_g^2 ist das zugehörige Risiko, denn mit den Portfoliogewichten des globalen varianzminimalen Portfolios, $\mathbf{w}_g^* = \boldsymbol{\Sigma}^{-1}\mathbf{1}/C$, gilt

$$\sigma_g^2 = \mathbf{w}_g^{*T} \boldsymbol{\Sigma} \mathbf{w}_g^* = \frac{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{1}}{C} = \frac{1}{C^2} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \frac{1}{C}.$$

Außerdem ist ψ^2 das Quadrat der Steigung der Asymptoten der Minimum-Varianz-Kurve. Die ursprünglichen Effizienzmengenkosten (A, B, C) lassen sich aus $(\psi^2, \mu_g, \sigma_g^2)$ durch

$$A = \frac{\mu_g}{\sigma_g^2}, \quad B = \psi^2 + \frac{\mu_g^2}{\sigma_g^2}, \quad C = \frac{1}{\sigma_g^2}$$

wieder herleiten. Dadurch können die Ergebnisse, die man mit $(\psi^2, \mu_g, \sigma_g^2)$ erhält, auf (A, B, C) übertragen werden.

Da die Effizienzmengenkosten von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ abhängen, müssen sie in der Praxis ebenfalls geschätzt werden. Sind $\hat{\boldsymbol{\mu}}$ und $\hat{\boldsymbol{\Sigma}}$ die empirischen Schätzer für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$, so sind die empirischen Effizienzmengenkosten

$$\hat{A} = \hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}, \quad \hat{B} = \hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}, \quad \hat{C} = \mathbf{1}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}$$

bzw.

$$\hat{\psi}^2 = \hat{B} - \frac{\hat{A}^2}{\hat{C}}, \quad \hat{\mu}_g = \frac{\hat{A}}{\hat{C}}, \quad \hat{\sigma}_g^2 = \frac{1}{\hat{C}}.$$

Diese sind aufgrund der Zufälligkeit von $\hat{\boldsymbol{\mu}}$ und $\hat{\boldsymbol{\Sigma}}$ ebenfalls Zufallsvariablen und besitzen somit eine Wahrscheinlichkeitsverteilung. Die gemeinsame Verteilung von $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$ liefert der folgende Satz von Kan und Smith (2008):

Satz 4.1.1 *Bezeichne $F_{m_{df}, n_{df}}(\delta)$ die nichtzentrale F -Verteilung mit m_{df} und n_{df} Freiheitsgraden und Nichtzentralitätsparameter δ . Sei $s \sim (m-1)/(n-m+1)F_{m-1, n-m+1}(n\psi^2)$, $q \sim \chi_{n-m}^2$ und $x \sim \mathcal{N}(0, 1)$ und seien s, q und x unabhängig voneinander. Dann gelten für $\hat{\psi}^2, \hat{\mu}_g$ und $\hat{\sigma}_g^2$ folgende Verteilungsgleichheiten:*

$$\hat{\psi}^2 = s, \quad \hat{\mu}_g = \mu_g + \left(\frac{1+s}{n}\right)^{\frac{1}{2}} \sigma_g x, \quad \hat{\sigma}_g^2 = \frac{\sigma_g^2 q}{n}.$$

Falls $\hat{\psi}^2$ gegeben ist, gilt weiterhin $\hat{\mu}_g \sim \mathcal{N}(\mu_g, \sigma_g^2(1 + \hat{\psi}^2)/n)$. Die gemeinsame Dichte von $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$ ist

$$f(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2) = f(\hat{\psi}^2, \hat{\mu}_g) f(\hat{\sigma}_g^2) = f(\hat{\psi}^2) f(\hat{\mu}_g | \hat{\psi}^2) f(\hat{\sigma}_g^2).$$

Die drei empirischen Effizienzmengenkosten sind gegeben durch $\hat{A} = \hat{\mu}_g / \hat{\sigma}_g^2$, $\hat{B} = \hat{\psi}^2 + \hat{\mu}_g^2 / \hat{\sigma}_g^2$ und $\hat{C} = 1 / \hat{\sigma}_g^2$.

Aus Satz 4.1.1 folgt, dass die gemeinsame Verteilung von $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$ allein durch die wahren Effizienzmengenkosten $(\psi^2, \mu_g, \sigma_g^2)$ und m und n bestimmt wird. Insbesondere werden die wahren Parameter $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ nicht benötigt, um die empirischen Effizienzmengenkosten zu berechnen. Möchte man also $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$ simulieren, so genügt es, die wahren Konstanten zu spezifizieren und Realisationen der drei unabhängigen Zufallsvariablen s, q und x zu generieren. Diese Methode ist vor allem für große n und m einfacher und schneller als die Bestimmung von $(\hat{A}, \hat{B}, \hat{C})$ mithilfe der empirischen Parameter $\hat{\boldsymbol{\mu}}$ und $\hat{\boldsymbol{\Sigma}}$. Für die zweite Methode benötigt man entweder historische Renditen, die man gegebenenfalls selbst generieren muss, oder man simuliert Realisationen von $\hat{\boldsymbol{\mu}}$ und $\hat{\boldsymbol{\Sigma}}$ mittels (3.2) und (3.3). Die Berechnung der empirischen Effizienzmengenkosten hängt somit von n und m ab und kann sogar die Kenntnis der wahren Parameter $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ verlangen. Somit liefert Satz 4.1.1 eine effiziente Methode zur Simulation der empirischen Effizienzmengenkosten. Außerdem ergibt sich aus dem Satz, dass $(\hat{\psi}^2, \hat{\mu}_g)$ und $\hat{\sigma}_g^2$ unabhängig sind.

Neben der gemeinsamen Verteilung der Effizienzmengenkosten sind für ihre Untersuchung auch ihre Erwartungswerte, Varianzen und Kovarianzen von Bedeutung. In den folgenden beiden Lemmata von Kan und Smith (2008) werden diese Parameter sowie unverzerrte Schätzer der wahren Konstanten vorgestellt.

Lemma 4.1.2 *Die Erwartungswerte von \hat{A}, \hat{B} und \hat{C} sind*

$$E[\hat{A}] = \frac{nA}{n-m-2}, \quad E[\hat{B}] = \frac{m+nB}{n-m-2}, \quad E[\hat{C}] = \frac{nC}{n-m-2}.$$

Die unverzerrten Schätzer für A, B und C sind

$$\hat{A}_u = \frac{(n-m-2)\hat{A}}{n}, \quad \hat{B}_u = \frac{(n-m-2)\hat{B} - m}{n}, \quad \hat{C}_u = \frac{(n-m-2)\hat{C}}{n}.$$

Die Kovarianzmatrix von \hat{A}, \hat{B} und \hat{C} ist gegeben durch

$$\begin{aligned} \text{Var}(\hat{A}) &= \frac{n^2 \left(BC + \binom{n-2}{n} C + \left(\frac{n-m}{n-m-2} \right) A^2 \right)}{(n-m-1)(n-m-2)(n-m-4)} \\ \text{Var}(\hat{B}) &= \frac{2n^2 B^2 + 2(n-2)(m+2nB)}{(n-m-2)^2(n-m-4)} \\ \text{Var}(\hat{C}) &= \frac{2n^2 C^2}{(n-m-2)^2(n-m-4)} \\ \text{Cov}(\hat{A}, \hat{B}) &= \frac{2n^2 AB + 2(n-2)nA}{(n-m-2)^2(n-m-4)} \\ \text{Cov}(\hat{A}, \hat{C}) &= \frac{2n^2 AC}{(n-m-2)^2(n-m-4)} \\ \text{Cov}(\hat{B}, \hat{C}) &= \frac{2n^2 \left(BC + \binom{n-2}{n} C + (n-m-2)A^2 \right)}{(n-m-1)(n-m-2)^2(n-m-4)}. \end{aligned}$$

Lemma 4.1.3 Die Erwartungswerte von $\widehat{\psi}^2$, $\widehat{\mu}_g$ und $\widehat{\sigma}_g^2$ sind

$$E[\widehat{\psi}^2] = \frac{m-1+n\psi^2}{n-m-1}, \quad E[\widehat{\mu}_g] = \mu_g, \quad E[\widehat{\sigma}_g^2] = \frac{(n-m)\sigma_g^2}{n}.$$

Die unverzerrten Schätzer für ψ^2 , μ_g und σ_g^2 sind

$$\widehat{\psi}_u^2 = \frac{(n-m-1)\widehat{\psi}^2 - (m-1)}{n}, \quad \widehat{\mu}_{gu} = \widehat{\mu}_g, \quad \widehat{\sigma}_{gu}^2 = \frac{n\widehat{\sigma}_g^2}{n-m}.$$

Die Kovarianzmatrix von $\widehat{\psi}^2$, $\widehat{\mu}_g$ und $\widehat{\sigma}_g^2$ ist gegeben durch

$$\begin{aligned} \text{Var}(\widehat{\psi}^2) &= \frac{2n^2\psi^4 + 2(n-2)(m-1 + 2n\psi^2)}{(n-m-1)^2(n-m-3)} \\ \text{Var}(\widehat{\mu}_g) &= \frac{(n(1+\psi^2) - 2)\sigma_g^2}{n(n-m-1)} \\ \text{Var}(\widehat{\sigma}_g^2) &= \frac{2(n-m)\sigma_g^4}{n^2} \\ \text{Cov}(\widehat{\psi}^2, \widehat{\mu}_g) &= \text{Cov}(\widehat{\psi}^2, \widehat{\sigma}_g^2) = \text{Cov}(\widehat{\mu}_g, \widehat{\sigma}_g^2) = 0. \end{aligned}$$

4.2. Die vier Minimum-Varianz-Kurven

Für die Untersuchung der Auswirkung der empirischen Schätzung von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ auf die Annäherung der Minimum-Varianz-Kurve sind vier verschiedene Kurven von Bedeutung, die in diesem Abschnitt vorgestellt und im Folgenden näher untersucht werden.

Aus Satz 2.2.4 ist bekannt, dass das Risiko des Minimum-Varianz-Portfolios, das eine bestimmte erwartete Zielrendite μ_* liefert, gegeben ist durch

$$\sigma^2(\mu_*) = \frac{C\mu_*^2 - 2A\mu_* + B}{BC - A^2} = (\mu_*, 1) \mathbf{M}^{-1} (\mu_*, 1)^T,$$

wobei \mathbf{M} die Matrix mit den Effizienzmengenkennkonstanten ist, die in (4.1) definiert wird. Mit den alternativen Effizienzmengenkennkonstanten $(\psi^2, \mu_g, \sigma_g^2)$ lässt sich $\sigma^2(\mu_*)$ auch schreiben als

$$\sigma^2(\mu_*) = \frac{\frac{1}{\sigma_g^2}\mu_*^2 - 2\frac{\mu_g}{\sigma_g^2}\mu_* + \psi^2 + \frac{\mu_g^2}{\sigma_g^2}}{\left(\psi^2 + \frac{\mu_g^2}{\sigma_g^2}\right)\frac{1}{\sigma_g^2} - \frac{\mu_g^2}{\sigma_g^4}} = \frac{\mu_*^2 - 2\mu_g\mu_* + \sigma_g^2\psi^2 + \mu_g^2}{\psi^2} = \sigma_g^2 + \frac{(\mu_* - \mu_g)^2}{\psi^2}.$$

Im Folgenden wird

$$\sigma_p^2 := \sigma_g^2 + \frac{(\mu_* - \mu_g)^2}{\psi^2} \tag{4.3}$$

als *wahre Minimum-Varianz-Kurve* bezeichnet, da sie mit den wahren Parametern $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ bzw. mit den sich daraus ergebenden Effizienzmengenkennkonstanten bestimmt wird. Das p

im Index von σ_p^2 bezieht sich auf die englische Bezeichnung *population minimum-variance frontier*, die Kan und Smith (2008) für diese Kurve verwenden.

Da die wahren Parameter $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ in der Praxis unbekannt sind, werden sie typischerweise durch ihre empirischen Schätzer ersetzt. Löst man das Minimum-Varianz-Problem mit diesen Schätzern, erhält man analog zu (4.2) die Lösung

$$\widehat{\mathbf{w}}^* = \widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\mu}}, \mathbf{1}) \widehat{\mathbf{M}}^{-1}(\mu_*, 1)^T \quad (4.4)$$

mit

$$\widehat{\mathbf{M}} = \begin{pmatrix} \widehat{B} & \widehat{A} \\ \widehat{A} & \widehat{C} \end{pmatrix}.$$

Da man keinerlei Kenntnis über das wahre Risiko $\boldsymbol{\Sigma}$ besitzt, lässt sich das Risiko eines empirischen Minimum-Varianz-Portfolios in Abhängigkeit von μ_* annähern durch

$$\widehat{\sigma}_p^2 = \widehat{\mathbf{w}}^{*T} \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{w}}^* = \frac{\widehat{C}\mu_*^2 - 2\widehat{A}\mu_* + \widehat{B}}{\widehat{D}} = (\mu_*, 1) \widehat{\mathbf{M}}^{-1}(\mu_*, 1)^T = \widehat{\sigma}_g^2 + \frac{(\mu_* - \widehat{\mu}_g)^2}{\widehat{\psi}^2}. \quad (4.5)$$

Diese Varianz wird auch als *in-sample* Varianz bezeichnet, da sie allein auf Grundlage der Stichprobe, also der historischen Renditen, bestimmt wird. Im Folgenden wird (4.5) als *empirische Minimum-Varianz-Kurve* bezeichnet.

Mit dem Erwartungswert von $\widehat{\sigma}_p^2$, $\overline{\sigma}_p^2 = E[\widehat{\sigma}_p^2]$, lässt sich die durchschnittliche Lage der empirischen Minimum-Varianz-Kurve zur wahren Kurve bestimmen. Diese Kurve wird als *mittlere empirische Minimum-Varianz-Kurve* oder *in-sample Kurve* bezeichnet.

Die vierte Kurve ist die *out-of-sample Kurve*. Mit ihr lässt sich untersuchen, wie ein empirisches Minimum-Varianz-Portfolio unter den tatsächlichen Gegebenheiten im Mittel abschneidet. Die *out-of-sample* Parameter sind die Zufallsvariablen

$$\widetilde{\mu}_p = \widehat{\mathbf{w}}^{*T} \boldsymbol{\mu}, \quad \widetilde{\sigma}_p^2 = \widehat{\mathbf{w}}^{*T} \boldsymbol{\Sigma} \widehat{\mathbf{w}}^*,$$

also die erwartete Rendite und das Risiko, die ein empirisch geschätztes Minimum-Varianz-Portfolio mit den wahren Parametern liefert. Mit ihren Erwartungswerten

$$\mu_p = E[\widetilde{\mu}_p], \quad \sigma_p^2 = E[\widetilde{\sigma}_p^2] \quad (4.6)$$

erhält man schließlich die *out-of-sample Kurve*. Für Investoren ist diese Kurve besonders wichtig, da sie eine realistische Prognose der *out-of-sample* Performance eines empirischen Minimum-Varianz-Portfolios bietet.

Der folgende Satz fasst die vier Kurven, die in den folgenden Abschnitten analysiert werden sollen, noch einmal zusammen.

Satz 4.2.1

- Die Beziehung von μ_* zu σ_p^2 wird als wahre Minimum-Varianz-Kurve bezeichnet.
- Die Beziehung von μ_* zu $\widehat{\sigma}_p^2$ wird empirische Minimum-Varianz-Kurve genannt.

- Die Beziehung von μ_* zu $\hat{\sigma}_p^2 = E[\hat{\sigma}_p^2]$ ist die mittlere empirische Minimum-Varianz-Kurve oder in-sample Kurve.
- Die Beziehung von $\underline{\mu}_p = E[\tilde{\mu}_p]$ zu $\underline{\sigma}_p^2 = E[\tilde{\sigma}_p^2]$, wobei $\tilde{\mu}_p$ und $\tilde{\sigma}_p^2$ die out-of-sample erwartete Rendite und das out-of-sample Risiko des empirischen Minimum-Varianz-Portfolios sind, heißt out-of-sample Kurve.

4.3. In-sample Performance von empirischen Minimum-Varianz-Portfolios

4.3.1. Verteilung der empirischen Minimum-Varianz-Kurve

Mit

$$\delta = \frac{\mu_* - \mu_g}{\sigma_g} \quad (4.7)$$

lässt sich das Risiko σ_p^2 anstatt durch $(\psi^2, \mu_g, \sigma_g^2)$ auch durch $(\psi^2, \sigma_g^2, \delta)$ beschreiben, denn durch Umformen von (4.3) erhält man

$$\sigma_p^2 = \sigma_g^2 \left(1 + \frac{\delta^2}{\psi^2} \right).$$

Satz 4.3.2 zeigt, dass die Verteilung von $\hat{\sigma}_p^2$ durch $(\psi^2, \sigma_g^2, \delta)$ und m und n bestimmt wird. Dem Beweis des Satzes dient das folgende Lemma. Dessen Ergebnisse werden auch für spätere Sätze benötigt.

Lemma 4.3.1 Sei $\mathbf{X} = \Sigma^{-\frac{1}{2}}(\hat{\boldsymbol{\mu}}, \mathbf{1})$ und $\mathbf{W} = n\Sigma^{-\frac{1}{2}}\hat{\Sigma}\Sigma^{-\frac{1}{2}}$. Weiterhin sei $\mathbf{Q} \in \mathbb{R}^{m \times (m-2)}$ eine Matrix mit orthonormalen Spalten, die orthogonal zu \mathbf{X} ist. Außerdem sei

$$\tilde{\mathbf{w}} = \Sigma^{-1}(\hat{\boldsymbol{\mu}}, \mathbf{1})\tilde{\mathbf{M}}^{-1}(\mu_*, 1)^T \quad (4.8)$$

mit $\tilde{\mathbf{M}} = (\hat{\boldsymbol{\mu}}, \mathbf{1})^T \Sigma^{-1}(\hat{\boldsymbol{\mu}}, \mathbf{1})$. Dann ist die out-of-sample erwartete Rendite $\tilde{\mu}_p$ eines Minimum-Varianz-Portfolios mit erwarteter Zielrendite μ_* gegeben durch

$$\tilde{\mu}_p = \hat{\mathbf{w}}^{*T} \boldsymbol{\mu} = \check{\mu}_p + (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \Sigma^{-\frac{1}{2}} \boldsymbol{\mu}$$

mit

$$\check{\mu}_p = \tilde{\mathbf{w}}^T \boldsymbol{\mu} \quad (4.9)$$

und die out-of-sample Varianz des Portfolios ist gegeben durch

$$\begin{aligned} \tilde{\sigma}_p^2 &= \hat{\mathbf{w}}^{*T} \Sigma \hat{\mathbf{w}}^* \\ &= \check{\sigma}_p^2 + (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) (\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T \end{aligned}$$

mit

$$\check{\sigma}_p^2 = (\mu_*, 1) \tilde{\mathbf{M}}^{-1} (\mu_*, 1)^T = \tilde{\mathbf{w}}^T \Sigma \tilde{\mathbf{w}} \quad (4.10)$$

(vgl. Kan und Smith (2008)).

Die Verteilung der *in-sample* Varianz $\hat{\sigma}_p^2$ wird im folgenden Satz von Kan und Smith (2008) vorgestellt.

Satz 4.3.2 Sei $u \sim \chi_{m-1}^2(n\psi^2)$, $v \sim \chi_{n-m+1}^2$ und $\tilde{y} \sim \mathcal{N}(\sqrt{n}\delta, 1)$, wobei u, v und \tilde{y} unabhängig voneinander seien und δ in (4.7) definiert wird. Für ein empirisches Minimum-Varianz-Portfolio mit erwarteter Zielrendite μ_* ist die Verteilung der empirischen Varianz $\hat{\sigma}_p^2$ gegeben durch

$$\hat{\sigma}_p^2 = \frac{\check{\sigma}_p^2 v}{n}$$

mit

$$\check{\sigma}_p^2 = \sigma_g^2 \left(1 + \frac{\tilde{y}^2}{u} \right).$$

Nach Satz 4.3.2 hängt die Verteilung der *in-sample* Varianz $\hat{\sigma}_p^2$ von den wahren Parametern $(\psi^2, \sigma_g^2, \delta)$ und von m und n ab. Wie die wahre Minimum-Varianz-Kurve ist auch die empirische Kurve eine quadratische Funktion in μ_* . Bei $\check{\sigma}_p^2$ handelt es sich um das Risiko eines Minimum-Varianz-Portfolios unter der Annahme, dass Σ bekannt ist. Dies folgt direkt aus der Darstellung $\check{\sigma}_p^2 = \tilde{\mathbf{w}}^T \Sigma \tilde{\mathbf{w}}$ aus Lemma 4.3.1. $\check{\sigma}_p^2$ ist also eine Funktion in $\hat{\boldsymbol{\mu}}$, was sich auch daraus herleiten lässt, dass die Zufälligkeit von $\check{\sigma}_p^2$ als Funktion in \tilde{y} und u , die wiederum beide Funktionen in $\hat{\boldsymbol{\mu}}$ sind, allein von $\hat{\boldsymbol{\mu}}$ stammt. Da jedoch auch das Risiko der Assets unbekannt ist, wird diese Unsicherheit in $\hat{\sigma}_p^2$ durch die Zufallsvariable v berücksichtigt, deren Zufälligkeit nach dem Beweis von Satz 4.3.2 von $\hat{\Sigma}$ stammt. Letztendlich nimmt die Schwankung von $\hat{\sigma}_p^2$ mit wachsendem n ab, da für steigende Werte für n auch die empirischen Schätzer als Maximum-Likelihood-Schätzer der wahren Parameter immer besser werden. Für eine Simulation der empirischen Minimum-Varianz-Kurve müssen nach Satz 4.3.2 lediglich $(\psi^2, \mu_g, \sigma_g^2)$ spezifiziert werden, womit Realisationen der Zufallsvariablen u, v und \tilde{y} generiert werden können, mit denen man schließlich $\hat{\sigma}_p^2$ bestimmen kann. Damit lassen sich meist aufwändigere Berechnungen umgehen, bei denen zunächst aus m mal n Daten von vergangenen Renditen $\hat{\boldsymbol{\mu}}$ und $\hat{\Sigma}$ bestimmt werden müssen, um mit diesen Schätzern Minimum-Varianz-Portfolios zu ermitteln. Noch bedeutender ist jedoch, dass nach Satz 4.3.2 die Verteilung von $\hat{\sigma}_p^2$ zwar von den wahren Effizienzmengenkonstanten, nicht aber von den wahren Parametern $\boldsymbol{\mu}$ und Σ abhängt, sodass die Verteilung der *in-sample* Varianz jeweils für alle $(\boldsymbol{\mu}, \Sigma)$ -Kombinationen gültig ist, mit denen man die entsprechenden Werte der Effizienzmengenkonstanten erhält.

Für die Bestimmung des Erwartungswerts und der Varianz von $\hat{\sigma}_p^2$ sowie für einige spätere Ergebnisse führen Kan und Smith (2008) das Integral

$$\phi = \frac{n\psi^2}{2} \int_0^1 e^{\frac{n\psi^2(y-1)}{2}} y^{\frac{m-3}{2}} dy \quad \text{für } m \geq 2 \quad (4.11)$$

ein. Man erhält dieses Integral aus der konfluenten hypergeometrischen Funktion:

Lemma 4.3.3 Für $m \geq 2$ gilt

$$\phi = \frac{n\psi^2 e^{-\frac{n\psi^2}{2}}}{m-1} {}_1F_1\left(\frac{m-1}{2}, \frac{m+1}{2}; \frac{n\psi^2}{2}\right) = \frac{n\psi^2}{m-1} {}_1F_1\left(1, \frac{m+1}{2}; -\frac{n\psi^2}{2}\right). \quad (4.12)$$

Mit den Ergebnissen aus dem folgenden Lemma von Kan und Smith (2008) können schließlich der Erwartungswert und die Varianz von $\hat{\sigma}_p^2$ bestimmt werden. Im Lemma ist \mathbf{z} die im Beweis von Satz 4.1.1 definierte Zufallsvariable, also $\mathbf{z} = \sqrt{n}\mathbf{P}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}_z, \mathbf{I}_{m-1})$ mit $\boldsymbol{\mu}_z = \sqrt{n}\mathbf{P}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}$ und $\boldsymbol{\mu}_z^T\boldsymbol{\mu}_z = n\psi^2$. Dabei ist $\mathbf{P} \in \mathbb{R}^{m \times (m-1)}$ eine Matrix mit orthonormalen Spalten, die orthogonal zu $\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{1}$ sind. Außerdem ist $u = \mathbf{z}^T\mathbf{z} \sim \chi_{m-1}^2(n\psi^2)$, was im Beweis von Satz 4.1.1 hergeleitet wird.

Lemma 4.3.4 *Für die Zufallsvariable \mathbf{z} gilt*

$$E \left[\frac{1}{\mathbf{z}^T\mathbf{z}} \right] = \frac{1-\phi}{m-3} \quad \text{für } m > 3, \quad (4.13)$$

$$E \left[\frac{1}{(\mathbf{z}^T\mathbf{z})^2} \right] = \frac{(m-5)\phi - n\psi^2(1-\phi) + 2}{2(m-3)(m-5)} \quad \text{für } m > 5, \quad (4.14)$$

$$E \left[\frac{\mathbf{z}^T\boldsymbol{\mu}_z}{\mathbf{z}^T\mathbf{z}} \right] = \phi \quad \text{für } m > 2, \quad (4.15)$$

$$E \left[\frac{\mathbf{z}^T\boldsymbol{\mu}_z}{(\mathbf{z}^T\mathbf{z})^2} \right] = \frac{n\psi^2(1-\phi)}{2(m-3)} - \frac{\phi}{2} \quad \text{für } m > 3, \quad (4.16)$$

$$E \left[\frac{(\mathbf{z}^T\boldsymbol{\mu}_z)^2}{\mathbf{z}^T\mathbf{z}} \right] = n\psi^2 - (m-2)\phi \quad \text{für } m > 1, \quad (4.17)$$

$$E \left[\frac{(\mathbf{z}^T\boldsymbol{\mu}_z)^2}{(\mathbf{z}^T\mathbf{z})^2} \right] = \frac{(m-2)\phi}{2} - \frac{n\psi^2(m-4)(1-\phi)}{2(m-3)} \quad \text{für } m > 3. \quad (4.18)$$

Damit können nun der Erwartungswert und die Varianz von $\hat{\sigma}_p^2$ bestimmt werden. Sie werden im folgenden Satz von Kan und Smith (2008) vorgestellt. Für den Satz werden $\check{\sigma}_p^2$ aus Satz 4.3.2 sowie $\delta = (\mu_* - \mu_g)/\sigma_g$ aus (4.7) benötigt.

Satz 4.3.5 *Die erwartete in-sample Varianz eines empirischen Minimum-Varianz-Portfolios mit erwarteter Zielrendite μ_* existiert genau dann, wenn $m > 3$ ist. Ist dies erfüllt, so ist der Erwartungswert von $\hat{\sigma}_p^2$ gegeben durch*

$$\bar{\sigma}_p^2 = E[\hat{\sigma}_p^2] = \left(\frac{n-m+1}{n} \right) E[\check{\sigma}_p^2], \quad (4.19)$$

wobei $E[\check{\sigma}_p^2] = \sigma_g^2(1 + hE[u^{-1}])$ mit $h = n\delta^2 + 1$ ist. Weiterhin existiert die Varianz von $\hat{\sigma}_p^2$ genau dann, wenn $m > 5$ ist. Ist dies erfüllt, so ist die Varianz von $\hat{\sigma}_p^2$ gegeben durch

$$\text{Var}(\hat{\sigma}_p^2) = \frac{(n-m+1) \left((n-m+3)\text{Var}(\check{\sigma}_p^2) + 2E[\check{\sigma}_p^2]^2 \right)}{n^2},$$

wobei $\text{Var}(\check{\sigma}_p^2) = \sigma_g^4((h^2 + 4h - 2)E[u^{-2}] - (hE[u^{-1}])^2)$ ist. $E[u^{-1}]$ und $E[u^{-2}]$ sind in (4.13) und (4.14) gegeben.

Der Erwartungswert von $\hat{\sigma}_p^2$ gibt Aufschluss über die durchschnittliche Lage der empirischen Minimum-Varianz-Kurve, während die Varianz angibt, wie stark die empirische Minimum-Varianz-Kurve um diese mittlere Kurve streut. Nach Satz 4.3.5 erhält man $\bar{\sigma}_p^2$, wenn man das mittlere Risiko eines Minimum-Varianz-Portfolios unter der Annahme, dass Σ bekannt ist, um den Faktor $(n - m + 1)/n$ verringert. Je kleiner das Verhältnis n/m ist, umso kleiner ist der Erwartungswert von $\hat{\sigma}_p^2$ im Vergleich zu dem von $\bar{\sigma}_p^2$.

Wie in der Simulationsstudie in Kapitel 5.3 gezeigt wird, liegt die *in-sample* Kurve stets links von der wahren Minimum-Varianz-Kurve. Je weiter man sich vom globalen varianz-minimalen Portfolio entfernt, umso schlechter ist die Annäherung an die wahre Kurve bzw. umso höher ist die Schwankung der empirischen Minimum-Varianz-Kurve. Kan und Smith (2008) identifizieren den Bias des Schätzers $1/\hat{\psi}^2$ von $1/\psi^2$ als Hauptursache für die starke Schwankung der empirischen Minimum-Varianz-Kurve. Aus diesem Grund schlagen sie einen verbesserten empirischen Schätzer der Minimum-Varianz-Kurve vor. Da sie jedoch keinen unverzerrten Schätzer $1/\hat{\psi}_u^2$ für $1/\psi^2$ fanden, mit dem man den unverzerrten Schätzer

$$\hat{\sigma}_{pu}^2 = \frac{n\hat{\sigma}_g^2}{n-m} + \frac{1}{\hat{\psi}_u^2} \left((\mu_* - \hat{\mu}_g)^2 - \frac{\hat{\sigma}_g^2(1 + \hat{\psi}^2)}{n-m} \right)$$

für σ_p^2 bilden kann, haben sie den Schätzer

$$\frac{1}{\hat{\psi}_a^2} = \frac{n \int_0^z u^{\frac{n-m-1}{2}} (1-u)^{\frac{m-5}{2}} du}{2z^{\frac{n-m-1}{2}} (1-z)^{\frac{m-3}{2}}}$$

mit $z = 1/(1 + \hat{\psi}^2)$ für $1/\psi^2$ gewählt, der einen geringeren Bias als $1/\hat{\psi}^2$ aufweist. Der Erwartungswert von $1/\hat{\psi}_a^2$ ist nach Kan und Smith (2008)

$$E \left[\frac{1}{\hat{\psi}_a^2} \right] = \frac{1 - e^{-\frac{n\psi^2}{2}}}{\psi^2}.$$

Mit $1/\hat{\psi}_a^2$ lässt sich der verbesserte Schätzer

$$\hat{\sigma}_{pa}^2 = \frac{n\hat{\sigma}_g^2}{n-m} + \max \left(\frac{1}{\hat{\psi}_a^2} \left((\mu_* - \hat{\mu}_g)^2 - \frac{\hat{\sigma}_g^2(1 + \hat{\psi}^2)}{n-m} \right), -\frac{n\hat{\sigma}_g^2}{n-m} \right)$$

für σ_p^2 definieren, wobei berücksichtigt wird, dass der Schätzer nicht negativ sein kann. Falls $((\mu_* - \hat{\mu}_g)^2 - \hat{\sigma}_g^2(1 + \hat{\psi}^2))/(n-m) > -n\hat{\sigma}_g^2/(n-m)$ ist, besitzt $\hat{\sigma}_{pa}^2$ den Erwartungswert

$$\begin{aligned} E[\hat{\sigma}_{pa}^2] &= E \left[\frac{n\hat{\sigma}_g^2}{n-m} + \frac{1}{\hat{\psi}_a^2} \left((\mu_* - \hat{\mu}_g)^2 - \frac{\hat{\sigma}_g^2(1 + \hat{\psi}^2)}{n-m} \right) \right] \\ &= \sigma_g^2 + E \left[E \left[\frac{1}{\hat{\psi}_a^2} (\mu_* - \hat{\mu}_g)^2 \mid \hat{\psi}^2 \right] \right] - E \left[E \left[\frac{1}{\hat{\psi}_a^2} \frac{\hat{\sigma}_g^2(1 + \hat{\psi}^2)}{n-m} \mid \hat{\psi}^2 \right] \right] \\ &= \sigma_g^2 + E \left[\frac{1}{\hat{\psi}_a^2} \left(\mu_*^2 - 2\mu_*\mu_g + \mu_g^2 + \frac{\sigma_g^2(1 + \hat{\psi}^2)}{n} \right) \right] - \frac{\sigma_g^2}{n} E \left[\frac{1 + \hat{\psi}^2}{\hat{\psi}_a^2} \right] \end{aligned}$$

$$= \sigma_g^2 + (\mu_* - \mu_g)^2 \cdot \frac{1 - e^{-\frac{n\psi^2}{2}}}{\psi^2} = \sigma_p^2 - (\mu_* - \mu_g)^2 e^{-\frac{n\psi^2}{2}}, \quad (4.20)$$

sonst ist der Erwartungswert gleich null. Da von dem wahren Risiko σ_p^2 des Minimum-Varianz-Portfolios der Term $(\mu_* - \mu_g)^2 e^{-\frac{n\psi^2}{2}}$ abgezogen wird, ist $\hat{\sigma}_{pa}^2$ im Mittel kleiner als σ_p^2 . Je näher die erwartete Zielrendite an der erwarteten Rendite des globalen varianzminimalen Portfolios liegt, umso geringer ist die Abweichung des geschätzten vom wahren Risiko. Zudem wird mit wachsendem n die Annäherung der wahren Minimum-Varianz-Kurve durch $\hat{\sigma}_{pa}^2$ im Mittel immer besser. Dass $\hat{\sigma}_{pa}^2$ im Schnitt tatsächlich besser abschneidet als $\hat{\sigma}_p^2$, ist in der Simulationsstudie in Kapitel 5.3 zu sehen. Allerdings ist die Schwankung von $\hat{\sigma}_{pa}^2$ höher als die von $\hat{\sigma}_p^2$.

4.3.2. Verteilung der Gewichte eines empirischen Minimum-Varianz-Portfolios

Nach (4.4) lassen sich die Gewichte eines empirischen Minimum-Varianz-Portfolios darstellen als

$$\hat{\mathbf{w}}^* = \hat{\Sigma}^{-1}(\hat{\boldsymbol{\mu}}, \mathbf{1}) \hat{\mathbf{M}}^{-1}(\mu_*, 1)^T.$$

Aufgrund der Zufälligkeit der Schätzer $\hat{\boldsymbol{\mu}}$ und $\hat{\Sigma}$ handelt es sich auch bei $\hat{\mathbf{w}}^*$ um eine Zufallsvariable. Deren Verteilung wird im folgenden Satz von Kan und Smith (2008) vorgestellt.

Satz 4.3.6 Sei $\widetilde{\mathbf{M}} = (\hat{\boldsymbol{\mu}}, \mathbf{1})^T \Sigma^{-1}(\hat{\boldsymbol{\mu}}, \mathbf{1})$, $\check{\sigma}_p^2 = (\mu_*, 1) \widetilde{\mathbf{M}}^{-1}(\mu_*, 1)^T$, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}_{m-2}, \mathbf{I}_{m-2})$, $\tilde{r} \sim \chi_{n-m+2}^2$ und $\mathbf{Q} \in \mathbb{R}^{m \times (m-2)}$ eine Matrix mit orthonormalen Spalten, die orthogonal zu $\Sigma^{-\frac{1}{2}}(\hat{\boldsymbol{\mu}}, \mathbf{1})$ ist, und seien \mathbf{x} , \tilde{r} und $\hat{\boldsymbol{\mu}}$ unabhängig voneinander. Die Gewichte des empirischen Minimum-Varianz-Portfolios mit erwarteter Zielrendite μ_* können dargestellt werden als

$$\hat{\mathbf{w}}^* = \tilde{\mathbf{w}} + \Sigma^{-\frac{1}{2}} \mathbf{Q} \frac{\check{\sigma}_p \mathbf{x}}{\tilde{r}^{\frac{1}{2}}} \quad (4.21)$$

mit $\tilde{\mathbf{w}} = \Sigma^{-1}(\hat{\boldsymbol{\mu}}, \mathbf{1}) \widetilde{\mathbf{M}}^{-1}(\mu_*, 1)^T$. Wenn $m > 2$ ist, existiert der Erwartungswert von $\hat{\mathbf{w}}^*$, der dann gegeben ist durch

$$E[\hat{\mathbf{w}}^*] = \mathbf{w}_g + \phi(\mathbf{w}^* - \mathbf{w}_g),$$

wobei \mathbf{w}^* die Gewichte des wahren Minimum-Varianz-Portfolios sind, ϕ das in (4.11) definierte Integral ist und $\mathbf{w}_g = \Sigma^{-1} \mathbf{1} / (\mathbf{1}^T \Sigma^{-1} \mathbf{1})$ die Gewichte des globalen varianzminimalen Portfolios sind. Weiterhin existiert die Kovarianzmatrix von $\hat{\mathbf{w}}^*$, wenn $m > 3$ ist, die dann gegeben ist durch

$$\begin{aligned} \text{Var}(\hat{\mathbf{w}}^*) &= \left(\left(\frac{n-m-1}{n-m} \right) \sigma_g^2 h \left(\frac{1-\phi}{2(m-3)} - \frac{\phi}{2n\psi^2} \right) + \frac{E[\check{\sigma}_p^2]}{n-m} - \frac{\sigma_g^2 \phi}{(n-m)n\psi^2} \right) \Sigma_a \\ &+ \left(\left(\frac{n-m-1}{n-m} \right) \sigma_g^2 h \left(\frac{(m-1)\phi}{2n\psi^2} - \frac{1-\phi}{2} \right) - \frac{\sigma_g^2}{n-m} \left(1 - \frac{(m-1)\phi}{n\psi^2} \right) \right) \Sigma_b \\ &- \phi^2 (\mathbf{w}^* - \mathbf{w}_g)(\mathbf{w}^* - \mathbf{w}_g)^T, \end{aligned}$$

wobei $h = n\delta^2 + 1$ ist, $E[\hat{\sigma}_p^2]$ in Satz 4.3.5 definiert wird und Σ_a und Σ_b gegeben sind durch

$$\begin{aligned}\Sigma_a &= \Sigma^{-1} - \frac{\Sigma^{-1}\mathbf{1}\mathbf{1}^T\Sigma^{-1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}, \\ \Sigma_b &= \frac{\Sigma^{-1}(\boldsymbol{\mu} - \mu_g\mathbf{1})(\boldsymbol{\mu} - \mu_g\mathbf{1})^T\Sigma^{-1}}{(\boldsymbol{\mu} - \mu_g\mathbf{1})^T\Sigma^{-1}(\boldsymbol{\mu} - \mu_g\mathbf{1})}.\end{aligned}$$

Da nach Kan und Smith (2008) das Integral ϕ für $m \geq 3$ nur Werte zwischen 0 und 1 annimmt, liegen die Gewichte des empirischen Minimum-Varianz-Portfolios im Mittel zwischen den Gewichten des globalen varianzminimalen Portfolios und des wahren Minimum-Varianz-Portfolios. Je größer ϕ ist, umso näher liegen die geschätzten Portfoliogewichte im Durchschnitt an den Gewichten des wahren Minimum-Varianz-Portfolios.

4.4. *Out-of-sample* Performance von empirischen Minimum-Varianz-Portfolios

Bei der Untersuchung der *in-sample* Performance wird lediglich die Verteilung von $\hat{\sigma}_p^2$ betrachtet, da die Gewichte $\hat{\mathbf{w}}^*$ des empirischen Minimum-Varianz-Portfolios stets so bestimmt werden, dass für den *in-sample* Erwartungswert $\hat{\mathbf{w}}^{*T}\hat{\boldsymbol{\mu}} = \mu_*$ gilt und dieser somit keine Zufälligkeit aufweist. Dagegen sind bei den *out-of-sample* Parametern sowohl die Varianz $\tilde{\sigma}_p^2 = \hat{\mathbf{w}}^{*T}\Sigma\hat{\mathbf{w}}^*$ als auch der Erwartungswert $\tilde{\mu}_p = \hat{\mathbf{w}}^{*T}\boldsymbol{\mu}$ aufgrund der Zufälligkeit von $\hat{\mathbf{w}}^*$ zufällig. Daher werden in diesem Abschnitt die Verteilung und die Momente beider Variablen hergeleitet.

Satz 4.4.1 Seien $\tilde{x} \sim \mathcal{N}(0, 1)$, $\tilde{q} \sim \chi_{m-3}^2$, $\tilde{r} \sim \chi_{n-m+2}^2$, $\tilde{u} \sim \chi_{m-2}^2$, $\tilde{y} \sim \mathcal{N}(\sqrt{n}\delta, 1)$ und $\tilde{z} \sim \mathcal{N}(\sqrt{n}\psi, 1)$ und seien diese Zufallsvariablen unabhängig voneinander. Sei weiter $u = \tilde{z}^2 + \tilde{u} \sim \chi_{m-1}^2(n\psi^2)$ und $\tilde{t} = \tilde{x}/\tilde{r}^{\frac{1}{2}}$. Der *out-of-sample* Erwartungswert und die *out-of-sample* Varianz des empirischen Minimum-Varianz-Portfolios mit erwarteter Zielrendite μ_* lassen sich darstellen als

$$\tilde{\mu}_p = \mu_g + \psi \left(\sigma_g \frac{\tilde{z}}{u} \tilde{y} + \hat{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \right)$$

und

$$\tilde{\sigma}_p^2 = \hat{\sigma}_p^2 \left(1 + \frac{\tilde{x}^2 + \tilde{q}}{\tilde{r}} \right),$$

wobei $\hat{\sigma}_p^2$ in Satz 4.3.2 definiert wird (Kan und Smith (2008)).

Aus dem Beweis von Satz 4.4.1 geht hervor, dass \tilde{u}, \tilde{y} und \tilde{z} Funktionen in $\hat{\boldsymbol{\mu}}$ sind, während es sich bei \tilde{x}, \tilde{q} und \tilde{r} um Funktionen in $\hat{\Sigma}$ handelt. Die *out-of-sample* erwartete Rendite setzt sich zusammen aus der erwarteten Rendite des globalen varianzminimalen Portfolios μ_g , dem Term $\psi\sigma_g\tilde{z}/u\tilde{y}$, dessen Zufälligkeit allein von $\hat{\boldsymbol{\mu}}$ stammt, und dem Term $\psi\hat{\sigma}_p(\tilde{u}/u)^{\frac{1}{2}}\tilde{t}$, dessen Zufälligkeit sowohl von $\hat{\boldsymbol{\mu}}$ als auch von $\hat{\Sigma}$ herrührt. Falls Σ bekannt

und $\boldsymbol{\mu}$ unbekannt ist, entfallen für $\tilde{\mu}_p$ und $\tilde{\sigma}_p^2$ die Terme mit \tilde{x}, \tilde{q} und \tilde{r} , sodass für ihre Verteilungen dann

$$\begin{aligned}\tilde{\mu}_p &= \mu_g + \psi \sigma_g \frac{\tilde{z}}{u} \tilde{y}, \\ \tilde{\sigma}_p^2 &= \tilde{\sigma}_p^2\end{aligned}$$

gilt. Daran ist zu erkennen, dass sich die Unsicherheit von $\hat{\boldsymbol{\Sigma}}$ auf die Unsicherheit der *out-of-sample* Parameter überträgt. Allerdings sind diese Auswirkungen für $\tilde{\mu}_p$ im Mittel gleich 0, da \tilde{t} unabhängig von \tilde{y}, \tilde{u} und u ist und $E[\tilde{t}] = 0$ ist. Das Schätzrisiko von $\hat{\boldsymbol{\Sigma}}$ erhöht die *out-of-sample* Varianz $\tilde{\sigma}_p^2$ stets um den Faktor $1 + (\tilde{x}^2 + \tilde{q})/\tilde{r}$, was sich negativ auf die *out-of-sample* Performance des empirischen Minimum-Varianz-Portfolios auswirkt. Falls nun umgekehrt $\boldsymbol{\mu}$ bekannt und $\boldsymbol{\Sigma}$ unbekannt ist, gilt mit den Definitionen der Zufallsvariablen aus dem Beweis von Satz 4.4.1 und mit (4.27)

$$\begin{aligned}\tilde{\sigma}_p^2 &= \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} = \sigma_p^2, \\ u &= \mathbf{z}^T \mathbf{z} = n \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} = n \left(B - \frac{A^2}{C} \right) = n \psi^2, \\ \tilde{y} &= \frac{\sqrt{n} \mu_*}{\sigma_g} - \frac{\sqrt{n} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\sqrt{C}} = \frac{\sqrt{n} \mu_*}{\sigma_g} - \frac{\sqrt{n} A}{\sqrt{C}}, \\ \tilde{z} &= \frac{\mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n} \psi} = \frac{n \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}}{\sqrt{n} \psi} = \frac{n \left(B - \frac{A^2}{C} \right)}{\sqrt{n} \psi} = \frac{n \psi^2}{\sqrt{n} \psi}, \\ \tilde{u} &= u - \tilde{z}^2 = 0,\end{aligned}$$

sodass sich die Verteilungen von $\tilde{\mu}_p$ und $\tilde{\sigma}_p^2$ vereinfachen zu

$$\begin{aligned}\tilde{\mu}_p &= \mu_g + \psi \sigma_g \frac{n \psi^2 / (\sqrt{n} \psi)}{n \psi^2} \left(\frac{\sqrt{n} \mu_*}{\sigma_g} - \frac{\sqrt{n} A}{\sqrt{C}} \right) = \mu_g + \mu_* - \frac{A}{C} = \mu_*, \\ \tilde{\sigma}_p^2 &= \sigma_p^2 \left(1 + \frac{\tilde{x}^2 + \tilde{q}}{\tilde{r}} \right).\end{aligned}$$

Die *out-of-sample* erwartete Rendite entspricht der erwarteten Zielrendite μ_* , wenn $\boldsymbol{\mu}$ bekannt ist. Lediglich die *out-of-sample* Varianz ist in diesem Fall zufällig.

Mit den Erwartungswerten von $\tilde{\mu}_p$ und $\tilde{\sigma}_p^2$ lassen sich die durchschnittliche *out-of-sample* erwartete Rendite und Varianz eines empirischen Minimum-Varianz-Portfolios identifizieren. Ihre Erwartungswerte sowie die Kovarianzmatrix von $(\tilde{\mu}_p, \tilde{\sigma}_p^2, \hat{\sigma}_p^2)$ werden im folgenden Satz von Kan und Smith (2008) vorgestellt.

Satz 4.4.2 *Die mittlere out-of-sample erwartete Rendite des empirischen Minimum-Varianz-Portfolios mit erwarteter Zielrendite μ_* existiert genau dann, wenn $m > 2$ ist. Ist dies erfüllt, so ist der Erwartungswert von $\tilde{\mu}_p$ gegeben durch*

$$\underline{\mu}_p = E[\tilde{\mu}_p] = \mu_* - (1 - \phi)(\mu_* - \mu_g), \quad (4.22)$$

wobei ϕ in (4.11) definiert wird. Weiterhin existiert die erwartete *out-of-sample* Varianz des empirischen Minimum-Varianz-Portfolios genau dann, wenn $m > 3$ ist. Ist dies erfüllt, so ist der Erwartungswert von $\hat{\sigma}_p^2$ gegeben durch

$$\underline{\sigma}_p^2 = E[\hat{\sigma}_p^2] = \left(\frac{n-2}{n-m} \right) E[\check{\sigma}_p^2]. \quad (4.23)$$

Des Weiteren existieren $\text{Var}(\tilde{\mu}_p)$ und $\text{Cov}(\tilde{\mu}_p, \tilde{\sigma}_p^2)$ genau dann, wenn $m > 3$ ist, und $\text{Var}(\tilde{\sigma}_p^2)$ existiert genau dann, wenn $m > 5$ ist. Falls sie existieren, sind die Varianzen und die Kovarianz gegeben durch

$$\begin{aligned} \text{Var}(\tilde{\mu}_p) &= \frac{\psi^2 E[\check{\sigma}_p^2] - \frac{\sigma_g^2}{n} \left(E \left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{\mathbf{z}^T \mathbf{z}} \right] + h E \left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{(\mathbf{z}^T \mathbf{z})^2} \right] \right)}{n-m} + \frac{\sigma_g^2 h}{n} E \left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{(\mathbf{z}^T \mathbf{z})^2} \right] \\ &\quad - \phi^2 (\mu_* - \mu_g)^2, \\ \text{Var}(\tilde{\sigma}_p^2) &= \frac{n-2}{(n-m)(n-m-2)} \left((n-4) \text{Var}(\check{\sigma}_p^2) + \frac{2(m-2) E[\check{\sigma}_p^2]^2}{n-m} \right), \\ \text{Cov}(\tilde{\mu}_p, \tilde{\sigma}_p^2) &= \frac{n-2}{n-m} (\mu_* - \mu_g) \sigma_g^2 \left((h+2) E \left[\frac{\mathbf{z}^T \boldsymbol{\mu}_z}{(\mathbf{z}^T \mathbf{z})^2} \right] - \frac{h\phi(1-\phi)}{m-3} \right), \end{aligned}$$

wobei $E[(\mathbf{z}^T \boldsymbol{\mu}_z)^2 / (\mathbf{z}^T \mathbf{z})]$, $E[\mathbf{z}^T \boldsymbol{\mu}_z / (\mathbf{z}^T \mathbf{z})^2]$ und $E[(\mathbf{z}^T \boldsymbol{\mu}_z)^2 / (\mathbf{z}^T \mathbf{z})^2]$ in Lemma 4.3.4 und h , $E[\check{\sigma}_p^2]$ und $\text{Var}(\check{\sigma}_p^2)$ in Satz 4.3.5 bestimmt werden. Die Kovarianzen zwischen der *in-sample* Varianz $\hat{\sigma}_p^2$ und $\tilde{\mu}_p$ bzw. $\tilde{\sigma}_p^2$ sind gegeben durch

$$\begin{aligned} \text{Cov}(\tilde{\mu}_p, \hat{\sigma}_p^2) &= \left(\frac{(n-m+1)(n-m)}{n(n-2)} \right) \text{Cov}(\tilde{\mu}_p, \tilde{\sigma}_p^2), \\ \text{Cov}(\hat{\sigma}_p^2, \tilde{\sigma}_p^2) &= \left(\frac{(n-m+1)(n-2)}{n(n-m)} \right) \text{Var}(\check{\sigma}_p^2). \end{aligned}$$

Ein Investor, der ein effizientes Portfolio halten möchte, von dem er eine bestimmte Zielrendite $\mu_* > \mu_g$ erwartet, und für die Ermittlung der Portfoliogewichte die erwartete Rendite und das Risiko empirisch schätzt, wird nach Satz 4.4.2 mit seinem geschätzt effizienten Portfolio im Mittel eine erwartete Rendite erhalten, die niedriger als μ_* ist. Dies folgt daraus, dass $0 < \phi < 1$ für $m \geq 3$ ist, sodass nach (4.22) die mittlere *out-of-sample* erwartete Rendite zwischen μ_g und μ_* liegt. Dabei liegt sie umso näher an der erwarteten Zielrendite, je höher ϕ ist. Die mittlere *out-of-sample* Varianz entspricht dem mit dem Faktor $(n-2)/(n-m)$ vergrößerten Erwartungswert des Portfoliorisikos, wenn $\boldsymbol{\Sigma}$ bekannt ist. Je kleiner das Verhältnis n/m ist, umso größer wird der Faktor. Die Umformung

$$\underline{\sigma}_p^2 = \frac{n-2}{n-m} E[\check{\sigma}_p^2] = \frac{(n-2)n}{(n-m)(n-m+1)} \bar{\sigma}_p^2.$$

mittels (4.19) ermöglicht den direkten Vergleich der *in-sample* mit der *out-of-sample* Varianz. Aufgrund des Faktors $(n-2)n/((n-m)(n-m+1))$ ist die *out-of-sample* Varianz

im Schnitt höher als die *in-sample* Varianz. Je kleiner die Differenz zwischen n und m ist, umso größer wird der Faktor, um den die mittlere *out-of-sample* Varianz die mittlere *in-sample* Varianz übersteigt. Dies hat zur Folge, dass das Risiko eines empirischen Minimum-Varianz-Portfolios im Mittel größer sein wird als von der *in-sample* Kurve vorhergesagt. Der Grund dafür, dass die *out-of-sample* erwartete Rendite eines empirischen Minimum-Varianz-Portfolios kleiner als seine *in-sample* erwartete Rendite und seine *out-of-sample* Varianz größer als seine *in-sample* Varianz ist, liegt darin, dass bei der Bestimmung der Portfoliogewichte durch Lösen des Minimum-Varianz-Problems mit empirisch geschätzten Parametern Assets, deren erwartete Rendite überschätzt bzw. deren Risiko unterschätzt wird, mit einem zu hohen Gewicht in das Portfolio eingehen. Dadurch erhält das Portfolio eine hohe *in-sample* erwartete Rendite und eine niedrige *in-sample* Varianz. Da aber die überbewerteten Assets zu stark und die unterbewerteten Assets zu gering gewichtet werden, fallen die *out-of-sample* Parameter des empirischen Minimum-Varianz-Portfolios dagegen weniger optimistisch aus als ihre *in-sample* Pendanten.

Als Schätzer für $\underline{\mu}_p$ und $\underline{\sigma}_p^2$ sind die *in-sample* Parameter somit nicht gut geeignet. Der folgende Satz von Kan und Smith (2008) liefert dafür die unverzerrten Schätzer von $\underline{\mu}_p$ und $\underline{\sigma}_p^2$.

Satz 4.4.3 Die unverzerrten Schätzer von $\underline{\mu}_p$ und $\underline{\sigma}_p^2$ sind gegeben durch

$$\hat{\mu}_{pu} = \mu_* - \frac{m-3}{(n-m+1)\hat{\psi}^2} (\mu_* - \hat{\mu}_g) \quad \text{für } m > 3$$

und

$$\hat{\sigma}_{pu}^2 = \frac{(n-2)n}{(n-m)(n-m+1)} \hat{\sigma}_p^2 \quad \text{für } m > 5.$$

Da die erwartete Zielrendite μ_* als Schätzer für die mittlere *out-of-sample* erwartete Rendite zu hoch ist, wird sie nach Satz 4.4.3 gegen $\hat{\mu}_g$ geschrumpft, um einen unverzerrten Schätzer für $\underline{\mu}_p$ zu erhalten. Weiterhin wird die *in-sample* Varianz durch den Faktor $(n-2)n/((n-m)(n-m+1))$ vergrößert, da $\hat{\sigma}_p^2$ als Schätzer für $\underline{\sigma}_p^2$ im Mittel zu optimistisch ist.

4.5. Beweise

Beweis von Satz 4.1.1

Definiere die beiden Matrizen

$$\widetilde{\mathbf{M}} = (\hat{\boldsymbol{\mu}}, \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}), \quad (4.24)$$

$$\widehat{\mathbf{M}} = (\hat{\boldsymbol{\mu}}, \mathbf{1})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) = \begin{pmatrix} \widehat{B} & \widehat{A} \\ \widehat{A} & \widehat{C} \end{pmatrix}. \quad (4.25)$$

Für

$$\widehat{\mathbf{M}}^{-1} = \frac{1}{\widehat{B}\widehat{C} - \widehat{A}^2} \begin{pmatrix} \widehat{C} & -\widehat{A} \\ -\widehat{A} & \widehat{B} \end{pmatrix} = \begin{pmatrix} \frac{1}{\widehat{B} - \frac{\widehat{A}^2}{\widehat{C}}} & -\frac{\widehat{A}}{\widehat{C}} \cdot \frac{1}{\widehat{B} - \frac{\widehat{A}^2}{\widehat{C}}} \\ -\frac{\widehat{A}}{\widehat{C}} \cdot \frac{1}{\widehat{B} - \frac{\widehat{A}^2}{\widehat{C}}} & \frac{(\widehat{B}\widehat{C} - \widehat{A}^2) + \widehat{A}^2}{\widehat{C}(\widehat{B}\widehat{C} - \widehat{A}^2)} \end{pmatrix} = \begin{pmatrix} \frac{1}{\widehat{\psi}^2} & -\frac{\widehat{\mu}_g}{\widehat{\psi}^2} \\ -\frac{\widehat{\mu}_g}{\widehat{\psi}^2} & \widehat{\sigma}_g^2 + \frac{\widehat{\mu}_g^2}{\widehat{\psi}^2} \end{pmatrix}$$

gilt nach Theorem 3.2.11 von Muirhead (1982) unter der Annahme, dass $\widehat{\boldsymbol{\mu}}$ gegeben ist,

$$\widehat{\mathbf{M}}^{-1} \sim \mathcal{W}_2 \left(n - m + 1, \widetilde{\mathbf{M}}^{-1}/n \right).$$

Um die Matrix $\widetilde{\mathbf{M}}^{-1}/n$ geeignet umzuschreiben, definiere die beiden Zufallsvariablen

$$y = \frac{\sqrt{n} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \widehat{\boldsymbol{\mu}}}{\sqrt{C}} \in \mathbb{R} \quad \text{und} \quad \mathbf{z} = \sqrt{n} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \widehat{\boldsymbol{\mu}} \in \mathbb{R}^{m-1}. \quad (4.26)$$

Dabei sei $\mathbf{P} \in \mathbb{R}^{m \times (m-1)}$ eine Matrix mit orthonormalen Spalten, d.h. es gilt $\mathbf{P}^T \mathbf{P} = \mathbf{I}_{m-1}$. Die Spalten von \mathbf{P} seien orthogonal zum Vektor $\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1}$, es gilt also $\mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1} = \mathbf{0}_{m-1}$, wobei $\mathbf{0}_{m-1}$ der $(m-1)$ -dimensionale Nullvektor sei. Nach Koecher (2003) bilden die Spalten einer reellen $m \times m$ Matrix \mathbf{H} genau dann eine Orthonormalbasis des \mathbb{R}^m , wenn $\mathbf{H}^T \mathbf{H} = \mathbf{I}_m$ oder äquivalent dazu $\mathbf{H} \mathbf{H}^T = \mathbf{I}_m$ ist. Somit bilden die Spalten der Matrix $(\mathbf{P}, \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1} / \sqrt{C})$ eine Orthonormalbasis des \mathbb{R}^m , da

$$\left(\mathbf{P}, \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1}}{\sqrt{C}} \right)^T \left(\mathbf{P}, \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1}}{\sqrt{C}} \right) = \begin{pmatrix} \mathbf{P}^T \mathbf{P} & \mathbf{P}^T \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1}}{\sqrt{C}} \\ \frac{\mathbf{1}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P}}{\sqrt{C}} & \frac{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}{C} \end{pmatrix} = \mathbf{I}_m$$

wegen $\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = C$ gilt. Dies ist nach Koecher (2003) äquivalent zu

$$\left(\mathbf{P}, \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1}}{\sqrt{C}} \right) \left(\mathbf{P}, \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1}}{\sqrt{C}} \right)^T = \mathbf{P} \mathbf{P}^T + \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}}{C} = \mathbf{I}_m. \quad (4.27)$$

Die Zufälligkeit von y und \mathbf{z} in (4.26) stammt allein von $\widehat{\boldsymbol{\mu}}$, bei dem es sich nach (3.2) um eine $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$ -verteilte Zufallsvariable handelt. Somit sind y und \mathbf{z} ebenfalls normalverteilt. Für ihre Erwartungswerte gilt

$$\begin{aligned} E[y] &= \frac{\sqrt{n} \mathbf{1}^T \boldsymbol{\Sigma}^{-1}}{\sqrt{C}} E[\widehat{\boldsymbol{\mu}}] = \frac{\sqrt{n} A}{\sqrt{C}} = \sqrt{n} \frac{\mu_g}{\sigma_g^2} \cdot \sigma_g = \sqrt{n} \frac{\mu_g}{\sigma_g} =: \mu_y, \\ E[\mathbf{z}] &= \sqrt{n} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} E[\widehat{\boldsymbol{\mu}}] = \sqrt{n} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} =: \boldsymbol{\mu}_z \end{aligned}$$

und für ihre Varianzen

$$\begin{aligned} \text{Var}(y) &= \frac{n}{C} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \text{Var}(\widehat{\boldsymbol{\mu}}) \boldsymbol{\Sigma}^{-1} \mathbf{1} = \frac{1}{C} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \mathbf{1} = 1, \\ \text{Var}(\mathbf{z}) &= n \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \text{Var}(\widehat{\boldsymbol{\mu}}) \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} = \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} = \mathbf{I}_{m-1}. \end{aligned}$$

Also ist

$$y \sim \mathcal{N}(\mu_y, 1) \quad \text{und} \quad \mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}_z, \mathbf{I}_{m-1}).$$

Wegen $\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$ und $\sqrt{n}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}/\sqrt{C}\boldsymbol{\Sigma}/n\sqrt{n}\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{P} = \mathbf{0}_{m-1}^T$ sind y und \mathbf{z} nach Satz C.1.3 im Anhang unabhängig voneinander. Mit diesen beiden Zufallsvariablen und mit (4.27) gilt nun

$$n\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} = \sqrt{nC}\frac{\sqrt{n}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{\sqrt{C}} = \sqrt{nC}y$$

und

$$\begin{aligned} n\hat{\boldsymbol{\mu}}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} &= n\hat{\boldsymbol{\mu}}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mathbf{P}\mathbf{P}^T + \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}}{C}\right)\boldsymbol{\Sigma}^{-\frac{1}{2}}\hat{\boldsymbol{\mu}} \\ &= n\hat{\boldsymbol{\mu}}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{P}\mathbf{P}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\hat{\boldsymbol{\mu}} + \frac{n\hat{\boldsymbol{\mu}}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{C} \\ &= \mathbf{z}^T\mathbf{z} + y^2. \end{aligned}$$

Damit kann $n\widetilde{\mathbf{M}}$ geschrieben werden als

$$n\widetilde{\mathbf{M}} = \begin{pmatrix} n\hat{\boldsymbol{\mu}}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} & n\hat{\boldsymbol{\mu}}^T\boldsymbol{\Sigma}^{-1}\mathbf{1} \\ n\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} & n\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1} \end{pmatrix} = \begin{pmatrix} y^2 + u & \sqrt{nC}y \\ \sqrt{nC}y & nC \end{pmatrix},$$

wobei $u = \mathbf{z}^T\mathbf{z}$ ist. Wegen $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}_z, \mathbf{I}_{m-1})$ ist u nichtzentral χ^2 -verteilt. Der Nichtzentralitätsparameter $\boldsymbol{\mu}_z^T\boldsymbol{\mu}_z$ lässt sich mithilfe von (4.27) berechnen. Es gilt

$$\begin{aligned} \boldsymbol{\mu}_z^T\boldsymbol{\mu}_z &= n\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{P}\mathbf{P}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} \\ &= n\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mathbf{I}_m - \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}}{C}\right)\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} \\ &= n\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{n\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{C} \\ &= n\left(B - \frac{A^2}{C}\right) = n\psi^2. \end{aligned} \tag{4.28}$$

Also ist $u \sim \chi_{m-1}^2(n\psi^2)$. Die Inverse von $n\widetilde{\mathbf{M}}$ ist

$$\begin{aligned} \frac{\widetilde{\mathbf{M}}^{-1}}{n} &= \frac{1}{(y^2 + u)nC - nCy^2} \begin{pmatrix} nC & -\sqrt{nC}y \\ -\sqrt{nC}y & y^2 + u \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{u} & -\frac{y}{\sqrt{nCu}} \\ -\frac{y}{\sqrt{nCu}} & \frac{y^2 + u}{nC} \end{pmatrix} = \begin{pmatrix} \frac{1}{u} & -\frac{\sigma_g y}{\sqrt{nu}} \\ -\frac{\sigma_g y}{\sqrt{nu}} & \frac{\sigma_g^2}{n} \left(1 + \frac{y^2}{u}\right) \end{pmatrix}. \end{aligned} \tag{4.29}$$

Mit Theorem 3.2.10 und Korollar 3.2.6 von Muirhead (1982) gilt unter der Annahme, dass u gegeben ist

$$\frac{1}{\widehat{\psi}^2} = \widehat{\mathbf{M}}_{11}^{-1} \sim \mathcal{W}_1(n - m + 1, 1/u) \tag{4.30}$$

und

$$\hat{\sigma}_g^2 = \widehat{\mathbf{M}}_{22}^{-1} - \widehat{\mathbf{M}}_{21}^{-1} \left(\widehat{\mathbf{M}}_{11}^{-1} \right)^{-1} \widehat{\mathbf{M}}_{12}^{-1} \sim \mathcal{W}_1(n-m, \sigma_g^2/n) \quad (4.31)$$

und dass $\hat{\psi}^2$ und $\hat{\sigma}_g^2$ unabhängig voneinander sind. Aus (4.30) folgt $v = u/\hat{\psi}^2 \sim \chi_{n-m+1}^2$. Da die unbedingte Verteilung von v der auf u bedingten Verteilung entspricht, ist v unabhängig von u . Deswegen ist $\hat{\psi}^2 = u/v =: s$ nichtzentral F -verteilt, also

$$s = \frac{(m-1) \frac{u}{m-1}}{(n-m+1) \frac{v}{n-m+1}} = \frac{m-1}{n-m+1} s^*, \quad \text{mit } s^* \sim F_{m-1, n-m+1}(n\psi^2)$$

bzw.

$$\hat{\psi}^2 = s \sim \frac{m-1}{n-m+1} F_{m-1, n-m+1}(n\psi^2). \quad (4.32)$$

Aus (4.31) folgt $\hat{\sigma}_g^2 = \sigma_g^2 q/n$ mit $q \sim \chi_{n-m}^2$ und s und q sind unabhängig voneinander. Unter der Annahme, dass $\widehat{\mathbf{M}}_{11}^{-1}$ bzw. $\hat{\psi}^2$ gegeben ist, gilt nach Theorem 3.2.10 von Muirhead (1982) außerdem

$$\hat{\mu}_g = -\widehat{\mathbf{M}}_{21}^{-1} \left(\widehat{\mathbf{M}}_{11}^{-1} \right)^{-1} \sim \mathcal{N}(\sigma_g y / \sqrt{n}, \sigma_g^2 \hat{\psi}^2 / n).$$

Somit kann $\hat{\mu}_g$ gegeben $\hat{\psi}^2$ geschrieben werden als

$$\hat{\mu}_g = \frac{\sigma_g (y + \hat{\psi} y_1)}{\sqrt{n}},$$

wobei $y_1 \sim \mathcal{N}(0, 1)$ unabhängig von $\hat{\psi}^2$ ist. Mit $x := (y + \hat{\psi} y_1 - \sqrt{n} \frac{\mu_g}{\sigma_g}) / (1 + \hat{\psi}^2)^{\frac{1}{2}}$ kann man $\hat{\mu}_g$ umschreiben als

$$\hat{\mu}_g = \mu_g + \left(\frac{1 + \hat{\psi}^2}{n} \right)^{\frac{1}{2}} \sigma_g x.$$

Gegeben $\hat{\psi}^2$ ist x wegen y und y_1 ebenfalls normalverteilt mit den Parametern

$$\begin{aligned} E[x] &= \frac{\mu_y + \hat{\psi} E[y_1] - \sqrt{n} \frac{\mu_g}{\sigma_g}}{\sqrt{1 + \hat{\psi}^2}} = \frac{\sqrt{n} \frac{\mu_g}{\sigma_g} + 0 - \sqrt{n} \frac{\mu_g}{\sigma_g}}{\sqrt{1 + \hat{\psi}^2}} = 0 \\ \text{Var}(x) &= \frac{\text{Var}(y) + \hat{\psi}^2 \text{Var}(y_1)}{1 + \hat{\psi}^2} = \frac{1 + \hat{\psi}^2 \cdot 1}{1 + \hat{\psi}^2} = 1, \end{aligned}$$

also $x \sim \mathcal{N}(0, 1)$. Diese Verteilung ist unabhängig von s und q . Somit ist x unabhängig von s und q und $\hat{\sigma}_g^2$ ist unabhängig von $(\hat{\psi}^2, \hat{\mu}_g)$. Gegeben $\hat{\psi}^2$ gilt also

$$\hat{\mu}_g \sim \mathcal{N}\left(\mu_g, \sigma_g^2 (1 + \hat{\psi}^2) / n\right). \quad (4.33)$$

Für die gemeinsame Dichte von $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$ gilt schließlich

$$f(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2) = f(\hat{\psi}^2, \hat{\mu}_g) f(\hat{\sigma}_g^2) = f(\hat{\psi}^2) f(\hat{\mu}_g | \hat{\psi}^2) f(\hat{\sigma}_g^2)$$

mit den entsprechenden Dichten der Verteilungen aus (4.31), (4.32) und (4.33). \square

Beweis von Lemma 4.1.2

Für zwei konstante Vektoren $\mathbf{e}, \mathbf{f} \in \mathbb{R}^m$ gilt nach Theorem 3.2 von Haff (1979)

$$E \left[\mathbf{e}^T \widehat{\Sigma}^{-1} \mathbf{f} \right] = \frac{n (\mathbf{e}^T \Sigma^{-1} \mathbf{f})}{n - m - 2} \quad (4.34)$$

$$E \left[\left(\mathbf{e}^T \widehat{\Sigma}^{-1} \mathbf{e} \right) \left(\mathbf{e}^T \widehat{\Sigma}^{-1} \mathbf{f} \right) \right] = \frac{n^2 (\mathbf{e}^T \Sigma^{-1} \mathbf{e}) (\mathbf{e}^T \Sigma^{-1} \mathbf{f})}{(n - m - 2)(n - m - 4)} \quad (4.35)$$

$$E \left[\left(\mathbf{e}^T \widehat{\Sigma}^{-1} \mathbf{f} \right)^2 \right] = \frac{n^2 \left((\mathbf{e}^T \Sigma^{-1} \mathbf{e}) (\mathbf{f}^T \Sigma^{-1} \mathbf{f}) + (n - m - 2) (\mathbf{e}^T \Sigma^{-1} \mathbf{f})^2 \right)}{(n - m - 1)(n - m - 2)(n - m - 4)} \quad (4.36)$$

$$E \left[\left(\mathbf{e}^T \widehat{\Sigma}^{-1} \mathbf{e} \right) \left(\mathbf{f}^T \widehat{\Sigma}^{-1} \mathbf{f} \right) \right] = \frac{n^2 \left((n - m - 3) (\mathbf{e}^T \Sigma^{-1} \mathbf{e}) (\mathbf{f}^T \Sigma^{-1} \mathbf{f}) + 2 (\mathbf{e}^T \Sigma^{-1} \mathbf{f})^2 \right)}{(n - m - 1)(n - m - 2)(n - m - 4)}. \quad (4.37)$$

Durch das Ersetzen von \mathbf{e} und \mathbf{f} durch $\widehat{\boldsymbol{\mu}}$ bzw. $\mathbf{1}$ können die Erwartungswerte von \widehat{A} , \widehat{B} und \widehat{C} unter Ausnutzen der Unabhängigkeit von $\widehat{\boldsymbol{\mu}}$ und $\widehat{\Sigma}$ mithilfe von (4.34) - (4.37) bestimmt werden. Setzt man $\mathbf{e} = \widehat{\boldsymbol{\mu}}$ und $\mathbf{f} = \mathbf{1}$, erhält man mit (4.34)

$$E \left[\widehat{A} \right] = E \left[\widehat{\boldsymbol{\mu}}^T \widehat{\Sigma}^{-1} \mathbf{1} \right] = E \left[E \left[\widehat{\boldsymbol{\mu}}^T \widehat{\Sigma}^{-1} \mathbf{1} \mid \widehat{\boldsymbol{\mu}} \right] \right] = \frac{E \left[n \widehat{\boldsymbol{\mu}}^T \Sigma^{-1} \mathbf{1} \right]}{n - m - 2} = \frac{nA}{n - m - 2}.$$

Mit $\mathbf{e} = \mathbf{f} = \widehat{\boldsymbol{\mu}}$ in (4.34) folgt für den Erwartungswert von \widehat{B}

$$E \left[\widehat{B} \right] = E \left[\widehat{\boldsymbol{\mu}}^T \widehat{\Sigma}^{-1} \widehat{\boldsymbol{\mu}} \right] = E \left[E \left[\widehat{\boldsymbol{\mu}}^T \widehat{\Sigma}^{-1} \widehat{\boldsymbol{\mu}} \mid \widehat{\boldsymbol{\mu}} \right] \right] = \frac{E \left[n \widehat{\boldsymbol{\mu}}^T \Sigma^{-1} \widehat{\boldsymbol{\mu}} \right]}{n - m - 2} = \frac{m + nB}{n - m - 2},$$

wobei die letzte Gleichheit gilt, da wegen $\widehat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma/n)$ die Zufallsvariable $n \widehat{\boldsymbol{\mu}}^T \Sigma^{-1} \widehat{\boldsymbol{\mu}}$ $\chi_m^2(nB)$ -verteilt ist und somit den Erwartungswert $m + nB$ besitzt. Für den Erwartungswert von \widehat{C} gilt

$$E \left[\widehat{C} \right] = E \left[\mathbf{1}^T \widehat{\Sigma}^{-1} \mathbf{1} \right] = \frac{n \mathbf{1}^T \Sigma^{-1} \mathbf{1}}{n - m - 2} = \frac{nC}{n - m - 2}.$$

Setzt man in (4.36) $\mathbf{e} = \widehat{\boldsymbol{\mu}}$ und $\mathbf{f} = \mathbf{1}$, erhält man

$$\begin{aligned} E \left[\widehat{A}^2 \right] &= E \left[\left(\widehat{\boldsymbol{\mu}}^T \widehat{\Sigma}^{-1} \mathbf{1} \right)^2 \right] = E \left[E \left[\left(\widehat{\boldsymbol{\mu}}^T \widehat{\Sigma}^{-1} \mathbf{1} \right)^2 \mid \widehat{\boldsymbol{\mu}} \right] \right] \\ &= \frac{E \left[n^2 \left(\left(\widehat{\boldsymbol{\mu}}^T \Sigma^{-1} \widehat{\boldsymbol{\mu}} \right) \left(\mathbf{1}^T \Sigma^{-1} \mathbf{1} \right) + (n - m - 2) \left(\widehat{\boldsymbol{\mu}}^T \Sigma^{-1} \mathbf{1} \right)^2 \right) \right]}{(n - m - 1)(n - m - 2)(n - m - 4)} \\ &= \frac{nCE \left[n \widehat{\boldsymbol{\mu}}^T \Sigma^{-1} \widehat{\boldsymbol{\mu}} \right] + n^2(n - m - 2) E \left[\left(\widehat{\boldsymbol{\mu}}^T \Sigma^{-1} \mathbf{1} \right)^2 \right]}{(n - m - 1)(n - m - 2)(n - m - 4)}. \end{aligned}$$

Wegen $\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$ ist auch $\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$ normalverteilt mit Erwartungswert $\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = A$ und Varianz $\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} / n \boldsymbol{\Sigma}^{-1} \mathbf{1} = C/n$. Das zweite Moment einer normalverteilten Zufallsvariablen ist die Summe aus dem Quadrat ihres Erwartungswertes und ihrer Varianz. Also ist

$$E \left[\widehat{A}^2 \right] = \frac{nC(m+nB) + n^2(n-m-2)(A^2 + \frac{1}{n}C)}{(n-m-1)(n-m-2)(n-m-4)} = \frac{n^2(BC + \frac{n-2}{n}C + (n-m-2)A^2)}{(n-m-1)(n-m-2)(n-m-4)}.$$

Somit gilt für die Varianz von \widehat{A}

$$\begin{aligned} \text{Var}(\widehat{A}) &= E \left[\widehat{A}^2 \right] - \left(E \left[\widehat{A} \right] \right)^2 = \frac{n^2(BC + \frac{n-2}{n}C + (n-m-2)A^2)}{(n-m-1)(n-m-2)(n-m-4)} - \frac{n^2 A^2}{(n-m-2)^2} \\ &= \frac{n^2 \left(BC + \frac{n-2}{n}C + \left(\frac{(n-m-2)^2 - (n-m-1)(n-m-4)}{(n-m-2)} \right) A^2 \right)}{(n-m-1)(n-m-2)(n-m-4)} \\ &= \frac{n^2 \left(BC + \frac{n-2}{n}C + \frac{n-m}{n-m-2} A^2 \right)}{(n-m-1)(n-m-2)(n-m-4)}. \end{aligned}$$

Setzt man in (4.35) $\mathbf{e} = \mathbf{f} = \hat{\boldsymbol{\mu}}$, erhält man

$$\begin{aligned} E \left[\widehat{B}^2 \right] &= E \left[\left(\hat{\boldsymbol{\mu}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}} \right)^2 \right] = E \left[E \left[\left(\hat{\boldsymbol{\mu}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}} \right)^2 \mid \hat{\boldsymbol{\mu}} \right] \right] \\ &= \frac{E \left[\left(n \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \right)^2 \right]}{(n-m-2)(n-m-4)} = \frac{(m+nB)^2 + 2(m+2nB)}{(n-m-2)(n-m-4)}, \end{aligned}$$

wobei die letzte Gleichung gilt, da das zweite Moment einer $\chi_m^2(nB)$ -verteilten Zufallsvariablen gleich $(m+nB)^2 + 2(m+2nB)$ ist. Somit gilt für die Varianz von \widehat{B}

$$\begin{aligned} \text{Var}(\widehat{B}) &= E \left[\widehat{B}^2 \right] - \left(E \left[\widehat{B} \right] \right)^2 = \frac{(m+nB)^2 + 2(m+2nB)}{(n-m-2)(n-m-4)} - \frac{(m+nB)^2}{(n-m-2)^2} \\ &= \frac{2(m+nB)^2 + 2(n-m-2)(m+2nB)}{(n-m-2)^2(n-m-4)} \\ &= \frac{2m^2 + 4mnB + 2n^2B^2 + 2nm + 4n^2B - 2m^2 - 4mnB - 4m - 8nB}{(n-m-2)^2(n-m-4)} \\ &= \frac{2n^2B^2 + 2nm + 4n^2B - 4m - 8nB}{(n-m-2)^2(n-m-4)} = \frac{2n^2B^2 + 2(n-2)(m+2nB)}{(n-m-2)^2(n-m-4)}. \end{aligned}$$

Setzt man in (4.35) $\mathbf{e} = \mathbf{f} = \mathbf{1}$, erhält man

$$E \left[\widehat{C}^2 \right] = E \left[\left(\mathbf{1}^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1} \right)^2 \right] = \frac{n^2 (\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^2}{(n-m-2)(n-m-4)} = \frac{n^2 C^2}{(n-m-2)(n-m-4)},$$

sodass für die Varianz von \widehat{C} gilt

$$\begin{aligned} \text{Var}(\widehat{C}) &= E[\widehat{C}^2] - (E[\widehat{C}])^2 = \frac{(n-m-2)n^2C^2 - (n-m-4)n^2C^2}{(n-m-2)^2(n-m-4)} \\ &= \frac{2n^2C^2}{(n-m-2)^2(n-m-4)}. \end{aligned}$$

Für die Kovarianz zwischen \widehat{A} und \widehat{B} muss zunächst $E[\widehat{A}\widehat{B}]$ bestimmt werden. Mit $\mathbf{e} = \widehat{\boldsymbol{\mu}}$ und $\mathbf{f} = \mathbf{1}$ in (4.35) erhält man

$$\begin{aligned} E[\widehat{A}\widehat{B}] &= E\left[\left(\widehat{\boldsymbol{\mu}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}\right) \left(\widehat{\boldsymbol{\mu}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}\right)\right] \\ &= E\left[E\left[\left(\widehat{\boldsymbol{\mu}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}\right) \left(\widehat{\boldsymbol{\mu}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}\right) \mid \widehat{\boldsymbol{\mu}}\right]\right] \\ &= \frac{E\left[\left(n\widehat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \widehat{\boldsymbol{\mu}}\right) \left(n\widehat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}\right)\right]}{(n-m-2)(n-m-4)} \\ &= \frac{(m+2+nB)nA}{(n-m-2)(n-m-4)}. \end{aligned}$$

Die letzte Gleichheit folgt daraus, dass für eine Zufallsvariable $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_y, \mathbf{I}_m)$ gilt, dass $E[(\mathbf{y}^T \mathbf{y})(\mathbf{y}^T \mathbf{h})] = (m+2 + \boldsymbol{\mu}_y^T \boldsymbol{\mu}_y)(\boldsymbol{\mu}_y^T \mathbf{h})$ für einen beliebigen konstanten Vektor \mathbf{h} ist. Wählt man $\mathbf{y} = \sqrt{n}\boldsymbol{\Sigma}^{-\frac{1}{2}}\widehat{\boldsymbol{\mu}}$, so gilt mit $\mathbf{h} = \sqrt{n}\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{1}$

$$\begin{aligned} E\left[\left(n\widehat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \widehat{\boldsymbol{\mu}}\right) \left(n\widehat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}\right)\right] &= E[(\mathbf{y}^T \mathbf{y})(\mathbf{y}^T \mathbf{h})] \\ &= (m+2 + n\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})(n\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}) \\ &= (m+2 + nB)nA. \end{aligned}$$

Also ist

$$\begin{aligned} \text{Cov}(\widehat{A}, \widehat{B}) &= E[\widehat{A}\widehat{B}] - E[\widehat{A}]E[\widehat{B}] \\ &= \frac{(m+2+nB)nA}{(n-m-2)(n-m-4)} - \frac{nA}{n-m-2} \cdot \frac{m+nB}{n-m-2} \\ &= \frac{(n-m-2)(m+2+nB)nA - (n-m-4)mnA - (n-m-4)n^2AB}{(n-m-2)^2(n-m-4)} \\ &= \frac{2n^2AB + 2(n-2)nA}{(n-m-2)^2(n-m-4)}. \end{aligned}$$

Für die Kovarianz zwischen \widehat{A} und \widehat{C} bestimmt man zunächst $E[\widehat{A}\widehat{C}]$. Mit $\mathbf{e} = \mathbf{1}$ und $\mathbf{f} = \widehat{\boldsymbol{\mu}}$ in (4.35) erhält man

$$\begin{aligned} E[\widehat{A}\widehat{C}] &= E\left[\left(\mathbf{1}^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}\right) \left(\widehat{\boldsymbol{\mu}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}\right)\right] \\ &= E\left[E\left[\left(\mathbf{1}^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}\right) \left(\widehat{\boldsymbol{\mu}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}\right) \mid \widehat{\boldsymbol{\mu}}\right]\right] \end{aligned}$$

$$\begin{aligned}
 &= \frac{E \left[(n\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}) \left(n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right) \right]}{(n-m-2)(n-m-4)} \\
 &= \frac{nCE \left[n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right]}{(n-m-2)(n-m-4)} \\
 &= \frac{n^2 AC}{(n-m-2)(n-m-4)}.
 \end{aligned}$$

Somit ist

$$\begin{aligned}
 \text{Cov} \left(\hat{A}, \hat{C} \right) &= E \left[\hat{A} \hat{C} \right] - E \left[\hat{A} \right] E \left[\hat{C} \right] \\
 &= \frac{n^2 AC}{(n-m-2)(n-m-4)} - \frac{nA}{n-m-2} \cdot \frac{nC}{n-m-2} \\
 &= \frac{(n-m-2)n^2 AC - (n-m-4)n^2 AC}{(n-m-2)^2(n-m-4)} \\
 &= \frac{2n^2 AC}{(n-m-2)^2(n-m-4)}.
 \end{aligned}$$

Schließlich gilt mit $\mathbf{e} = \hat{\boldsymbol{\mu}}$ und $\mathbf{f} = \mathbf{1}$ in (4.37)

$$\begin{aligned}
 E \left[\hat{B} \hat{C} \right] &= E \left[\left(\hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}} \right) \left(\mathbf{1}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1} \right) \right] \\
 &= E \left[E \left[\left(\hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}} \right) \left(\mathbf{1}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1} \right) \mid \hat{\boldsymbol{\mu}} \right] \right] \\
 &= \frac{E \left[n^2 \left((n-m-3) \left(\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \right) \left(\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right) + 2 \left(\hat{\boldsymbol{\mu}} \boldsymbol{\Sigma}^{-1} \mathbf{1} \right)^2 \right) \right]}{(n-m-1)(n-m-2)(n-m-4)} \\
 &= \frac{n(n-m-3)CE \left[n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \right] + 2n^2 E \left[\left(\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right)^2 \right]}{(n-m-1)(n-m-2)(n-m-4)} \\
 &= \frac{n(n-m-3)C(m+nB) + 2n^2 \left(A^2 + \frac{1}{n}C \right)}{(n-m-1)(n-m-2)(n-m-4)}
 \end{aligned}$$

und somit

$$\begin{aligned}
 \text{Cov} \left(\hat{B}, \hat{C} \right) &= E \left[\hat{B} \hat{C} \right] - E \left[\hat{B} \right] E \left[\hat{C} \right] \\
 &= \frac{n(n-m-3)C(m+nB) + 2n^2 \left(A^2 + \frac{1}{n}C \right)}{(n-m-1)(n-m-2)(n-m-4)} - \frac{nC(m+nB)}{(n-m-2)^2} \\
 &= \frac{\left((n-m-2)(n-m-3) - (n-m-1)(n-m-4) \right) nC(m+nB)}{(n-m-1)(n-m-2)^2(n-m-4)} \\
 &\quad + \frac{2n^2(n-m-2) \left(A^2 + \frac{1}{n}C \right)}{(n-m-1)(n-m-2)^2(n-m-4)} \\
 &= \frac{2nC(m+nB) + 2n^2(n-m-2) \left(A^2 + \frac{1}{n}C \right)}{(n-m-1)(n-m-2)^2(n-m-4)}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{2n^2 \left(\frac{Cm}{n} + BC + (n-m-2)A^2 + \frac{n-m-2}{n}C \right)}{(n-m-1)(n-m-2)^2(n-m-4)} \\
 &= \frac{2n^2 \left(BC + \frac{n-2}{n}C + (n-m-2)A^2 \right)}{(n-m-1)(n-m-2)^2(n-m-4)}.
 \end{aligned}$$

Dass es sich bei \hat{A}_u, \hat{B}_u und \hat{C}_u um unverzerzte Schätzer von A, B und C handelt, lässt sich durch die Bestimmung ihrer Erwartungswerte leicht überprüfen. \square

Beweis von Lemma 4.1.3

Im Beweis von Satz 4.1.1 wurde gezeigt, dass $\hat{\psi}^2 = u/v$ mit $u \sim \chi_{m-1}^2(n\psi^2)$ und $v \sim \chi_{n-m+1}^2$ ist, wobei u und v unabhängig voneinander sind. Der Erwartungswert und das zweite Moment der nichtzentral χ^2 -verteilten Zufallsvariable u sind

$$E[u] = m - 1 + n\psi^2, \quad E[u^2] = (m - 1 + n\psi^2)^2 + 2(m - 1 + 2n\psi^2).$$

Für v gilt

$$E[v^{-1}] = \frac{1}{n - m - 1}, \quad E[v^{-2}] = \frac{1}{(n - m - 1)(n - m - 3)}.$$

Damit lassen sich der Erwartungswert und die Varianz von $\hat{\psi}^2$ berechnen. Es ist

$$E[\hat{\psi}^2] = E[u] E[v^{-1}] = \frac{m - 1 + n\psi^2}{n - m - 1},$$

$$E[\hat{\psi}^4] = E[u^2] E[v^{-2}] = \frac{(m - 1 + n\psi^2)^2 + 2(m - 1 + 2n\psi^2)}{(n - m - 1)(n - m - 3)}$$

und somit

$$\begin{aligned}
 \text{Var}(\hat{\psi}^2) &= E[\hat{\psi}^4] - \left(E[\hat{\psi}^2]\right)^2 \\
 &= \frac{(m - 1 + n\psi^2)^2 + 2(m - 1 + 2n\psi^2)}{(n - m - 1)(n - m - 3)} - \frac{(m - 1 + n\psi^2)^2}{(n - m - 1)^2} \\
 &= \frac{(n - m - 1 - (n - m - 3))(m - 1 + n\psi^2)^2 + 2(n - m - 1)(m - 1 + 2n\psi^2)}{(n - m - 1)^2(n - m - 3)} \\
 &= \frac{2((m - 1)^2 + 2(m - 1)n\psi^2 + n^2\psi^4) + 2(n - m - 1)(m - 1 + 2n\psi^2)}{(n - m - 1)^2(n - m - 3)} \\
 &= \frac{2n^2\psi^4 + 2(m - 1)(m - 1 + 2n\psi^2) + 2(n - m - 1)(m - 1 + 2n\psi^2)}{(n - m - 1)^2(n - m - 3)} \\
 &= \frac{2n^2\psi^4 + 2(n - 2)(m - 1 + 2n\psi^2)}{(n - m - 1)^2(n - m - 3)}.
 \end{aligned}$$

Nach Satz 4.1.1 gilt für $\hat{\mu}_g$

$$E[\hat{\mu}_g] = E\left[E[\hat{\mu}_g \mid \hat{\psi}^2]\right] = \mu_g$$

und

$$\begin{aligned}
 \text{Var}(\hat{\mu}_g) &= E \left[\text{Var}(\hat{\mu}_g \mid \hat{\psi}^2) \right] + \text{Var} \left(E[\hat{\mu}_g \mid \hat{\psi}^2] \right) \\
 &= E \left[\frac{\sigma_g^2 (1 + \hat{\psi}^2)}{n} \right] + \text{Var}(\mu_g) \\
 &= \frac{\sigma_g^2}{n} \left(1 + E[\hat{\psi}^2] \right) = \frac{\sigma_g^2}{n} \left(1 + \frac{m-1 + n\psi^2}{n-m-1} \right) \\
 &= \frac{\sigma_g^2}{n} \left(\frac{n-m-1 + m-1 + n\psi^2}{n-m-1} \right) = \frac{(n(1+\psi^2) - 2) \sigma_g^2}{n(n-m-1)}.
 \end{aligned}$$

Außerdem gilt nach Satz 4.1.1 für $\hat{\sigma}_g^2$ die Darstellung $\hat{\sigma}_g^2 = \sigma_g^2 q/n$ mit $q \sim \chi_{n-m}^2$. Mit dem Erwartungswert und der Varianz von q ,

$$E[q] = n - m, \quad \text{Var}(q) = 2(n - m),$$

lassen sich der Erwartungswert und die Varianz von $\hat{\sigma}_g^2$ bestimmen. Es gilt

$$E[\hat{\sigma}_g^2] = E \left[\frac{\sigma_g^2 q}{n} \right] = \frac{(n-m)\sigma_g^2}{n}$$

und

$$\text{Var}(\hat{\sigma}_g^2) = \text{Var} \left(\frac{\sigma_g^2 q}{n} \right) = \text{Var}(q) \frac{\sigma_g^4}{n^2} = \frac{2(n-m)\sigma_g^4}{n^2}.$$

Für die Kovarianz zwischen $\hat{\psi}^2$ und $\hat{\mu}_g$ gilt

$$\begin{aligned}
 \text{Cov}(\hat{\psi}^2, \hat{\mu}_g) &= E[\hat{\psi}^2 \hat{\mu}_g] - E[\hat{\psi}^2] E[\hat{\mu}_g] \\
 &= E[\hat{\psi}^2 E[\hat{\mu}_g \mid \hat{\psi}^2]] - E[\hat{\psi}^2] E[E[\hat{\mu}_g \mid \hat{\psi}^2]] \\
 &= E[\hat{\psi}^2 \mu_g] - E[\hat{\psi}^2] E[\mu_g] \\
 &= E[\hat{\psi}^2] \mu_g - E[\hat{\psi}^2] \mu_g = 0.
 \end{aligned}$$

Aufgrund der Unabhängigkeit von $\hat{\sigma}_g^2$ und $\hat{\psi}^2$ bzw. $\hat{\mu}_g$ sind die jeweiligen Kovarianzen ebenfalls gleich null.

Dass es sich bei $\hat{\psi}_u^2, \hat{\mu}_{gu}$ und $\hat{\sigma}_{gu}^2$ um unverzerrte Schätzer von ψ^2, μ_g und σ_g^2 handelt, lässt sich durch die Bestimmung ihrer Erwartungswerte leicht überprüfen. \square

Beweis von Lemma 4.3.1

Im Lemma werden $\mathbf{X} = \Sigma^{-\frac{1}{2}}(\hat{\boldsymbol{\mu}}, \mathbf{1})$ und $\mathbf{W} = n\Sigma^{-\frac{1}{2}}\hat{\Sigma}\Sigma^{-\frac{1}{2}}$ definiert. Es ist $\mathbf{X}^T \mathbf{X} = \widetilde{\mathbf{M}}$, wobei $\widetilde{\mathbf{M}}$ in (4.24) gegeben ist, und es gilt $\mathbf{W} \sim \mathcal{W}_m(n-1, \mathbf{I}_m)$. Aufgrund der Unabhängigkeit von $\hat{\boldsymbol{\mu}}$ und $\hat{\Sigma}$ sind auch \mathbf{X} und \mathbf{W} unabhängig. Mit ihnen können folgende

Umformungen vorgenommen werden:

$$\begin{aligned}\frac{1}{n}\widehat{\mathbf{M}} &= \frac{1}{n}(\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\mu}}, \mathbf{1}) = \frac{1}{n}(\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Sigma}^{\frac{1}{2}} \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{\Sigma}^{-\frac{1}{2}}(\widehat{\boldsymbol{\mu}}, \mathbf{1}) = \mathbf{X}^T \mathbf{W}^{-1} \mathbf{X}, \\ \frac{1}{n}(\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu} &= \frac{1}{n}(\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Sigma}^{\frac{1}{2}} \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} = \mathbf{X}^T \mathbf{W}^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}\end{aligned}$$

und

$$\begin{aligned}\frac{1}{n^2}(\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\mu}}, \mathbf{1}) &= \frac{1}{n^2}(\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Sigma}^{\frac{1}{2}} \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{\Sigma}^{\frac{1}{2}} \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{\Sigma}^{-\frac{1}{2}}(\widehat{\boldsymbol{\mu}}, \mathbf{1}) \\ &= \mathbf{X}^T \mathbf{W}^{-2} \mathbf{X}.\end{aligned}$$

Somit gilt

$$\widehat{\mathbf{M}}^{-1}(\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu} = (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}, \quad (4.38)$$

$$\widehat{\mathbf{M}}^{-1}(\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\mu}}, \mathbf{1}) \widehat{\mathbf{M}}^{-1} = (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-2} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \quad (4.39)$$

Da $\mathbf{Q} \in \mathbb{R}^{m \times (m-2)}$ eine Matrix mit orthonormalen Spalten ist, die orthogonal zu \mathbf{X} ist, d.h. es gilt $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_{m-2}$ und $\mathbf{Q}^T \mathbf{X} = \mathbf{0}_{(m-2) \times 2}$, erhält man mit $(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}}, \mathbf{Q})$ eine Orthonormalbasis des \mathbb{R}^m , denn es gilt

$$\begin{aligned}(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}}, \mathbf{Q})^T (\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}}, \mathbf{Q}) &= \begin{pmatrix} (\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}} & (\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}} \mathbf{X}^T \mathbf{Q} \\ \mathbf{Q}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}} & \mathbf{Q}^T \mathbf{Q} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_{2 \times (m-2)} \\ \mathbf{0}_{(m-2) \times 2} & \mathbf{I}_{(m-2)} \end{pmatrix} = \mathbf{I}_m.\end{aligned}$$

Dies ist nach Koecher (2003) äquivalent zu

$$(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}}, \mathbf{Q})(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}}, \mathbf{Q})^T = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{Q} \mathbf{Q}^T = \mathbf{I}_m. \quad (4.40)$$

Mit (4.40) folgt

$$\begin{aligned}(\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} &= (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} (\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{Q} \mathbf{Q}^T) \\ &= (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \\ &\quad + (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q} \mathbf{Q}^T \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T.\end{aligned} \quad (4.41)$$

Mit (4.38), (4.41) und (4.4) erhält man die *out-of-sample* erwartete Rendite

$$\begin{aligned}\widehat{\mathbf{w}}^{*T} \boldsymbol{\mu} &= (\mu_*, 1) \widehat{\mathbf{M}}^{-1}(\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu} \\ &= (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} \\ &= (\mu_*, 1) \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \right) \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} \\ &= \check{\mu}_p + (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}\end{aligned}$$

mit

$$\check{\mu}_p = (\mu_*, 1) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-\frac{1}{2}} \boldsymbol{\mu} = (\mu_*, 1) \widetilde{\mathbf{M}}^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \Sigma^{-1} \boldsymbol{\mu} = \widetilde{\mathbf{w}}^T \boldsymbol{\mu}$$

und

$$\widetilde{\mathbf{w}} = \Sigma^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1}) \widetilde{\mathbf{M}}^{-1} (\mu_*, 1)^T.$$

Mit (4.39), (4.41) und (4.4) erhält man die *out-of-sample* Varianz

$$\begin{aligned} \widehat{\mathbf{w}}^{*T} \Sigma \widehat{\mathbf{w}}^* &= (\mu_*, 1) \widehat{\mathbf{M}}^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \widehat{\Sigma}^{-1} \Sigma \widehat{\Sigma}^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1}) \widehat{\mathbf{M}}^{-1} (\mu_*, 1)^T \\ &= (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-2} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T \\ &= (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} \mathbf{W}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T \\ &= (\mu_*, 1) \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \right) \\ &\quad \times \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \right)^T (\mu_*, 1)^T \\ &= (\mu_*, 1) \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \right) \\ &\quad \times \left(\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} + \mathbf{Q} (\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \right) (\mu_*, 1)^T \\ &= (\mu_*, 1) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} (\mu_*, 1)^T \\ &\quad + (\mu_*, 1) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Q} (\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T \\ &\quad + (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} (\mu_*, 1)^T \\ &\quad + (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \\ &\quad \quad \times \mathbf{Q} (\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T \\ &= (\mu_*, 1) \widetilde{\mathbf{M}}^{-1} (\mu_*, 1)^T \\ &\quad + (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) (\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T \\ &= \check{\sigma}_p^2 + (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) (\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T \end{aligned}$$

mit

$$\begin{aligned} \check{\sigma}_p^2 &= (\mu_*, 1) \widetilde{\mathbf{M}}^{-1} (\mu_*, 1)^T \\ &= (\mu_*, 1) \widetilde{\mathbf{M}}^{-1} \widetilde{\mathbf{M}} \widetilde{\mathbf{M}}^{-1} (\mu_*, 1)^T \\ &= (\mu_*, 1) \widetilde{\mathbf{M}}^{-1} \mathbf{X}^T \mathbf{X} \widetilde{\mathbf{M}}^{-1} (\mu_*, 1)^T \\ &= (\mu_*, 1) \widetilde{\mathbf{M}}^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} (\widehat{\boldsymbol{\mu}}, \mathbf{1}) \widetilde{\mathbf{M}}^{-1} (\mu_*, 1)^T \\ &= (\mu_*, 1) \widetilde{\mathbf{M}}^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \Sigma^{-1} \Sigma \Sigma^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1}) \widetilde{\mathbf{M}}^{-1} (\mu_*, 1)^T \\ &= \widetilde{\mathbf{w}}^T \Sigma \widetilde{\mathbf{w}}. \end{aligned}$$

□

Beweis von Satz 4.3.2

Seien \mathbf{X} , \mathbf{W} und \mathbf{Q} wie in Lemma 4.3.1. Aus dem Beweis dieses Lemmas ist bekannt, dass $(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}}, \mathbf{Q})$ eine Orthonormalbasis des \mathbb{R}^m ist. Sei

$$\begin{aligned} \mathbf{V} &= \left(\left(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}}, \mathbf{Q} \right)^T \mathbf{W}^{-1} \left(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}}, \mathbf{Q} \right) \right)^{-1} \\ &= \begin{pmatrix} \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}}, \mathbf{Q} \end{pmatrix}^T \mathbf{W} \begin{pmatrix} \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}}, \mathbf{Q} \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}} \mathbf{X}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}} & (\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}} \mathbf{X}^T \mathbf{W} \mathbf{Q} \\ \mathbf{Q}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}} & \mathbf{Q}^T \mathbf{W} \mathbf{Q} \end{pmatrix}. \end{aligned} \quad (4.42)$$

Mit Theorem 3.2.11 von Muirhead (1982) lässt sich zeigen, dass bedingt auf \mathbf{X}

$$\mathbf{V} \sim \mathcal{W}_m(n-1, \mathbf{I}_m)$$

gilt. Anhand der Darstellung von \mathbf{V} in (4.42) lässt sich erkennen, dass die Matrix aus vier Blöcken besteht. Es sei \mathbf{V}_{ij} der Block in der i -ten Zeile und j -ten Spalte. Mit (4.41) und der Orthogonalität von \mathbf{Q} zu \mathbf{X} gilt

$$\begin{aligned} &\mathbf{V}_{11} - \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21} \\ &= (\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}} \mathbf{X}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}} - (\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}} \mathbf{X}^T \mathbf{W} \mathbf{Q} (\mathbf{Q}^T \mathbf{W} \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}} \\ &= (\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \left((\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1} \right. \\ &\quad \left. - (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Q} (\mathbf{Q}^T \mathbf{W} \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1} \right) (\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \\ &= (\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \left(\left((\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} - (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \right) \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1} \right. \\ &\quad \left. - \left((\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} - (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \right) \right. \\ &\quad \left. \times \mathbf{W} \mathbf{Q} (\mathbf{Q}^T \mathbf{W} \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1} \right) (\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \\ &= (\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \left((\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} - (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1} \right. \\ &\quad \left. + (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1} \right) (\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \\ &= (\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \\ &= n \widetilde{\mathbf{M}}^{\frac{1}{2}} \widehat{\mathbf{M}}^{-1} \widetilde{\mathbf{M}}^{\frac{1}{2}}. \end{aligned}$$

Der letzte Schritt folgt dabei mit dem Beweis von Lemma 4.3.1. Mit Theorem 3.2.10 von Muirhead (1982) folgt, dass $n \widetilde{\mathbf{M}}^{\frac{1}{2}} \widehat{\mathbf{M}}^{-1} \widetilde{\mathbf{M}}^{\frac{1}{2}} \sim \mathcal{W}_2(n-m+1, \mathbf{I}_2)$ ist. Nach der Definition der Wishart-Verteilung bedeutet das, dass sich $n \widetilde{\mathbf{M}}^{\frac{1}{2}} \widehat{\mathbf{M}}^{-1} \widetilde{\mathbf{M}}^{\frac{1}{2}}$ als Produkt $\mathbf{J}^T \mathbf{J}$ darstellen

lässt, wobei \mathbf{J} eine $((n - m + 1) \times 2)$ -Matrix ist, deren Zeilen jeweils $\mathcal{N}(0, \mathbf{I}_2)$ -verteilt sind. Für

$$\hat{\sigma}_p^2 = (\mu_*, 1) \widehat{\mathbf{M}}^{-1} (\mu_*, 1)^T = \frac{1}{n} (\mu_*, 1) \widetilde{\mathbf{M}}^{-\frac{1}{2}} \left(n \widetilde{\mathbf{M}}^{\frac{1}{2}} \widehat{\mathbf{M}}^{-1} \widetilde{\mathbf{M}}^{\frac{1}{2}} \right) \widetilde{\mathbf{M}}^{-\frac{1}{2}} (\mu_*, 1)^T$$

gilt dann, dass $\hat{\sigma}_p^2 = \bar{\mathbf{J}}^T \bar{\mathbf{J}}$ mit $\bar{\mathbf{J}} = \frac{1}{\sqrt{n}} \mathbf{J} \widetilde{\mathbf{M}}^{-\frac{1}{2}} (\mu_*, 1)^T$ ist. Dabei ist $\bar{\mathbf{J}}$ eine $((n - m + 1) \times 1)$ -Matrix, deren Zeilen $\mathcal{N}(0, \check{\sigma}_p^2/n)$ -verteilt sind, wobei $\check{\sigma}_p^2$ in (4.10) definiert ist. Somit ist $\hat{\sigma}_p^2 \sim \mathcal{W}_1(n - m + 1, \check{\sigma}_p^2/n)$ und $n\hat{\sigma}_p^2/\check{\sigma}_p^2 = v \sim \chi_{n-m+1}^2$, wobei v unabhängig von $\check{\sigma}_p^2$ ist, da die unbedingte Verteilung von v der auf $\check{\sigma}_p^2$ bedingten Verteilung entspricht. Für $\check{\sigma}_p^2$ gibt es mit (4.29) die alternative Darstellung

$$\begin{aligned} \check{\sigma}_p^2 &= (\mu_*, 1) \widetilde{\mathbf{M}}^{-1} (\mu_*, 1)^T \\ &= (\mu_*, 1) n \begin{pmatrix} \frac{1}{u} & -\frac{\sigma_g y}{\sqrt{nu}} \\ -\frac{\sigma_g y}{\sqrt{nu}} & \frac{\sigma_g^2}{n} \left(1 + \frac{y^2}{u}\right) \end{pmatrix} (\mu_*, 1)^T \\ &= \left(\frac{n\mu_*}{u} - \frac{\sqrt{n}\sigma_g y}{u}, \sigma_g^2 \left(1 + \frac{y^2}{u}\right) - \frac{\sqrt{n}\sigma_g y \mu_*}{u} \right) (\mu_*, 1)^T \\ &= \frac{n\mu_*^2}{u} - \frac{\sqrt{n}\sigma_g y \mu_*}{u} + \sigma_g^2 \left(1 + \frac{y^2}{u}\right) - \frac{\sqrt{n}\sigma_g y \mu_*}{u} \\ &= \sigma_g^2 \left(1 + \frac{\left(\frac{\sqrt{n}\mu_*}{\sigma_g} - y\right)^2}{u}\right) = \sigma_g^2 \left(1 + \frac{\tilde{y}^2}{u}\right) \end{aligned}$$

mit $\tilde{y} = \sqrt{n}\mu_*/\sigma_g - y \sim \mathcal{N}(\sqrt{n}\delta, 1)$. Außerdem sind \tilde{y} und u unabhängig voneinander. \square

Beweis von Lemma 4.3.3

Für die konfluente hypergeometrische Funktion ${}_1F_1(a, b; c)$ gilt die Integraldarstellung

$${}_1F_1(a, b; c) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{cy} y^{a-1} (1-y)^{b-a-1} dy$$

für $0 < \Re a < \Re b$ (siehe z.B. Tricomi (1955)). Somit ist

$${}_1F_1\left(\frac{m-1}{2}, \frac{m+1}{2}; \frac{n\psi^2}{2}\right) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)\Gamma(1)} \int_0^1 e^{\frac{n\psi^2}{2} y} y^{\frac{m-3}{2}} dy$$

für $m \geq 2$. Für die Gammafunktionen gilt

$$\Gamma\left(\frac{m+1}{2}\right) = \begin{cases} \left(\frac{m+1}{2} - 1\right)!, & \text{falls } m \text{ ungerade} \\ \Gamma\left(\frac{m}{2} + \frac{1}{2}\right) = \frac{m!}{\left(\frac{m}{2}\right)! 4^{\frac{m}{2}}} \sqrt{\pi}, & \text{falls } m \text{ gerade,} \end{cases}$$

$$\Gamma\left(\frac{m-1}{2}\right) = \begin{cases} \left(\frac{m-1}{2} - 1\right)!, & \text{falls } m \text{ ungerade} \\ \Gamma\left(\left(\frac{m}{2} - 1\right) + \frac{1}{2}\right) = \frac{(2\left(\frac{m}{2} - 1\right))!}{\left(\frac{m}{2} - 1\right)! 4^{\frac{m}{2}-1}} \sqrt{\pi}, & \text{falls } m \text{ gerade} \end{cases}$$

und

$$\Gamma(1) = 1.$$

Somit ist

$$\begin{aligned} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)\Gamma(1)} &= \begin{cases} \frac{\left(\frac{m+1}{2}-1\right)!}{\left(\frac{m-1}{2}-1\right)!}, & \text{falls } m \text{ ungerade} \\ \frac{m!}{\left(\frac{m}{2}\right)!4^{\frac{m}{2}}} \cdot \frac{\left(\frac{m}{2}-1\right)!4^{\frac{m}{2}-1}}{\left(2\left(\frac{m}{2}-1\right)\right)!}, & \text{falls } m \text{ gerade} \end{cases} \\ &= \begin{cases} \frac{m+1}{2} - 1, & \text{falls } m \text{ ungerade} \\ \frac{m(m-1)}{4^{\frac{m}{2}}}, & \text{falls } m \text{ gerade} \end{cases} = \frac{m-1}{2}. \end{aligned}$$

Damit gilt schließlich

$$\frac{n\psi^2 e^{-\frac{n\psi^2}{2}}}{m-1} {}_1F_1\left(\frac{m-1}{2}, \frac{m+1}{2}; \frac{n\psi^2}{2}\right) = \frac{n\psi^2}{2} \int_0^1 e^{\frac{n\psi^2}{2}(y-1)} y^{\frac{m-3}{2}} dy.$$

Mit der Transformationsformel von Kummer

$$e^c {}_1F_1(b-a, b; -c) = {}_1F_1(a, b; c),$$

(siehe z.B. Bock, Judge und Yancey (1984)), folgt direkt die zweite Identität in (4.12). \square

Beweis von Lemma 4.3.4

Alle Identitäten aus dem Lemma können mit Lemma 1 von Sawa (1972) berechnet werden, wobei im Folgenden lediglich (4.18) bewiesen werden soll. Für den Beweis der übrigen Identitäten liefern Kan und Smith (2008) hilfreiche Ansätze.

Sei $x_1 = \mathbf{z}^T \mathbf{z}$ und $x_2 = \mathbf{z}^T \boldsymbol{\mu}_z$. Die gemeinsame momenterzeugende Funktion von x_1 und x_2 ist

$$\begin{aligned} \varphi(\theta_1, \theta_2) &= E[\exp(\theta_1 x_1 + \theta_2 x_2)] \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{m-1}{2}}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_z)^T(\mathbf{z}-\boldsymbol{\mu}_z) + \theta_1 \mathbf{z}^T \mathbf{z} + \theta_2 \mathbf{z}^T \boldsymbol{\mu}_z} dz_1 \dots dz_{m-1} \\ &= \frac{1}{(1-2\theta_1)^{\frac{m-1}{2}}} e^{\frac{1}{2}\left(\frac{(1+\theta_2)^2}{1-2\theta_1}-1\right)\boldsymbol{\mu}_z^T \boldsymbol{\mu}_z}. \end{aligned}$$

Mit Lemma 1 von Sawa (1972) gilt

$$E\left[\frac{x_2^2}{x_1^2}\right] = \int_{-\infty}^0 -\theta_1 \frac{\partial^2 \varphi(\theta_1, \theta_2)}{\partial \theta_2^2} \Big|_{\theta_2=0} d\theta_1. \quad (4.43)$$

Die erste partielle Ableitung von $\varphi(\theta_1, \theta_2)$ nach θ_2 ist

$$\begin{aligned} \frac{\partial \varphi(\theta_1, \theta_2)}{\partial \theta_2} &= \frac{1}{(1-2\theta_1)^{\frac{m-1}{2}}} e^{\frac{1}{2}\left(\frac{(1+\theta_2)^2}{1-2\theta_1}-1\right)\boldsymbol{\mu}_z^T \boldsymbol{\mu}_z} \frac{1+\theta_2}{1-2\theta_1} \boldsymbol{\mu}_z^T \boldsymbol{\mu}_z \\ &= \frac{2\theta}{(1-2\theta_1)^{\frac{m-1}{2}+1}} e^{\left(\frac{(1+\theta_2)^2}{1-2\theta_1}-1\right)\theta} \cdot (1+\theta_2) \end{aligned}$$

mit $\theta = (\boldsymbol{\mu}_z^T \boldsymbol{\mu}_z)/2 = n\psi^2/2$ (siehe (4.28)). Leitet man $\varphi(\theta_1, \theta_2)$ ein zweites mal nach θ_2 ab, erhält man

$$\begin{aligned} \frac{\partial^2 \varphi(\theta_1, \theta_2)}{\partial \theta_2^2} &= \frac{2\theta}{(1-2\theta_1)^{\frac{m-1}{2}+1}} e^{\left(\frac{(1+\theta_2)^2}{1-2\theta_1}-1\right)\theta} \cdot \frac{2\theta(1+\theta_2)}{1-2\theta_1} \cdot (1+\theta_2) \\ &\quad + \frac{2\theta}{(1-2\theta_1)^{\frac{m-1}{2}+1}} e^{\left(\frac{(1+\theta_2)^2}{1-2\theta_1}-1\right)\theta} \\ &= \frac{2\theta}{(1-2\theta_1)^{\frac{m-1}{2}+1}} e^{\left(\frac{(1+\theta_2)^2}{1-2\theta_1}-1\right)\theta} \cdot \left(1 + \frac{2\theta(1+\theta_2)^2}{1-2\theta_1}\right). \end{aligned}$$

Somit ist

$$\left. \frac{\partial^2 \varphi(\theta_1, \theta_2)}{\partial \theta_2^2} \right|_{\theta_2=0} = \frac{2\theta}{(1-2\theta_1)^{\frac{m-1}{2}+1}} e^{\left(\frac{1}{1-2\theta_1}-1\right)\theta} \cdot \left(1 + \frac{2\theta}{1-2\theta_1}\right).$$

Damit lässt sich (4.43) schreiben als

$$E \begin{bmatrix} x_2^2 \\ x_1^2 \end{bmatrix} = \int_{-\infty}^0 -\theta_1 \frac{2\theta}{(1-2\theta_1)^{\frac{m-1}{2}+1}} e^{\left(\frac{1}{1-2\theta_1}-1\right)\theta} \cdot \left(1 + \frac{2\theta}{1-2\theta_1}\right) d\theta_1.$$

Setzt man $y = 1/(1-2\theta_1)$, erhält man

$$\begin{aligned} E \begin{bmatrix} x_2^2 \\ x_1^2 \end{bmatrix} &= \int_0^1 -\left(\frac{1}{2} - \frac{1}{2y}\right) 2\theta y^{\frac{m-1}{2}+1} \cdot e^{(y-1)\theta} \cdot (1+2\theta y) \cdot \frac{1}{2} y^{-2} dy \\ &= \frac{\theta}{2} \int_0^1 \left(-y^{\frac{m-1}{2}+1} \cdot e^{(y-1)\theta} \cdot (1+2\theta y)y^{-2} + y^{\frac{m-1}{2}} \cdot e^{(y-1)\theta} \cdot (1+2\theta y)y^{-2}\right) dy \\ &= \frac{\theta}{2} \int_0^1 \left(-y^{\frac{m-3}{2}} - 2\theta y^{\frac{m-1}{2}} + y^{\frac{m-5}{2}} + 2\theta y^{\frac{m-3}{2}}\right) e^{(y-1)\theta} dy \\ &= \frac{\theta}{2} \int_0^1 \left(y^{\frac{m-5}{2}} - (1-2\theta)y^{\frac{m-3}{2}} - 2\theta y^{\frac{m-1}{2}}\right) e^{(y-1)\theta} dy \\ &= \frac{\theta}{2} \int_0^1 y^{\frac{m-5}{2}} e^{(y-1)\theta} dy - \frac{1-2\theta}{2} \theta \int_0^1 y^{\frac{m-3}{2}} e^{(y-1)\theta} dy - \theta^2 \int_0^1 y^{\frac{m-1}{2}} e^{(y-1)\theta} dy. \end{aligned}$$

Setzt man im mittleren Term $n\psi^2/2$ für θ ein, erhält man

$$\begin{aligned} E \begin{bmatrix} x_2^2 \\ x_1^2 \end{bmatrix} &= \frac{\theta}{2} \int_0^1 y^{\frac{m-5}{2}} e^{(y-1)\theta} dy - \frac{1-2\theta}{2} \frac{n\psi^2}{2} \int_0^1 y^{\frac{m-3}{2}} e^{\frac{n\psi^2(y-1)}{2}\theta} dy - \theta^2 \int_0^1 y^{\frac{m-1}{2}} e^{(y-1)\theta} dy \\ &= \frac{\theta}{2} \int_0^1 y^{\frac{m-5}{2}} e^{(y-1)\theta} dy - \frac{1-2\theta}{2} \phi - \theta^2 \int_0^1 y^{\frac{m-1}{2}} e^{(y-1)\theta} dy. \end{aligned} \quad (4.44)$$

Für den ersten Term in (4.44) gilt mittels partieller Integration

$$\begin{aligned}
 \frac{\theta}{2} \int_0^1 y^{\frac{m-5}{2}} e^{(y-1)\theta} dy &= \frac{\theta}{2} \left(\left[\frac{2}{m-3} y^{\frac{m-3}{2}} e^{(y-1)\theta} \right]_0^1 - \int_0^1 \frac{2}{m-3} y^{\frac{m-3}{2}} \cdot \theta e^{(y-1)\theta} dy \right) \\
 &= \frac{\theta}{m-3} - \frac{\theta^2}{m-3} \int_0^1 y^{\frac{m-3}{2}} \cdot e^{(y-1)\theta} dy \\
 &= \frac{\theta}{m-3} - \frac{\theta}{m-3} \phi = \frac{\theta(1-\phi)}{m-3},
 \end{aligned}$$

wobei $m > 3$ sein muss. Analog dazu gilt für den letzten Term in (4.44)

$$\begin{aligned}
 \theta^2 \int_0^1 y^{\frac{m-1}{2}} e^{(y-1)\theta} dy &= \theta^2 \left(\left[y^{\frac{m-1}{2}} \cdot \frac{1}{\theta} \cdot e^{(y-1)\theta} \right]_0^1 - \int_0^1 \frac{m-1}{2} y^{\frac{m-3}{2}} \cdot \frac{1}{\theta} \cdot e^{(y-1)\theta} dy \right) \\
 &= \theta - \frac{(m-1)\phi}{2}.
 \end{aligned}$$

Damit folgt schließlich

$$\begin{aligned}
 E \left[\frac{x_2^2}{x_1^2} \right] &= \frac{\theta(1-\phi)}{m-3} - \frac{(1-2\theta)\phi}{2} - \theta + \frac{(m-1)\phi}{2} \\
 &= \frac{n\psi^2(1-\phi)}{2(m-3)} - \frac{(1-n\psi^2)\phi}{2} - \frac{n\psi^2}{2} + \frac{(m-1)\phi}{2} \\
 &= \frac{n\psi^2 - n\psi^2\phi}{2(m-3)} - \frac{\phi}{2} + \frac{(m-3)n\psi^2\phi}{2(m-3)} - \frac{(m-3)n\psi^2}{2(m-3)} + \frac{m\phi}{2} - \frac{\phi}{2} \\
 &= \frac{n\psi^2 - n\psi^2\phi + mn\psi^2\phi - 3n\psi^2\phi - mn\psi^2 + 3n\psi^2}{2(m-3)} + \frac{m\phi - 2\phi}{2} \\
 &= \frac{(n\psi^2 - mn\psi^2 + 3n\psi^2)(1-\phi)}{2(m-3)} + \frac{(m-2)\phi}{2} \\
 &= \frac{n\psi^2(4-m)(1-\phi)}{2(m-3)} + \frac{(m-2)\phi}{2} \\
 &= \frac{(m-2)\phi}{2} - \frac{n\psi^2(m-4)(1-\phi)}{2(m-3)}.
 \end{aligned}$$

□

Beweis von Satz 4.3.5

Im Beweis von Satz 4.3.2 wird gezeigt, dass $n\hat{\sigma}_p^2/\check{\sigma}_p^2 = v \sim \chi_{n-m+1}^2$ ist und v und $\check{\sigma}_p^2$ unabhängig sind. Damit gilt

$$\bar{\sigma}_p^2 = E[\hat{\sigma}_p^2] = E \left[\frac{v\check{\sigma}_p^2}{n} \right] = E \left[\frac{v}{n} \right] E[\check{\sigma}_p^2] = \frac{n-m+1}{n} E[\check{\sigma}_p^2].$$

Nach Satz 4.3.2 ist $\check{\sigma}_p^2 = \sigma_g^2(1 + \tilde{y}^2/u)$ mit $\tilde{y} \sim \mathcal{N}(\sqrt{n}\delta, 1)$ und $u \sim \chi_{m-1}^2(n\psi^2)$, wobei \tilde{y} und u unabhängig sind. Mit

$$E[\tilde{y}^2] = 1 + n\delta^2 =: h$$

gilt somit

$$E[\check{\sigma}_p^2] = E\left[\sigma_g^2\left(1 + \frac{\tilde{y}^2}{u}\right)\right] = \sigma_g^2(1 + E[u^{-1}]E[\tilde{y}^2]) = \sigma_g^2(1 + hE[u^{-1}]).$$

Für die Varianz von $\hat{\sigma}_p^2$ kann man ausnutzen, dass

$$E[v^2] = (n - m + 1)(n - m + 3) \quad \text{und} \quad \text{Var}(v) = 2(n - m + 1)$$

ist. Damit gilt

$$\begin{aligned} \text{Var}(\hat{\sigma}_p^2) &= E\left[\text{Var}(\hat{\sigma}_p^2 \mid v)\right] + \text{Var}\left(E[\hat{\sigma}_p^2 \mid v]\right) \\ &= E\left[\text{Var}\left(\frac{v\check{\sigma}_p^2}{n} \mid v\right)\right] + \text{Var}\left(E\left[\frac{v\check{\sigma}_p^2}{n} \mid v\right]\right) \\ &= \frac{E[v^2]}{n^2}\text{Var}(\check{\sigma}_p^2) + \text{Var}(v)\frac{E[\check{\sigma}_p^2]^2}{n^2} \\ &= \frac{(n - m + 1)(n - m + 3)\text{Var}(\check{\sigma}_p^2)}{n^2} + \frac{2(n - m + 1)E[\check{\sigma}_p^2]^2}{n^2} \\ &= \frac{(n - m + 1)\left((n - m + 3)\text{Var}(\check{\sigma}_p^2) + 2E[\check{\sigma}_p^2]^2\right)}{n^2}. \end{aligned}$$

Um die Varianz von $\check{\sigma}_p^2$ zu bestimmen, lässt sich ausnutzen, dass

$$\begin{aligned} \text{Var}(\tilde{y}^2) &= E[\tilde{y}^4] - (E[\tilde{y}^2])^2 = n^2\delta^4 + 6n\delta^2 + 3 - (n\delta^2 + 1)^2 \\ &= 2(2n\delta^2 + 1) = 2(2h - 1) \end{aligned}$$

ist. Damit gilt

$$\begin{aligned} \text{Var}(\check{\sigma}_p^2) &= E\left[\text{Var}(\check{\sigma}_p^2 \mid u)\right] + \text{Var}\left(E[\check{\sigma}_p^2 \mid u]\right) \\ &= E\left[\text{Var}\left(\sigma_g^2\left(1 + \frac{\tilde{y}^2}{u}\right) \mid u\right)\right] + \text{Var}\left(E\left[\sigma_g^2\left(1 + \frac{\tilde{y}^2}{u}\right) \mid u\right]\right) \\ &= \sigma_g^4 E[u^{-2}]\text{Var}(\tilde{y}^2) + \sigma_g^4 (E[\tilde{y}^2])^2 \text{Var}(u^{-1}) \\ &= \sigma_g^4 \cdot 2(2h - 1)E[u^{-2}] + \sigma_g^4 h^2 \text{Var}(u^{-1}) \\ &= \sigma_g^4 \left(2(2h - 1)E[u^{-2}] + h^2 (E[u^{-2}] - E[u^{-1}]^2)\right) \\ &= \sigma_g^4 \left((h^2 + 4h - 2)E[u^{-2}] - (hE[u^{-1}])^2\right). \end{aligned}$$

Nach Johnson, Kotz und Balakrishnan (1995) existiert $E[u^{-1}]$ genau dann, wenn $m > 3$ ist, und $E[u^{-2}]$ existiert genau dann, wenn $m > 5$ ist. Da $h > 0$ ist, existiert auch $E[\hat{\sigma}_p^2]$ genau dann, wenn $m > 3$, und $\text{Var}(\hat{\sigma}_p^2)$ genau dann, wenn $m > 5$ ist. \square

Beweis von Satz 4.3.6

Es seien $\mathbf{X} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\hat{\boldsymbol{\mu}}, \mathbf{1})$ und $\mathbf{W} = n\boldsymbol{\Sigma}^{-\frac{1}{2}}\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-\frac{1}{2}} \sim \mathcal{W}_m(n-1, \mathbf{I}_m)$ wie in Lemma 4.3.1. Aufgrund der Unabhängigkeit von $\hat{\boldsymbol{\mu}}$ und $\hat{\boldsymbol{\Sigma}}$ sind auch \mathbf{X} und \mathbf{W} unabhängig. Nach Voraussetzung ist \mathbf{Q} orthogonal zu \mathbf{X} . In Lemma 4.3.1 wird die Identität

$$\hat{\mathbf{w}}^{*T}\boldsymbol{\mu} = \tilde{\mathbf{w}}^T\boldsymbol{\mu} + (\mu_*, 1) (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1} (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{Q}) \mathbf{Q}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}$$

mit

$$\tilde{\mathbf{w}} = \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\mu}}, \mathbf{1})\tilde{\mathbf{M}}^{-1}(\mu_*, 1)^T$$

gezeigt. Somit ist

$$\hat{\mathbf{w}}^* = \tilde{\mathbf{w}} + \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Q}(\mathbf{Q}^T\mathbf{W}^{-1}\mathbf{X}) (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1} (\mu_*, 1)^T.$$

Für die Matrix \mathbf{V} in (4.42) aus dem Beweis von Satz 4.3.2 gilt mit (4.41) und der Orthogonalität von \mathbf{Q} zu \mathbf{X}

$$\begin{aligned} -\mathbf{V}_{22}^{-1}\mathbf{V}_{21} &= -(\mathbf{Q}^T\mathbf{W}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{W}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-\frac{1}{2}} \\ &= -(\mathbf{Q}^T\mathbf{W}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{W}\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right)^T(\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \\ &= -(\mathbf{Q}^T\mathbf{W}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{W}\left((\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}\right. \\ &\quad \left.-(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{Q})\mathbf{Q}^T\right)^T(\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \\ &= \left(-(\mathbf{Q}^T\mathbf{W}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{W}\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\right. \\ &\quad \left.+(\mathbf{Q}^T\mathbf{W}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{W}\mathbf{Q}\mathbf{Q}^T\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\right)(\mathbf{X}^T\mathbf{X})^{\frac{1}{2}} \\ &= \mathbf{Q}^T\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\tilde{\mathbf{M}}^{\frac{1}{2}}. \end{aligned}$$

Dabei ist $\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$ unabhängig von $\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$ (folgt aus Theorem 3.2.10 von Muirhead (1982)) und von $v = n\hat{\sigma}_p^2/\check{\sigma}_p^2$. Dickey (1967) zeigt, dass

$$\mathbf{V}_{22}^{-1}\mathbf{V}_{21} \sim \mathcal{T}_{n-m+2}(\mathbf{I}_{m-2}, \mathbf{I}_2, n+1)$$

gilt. Nach der Definition der Matrix- t -Verteilung lässt sich $\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$ somit als Produkt $(\mathbf{U}^{\frac{1}{2}T})^{-1}\mathbf{Z}$ mit $\mathbf{U} \sim \mathcal{W}_{m-2}(n-1, \mathbf{I}_{m-2})$ schreiben, wobei \mathbf{Z} eine Matrix ist, deren Zeilenvektoren unabhängig $\mathcal{N}(\mathbf{0}_2, \mathbf{I}_2)$ -verteilt sind. Multipliziert man $\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$ mit $-\tilde{\mathbf{M}}^{-\frac{1}{2}}(\mu_*, 1)^T$ so erhält man aus \mathbf{Z} eine Matrix \mathbf{Z}^* , deren Zeilen unabhängig $\mathcal{N}(0, \check{\sigma}_p^2)$ -verteilt sind, da $(\mu_*, 1)\tilde{\mathbf{M}}^{-1}(\mu_*, 1)^T = \check{\sigma}_p^2$ ist. Somit ist dann

$$\mathbf{Y} := \mathbf{Q}^T\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}(\mu_*, 1)^T = -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\tilde{\mathbf{M}}^{-\frac{1}{2}}(\mu_*, 1)^T \sim \mathcal{T}_{n-m+2}(\mathbf{I}_{m-2}, \check{\sigma}_p^2, n).$$

Da $\check{\sigma}_p^2$ eindimensional ist, ist \mathbf{Y} nach Dawid (1981) eine mit $(n-m+2)^{-\frac{1}{2}}$ skalierte multivariat t -verteilte Zufallsvariable. Dividiert man \mathbf{Y} durch $\check{\sigma}_p$ wird aus \mathbf{Z}^* eine Matrix

mit $\mathcal{N}(0, 1)$ -verteilten Zeilen. Durch die Multiplikation mit $(n - m + 2)^{\frac{1}{2}}$ erhält man mit $(n - m + 2)^{\frac{1}{2}} \mathbf{Y} / \check{\sigma}_p$ schließlich eine standard multivariat t -verteilte Zufallsvariable mit $n - m + 2$ Freiheitsgraden. Daher lässt sich \mathbf{Y} folgendermaßen umschreiben:

$$\frac{(n - m + 2)^{\frac{1}{2}} \mathbf{Y}}{\check{\sigma}_p} = \mathbf{x} \left(\frac{\tilde{r}}{n - m + 2} \right)^{-\frac{1}{2}} \Leftrightarrow \mathbf{Y} = \frac{\check{\sigma}_p \mathbf{x}}{\tilde{r}^{\frac{1}{2}}}$$

mit $\mathbf{x} \sim \mathcal{N}(\mathbf{0}_{m-2}, \mathbf{I}_{m-2})$ und $\tilde{r} \sim \chi_{n-m+2}^2$, wobei \mathbf{x} und \tilde{r} unabhängig voneinander sind. Damit erhält man die gewünschte Darstellung der Gewichte des Minimum-Varianz-Portfolios

$$\hat{\mathbf{w}}^* = \tilde{\mathbf{w}} + \Sigma^{-\frac{1}{2}} \mathbf{Q} \mathbf{Y} = \tilde{\mathbf{w}} + \Sigma^{-\frac{1}{2}} \mathbf{Q} \frac{\check{\sigma}_p \mathbf{x}}{\tilde{r}^{\frac{1}{2}}}.$$

Für den Erwartungswert von $\hat{\mathbf{w}}^*$ gilt

$$E[\hat{\mathbf{w}}^*] = E\left[\tilde{\mathbf{w}} + \Sigma^{-\frac{1}{2}} \mathbf{Q} \frac{\check{\sigma}_p \mathbf{x}}{\tilde{r}^{\frac{1}{2}}}\right] = E[\tilde{\mathbf{w}}],$$

da $E[\mathbf{x}] = \mathbf{0}_{m-2}$ ist und \mathbf{x}, \tilde{r} und $\hat{\boldsymbol{\mu}}$ unabhängig sind. Um den Erwartungswert von $\tilde{\mathbf{w}}$ zu bestimmen, wird die Zufallsvariable zunächst geeignet umgeformt. Mit der Darstellung von $\tilde{\mathbf{M}}^{-1}$ in (4.29) gilt

$$\begin{aligned} \tilde{\mathbf{w}} &= \Sigma^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) \tilde{\mathbf{M}}^{-1} (\mu_*, 1)^T \\ &= n \Sigma^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) \begin{pmatrix} \frac{1}{u} & -\frac{\sigma_g y}{\sqrt{nu}} \\ -\frac{\sigma_g y}{\sqrt{nu}} & \frac{\sigma_g^2}{n} \left(1 + \frac{y^2}{u}\right) \end{pmatrix} \begin{pmatrix} \mu_* \\ 1 \end{pmatrix} \\ &= n \Sigma^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) \begin{pmatrix} \frac{1}{u} \left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y\right) \\ \frac{\sigma_g^2}{n} - \frac{\sigma_g y}{\sqrt{nu}} \left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y\right) \end{pmatrix} \\ &= n \Sigma^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) \begin{pmatrix} 0 \\ \frac{\sigma_g^2}{n} \end{pmatrix} + n \Sigma^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) \begin{pmatrix} \frac{1}{u} \left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y\right) \\ -\frac{\sigma_g y}{\sqrt{nu}} \left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y\right) \end{pmatrix} \\ &= \sigma_g^2 \Sigma^{-1} \mathbf{1} + \left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y\right) \frac{n}{u} \Sigma^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) \begin{pmatrix} 1 \\ -\frac{\sigma_g y}{\sqrt{n}} \end{pmatrix} \\ &= \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} + \left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y\right) \frac{n}{u} \Sigma^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) \begin{pmatrix} 1 \\ -\frac{\sigma_g y}{\sqrt{n}} \end{pmatrix} \\ &= \mathbf{w}_g + \left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y\right) \frac{n}{u} \Sigma^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) \begin{pmatrix} 1 \\ -\frac{\sigma_g y}{\sqrt{n}} \end{pmatrix}. \end{aligned}$$

Wegen

$$\begin{aligned}
 \Sigma^{-1}(\hat{\boldsymbol{\mu}}, \mathbf{1}) \begin{pmatrix} 1 \\ -\frac{\sigma_g y}{\sqrt{n}} \end{pmatrix} &= \Sigma^{-1} \left(\hat{\boldsymbol{\mu}} - \frac{\sigma_g y}{\sqrt{n}} \mathbf{1} \right) \\
 &= \Sigma^{-1} \left(\hat{\boldsymbol{\mu}} - \frac{(\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-\frac{1}{2}} (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-\frac{1}{2}} \sqrt{n} \mathbf{1}^T \Sigma^{-1} \hat{\boldsymbol{\mu}}}{\sqrt{n}} \mathbf{1} \right) \\
 &= \Sigma^{-\frac{1}{2}} \left(\mathbf{I}_m - \Sigma^{-\frac{1}{2}} \mathbf{1} (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \right) \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}}
 \end{aligned}$$

lässt sich $\tilde{\mathbf{w}}$ weiter umformen durch

$$\tilde{\mathbf{w}} = \mathbf{w}_g + \left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y \right) \frac{n \Sigma^{-\frac{1}{2}} \left(\mathbf{I}_m - \Sigma^{-\frac{1}{2}} \mathbf{1} (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \right) \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}}}{u}.$$

Nach (4.27) gilt für die $(m \times (m-1))$ -Matrix \mathbf{P} aus dem Beweis von Satz 4.1.1, deren Spalten orthogonal zu $\Sigma^{-\frac{1}{2}} \mathbf{1}$ sind,

$$\mathbf{P} \mathbf{P}^T = \mathbf{I}_m - \Sigma^{-\frac{1}{2}} \mathbf{1} (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-\frac{1}{2}},$$

sodass sich $\tilde{\mathbf{w}}$ schließlich schreiben lässt als

$$\begin{aligned}
 \tilde{\mathbf{w}} &= \mathbf{w}_g + \left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y \right) \frac{n \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}}}{u} \\
 &= \mathbf{w}_g + \sigma_g \left(\frac{\sqrt{n} \mu_*}{\sigma_g} - y \right) \frac{\sqrt{n} \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}}}{u} \\
 &= \mathbf{w}_g + \sigma_g \tilde{y} \frac{\sqrt{n} \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}}}{u} \\
 &= \mathbf{w}_g + \sigma_g \tilde{y} \frac{\Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{z}}{u}. \tag{4.45}
 \end{aligned}$$

Dabei sind $\mathbf{z} = \sqrt{n} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \sim \mathcal{N}(\sqrt{n} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \boldsymbol{\mu}, \mathbf{I}_{m-1})$ und $u = \mathbf{z}^T \mathbf{z}$ im Beweis von Satz 4.1.1 und $\tilde{y} = \sqrt{n} \mu_* / \sigma_g - y$ im Beweis von Satz 4.3.2 definierte Zufallsvariablen, wobei \tilde{y} und \mathbf{z} unabhängig sind. Zur Bestimmung des Erwartungswerts und der Kovarianzmatrix von $\hat{\mathbf{w}}^*$ werden folgende Identitäten benötigt:

$$E \left[\frac{\mathbf{z}}{\mathbf{z}^T \mathbf{z}} \right] = \frac{\phi}{n \psi^2} \boldsymbol{\mu}_z \quad \text{für } m > 2, \tag{4.46}$$

$$E \left[\frac{\mathbf{z} \mathbf{z}^T}{\mathbf{z}^T \mathbf{z}} \right] = \frac{\phi}{n \psi^2} \mathbf{I}_{m-1} + \left(\frac{1}{n \psi^2} - \frac{(m-1)\phi}{n^2 \psi^4} \right) \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \quad \text{für } m > 1, \tag{4.47}$$

$$\begin{aligned}
 E \left[\frac{\mathbf{z} \mathbf{z}^T}{(\mathbf{z}^T \mathbf{z})^2} \right] &= \left(\frac{1-\phi}{2(m-3)} - \frac{\phi}{2n\psi^2} \right) \mathbf{I}_{m-1} \\
 &\quad + \left(\frac{(m-1)\phi}{2n^2\psi^4} - \frac{1-\phi}{2n\psi^2} \right) \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \quad \text{für } m > 3. \tag{4.48}
 \end{aligned}$$

Für den Beweis von (4.46) lässt sich wie für die Identitäten aus Lemma 4.3.4 Lemma 1 von Sawa (1972) verwenden. Für (4.47) und (4.48) leiten Kan und Smith (2008) mit Theorem 6 von Kan, Wang und Hillier (2007) die Identität

$$E \left[\frac{\mathbf{z}\mathbf{z}^T}{(\mathbf{z}^T\mathbf{z})^q} \right] = \frac{{}_1F_1\left(q, \frac{m+1}{2}; -\frac{n\psi^2}{2}\right)}{2^q \left(\frac{m+1}{2} - q\right)_q} \mathbf{I}_{m-1} + \frac{{}_1F_1\left(q, \frac{m+3}{2}; -\frac{n\psi^2}{2}\right)}{2^q \left(\frac{m+3}{2} - q\right)_q} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \quad \text{für } m > 2q - 1$$

mit $(a)_q := a(a+1)\cdots(a+q-1)$, $(a)_0 := 1$ her. Mit ${}_1F_1(0, b; c) = 1$, der Beziehung

$$\frac{{}_1F_1\left(1, \frac{m+1}{2}; -\frac{n\psi^2}{2}\right)}{m-1} = \frac{\phi}{n\psi^2},$$

die man aus (4.11) und (4.12) erhält, und mit den Differenzgleichungen

$${}_1F_1(a, b+1; c) = \frac{b}{c} ({}_1F_1(a, b; c) - {}_1F_1(a-1, b; c)) \quad (4.49)$$

$${}_1F_1(a+1, b; c) = \frac{b(a+c)}{ab} {}_1F_1(a, b; c) + \frac{c(a-b)}{ab} {}_1F_1(a, b+1; c) \quad (4.50)$$

$${}_1F_1(a+1, b; c) = \frac{1+a-b}{a} {}_1F_1(a, b; c) + \frac{b-1}{a} {}_1F_1(a, b-1; c) \quad (4.51)$$

von Abramowitz und Stegun (1965) gilt

$$\begin{aligned} \frac{{}_1F_1\left(1, \frac{m+3}{2}; -\frac{n\psi^2}{2}\right)}{m+1} &= -\frac{\frac{m+1}{n\psi^2}}{m+1} \left({}_1F_1\left(1, \frac{m+1}{2}; -\frac{n\psi^2}{2}\right) - {}_1F_1\left(0, \frac{m+1}{2}; -\frac{n\psi^2}{2}\right) \right) \\ &= -\frac{1}{n\psi^2} \left({}_1F_1\left(1, \frac{m+1}{2}; -\frac{n\psi^2}{2}\right) - 1 \right) \\ &= \frac{1}{n\psi^2} - \frac{(m-1)\phi}{n^2\psi^4}, \end{aligned}$$

$$\begin{aligned} \frac{{}_1F_1\left(2; \frac{m+1}{2}; -\frac{n\psi^2}{2}\right)}{(m-1)(m-3)} &= \frac{1}{(m-1)(m-3)} \left(\left(1 - \frac{n\psi^2}{2}\right) {}_1F_1\left(1, \frac{m+1}{2}; -\frac{n\psi^2}{2}\right) \right. \\ &\quad \left. + \left(\frac{n\psi^2}{2} - \frac{n\psi^2}{m+1}\right) {}_1F_1\left(1, \frac{m+3}{2}; -\frac{n\psi^2}{2}\right) \right) \\ &= \frac{1}{(m-1)(m-3)} \left(\left(1 - \frac{n\psi^2}{2}\right) (m-1) \frac{\phi}{n\psi^2} \right. \\ &\quad \left. + \left(\frac{n\psi^2}{2} - \frac{n\psi^2}{m+1}\right) \left(\frac{1}{n\psi^2} - \frac{(m-1)\phi}{n^2\psi^4}\right) (m+1) \right) \\ &= \frac{\phi}{n\psi^2(m-3)} - \frac{\phi}{2(m-3)} + \frac{m+1}{2(m-1)(m-3)} - \frac{(m+1)\phi}{2n\psi^2(m-3)} \\ &\quad - \frac{1}{(m-1)(m-3)} + \frac{\phi}{n\psi^2(m-3)} \end{aligned}$$

$$\begin{aligned}
 &= \frac{m+1-2-(m-1)\phi}{2(m-1)(m-3)} + \frac{4\phi-(m+1)\phi}{2n\psi^2(m-3)} \\
 &= \frac{1-\phi}{2(m-3)} - \frac{\phi}{2n\psi^2}
 \end{aligned}$$

und

$$\begin{aligned}
 \frac{{}_1F_1\left(2; \frac{m+3}{2}; -\frac{n\psi^2}{2}\right)}{(m-1)(m+1)} &= \frac{1}{(m-1)(m+1)} \left(\frac{1-m}{2} {}_1F_1\left(1; \frac{m+3}{2}; -\frac{n\psi^2}{2}\right) \right. \\
 &\quad \left. + \frac{m+1}{2} {}_1F_1\left(1; \frac{m+1}{2}; -\frac{n\psi^2}{2}\right) \right) \\
 &= \frac{1}{(m-1)(m+1)} \left(\frac{1-m}{2} \left(\frac{1}{n\psi^2} - \frac{(m-1)\phi}{n^2\psi^4} \right) (m+1) \right. \\
 &\quad \left. + \frac{m+1}{2} \frac{\phi}{n\psi^2} (m-1) \right) \\
 &= -\frac{1}{2n\psi^2} + \frac{(m-1)\phi}{2n^2\psi^2} + \frac{\phi}{2n\psi^2} \\
 &= \frac{(m-1)\phi}{2n^2\psi^4} - \frac{1-\phi}{2n\psi^2}
 \end{aligned}$$

für $m > 1$ bzw. $m > 3$. Damit können (4.47) und (4.48) berechnet werden. Es gilt

$$\begin{aligned}
 E \begin{bmatrix} \mathbf{z}\mathbf{z}^T \\ \mathbf{z}^T\mathbf{z} \end{bmatrix} &= \frac{{}_1F_1\left(1, \frac{m+1}{2}; -\frac{n\psi^2}{2}\right)}{2\left(\frac{m+1}{2}-1\right)} \mathbf{I}_{m-1} + \frac{{}_1F_1\left(1, \frac{m+3}{2}; -\frac{n\psi^2}{2}\right)}{2\left(\frac{m+3}{2}-1\right)} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \\
 &= \frac{\phi}{n\psi^2} \mathbf{I}_{m-1} + \left(\frac{1}{n\psi^2} - \frac{(m-1)\phi}{n^2\psi^4} \right) \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T
 \end{aligned}$$

und

$$\begin{aligned}
 E \begin{bmatrix} \mathbf{z}\mathbf{z}^T \\ (\mathbf{z}^T\mathbf{z})^2 \end{bmatrix} &= \frac{{}_1F_1\left(2, \frac{m+1}{2}; -\frac{n\psi^2}{2}\right)}{2^2\left(\frac{m+1}{2}-2\right)_2} \mathbf{I}_{m-1} + \frac{{}_1F_1\left(2, \frac{m+3}{2}; -\frac{n\psi^2}{2}\right)}{2^2\left(\frac{m+3}{2}-2\right)_2} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \\
 &= \left(\frac{1-\phi}{2(m-3)} - \frac{\phi}{2n\psi^2} \right) \mathbf{I}_{m-1} + \left(\frac{(m-1)\phi}{2n^2\psi^4} - \frac{1-\phi}{2n\psi^2} \right) \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T.
 \end{aligned}$$

Für den Erwartungswert der Gewichte des empirischen Minimum-Varianz-Portfolios gilt mit (4.45) aufgrund der Unabhängigkeit von \tilde{y} und \mathbf{z} und wegen $E[\tilde{y}] = \sqrt{n}\delta$ mit $\delta = (\mu_* - \mu_g)/\sigma_g$

$$\begin{aligned}
 E[\widehat{\mathbf{w}}^*] &= E[\tilde{\mathbf{w}}] = \mathbf{w}_g + \sigma_g \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} E[\tilde{y}] E \begin{bmatrix} \mathbf{z} \\ u \end{bmatrix} \\
 &= \mathbf{w}_g + (\mu_* - \mu_g) \sqrt{n} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} E \begin{bmatrix} \mathbf{z} \\ u \end{bmatrix}.
 \end{aligned}$$

Mit der Identität aus (4.46) und mit $\boldsymbol{\mu}_z = \sqrt{n}\mathbf{P}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}$ folgt für $m > 2$

$$\begin{aligned}
 \sqrt{n}\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{P}E\left[\frac{\mathbf{z}}{u}\right] &= \sqrt{n}\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{P}\frac{\phi}{n\psi^2}\boldsymbol{\mu}_z \\
 &= \frac{\phi}{\psi^2}\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{P}\mathbf{P}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} \\
 &= \frac{\phi}{\psi^2}\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mathbf{I}_m - \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{1}\mathbf{1}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}\right)\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} \\
 &= \frac{\phi}{\psi^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_g\mathbf{1}), \tag{4.52}
 \end{aligned}$$

wobei $\mu_g = A/C = \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}$ ist und für $\mathbf{P}\mathbf{P}^T$ (4.27) ausgenutzt wurde. Es gilt also

$$E[\widehat{\mathbf{w}}^*] = \mathbf{w}_g + (\mu_* - \mu_g)\frac{\phi}{\psi^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_g\mathbf{1})$$

für $m > 2$. Im Beweis von Satz 4.1.1 wurde die Inverse von $\widehat{\mathbf{M}}$ bestimmt. Das Ergebnis lässt sich auf die Inverse von \mathbf{M} übertragen, indem man anstelle der empirischen Effizienzmengekonstanten die wahren Konstanten einsetzt. Es ist also

$$\begin{aligned}
 \mathbf{w}^* &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}, \mathbf{1})\mathbf{M}^{-1}(\mu_*, \mathbf{1})^T \\
 &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}, \mathbf{1})\begin{pmatrix} \frac{1}{\psi^2} & -\frac{\mu_g}{\psi^2} \\ -\frac{\mu_g}{\psi^2} & \sigma_g^2 + \frac{\mu_g^2}{\psi^2} \end{pmatrix}(\mu_*, \mathbf{1})^T \\
 &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}, \mathbf{1})\begin{pmatrix} \frac{\mu_* - \mu_g}{\psi^2} \\ \sigma_g^2 - \mu_g\left(\frac{\mu_* - \mu_g}{\psi^2}\right) \end{pmatrix} \\
 &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}, \mathbf{1})\begin{pmatrix} 0 \\ \sigma_g^2 \end{pmatrix} + \frac{\mu_* - \mu_g}{\psi^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}, \mathbf{1})\begin{pmatrix} 1 \\ -\mu_g \end{pmatrix} \\
 &= \boldsymbol{\Sigma}^{-1}\mathbf{1}\sigma_g^2 + \frac{\mu_* - \mu_g}{\psi^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_g\mathbf{1}) \\
 &= \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} + \frac{\mu_* - \mu_g}{\psi^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_g\mathbf{1}) \\
 &= \mathbf{w}_g + \frac{\mu_* - \mu_g}{\psi^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_g\mathbf{1}).
 \end{aligned}$$

Damit gilt für $m > 2$ schließlich

$$E[\widehat{\mathbf{w}}^*] = \mathbf{w}_g + \phi\frac{\mu_* - \mu_g}{\psi^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_g\mathbf{1}) = \mathbf{w}_g + \phi(\mathbf{w}^* - \mathbf{w}_g). \tag{4.53}$$

Für die Varianz der Gewichte des empirischen Minimum-Varianz-Portfolios gilt

$$\begin{aligned}
 \text{Var}(\widehat{\mathbf{w}}^*) &= \text{Var}\left(\widetilde{\mathbf{w}} + \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Q}\frac{\check{\sigma}_p\mathbf{X}}{\check{r}^{\frac{1}{2}}}\right) \\
 &= \text{Var}(\widetilde{\mathbf{w}}) + \text{Var}\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Q}\frac{\check{\sigma}_p\mathbf{X}}{\check{r}^{\frac{1}{2}}}\right) + 2\text{Cov}\left(\widetilde{\mathbf{w}}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Q}\frac{\check{\sigma}_p\mathbf{X}}{\check{r}^{\frac{1}{2}}}\right).
 \end{aligned}$$

Da \mathbf{x} , \tilde{r} und $\hat{\boldsymbol{\mu}}$ unabhängig sind und der Erwartungswert von \mathbf{x} gleich null ist, gilt

$$\begin{aligned} \text{Cov} \left(\tilde{\mathbf{w}}, \Sigma^{-\frac{1}{2}} \mathbf{Q} \frac{\check{\sigma}_p \mathbf{x}}{\tilde{r}^{\frac{1}{2}}} \right) &= E \left[\tilde{\mathbf{w}} \frac{\check{\sigma}_p \mathbf{x}^T}{\tilde{r}^{\frac{1}{2}}} \mathbf{Q}^T \Sigma^{-\frac{1}{2}} \right] - E[\tilde{\mathbf{w}}] E \left[\frac{\check{\sigma}_p \mathbf{x}^T}{\tilde{r}^{\frac{1}{2}}} \mathbf{Q}^T \Sigma^{-\frac{1}{2}} \right] \\ &= E \left[\tilde{\mathbf{w}} \frac{\check{\sigma}_p}{\tilde{r}^{\frac{1}{2}}} \mathbf{Q}^T \Sigma^{-\frac{1}{2}} \right] E[\mathbf{x}^T] - E[\tilde{\mathbf{w}}] E \left[\frac{\check{\sigma}_p}{\tilde{r}^{\frac{1}{2}}} \mathbf{Q}^T \Sigma^{-\frac{1}{2}} \right] E[\mathbf{x}^T] \\ &= \mathbf{0}_{m \times m}, \end{aligned}$$

und

$$\begin{aligned} \text{Var} \left(\Sigma^{-\frac{1}{2}} \mathbf{Q} \frac{\check{\sigma}_p \mathbf{x}}{\tilde{r}^{\frac{1}{2}}} \right) &= \Sigma^{-\frac{1}{2}} E \left[\frac{\check{\sigma}_p^2 \mathbf{Q} \mathbf{x} \mathbf{x}^T \mathbf{Q}^T}{\tilde{r}} \right] \Sigma^{-\frac{1}{2}} - E \left[\Sigma^{-\frac{1}{2}} \mathbf{Q} \frac{\check{\sigma}_p \mathbf{x}}{\tilde{r}^{\frac{1}{2}}} \right] E \left[\Sigma^{-\frac{1}{2}} \mathbf{Q} \frac{\check{\sigma}_p \mathbf{x}}{\tilde{r}^{\frac{1}{2}}} \right]^T \\ &= \Sigma^{-\frac{1}{2}} E \left[\frac{\check{\sigma}_p^2 \mathbf{Q} \mathbf{x} \mathbf{x}^T \mathbf{Q}^T}{\tilde{r}} \right] \Sigma^{-\frac{1}{2}}. \end{aligned}$$

Somit ist

$$\text{Var}(\hat{\mathbf{w}}^*) = \text{Var}(\tilde{\mathbf{w}}) + \Sigma^{-\frac{1}{2}} E \left[\frac{\check{\sigma}_p^2 \mathbf{Q} \mathbf{x} \mathbf{x}^T \mathbf{Q}^T}{\tilde{r}} \right] \Sigma^{-\frac{1}{2}}. \quad (4.54)$$

Zunächst wird die Kovarianzmatrix von $\tilde{\mathbf{w}}$,

$$\text{Var}(\tilde{\mathbf{w}}) = E[\tilde{\mathbf{w}} \tilde{\mathbf{w}}^T] - E[\tilde{\mathbf{w}}] E[\tilde{\mathbf{w}}]^T,$$

betrachtet. Mit der Darstellung von $\tilde{\mathbf{w}}$ in (4.45) erhält man für $E[\tilde{\mathbf{w}} \tilde{\mathbf{w}}^T]$ aufgrund der Unabhängigkeit von y und \mathbf{z} und wegen $E[y] = \sqrt{n} \mu_g / \sigma_g$

$$\begin{aligned} E[\tilde{\mathbf{w}} \tilde{\mathbf{w}}^T] &= E \left[\left(\mathbf{w}_g + \sigma_g \tilde{y} \frac{\Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{z}}{u} \right) \left(\mathbf{w}_g + \sigma_g \tilde{y} \frac{\Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{z}}{u} \right)^T \right] \\ &= \mathbf{w}_g \mathbf{w}_g^T + \sigma_g E \left[\left(\frac{\sqrt{n} \mu_*}{\sigma_g} - y \right) \frac{\Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{w}_g^T}{u} \right] \\ &\quad + \sigma_g E \left[\left(\frac{\sqrt{n} \mu_*}{\sigma_g} - y \right) \frac{\mathbf{w}_g \mathbf{z}^T \mathbf{P}^T \Sigma^{-\frac{1}{2}}}{u} \right] \\ &\quad + \sigma_g^2 E \left[\left(\frac{\sqrt{n} \mu_*}{\sigma_g} - y \right)^2 \frac{\Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{z}^T \mathbf{P}^T \Sigma^{-\frac{1}{2}}}{u^2} \right] \\ &= \mathbf{w}_g \mathbf{w}_g^T + \mu_* E \left[\frac{\sqrt{n} \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{w}_g^T}{u} \right] - \sigma_g E[y] E \left[\frac{\Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{w}_g^T}{u} \right] \\ &\quad + \mu_* E \left[\frac{\sqrt{n} \mathbf{w}_g \mathbf{z}^T \mathbf{P}^T \Sigma^{-\frac{1}{2}}}{u} \right] - \sigma_g E[y] E \left[\frac{\mathbf{w}_g \mathbf{z}^T \mathbf{P}^T \Sigma^{-\frac{1}{2}}}{u} \right] \\ &\quad + E \left[\left(\mu_* - \frac{\sigma_g}{\sqrt{n}} y \right)^2 \right] E \left[\frac{n \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{z}^T \mathbf{P}^T \Sigma^{-\frac{1}{2}}}{u^2} \right] \end{aligned}$$

$$\begin{aligned}
 &= \mathbf{w}_g \mathbf{w}_g^T + (\mu_* - \mu_g) E \left[\frac{\sqrt{n} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{w}_g^T}{u} \right] + (\mu_* - \mu_g) E \left[\frac{\sqrt{n} \mathbf{w}_g \mathbf{z}^T \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}}{u} \right] \\
 &\quad + \left(\mu_*^2 - 2\mu_* \frac{\sigma_g}{\sqrt{n}} E[y] + \frac{\sigma_g^2}{n} E[y^2] \right) E \left[\frac{n \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{z}^T \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}}{u^2} \right] \\
 &= \mathbf{w}_g \mathbf{w}_g^T + (\mu_* - \mu_g) E \left[\frac{\sqrt{n} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{w}_g^T}{u} \right] + (\mu_* - \mu_g) E \left[\frac{\sqrt{n} \mathbf{w}_g \mathbf{z}^T \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}}{u} \right] \\
 &\quad + \left((\mu_* - \mu_g)^2 + \frac{\sigma_g^2}{n} \right) E \left[\frac{n \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{z}^T \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}}{u^2} \right].
 \end{aligned}$$

Unter Ausnutzung von (4.52) und (4.53) und wegen $\delta = (\mu_* - \mu_g)/\sigma_g$ und $h = n\delta^2 + 1$ lässt sich dieser Ausdruck umschreiben als

$$E[\tilde{\mathbf{w}} \tilde{\mathbf{w}}^T] = \mathbf{w}_g \mathbf{w}_g^T + \phi(\mathbf{w}^* - \mathbf{w}_g) \mathbf{w}_g^T + \phi \mathbf{w}_g (\mathbf{w}^* - \mathbf{w}_g)^T + \frac{\sigma_g^2 h}{n} E \left[\frac{n \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{z}^T \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}}{u^2} \right].$$

Da $E[\tilde{\mathbf{w}}] = E[\hat{\mathbf{w}}^*]$ gilt, folgt damit für die Varianz von $\tilde{\mathbf{w}}$

$$\begin{aligned}
 \text{Var}(\tilde{\mathbf{w}}) &= \mathbf{w}_g \mathbf{w}_g^T + \phi(\mathbf{w}^* - \mathbf{w}_g) \mathbf{w}_g^T + \phi \mathbf{w}_g (\mathbf{w}^* - \mathbf{w}_g)^T + \frac{\sigma_g^2 h}{n} E \left[\frac{n \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{z}^T \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}}{u^2} \right] \\
 &\quad - (\mathbf{w}_g + \phi(\mathbf{w}^* - \mathbf{w}_g)) (\mathbf{w}_g + \phi(\mathbf{w}^* - \mathbf{w}_g))^T \\
 &= \sigma_g^2 h E \left[\frac{\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{z} \mathbf{z}^T \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}}{u^2} \right] - \phi^2 (\mathbf{w}^* - \mathbf{w}_g) (\mathbf{w}^* - \mathbf{w}_g)^T \\
 &= \sigma_g^2 h \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} E \left[\frac{\mathbf{z} \mathbf{z}^T}{(\mathbf{z}^T \mathbf{z})^2} \right] \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} - \phi^2 (\mathbf{w}^* - \mathbf{w}_g) (\mathbf{w}^* - \mathbf{w}_g)^T.
 \end{aligned}$$

Für den zweiten Term des Ausdrucks für die Varianz von $\hat{\mathbf{w}}^*$ in (4.54) gilt wegen $E[\mathbf{x} \mathbf{x}^T] = \mathbf{I}_{m-2}$, $E[1/\tilde{r}] = 1/(n-m)$ und der Unabhängigkeit von \mathbf{x} , \tilde{r} und $\hat{\boldsymbol{\mu}}$

$$\boldsymbol{\Sigma}^{-\frac{1}{2}} E \left[\frac{\check{\sigma}_p^2 \mathbf{Q} \mathbf{x} \mathbf{x}^T \mathbf{Q}^T}{\tilde{r}} \right] \boldsymbol{\Sigma}^{-\frac{1}{2}} = \frac{1}{n-m} E \left[\check{\sigma}_p^2 \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{Q} \mathbf{Q}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \right].$$

Um $\mathbf{Q} \mathbf{Q}^T$ umzuschreiben, werden (4.27) und (4.40) ausgenutzt. Mit $\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} =: \tilde{A}$, $\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} =: \tilde{B}$ und $\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = C$ gilt

$$\begin{aligned}
 \mathbf{Q} \mathbf{Q}^T &= \mathbf{I}_m - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \\
 &= \mathbf{I}_m - \mathbf{X} (\tilde{\mathbf{M}})^{-1} \mathbf{X}^T \\
 &= \mathbf{I}_m - \left(\boldsymbol{\Sigma}^{-\frac{1}{2}} \hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1} \right) \cdot \frac{1}{\tilde{B}C - \tilde{A}^2} \begin{pmatrix} C & -\tilde{A} \\ -\tilde{A} & \tilde{B} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \\ \mathbf{1}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \end{pmatrix}
 \end{aligned}$$

$$\begin{aligned}
 &= \mathbf{I}_m - \frac{1}{\tilde{B}C - \tilde{A}^2} \left(\Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} C - \Sigma^{-\frac{1}{2}} \mathbf{1} \tilde{A}, \Sigma^{-\frac{1}{2}} \mathbf{1} \tilde{B} - \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \tilde{A} \right) \begin{pmatrix} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \\ \mathbf{1}^T \Sigma^{-\frac{1}{2}} \end{pmatrix} \\
 &= \mathbf{I}_m - \frac{1}{\tilde{B}C - \tilde{A}^2} \left(\Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} C - \Sigma^{-\frac{1}{2}} \mathbf{1} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \tilde{A} + \Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \tilde{B} \right. \\
 &\quad \left. - \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \tilde{A} \right) \\
 &= \mathbf{P} \mathbf{P}^T + \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}}}{C} - \frac{1}{\tilde{B}C - \tilde{A}^2} \left(\Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} C - \Sigma^{-\frac{1}{2}} \mathbf{1} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \tilde{A} \right. \\
 &\quad \left. + \Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \tilde{B} - \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \tilde{A} \right) \\
 &= \mathbf{P} \mathbf{P}^T + \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \left(\tilde{B} - \frac{\tilde{A}^2}{C} \right)}{\tilde{B}C - \tilde{A}^2} \\
 &\quad - \frac{\left(\Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} C - \Sigma^{-\frac{1}{2}} \mathbf{1} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \tilde{A} + \Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \tilde{B} - \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \tilde{A} \right)}{\tilde{B}C - \tilde{A}^2} \\
 &= \mathbf{P} \mathbf{P}^T - \frac{\Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} C - \Sigma^{-\frac{1}{2}} \mathbf{1} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \tilde{A} - \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \tilde{A} + \Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \frac{\tilde{A}^2}{C}}{\tilde{B}C - \tilde{A}^2} \\
 &= \mathbf{P} \mathbf{P}^T - \frac{\Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} - \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \tilde{A}}{C} - \frac{\Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \tilde{A}}{C} + \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}} \tilde{A}^2}{C^2}}{\tilde{B} - \frac{\tilde{A}^2}{C}} \\
 &= \mathbf{P} \mathbf{P}^T - \frac{\left(\Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} - \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-1} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}}}{C} \right) \left(\mathbf{I}_m - \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}}}{C} \right)}{\hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \left(\mathbf{I}_m - \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}}}{C} \right) \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}}} \\
 &= \mathbf{P} \mathbf{P}^T - \frac{\left(\mathbf{I}_m - \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}}}{C} \right) \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \left(\mathbf{I}_m - \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}}}{C} \right)}{\hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \left(\mathbf{I}_m - \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}}}{C} \right) \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}}} \\
 &= \mathbf{P} \mathbf{P}^T - \frac{\mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T}{\hat{\boldsymbol{\mu}}^T \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \hat{\boldsymbol{\mu}}} = \mathbf{P} \mathbf{P}^T - \frac{\mathbf{P} \mathbf{z} \mathbf{z}^T \mathbf{P}^T}{\mathbf{z}^T \mathbf{z}}.
 \end{aligned}$$

Mit dieser Umformung und wegen $\check{\sigma}_p^2 = \sigma_g^2(1 + \tilde{y}^2/u)$, wobei \tilde{y} und $u = \mathbf{z}^T \mathbf{z}$ unabhängig sind, gilt

$$\begin{aligned}
 E \left[\check{\sigma}_p^2 \Sigma^{-\frac{1}{2}} \mathbf{Q} \mathbf{Q}^T \Sigma^{-\frac{1}{2}} \right] &= E \left[\check{\sigma}_p^2 \Sigma^{-\frac{1}{2}} \left(\mathbf{P} \mathbf{P}^T - \frac{\mathbf{P} \mathbf{z} \mathbf{z}^T \mathbf{P}^T}{\mathbf{z}^T \mathbf{z}} \right) \Sigma^{-\frac{1}{2}} \right] \\
 &= E \left[\check{\sigma}_p^2 \right] \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}} - \sigma_g^2 \Sigma^{-\frac{1}{2}} \mathbf{P} E \left[\frac{\mathbf{z} \mathbf{z}^T}{\mathbf{z}^T \mathbf{z}} \right] \mathbf{P}^T \Sigma^{-\frac{1}{2}} \\
 &\quad - \sigma_g^2 h \Sigma^{-\frac{1}{2}} \mathbf{P} E \left[\frac{\mathbf{z} \mathbf{z}^T}{(\mathbf{z}^T \mathbf{z})^2} \right] \mathbf{P}^T \Sigma^{-\frac{1}{2}},
 \end{aligned}$$

wobei ausgenutzt wurde, dass $E[\tilde{y}^2] = h$ ist. Mit (4.47) und (4.48) folgt für $m > 3$ nun insgesamt

$$\begin{aligned}
 \text{Var}(\widehat{\mathbf{w}}^*) &= \text{Var}(\tilde{\mathbf{w}}) + \frac{E\left[\check{\sigma}_p^2 \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{Q} \mathbf{Q}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}\right]}{n-m} \\
 &= \sigma_g^2 h \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} E \left[\frac{\mathbf{z} \mathbf{z}^T}{(\mathbf{z}^T \mathbf{z})^2} \right] \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} - \phi^2 (\mathbf{w}^* - \mathbf{w}_g) (\mathbf{w}^* - \mathbf{w}_g)^T \\
 &\quad + \frac{1}{n-m} \left(E[\check{\sigma}_p^2] \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} - \sigma_g^2 \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} E \left[\frac{\mathbf{z} \mathbf{z}^T}{\mathbf{z}^T \mathbf{z}} \right] \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \right. \\
 &\quad \left. - \sigma_g^2 h \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} E \left[\frac{\mathbf{z} \mathbf{z}^T}{(\mathbf{z}^T \mathbf{z})^2} \right] \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \right) \\
 &= \frac{n-m-1}{n-m} \sigma_g^2 h \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} E \left[\frac{\mathbf{z} \mathbf{z}^T}{(\mathbf{z}^T \mathbf{z})^2} \right] \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} + \frac{E[\check{\sigma}_p^2]}{n-m} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \\
 &\quad - \frac{\sigma_g^2}{n-m} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} E \left[\frac{\mathbf{z} \mathbf{z}^T}{\mathbf{z}^T \mathbf{z}} \right] \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} - \phi^2 (\mathbf{w}^* - \mathbf{w}_g) (\mathbf{w}^* - \mathbf{w}_g)^T \\
 &= \frac{n-m-1}{n-m} \sigma_g^2 h \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \left(\left(\frac{1-\phi}{2(m-3)} - \frac{\phi}{2n\psi^2} \right) \mathbf{I}_{m-1} \right. \\
 &\quad \left. + \left(\frac{(m-1)\phi}{2n^2\psi^4} - \frac{1-\phi}{2n\psi^2} \right) \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \right) \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} + \frac{E[\check{\sigma}_p^2]}{n-m} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \\
 &\quad - \frac{\sigma_g^2}{n-m} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \left(\frac{\phi}{n\psi^2} \mathbf{I}_{m-1} + \left(\frac{1}{n\psi^2} - \frac{(m-1)\phi}{n^2\psi^4} \right) \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \right) \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \\
 &\quad - \phi^2 (\mathbf{w}^* - \mathbf{w}_g) (\mathbf{w}^* - \mathbf{w}_g)^T \\
 &= \frac{n-m-1}{n-m} \sigma_g^2 h \left(\frac{1-\phi}{2(m-3)} - \frac{\phi}{2n\psi^2} \right) \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \\
 &\quad + \frac{n-m-1}{n-m} \sigma_g^2 h \left(\frac{(m-1)\phi}{2n^2\psi^4} - \frac{1-\phi}{2n\psi^2} \right) \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \\
 &\quad + \frac{E[\check{\sigma}_p^2]}{n-m} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} - \frac{\sigma_g^2 \phi}{(n-m)n\psi^2} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \\
 &\quad - \frac{\sigma_g^2}{(n-m)} \left(\frac{1}{n\psi^2} - \frac{(m-1)\phi}{n^2\psi^4} \right) \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \\
 &\quad - \phi^2 (\mathbf{w}^* - \mathbf{w}_g) (\mathbf{w}^* - \mathbf{w}_g)^T
 \end{aligned}$$

$$\begin{aligned}
 &= \left(\frac{n-m-1}{n-m} \sigma_g^2 h \left(\frac{1-\phi}{2(m-3)} - \frac{\phi}{2n\psi^2} \right) + \frac{E[\check{\sigma}_p^2]}{n-m} - \frac{\sigma_g^2 \phi}{(n-m)n\psi^2} \right) \\
 &\quad \times \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \\
 &\quad + \left(\frac{n-m-1}{n-m} \sigma_g^2 h \left(\frac{(m-1)\phi}{2n\psi^2} - \frac{1-\phi}{2} \right) - \frac{\sigma_g^2}{n-m} \left(1 - \frac{(m-1)\phi}{n\psi^2} \right) \right) \\
 &\quad \times \frac{\Sigma^{-\frac{1}{2}} \mathbf{P} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \mathbf{P}^T \Sigma^{-\frac{1}{2}}}{n\psi^2} - \phi^2 (\mathbf{w}^* - \mathbf{w}_g)(\mathbf{w}^* - \mathbf{w}_g)^T.
 \end{aligned}$$

Mithilfe von (4.27) lässt sich $\Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}}$ folgendermaßen umschreiben:

$$\begin{aligned}
 \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}} &= \Sigma^{-\frac{1}{2}} \left(\mathbf{I}_m - \frac{\Sigma^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \Sigma^{-\frac{1}{2}}}{C} \right) \Sigma^{-\frac{1}{2}} \\
 &= \Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{1} \mathbf{1}^T \Sigma^{-1}}{C} = \Sigma_a.
 \end{aligned}$$

Mit diesem Ergebnis sowie mit $\boldsymbol{\mu}_z = \sqrt{n} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \boldsymbol{\mu}$ und der Umformung

$$\begin{aligned}
 \psi^2 &= \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{1} (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-1} \boldsymbol{\mu} \\
 &= \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} - \mu_g \mathbf{1}^T \Sigma^{-1} \boldsymbol{\mu} - \mu_g \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{1} + \mu_g \mathbf{1}^T \Sigma^{-1} \boldsymbol{\mu} \\
 &= \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} - \mu_g \mathbf{1}^T \Sigma^{-1} \boldsymbol{\mu} - \mu_g \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{1} + \mu_g \mu_g \mathbf{1}^T \Sigma^{-1} \mathbf{1} \\
 &= (\boldsymbol{\mu} - \mu_g \mathbf{1})^T \Sigma^{-1} (\boldsymbol{\mu} - \mu_g \mathbf{1})
 \end{aligned}$$

gilt für $\Sigma^{-\frac{1}{2}} \mathbf{P} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \mathbf{P}^T \Sigma^{-\frac{1}{2}} / (n\psi^2)$

$$\begin{aligned}
 \frac{\Sigma^{-\frac{1}{2}} \mathbf{P} \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T \mathbf{P}^T \Sigma^{-\frac{1}{2}}}{n\psi^2} &= \frac{n \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}} \boldsymbol{\mu} \boldsymbol{\mu}^T \Sigma^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \Sigma^{-\frac{1}{2}}}{n\psi^2} \\
 &= \frac{\left(\Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{1} \mathbf{1}^T \Sigma^{-1}}{C} \right) \boldsymbol{\mu} \boldsymbol{\mu}^T \left(\Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{1} \mathbf{1}^T \Sigma^{-1}}{C} \right)}{(\boldsymbol{\mu} - \mu_g \mathbf{1})^T \Sigma^{-1} (\boldsymbol{\mu} - \mu_g \mathbf{1})} \\
 &= \frac{\Sigma^{-1} (\boldsymbol{\mu} - \mu_g \mathbf{1}) (\boldsymbol{\mu} - \mu_g \mathbf{1})^T \Sigma^{-1}}{(\boldsymbol{\mu} - \mu_g \mathbf{1})^T \Sigma^{-1} (\boldsymbol{\mu} - \mu_g \mathbf{1})} = \Sigma_b.
 \end{aligned}$$

Damit erhält man schließlich die gewünschte Darstellung der Varianz von $\widehat{\mathbf{w}}^*$. \square

Beweis von Satz 4.4.1

Nach Lemma 4.3.1 gilt

$$\tilde{\mu}_p = \widehat{\mathbf{w}}^{*T} \boldsymbol{\mu} = \check{\mu}_p + (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) \mathbf{Q}^T \Sigma^{-\frac{1}{2}} \boldsymbol{\mu}$$

mit $\check{\mu}_p = \widetilde{\mathbf{w}}^T \boldsymbol{\mu} = (\mu_*, 1) \widetilde{\mathbf{M}}^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \Sigma^{-1} \boldsymbol{\mu}$ und

$$\begin{aligned}
 \widetilde{\sigma}_p^2 &= \widehat{\mathbf{w}}^{*T} \Sigma \widehat{\mathbf{w}}^* \\
 &= \check{\sigma}_p^2 + (\mu_*, 1) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{Q}) (\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T
 \end{aligned}$$

mit $\check{\sigma}_p^2 = \tilde{\mathbf{w}}^T \boldsymbol{\Sigma} \tilde{\mathbf{w}} = \sigma_g^2 (1 + \tilde{y}^2/u)$. Mit der im Beweis von Satz 4.3.6 definierten Zufallsvariablen

$$\mathbf{Y} = (\mathbf{Q}^T \mathbf{W}^{-1} \mathbf{X}) (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T = \frac{\check{\sigma}_p \mathbf{x}}{\tilde{r}^{\frac{1}{2}}}$$

mit $\mathbf{x} \sim \mathcal{N}(\mathbf{0}_{m-2}, \mathbf{I}_{m-2})$ und $\tilde{r} \sim \chi_{n-m+2}^2$, die unabhängig voneinander und unabhängig von $\hat{\boldsymbol{\mu}}$ sind, lassen sich $\tilde{\mu}_p$ und $\tilde{\sigma}_p^2$ schreiben als

$$\tilde{\mu}_p = \check{\mu}_p + \mathbf{Y}^T \mathbf{Q}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} = \check{\mu}_p + \frac{\mathbf{x}^T \mathbf{Q}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}}{\tilde{r}^{\frac{1}{2}}} \check{\sigma}_p$$

und

$$\tilde{\sigma}_p^2 = \check{\sigma}_p^2 + \mathbf{Y}^T \mathbf{Y} = \check{\sigma}_p^2 \left(1 + \frac{\mathbf{x}^T \mathbf{x}}{\tilde{r}} \right).$$

Definiere nun

$$\tilde{d} = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{Q} \mathbf{Q}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\mathbf{I}_m - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}$$

und

$$\tilde{x} = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{Q} \frac{\mathbf{x}}{\tilde{d}^{\frac{1}{2}}}.$$

Für den zweiten Ausdruck von \tilde{d} wurde ausgenutzt, dass $(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1}, \mathbf{Q})$ eine Orthonormalbasis des \mathbb{R}^m ist und somit $\mathbf{Q} \mathbf{Q}^T + \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \mathbf{I}_m$ gilt. Da $\mathbf{x} \sim \mathcal{N}(\mathbf{0}_{m-2}, \mathbf{I}_{m-2})$ ist, gilt $\tilde{x} \sim \mathcal{N}(0, 1)$. Für die Zufallsvariable

$$\tilde{q} = \mathbf{x}^T \mathbf{x} - \tilde{x}^2$$

gilt wegen $\mathbf{x}^T \mathbf{x} \sim \chi_{m-2}^2$ und $\tilde{x}^2 \sim \chi_1^2$, dass sie χ_{m-3}^2 -verteilt ist. Außerdem sind \tilde{x} und \tilde{q} aufgrund der Übereinstimmung von bedingter und unbedingter Verteilung unabhängig voneinander. Somit lassen sich $\tilde{\mu}_p$ und $\tilde{\sigma}_p^2$ weiter umschreiben als

$$\tilde{\mu}_p = \check{\mu}_p + \check{\sigma}_p \tilde{d}^{\frac{1}{2}} \frac{\tilde{x}}{\tilde{r}^{\frac{1}{2}}}$$

und

$$\tilde{\sigma}_p^2 = \check{\sigma}_p^2 \left(1 + \frac{\tilde{x}^2 + \tilde{q}}{\tilde{r}} \right).$$

Letzteres ist bereits die gewünschte Form von $\tilde{\sigma}_p^2$. Es seien $y = \sqrt{n} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} / \sqrt{C}$, $\mathbf{z} = \sqrt{n} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}_z, \mathbf{I}_{m-1})$ mit $\boldsymbol{\mu}_z = \sqrt{n} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}$, $u = \mathbf{z}^T \mathbf{z}$ und $\tilde{y} = \sqrt{n} \mu_* / \sigma_g - y$ wie in den Beweisen von Satz 4.1.1 und Satz 4.3.2. Nach (4.28) gilt $\boldsymbol{\mu}_z^T \boldsymbol{\mu}_z = n\psi^2$. Die Zufallsvariable

$$\tilde{z} = \frac{\mathbf{z}^T \boldsymbol{\mu}_z}{(\boldsymbol{\mu}_z^T \boldsymbol{\mu}_z)^{\frac{1}{2}}} = \frac{\mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n\psi}}$$

ist aufgrund der Normalverteilung von \mathbf{z} ebenfalls normalverteilt mit $E[\tilde{z}] = \sqrt{n}\psi$ und $Var(\tilde{z}) = 1$. Definiere

$$\tilde{u} = \mathbf{z}^T \left(\mathbf{I}_{m-1} - \boldsymbol{\mu}_z (\boldsymbol{\mu}_z^T \boldsymbol{\mu}_z)^{-1} \boldsymbol{\mu}_z^T \right) \mathbf{z} = u - \tilde{z}^2.$$

Wegen $u \sim \chi_{m-1}^2(n\psi^2)$ und $\tilde{z}^2 \sim \chi_1^2(n\psi^2)$ ist $\tilde{u} \sim \chi_{m-2}^2$. Um die gewünschte Form von $\check{\mu}_p$ zu erhalten, müssen $\check{\mu}_p$ und \tilde{d} geeignet umgeschrieben werden. Dafür lässt sich ausnutzen, dass sich $\mathbf{z}^T \boldsymbol{\mu}_z$ mithilfe von (4.27) folgendermaßen schreiben lässt:

$$\begin{aligned} \mathbf{z}^T \boldsymbol{\mu}_z &= n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{P} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} \\ &= n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\mathbf{I}_m - \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-\frac{1}{2}}}{C} \right) \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} \\ &= n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{C}. \end{aligned}$$

Mit der Darstellung von $\tilde{\mathbf{M}}^{-1}$ in (4.29) erhält man damit

$$\begin{aligned} \check{\mu}_p &= (\mu_*, 1) \tilde{\mathbf{M}}^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1})^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ &= (\mu_*, 1) \begin{pmatrix} \frac{n}{u} & -\frac{\sqrt{n}\sigma_g y}{u} \\ -\frac{\sqrt{n}\sigma_g y}{u} & \sigma_g^2 \left(1 + \frac{y^2}{u}\right) \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \end{pmatrix} \\ &= \left(\frac{n\mu_*}{u} - \frac{\sqrt{n}\sigma_g y}{u}, \sigma_g^2 \left(1 + \frac{y^2}{u}\right) - \frac{\sqrt{n}\sigma_g y \mu_*}{u} \right) \begin{pmatrix} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \end{pmatrix} \\ &= \frac{n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{u} \left(\mu_* - \frac{\sigma_g y}{\sqrt{n}} \right) + \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \left(\sigma_g^2 - \left(\mu_* - \frac{\sigma_g y}{\sqrt{n}} \right) \frac{\sqrt{n}\sigma_g y}{u} \right) \\ &= \mu_g + \left(\mu_* - \frac{\sigma_g y}{\sqrt{n}} \right) \left(\frac{n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{u} - \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \frac{\sqrt{n}\sigma_g y}{u} \right) \\ &= \mu_g + \frac{\sigma_g \tilde{y}}{\sqrt{n}} \left(\frac{n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{u} - \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \frac{\sqrt{n}\sigma_g \frac{\sqrt{n}\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}{\sqrt{C}}}{u} \right) \\ &= \mu_g + \frac{\sigma_g \tilde{y}}{\sqrt{n}} \left(\frac{\mathbf{z}^T \boldsymbol{\mu}_z + \frac{n\hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{C}}{u} - \frac{\frac{n\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}{C}}{u} \right) \\ &= \mu_g + \frac{\sigma_g \tilde{y}}{\sqrt{n}} \frac{\mathbf{z}^T \boldsymbol{\mu}_z}{u} = \mu_g + \frac{\psi \tilde{z}}{u} \sigma_g \tilde{y}. \end{aligned}$$

Dabei wurde ausgenutzt, dass $\boldsymbol{\Sigma}$ symmetrisch ist und $\sigma_g^2 = 1/C$ und $\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} / C = A/C = \mu_g$ gilt. Mit

$$\begin{aligned} (\mathbf{z}^T \boldsymbol{\mu}_z)^2 &= n^2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{2n^2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{C} \\ &\quad + \frac{n^2 \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}{C^2} \end{aligned}$$

gilt für \tilde{d}

$$\begin{aligned}
 \tilde{d} &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} \\
 &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1}) \tilde{\mathbf{M}}^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1})^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\
 &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - (\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}, \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}) \begin{pmatrix} \frac{n}{u} & -\frac{\sqrt{n}\sigma_g y}{u} \\ -\frac{\sqrt{n}\sigma_g y}{u} & \sigma_g^2 \left(1 + \frac{y^2}{u}\right) \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \end{pmatrix} \\
 &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\
 &\quad - \left(\frac{n \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}{u} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \frac{\sqrt{n}\sigma_g y}{u}, \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \sigma_g^2 \left(1 + \frac{y^2}{u}\right) - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \frac{\sqrt{n}\sigma_g y}{u} \right) \\
 &\quad \times \begin{pmatrix} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \end{pmatrix} \\
 &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{n \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{u} + \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \frac{\sqrt{n}\sigma_g y}{u} \\
 &\quad - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \sigma_g^2 \left(1 + \frac{y^2}{u}\right) + \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \frac{\sqrt{n}\sigma_g y}{u} \\
 &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{n \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{u} + 2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \frac{\sqrt{n}\sigma_g y}{u} \\
 &\quad - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \sigma_g^2 - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \sigma_g^2 \frac{y^2}{u} \\
 &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{nu} - \frac{2n \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{Cu} \\
 &\quad + \frac{n \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}{C^2 u} + 2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \frac{\sqrt{n}\sigma_g \sqrt{n} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}{\sqrt{Cu}} \\
 &\quad - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \sigma_g^2 - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \sigma_g^2 \frac{n \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}{Cu} \\
 &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{nu} - \frac{2n \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{Cu} \\
 &\quad + \frac{n \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}{C^2 u} + \frac{2n \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}}{Cu} \\
 &\quad - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \sigma_g^2 - \frac{n \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}{C^2 u} \\
 &= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{nu} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \sigma_g^2 \\
 &= \psi^2 - \frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{nu} = \psi^2 \left(1 - \frac{\tilde{z}^2}{u}\right) = \frac{\psi^2 \tilde{u}}{u},
 \end{aligned}$$

wobei ausgenutzt wurde, dass $\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \sigma_g^2 = B - A^2/C = \psi^2$ ist. Mit diesen beiden Ergebnissen erhält man schließlich die gewünschte Form von $\tilde{\mu}_p$. \square

Beweis von Satz 4.4.2

Mit der Zufallsvariablen $\tilde{z} = \mathbf{z}^T \boldsymbol{\mu}_z / (\sqrt{n}\psi)$ aus dem Beweis von Satz 4.4.1 gilt

$$\begin{aligned} E[\tilde{\mu}_p] &= E\left[\mu_g + \psi\sigma_g \frac{\tilde{z}}{u} \tilde{y} + \psi\check{\sigma}_p \left(\frac{\tilde{u}}{u}\right)^{\frac{1}{2}} \tilde{t}\right] = \mu_g + E\left[\frac{\mathbf{z}^T \boldsymbol{\mu}_z}{u} \cdot \frac{\sigma_g \tilde{y}}{\sqrt{n}}\right] + E\left[\psi\check{\sigma}_p \left(\frac{\tilde{u}}{u}\right)^{\frac{1}{2}} \tilde{t}\right] \\ &= \mu_g + E\left[\frac{\mathbf{z}^T \boldsymbol{\mu}_z}{u} \cdot \frac{\sigma_g \tilde{y}}{\sqrt{n}}\right], \end{aligned}$$

wobei die letzte Gleichheit gilt, weil \tilde{t} unabhängig von \tilde{u} , \tilde{y} und \tilde{z} ist und Erwartungswert 0 hat. Wegen $\tilde{y} = \sqrt{n}\mu_*/\sigma_g - y$, $\mu_y = \sqrt{n}\mu_g/\sigma_g$ und aufgrund der Unabhängigkeit von y und \mathbf{z} gilt weiter

$$\begin{aligned} E[\tilde{\mu}_p] &= \mu_g + E\left[\frac{\mathbf{z}^T \boldsymbol{\mu}_z}{u} \left(\mu_* - \frac{\sigma_g y}{\sqrt{n}}\right)\right] = \mu_g + E\left[\frac{\mathbf{z}^T \boldsymbol{\mu}_z}{u}\right] E\left[\mu_* - \frac{\sigma_g y}{\sqrt{n}}\right] \\ &= \mu_g + E\left[\frac{\mathbf{z}^T \boldsymbol{\mu}_z}{u}\right] (\mu_* - \mu_g) = \mu_g + \phi(\mu_* - \mu_g), \end{aligned}$$

wobei die letzte Gleichheit mit Lemma 4.3.4 folgt, aber nur genau dann gilt, wenn $m > 2$ ist. Durch Umstellen erhält man den gewünschten Ausdruck

$$E[\tilde{\mu}_p] = \mu_* - \mu_* + \phi\mu_* + \mu_g - \phi\mu_g = \mu_* - (1 - \phi)(\mu_* - \mu_g).$$

Für den Erwartungswert von $\check{\sigma}_g^2$ gilt

$$\begin{aligned} E[\check{\sigma}_g^2] &= E\left[\left(1 + \frac{\tilde{x}^2 + \tilde{q}}{\tilde{r}}\right) \check{\sigma}_p^2\right] = (1 + (E[\tilde{x}^2] + E[\tilde{q}]) E[\tilde{r}^{-1}]) E[\check{\sigma}_p^2] \\ &= \left(1 + (1 + m - 3) \cdot \frac{1}{n - m}\right) E[\check{\sigma}_p^2] = \left(\frac{n - m + m - 2}{n - m}\right) E[\check{\sigma}_p^2], \end{aligned}$$

da $\check{\sigma}_p^2$ als Funktion in \tilde{z} , \tilde{u} und \tilde{y} unabhängig von \tilde{x} , \tilde{q} und \tilde{r} ist. Für die Kovarianzmatrix soll in diesem Beweis lediglich die Varianz von $\tilde{\mu}_p$ bestimmt werden. Es gilt

$$\begin{aligned} Var(\tilde{\mu}_p) &= Var\left(\psi\sigma_g \frac{\tilde{z}}{u} \tilde{y} + \psi\check{\sigma}_p \left(\frac{\tilde{u}}{u}\right)^{\frac{1}{2}} \tilde{t}\right) \\ &= Var\left(\psi\sigma_g \frac{\tilde{z}}{u} \tilde{y}\right) + Var\left(\psi\check{\sigma}_p \left(\frac{\tilde{u}}{u}\right)^{\frac{1}{2}} \tilde{t}\right) + 2\psi^2\sigma_g Cov\left(\frac{\tilde{z}}{u} \tilde{y}, \check{\sigma}_p \left(\frac{\tilde{u}}{u}\right)^{\frac{1}{2}} \tilde{t}\right). \end{aligned}$$

Der erste Varianzterm lässt sich schreiben als

$$Var\left(\psi\sigma_g \frac{\tilde{z}}{u} \tilde{y}\right) = Var\left(\psi\sigma_g \tilde{y} \frac{\mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n}\psi u}\right) = Var\left(\frac{\sigma_g \tilde{y}}{\sqrt{n}} \frac{\mathbf{z}^T \boldsymbol{\mu}_z}{\mathbf{z}^T \mathbf{z}}\right)$$

und durch Umformen des zweiten Varianzterms erhält man für diesen

$$\begin{aligned}
 \text{Var} \left(\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \right) &= E \left[\text{Var} \left(\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \mid \tilde{t} \right) \right] + \text{Var} \left(E \left[\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \mid \tilde{t} \right] \right) \\
 &= E \left[\tilde{t}^2 \text{Var} \left(\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right) \right] + \text{Var} \left(\tilde{t} E \left[\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right] \right) \\
 &= \text{Var} \left(\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right) E [\tilde{t}^2] + \left(E \left[\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right] \right)^2 \text{Var} (\tilde{t}) \\
 &= \left(E \left[\psi^2 \check{\sigma}_p^2 \frac{\tilde{u}}{u} \right] - \left(E \left[\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right] \right)^2 \right) E [\tilde{t}^2] \\
 &\quad + \left(E \left[\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right] \right)^2 \left(E [\tilde{t}^2] - (E [\tilde{t}])^2 \right) \\
 &= E \left[\psi^2 \check{\sigma}_p^2 \frac{\tilde{u}}{u} \right] E [\tilde{t}^2] - \left(E \left[\psi \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right] \right)^2 (E [\tilde{t}])^2 \\
 &= \frac{1}{n-m} E \left[\psi^2 \check{\sigma}_p^2 \frac{\tilde{u}}{u} \right],
 \end{aligned}$$

wobei die letzte Gleichheit daraus folgt, dass zum einen der Erwartungswert von \tilde{t} gleich null ist und zum anderen

$$\tilde{t}^2 = \frac{\tilde{x}^2}{\tilde{r}} = \frac{1}{n-m+2} \frac{\tilde{x}^2}{\frac{\tilde{r}}{n-m+2}}$$

wegen $\tilde{x}^2 \sim \chi_1^2$ und $\tilde{r} \sim \chi_{n-m+2}^2$ F -verteilt ist und somit den Erwartungswert $1/(n-m)$ besitzt. Der Kovarianzterm fällt weg, da

$$\begin{aligned}
 \text{Cov} \left(\frac{\tilde{z}}{u} \tilde{y}, \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \right) &= E \left[\text{Cov} \left(\frac{\tilde{z}}{u} \tilde{y}, \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \mid \tilde{t} \right) \right] \\
 &\quad + \text{Cov} \left(E \left[\frac{\tilde{z}}{u} \tilde{y} \mid \tilde{t} \right], E \left[\check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \mid \tilde{t} \right] \right) \\
 &= E [\tilde{t}] \text{Cov} \left(\frac{\tilde{z}}{u} \tilde{y}, \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right) + E \left[E \left[\frac{\tilde{z}}{u} \tilde{y} \mid \tilde{t} \right] E \left[\check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \mid \tilde{t} \right] \right] \\
 &\quad - E \left[E \left[\frac{\tilde{z}}{u} \tilde{y} \mid \tilde{t} \right] \right] E \left[E \left[\check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \mid \tilde{t} \right] \right] \\
 &= E \left[\frac{\tilde{z}}{u} \tilde{y} \right] E \left[\check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right] E [\tilde{t}] - E \left[\frac{\tilde{z}}{u} \tilde{y} \right] E \left[\check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \right] E [\tilde{t}] \\
 &= 0
 \end{aligned}$$

ist. Somit folgt für die Varianz von $\tilde{\mu}_p$

$$\begin{aligned}
 \text{Var}(\tilde{\mu}_p) &= \text{Var}\left(\frac{\sigma_g \tilde{y} \mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n} \mathbf{z}^T \mathbf{z}}\right) + \frac{1}{n-m} E\left[\psi^2 \check{\sigma}_p^2 \frac{\tilde{u}}{u}\right] \\
 &= E\left[\text{Var}\left(\frac{\sigma_g \tilde{y} \mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n} \mathbf{z}^T \mathbf{z}} \mid \tilde{y}\right)\right] + \text{Var}\left(E\left[\frac{\sigma_g \tilde{y} \mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n} \mathbf{z}^T \mathbf{z}} \mid \tilde{y}\right]\right) + \frac{E\left[\psi^2 \check{\sigma}_p^2 \frac{u - \tilde{z}^2}{u}\right]}{n-m} \\
 &= E[\tilde{y}^2] \text{Var}\left(\frac{\sigma_g \mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n} \mathbf{z}^T \mathbf{z}}\right) + \text{Var}(\tilde{y}) \left(E\left[\frac{\sigma_g \mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n} \mathbf{z}^T \mathbf{z}}\right]\right)^2 + \frac{E\left[\psi^2 \check{\sigma}_p^2 \left(1 - \frac{\tilde{z}^2}{u}\right)\right]}{n-m} \\
 &= \left(n \cdot \frac{(\mu_* - \mu_g)^2}{\sigma_g^2} + 1\right) \text{Var}\left(\frac{\sigma_g \mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n} \mathbf{z}^T \mathbf{z}}\right) + \left(E\left[\frac{\sigma_g \mathbf{z}^T \boldsymbol{\mu}_z}{\sqrt{n} \mathbf{z}^T \mathbf{z}}\right]\right)^2 \\
 &\quad + \frac{\psi^2 E[\check{\sigma}_p^2] - E\left[\frac{\check{\sigma}_p^2 (\mathbf{z}^T \boldsymbol{\mu}_z)^2}{n \mathbf{z}^T \mathbf{z}}\right]}{n-m} \\
 &= (\mu_* - \mu_g)^2 \text{Var}\left(\frac{\mathbf{z}^T \boldsymbol{\mu}_z}{\mathbf{z}^T \mathbf{z}}\right) + E\left[\frac{\sigma_g^2 (\mathbf{z}^T \boldsymbol{\mu}_z)^2}{n (\mathbf{z}^T \mathbf{z})^2}\right] \\
 &\quad + \frac{\psi^2 E[\check{\sigma}_p^2] - E\left[\frac{\sigma_g^2 \left(1 + \frac{\tilde{y}^2}{u}\right) (\mathbf{z}^T \boldsymbol{\mu}_z)^2}{n \mathbf{z}^T \mathbf{z}}\right]}{n-m} \\
 &= (\mu_* - \mu_g)^2 \left(E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{(\mathbf{z}^T \mathbf{z})^2}\right] - \left(E\left[\frac{\mathbf{z}^T \boldsymbol{\mu}_z}{\mathbf{z}^T \mathbf{z}}\right]\right)^2\right) + \frac{\sigma_g^2}{n} E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{(\mathbf{z}^T \mathbf{z})^2}\right] \\
 &\quad + \frac{\psi^2 E[\check{\sigma}_p^2] - \frac{\sigma_g^2}{n} \left(E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{\mathbf{z}^T \mathbf{z}}\right] + h E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{(\mathbf{z}^T \mathbf{z})^2}\right]\right)}{n-m} \\
 &= \left((\mu_* - \mu_g)^2 + \frac{\sigma_g^2}{n}\right) E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{(\mathbf{z}^T \mathbf{z})^2}\right] - (\mu_* - \mu_g)^2 \phi^2 \\
 &\quad + \frac{\psi^2 E[\check{\sigma}_p^2] - \frac{\sigma_g^2}{n} \left(E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{\mathbf{z}^T \mathbf{z}}\right] + h E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{(\mathbf{z}^T \mathbf{z})^2}\right]\right)}{n-m} \\
 &= \frac{\sigma_g^2 h}{n} E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{(\mathbf{z}^T \mathbf{z})^2}\right] - (\mu_* - \mu_g)^2 \phi^2 \\
 &\quad + \frac{\psi^2 E[\check{\sigma}_p^2] - \frac{\sigma_g^2}{n} \left(E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{\mathbf{z}^T \mathbf{z}}\right] + h E\left[\frac{(\mathbf{z}^T \boldsymbol{\mu}_z)^2}{(\mathbf{z}^T \mathbf{z})^2}\right]\right)}{n-m},
 \end{aligned}$$

wobei im vorletzten Schritt (4.15) und im letzten Schritt die Umformung

$$(\mu_* - \mu_g)^2 + \frac{\sigma_g^2}{n} = \frac{\sigma_g^2}{n} \left(1 + n \frac{(\mu_* - \mu_g)^2}{\sigma_g^2}\right) = \frac{\sigma_g^2}{n} h$$

ausgenutzt wurde. Da $E[(\mathbf{z}^T \boldsymbol{\mu}_z)^2 / (\mathbf{z}^T \mathbf{z})]$ und $E[(\mathbf{z}^T \boldsymbol{\mu}_z)^2 / (\mathbf{z}^T \mathbf{z})^2]$ genau dann existieren, wenn $m > 1$ bzw. $m > 3$ ist, existiert auch $\text{Var}(\tilde{\mu}_p)$ genau dann, wenn $m > 3$ ist.

Die Varianz von $\tilde{\sigma}_p^2$ und die Kovarianzen lassen sich ebenfalls mithilfe der Zerlegungssätze für Varianz und Kovarianz bestimmen. \square

Beweis von Satz 4.4.3

Nach Satz 4.1.1 gilt $E[\hat{\mu}_g | \hat{\psi}^2] = \mu_g$. Im Beweis zu Satz 4.1.1 wurde gezeigt, dass $\hat{\psi}^2 = u/v$ ist, wobei $v \sim \chi_{n-m+1}^2$ unabhängig von $u = \mathbf{z}^T \mathbf{z}$ ist. Mit Lemma 4.3.4 gilt somit

$$E[\hat{\psi}^{-2}] = E[v] E[u^{-1}] = (n - m + 1) \frac{1 - \phi}{m - 3}.$$

Also ist

$$\begin{aligned} E[\hat{\mu}_{pu}] &= E\left[\mu_* - \frac{m-3}{(n-m+1)\hat{\psi}^2}(\mu_* - \hat{\mu}_g)\right] \\ &= E\left[E\left[\mu_* - \frac{m-3}{(n-m+1)\hat{\psi}^2}(\mu_* - \hat{\mu}_g) \mid \hat{\psi}^2\right]\right] \\ &= \mu_* - E\left[\frac{m-3}{n-m+1}\hat{\psi}^{-2}E[\mu_* - \hat{\mu}_g \mid \hat{\psi}^2]\right] \\ &= \mu_* - E\left[\frac{m-3}{n-m+1}\hat{\psi}^{-2}(\mu_* - \mu_g)\right] \\ &= \mu_* - \frac{m-3}{n-m+1} \cdot \frac{n-m+1}{m-3} (1-\phi)(\mu_* - \mu_g) \\ &= \mu_* - (1-\phi)(\mu_* - \mu_g) = \underline{\mu}_p. \end{aligned}$$

Mit (4.19) gilt

$$\begin{aligned} E[\hat{\sigma}_{pu}^2] &= \frac{(n-2)n}{(n-m)(n-m+1)} E[\hat{\sigma}_p^2] \\ &= \frac{(n-2)n}{(n-m)(n-m+1)} \cdot \frac{n-m+1}{n} E[\check{\sigma}_p^2] \\ &= \frac{n-2}{n-m} E[\check{\sigma}_p^2] = \underline{\sigma}_p^2. \end{aligned}$$

\square

5. Simulationsstudie

Die Simulationsstudie setzt sich aus drei Abschnitten zusammen. Zunächst wird die Auswirkung einer Leerverkaufsbeschränkung auf die Effizienzkurve untersucht. Als nächstes werden die geschätzten Effizienzkurven, die man erhält, wenn man das Mean-Variance-Optimierungsproblem mit den verschiedenen in Kapitel 3 vorgestellten Schätzern löst, miteinander verglichen. Anschließend werden die in Kapitel 4 hergeleiteten Ergebnisse für die Verteilung der empirischen Minimum-Varianz-Kurve empirisch überprüft.

5.1. Die Effizienzkurve mit und ohne Leerverkaufsrestriktion

Sind Leerverkäufe nicht zugelassen, ist die Konstruktion von Portfolios nur eingeschränkt möglich. Daher stimmen die Effizienzkurven im restringierten Fall (ohne Leerverkäufe) und im unrestringierten Fall (mit Leerverkäufen) im Allgemeinen nicht überein. Der zulässige Bereich des restringierten Mean-Variance-Optimierungsproblems (in ihm sind alle Portfoliogewichte enthalten, die die Nebenbedingungen des Optimierungsproblems erfüllen) ist eine Teilmenge des zulässigen Bereichs des unrestringierten Problems. Ohne die Nichtnegativitätsbeschränkung der Portfoliogewichte lassen sich also alle Portfolios konstruieren, die man auch mit der Restriktion bilden kann. Darüber hinaus besteht jedoch die Möglichkeit, Portfolios zu finden, welche die effizienten Portfolios aus dem restringierten Fall dominieren, die also zu einem bestimmten Risiko eine höhere Rendite erwarten lassen bzw. eine bestimmte erwartete Rendite unter einem geringeren Risiko liefern. Somit liegt die restringierte Effizienzkurve niemals oberhalb der unrestringierten Kurve.

Für die Untersuchung dieser Überlegungen werden exemplarisch Wertpapiere von drei im DAX notierten Unternehmen betrachtet, und zwar von BASF, Bayer und Merck. Die aus ihren vergangenen Monatsrenditen¹ von 1/2003-12/2012 berechneten empirischen Mittelwerte und ihre empirische Kovarianzmatrix werden als wahre Parameter $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ angenommen. Es ist

$$\boldsymbol{\mu} = (-0,01874850 \quad -0,01355396 \quad -0,01181897) \quad (5.1)$$

und

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0,008625638 & 0,002928710 & 0,001277821 \\ 0,002928710 & 0,006187523 & 0,002490726 \\ 0,001277821 & 0,002490726 & 0,005410025 \end{pmatrix}. \quad (5.2)$$

Dabei betreffen Komponenten mit (Zeilen-/Spalten-) Index 1 die Renditen von BASF und Komponenten mit Index 2 bzw. 3 beziehen sich auf die Renditen von Bayer bzw. Merck.

¹log>Returns, die aus um Dividenden und Splits bereinigten Schlusspreisen bestimmt wurden, Quelle: <http://de.finance.yahoo.com/>

Zum Beispiel entspricht der Eintrag in der ersten Zeile und der dritten Spalte von Σ der Kovarianz zwischen den Renditen von BASF und Merck. Mit $\boldsymbol{\mu}$ und Σ lassen sich die Effizienzmengenkennkonstanten (ψ, μ_g, σ_g) berechnen. Es gilt

$$\psi = 0,06634181 \quad \mu_g = -0,0140413 \quad \sigma_g = 0,05986762.$$

Mit ihnen lässt sich die unrestringierte Effizienzkurve bestimmen, indem man für nahe beieinanderliegende Werte für μ_* das Risiko $\sigma_p = \sqrt{\sigma_g^2 + (\mu_* - \mu_g)^2/\psi^2}$ berechnet und in ein (μ, σ) -Diagramm abträgt. Die daraus resultierende Kurve ist in Abbildung 5.1 als durchgezogene Linie dargestellt. Da eine Charakterisierung der restringierten Effizienzkur-

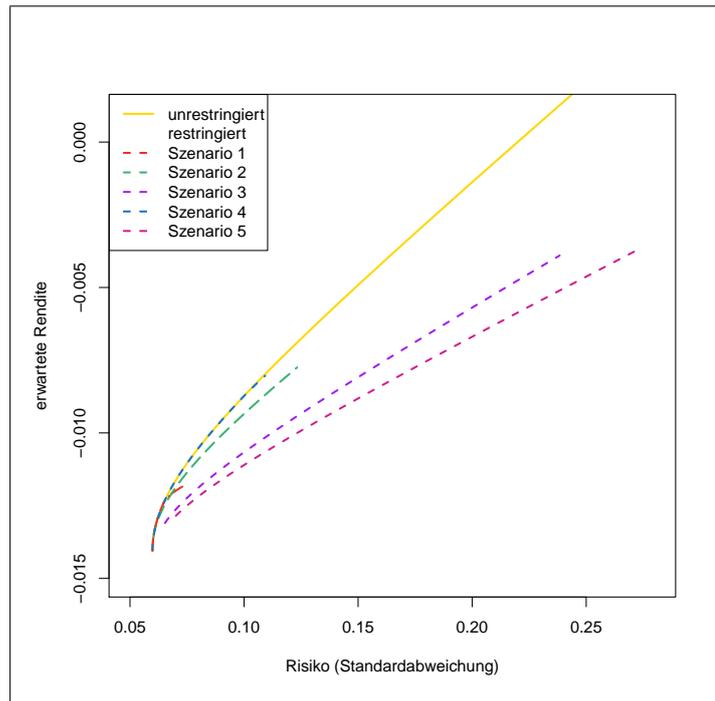


Abbildung 5.1.: Vergleich von Effizienzkurven mit und ohne Leerverkaufsrestriktion

ve durch die Effizienzmengenkennkonstanten nicht möglich ist, werden in diesem Fall die Portfoliogewichte \mathbf{w}^* in Abhängigkeit von μ_* numerisch ermittelt und mit ihnen das Risiko $\sigma_p = \sqrt{\mathbf{w}^{*T} \Sigma \mathbf{w}^*}$ bestimmt. Die Kurve, die man damit erhält, ist in Abbildung 5.1 als rote gestrichelte Linie dargestellt. Hierbei handelt es sich jedoch nur um eines von vielen möglichen Szenarien, wie die restringierte Effizienzkurve im Vergleich zur unrestringierten Kurve verlaufen kann. Vier weitere $(\boldsymbol{\mu}, \Sigma)$ -Kombinationen, die $\psi = 0,06634181$, $\mu_g = -0,0140413$ und $\sigma_g = 0,05986762$ liefern, sind in Tabelle 5.1 zusammengefasst. Die aus ihnen resultierenden restringierten Effizienzkurven sind ebenfalls in Abbildung 5.1 eingezeichnet. Die Abbildung bestätigt, dass die restringierten Kurven niemals oberhalb der unrestringierten Kurve liegen. Es fällt auf, dass sich die Kurven aus Szenario 3 und Szenario 5 wesentlich von den anderen Kurven unterscheiden: Die erwarteten Renditen der globalen varianzmini-

Szenario 2			
$\boldsymbol{\mu}$	-0,014640000	-0,007747502	-0,013585287
$\boldsymbol{\Sigma}$	0,014186923	0,007966406	0,003647577
	0,007966406	0,015192438	0,004298708
	0,003647577	0,004298708	0,003632876
Szenario 3			
$\boldsymbol{\mu}$	-0,011017000	-0,003900419	-0,013096469
$\boldsymbol{\Sigma}$	0,060136497	0,050894199	0,009345391
	0,050894199	0,056825493	0,009546907
	0,009345391	0,009546907	0,004266689
Szenario 4			
$\boldsymbol{\mu}$	-0,014520000	-0,016036762	-0,008034997
$\boldsymbol{\Sigma}$	0,009832422	0,001643740	0,001953354
	0,001643740	0,005240029	0,001201304
	0,001953354	0,001201304	0,011935059
Szenario 5			
$\boldsymbol{\mu}$	-0,010587000	-0,003617474	-0,012847038
$\boldsymbol{\Sigma}$	0,08029076	0,07085229	0,013150610
	0,07085229	0,07545406	0,013303181
	0,01315061	0,01330318	0,004914501

Tabelle 5.1.: Weitere $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ - Kombinationen, die für die Effizienzmengenkennkonstanten $\psi = 0,06634181$, $\mu_g = -0,0140413$ und $\sigma_g = 0,05986762$ liefern.

malen Portfolios sind deutlich höher als in den anderen Szenarien, die Kurven sind auf der Risiko-Achse nach rechts verschoben und insgesamt länger als die anderen restringierten Kurven. Bezeichnet $\boldsymbol{\mu}_i$, $i = 1, \dots, 5$, den Vektor mit den erwarteten Renditen in Szenario i , so gilt $\boldsymbol{\mu}_5 > \boldsymbol{\mu}_3 > \boldsymbol{\mu}_2$ (komponentenweise), was die höheren erwarteten Renditen der globalen varianzminimalen Portfolios sowie die Länge der Kurven erklärt. Denn mit Assets, deren erwarteten Renditen höher als bei anderen Assets sind, lassen sich Portfolios konstruieren, mit denen höhere erwartete Renditen erzielt werden können. Die Verschiebung in Richtung der Risiko-Achse folgt aus dem höheren Risiko der Renditen in den beiden Szenarien. Denn auch hier gilt $\boldsymbol{\Sigma}_5 > \boldsymbol{\Sigma}_3 > \boldsymbol{\Sigma}_2$, wobei $\boldsymbol{\Sigma}_i$ die Kovarianzmatrix der Renditen in Szenario i bezeichne und der Vergleich hierbei ebenfalls komponentenweise erfolgt. Weiterhin lässt sich erkennen, dass die Kurven von Szenario 1, Szenario 2 und Szenario 4 nahe beieinander liegen. Vergleicht man die mittleren quadratischen Abweichungen der Parameter aus den Szenarien 2 bis 5 von den Parametern aus Szenario 1, also

$$\frac{1}{3} \sum_{j=1}^3 (\mu_{ij} - \mu_{1j})^2 \quad \text{und} \quad \frac{1}{9} \sum_{j,k=1}^3 (\sigma_{ijk} - \sigma_{1jk})^2,$$

wobei μ_{ij} die erwartete Rendite von Asset j und σ_{ijk} die Kovarianz zwischen Asset j und Asset k in Szenario i sind, so erkennt man, dass die Abweichungen in Szenario 2 und Szenario 4 kleiner als in den beiden anderen Szenarien sind; siehe Tabelle 5.2. Da sich die Parameter in Szenario 1, Szenario 2 und Szenario 4 somit untereinander nur wenig unterscheiden, gibt es auch nur geringe Differenzen zwischen den Kurven. Dass die Kurven in diesen Szenarien sogar zum Teil auf der unrestringierten Kurve liegen, lässt sich damit begründen, dass

	Szenario 2	Szenario 3	Szenario 4	Szenario 5
Abweichung $\boldsymbol{\mu}$	1,790486e-05	5,153299e-05	1,278768e-05	5,546692e-05
Abweichung $\boldsymbol{\Sigma}$	2,041117e-05	0,001116667	5,830038e-06	0.002186329

Tabelle 5.2.: Mittlere quadratische Abweichungen der Parameter der Szenarien 2 bis 5 von den Parametern aus Szenario 1.

die zulässigen Bereiche der restringierten Probleme für kleine Werte von μ_* (ungefähr) mit dem zulässigen Bereich des unrestringierten Problems übereinstimmen. Sobald die Bereiche mit zunehmenden μ_* voneinander abweichen, neigen sich die restringierten Kurven von der unrestringierten Kurve weg.

5.2. Vergleich der Schätzmethode zur Approximation der wahren Effizienzkurve

In Kapitel 3 wurden verschiedene Methoden vorgestellt, mit denen $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ und \mathbf{w}^* geschätzt werden können. Im Folgenden soll untersucht werden, wie gut sich die wahre Effizienzkurve mithilfe empirischer Schätzer, der *resampling*-Methode und mit Bayes-Schätzern im Schnitt approximieren lässt. Da es in der Praxis von Bedeutung ist, wie die geschätzten effizienten Portfolios unter den tatsächlichen Bedingungen abschneiden, werden die Kurven mithilfe der *out-of-sample* Parameter $\tilde{\boldsymbol{\mu}} = \widehat{\mathbf{w}}^{*T} \boldsymbol{\mu}$ und $\tilde{\sigma}^2 = \widehat{\mathbf{w}}^{*T} \boldsymbol{\Sigma} \widehat{\mathbf{w}}^*$ bestimmt, wobei $\widehat{\mathbf{w}}^*$ die mit der jeweiligen Schätzmethode ermittelten Portfoliogewichte sind. Es wird sowohl der Fall, dass Leerverkäufe erlaubt sind, als auch der Fall, dass Leerverkäufe unzulässig sind, betrachtet.

Es werden dieselben Daten zugrunde gelegt, die Lai, Xing und Chen (2011) in ihrer Simulationsstudie verwenden. Somit sei $m = 4$, $n = 6$ und

$$\boldsymbol{\mu} = (2,42 \quad 1,88 \quad 1,58 \quad 3,47),$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1,17 & 0,79 & 0,84 & 1,61 \\ 0,79 & 0,82 & 0,61 & 1,23 \\ 0,84 & 0,61 & 1,37 & 1,35 \\ 1,61 & 1,23 & 1,35 & 2,86 \end{pmatrix}.$$

Die mittlere empirische Effizienzkurve erhält man im unrestringierten Fall, indem man die Erwartungswerte von $\tilde{\boldsymbol{\mu}}$ und $\tilde{\sigma}^2$ wie in (4.22) und (4.23) bestimmt und $\boldsymbol{\mu} = E[\tilde{\boldsymbol{\mu}}]$ gegen $\sigma = \sqrt{E[\tilde{\sigma}^2]}$ plottet. Da im restringierten Fall die Lösung des Optimierungsproblems nicht explizit angegeben werden kann, sondern numerisch angenähert werden muss, sind hier die Verteilungen der *out-of-sample* Parameter und somit insbesondere ihre Erwartungswerte unbekannt. Sie werden deshalb empirisch geschätzt. Dazu wird eine Stichprobe $\{\mathbf{r}_1, \dots, \mathbf{r}_n\}$ mit $\mathbf{r}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \dots, n$, generiert und mit ihrem empirischen Mittelwert und ihrer empirischen Kovarianzmatrix das Mean-Variance-Optimierungsproblem für verschiedene Werte für μ_* gelöst. Mit der Lösung $\widehat{\mathbf{w}}^*$ werden die *out-of-sample* Parameter $\tilde{\boldsymbol{\mu}} = \widehat{\mathbf{w}}^{*T} \boldsymbol{\mu}$ und $\tilde{\sigma} = \sqrt{\widehat{\mathbf{w}}^{*T} \boldsymbol{\Sigma} \widehat{\mathbf{w}}^*}$ bestimmt. Wiederholt man dieses Verfahren 500 Mal, bildet anschließend die Mittelwerte von $\tilde{\boldsymbol{\mu}}$ und $\tilde{\sigma}$ zu jeweils demselben Wert von μ_* und trägt diese

gegeneinander ab, erhält man eine Annäherung der mittleren empirischen Effizienzkurve im Fall ohne Leerverkäufe.

Für die Bestimmung der mittleren *resampled frontier* wird eine Stichprobe $\{\mathbf{r}_1, \dots, \mathbf{r}_n\}$ mit $\mathbf{r}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1 \dots, n$, generiert und ihr empirischer Mittelwert $\hat{\boldsymbol{\mu}}$ und ihre empirische Kovarianzmatrix $\hat{\boldsymbol{\Sigma}}$ bestimmt. Damit werden $B = 500$ Bootstrap-Stichproben der Größe n aus einer multivariaten $\mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ -Verteilung generiert. Für jede dieser Bootstrap-Stichproben wird der empirische Mittelwert und die empirische Kovarianzmatrix bestimmt und mit diesen Parametern das Mean-Variance-Optimierungsproblem für verschiedene Werte für μ_* gelöst, entweder numerisch oder wie in (4.4), je nachdem ob Leerverkäufe unzulässig oder erlaubt sind. Anschließend bestimmt man für jeden Wert von μ_* den Mittelwert $\bar{\mathbf{w}}$ der zugehörigen Gewichte der einzelnen Assets aus den verschiedenen Bootstrap-Stichproben. Damit lässt sich das *out-of-sample* Risiko und die *out-of-sample* erwartete Rendite mittels $\tilde{\sigma} = \sqrt{\bar{\mathbf{w}}^T \boldsymbol{\Sigma} \bar{\mathbf{w}}}$ und $\tilde{\mu} = \bar{\mathbf{w}}^T \boldsymbol{\mu}$ berechnen. Wiederholt man dieses Vorgehen 500 Mal und bildet den Durchschnitt der dabei resultierenden Kurven, erhält man eine Annäherung der mittleren *resampled frontier*.

Zur Berechnung der Bayes-Schätzer von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ müssen zunächst die Hyperparameter der A-priori-Verteilung spezifiziert werden. Wie bei Lai, Xing und Chen (2011) sei $\kappa = 5$, $\nu_0 = 10$,

$$\boldsymbol{\mu}_0 = (2,48 \quad 2,17 \quad 1,61 \quad 3,42)$$

und

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 3,37 & 2,04 & 0,32 & 1,59 \\ 2,04 & 4,22 & -0,05 & 3,02 \\ 0,32 & -0,05 & 2,75 & 1,08 \\ 1,59 & 3,02 & 1,08 & 8,43 \end{pmatrix}.$$

Generiert man eine Stichprobe $\{\mathbf{r}_1, \dots, \mathbf{r}_n\}$ mit $\mathbf{r}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1 \dots, n$, können die Bayes-Schätzer von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ wie in (3.9) und (3.10) bestimmt werden. Mit diesen Schätzern wird nun das Mean-Variance-Optimierungsproblem für verschiedene Werte für μ_* gelöst, entweder numerisch oder wie in (4.4). Mit der Lösung $\hat{\mathbf{w}}^*$ werden die *out-of-sample* Parameter $\tilde{\mu}$ und $\tilde{\sigma}$ bestimmt. Bildet man die Mittelwerte dieser Parameter aus 500 Wiederholungen dieses Verfahrens, erhält man eine Annäherung der mittleren Bayes'schen Kurve.

Die Ergebnisse sind in Abbildung 5.2 dargestellt. Es ist zu erkennen, dass die geschätzten Kurven unter der Annahme, dass Leerverkäufe erlaubt sind, viel stärker von der wahren Effizienzkurve abweichen als im restringierten Fall. Im unrestringierten Fall weisen die geschätzten Kurven auch untereinander deutliche Differenzen auf, während sie sich im restringierten Fall kaum unterscheiden. Mit den Bayes-Schätzern kann man in beiden Fällen für hohe erwartete Zielrenditen die wahre Effizienzkurve im Mittel am besten annähern. Die *resampled frontier* ist im unrestringierten Fall der empirischen Effizienzkurve vorzuziehen, im restringierten Fall stimmen die beiden Kurven jedoch beinahe überein. Es lässt sich schließlich feststellen, dass alle drei Methoden für die Approximation der Effizienzkurve im Schnitt nur ungenügend geeignet sind.

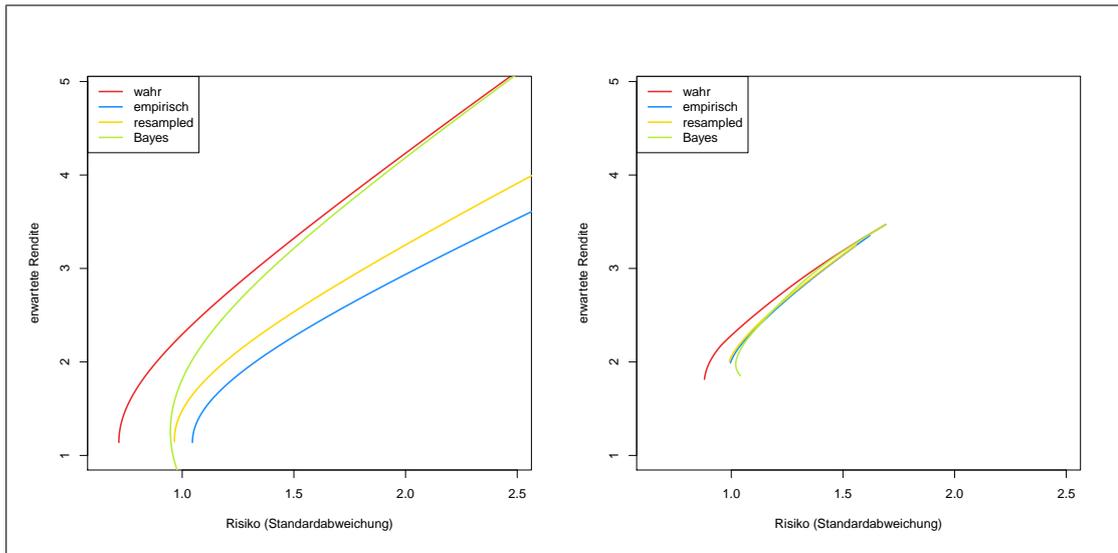


Abbildung 5.2.: Vergleich der Schätzmethode zur Approximation der wahren Effizienzkurve, links mit Leerverkäufen, rechts ohne Leerverkäufe.

5.3. Die empirische Minimum-Varianz-Kurve

Obwohl die empirischen Schätzer hinsichtlich der Annäherung der wahren Minimum-Varianz-Kurve nicht sehr gut abschneiden, wird in der Praxis dennoch häufig auf sie zurückgegriffen. Deswegen beschäftigte sich Kapitel 4 intensiv mit der Verteilung der empirischen Minimum-Varianz-Kurve. Dabei wurde zwischen vier Kurven unterschieden: Der wahren Minimum-Varianz-Kurve, der empirischen Minimum-Varianz-Kurve, der mittleren empirischen Minimum-Varianz-Kurve (kurz *in-sample* Kurve) und der *out-of-sample* Kurve. Diese wurden in Kapitel 4.2 vorgestellt und im weiteren Verlauf in Hinblick auf ihre Verteilungen analysiert.

Für die empirische Untersuchung der Resultate werden im Folgenden jeweils drei Simulationen durchgeführt: Um die Ergebnisse auf verschiedene Daten anzuwenden, werden Wertpapiere von je 14 im NASDAQ-100 und im DAX gelisteten Unternehmen betrachtet. Eine mögliche Auswirkung der Größe der Daten soll in einem dritten Durchlauf mit 25 Wertpapieren von im DAX notierten Unternehmen untersucht werden. Aus den um Dividenden und Splits bereinigten monatlichen Schlusspreisen der Wertpapiere von 1/2003-12/2012 werden die log>Returns berechnet. Ihre empirischen Mittelwerte und ihre empirische Kovarianzmatrix werden als wahre Parameter $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ angenommen. In Tabelle 5.3, die sich am Ende dieses Kapitels befindet, sind die für die Simulation verwendeten Wertpapiere sowie die Werte der Parameter für $m = 14$ zusammengefasst. Auf die Daten für den Fall $m = 25$ wird aus Platzgründen verzichtet.

Die unrestringierte *in-sample* Kurve erhält man, indem man $\bar{\sigma}_p^2 = E[\hat{\sigma}_p^2]$ wie in (4.19) für verschiedene Werte für μ_* bestimmt und $\bar{\sigma}_p = \sqrt{E[\hat{\sigma}_p^2]}$ in ein (μ, σ) -Diagramm einträgt.

Sind Leerverkäufe dagegen unzulässig, wird eine Stichprobe $\{\mathbf{r}_1, \dots, \mathbf{r}_n\}$ mit $\mathbf{r}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \dots, n$, generiert und mit ihrem empirischen Mittelwert $\boldsymbol{\mu}$ und ihrer empirischen Kovarianzmatrix $\widehat{\boldsymbol{\Sigma}}$ das Minimum-Varianz-Problem für verschiedene Werte für μ_* numerisch gelöst. Mit der Lösung $\widehat{\mathbf{w}}^*$ wird das Risiko mittels $\widehat{\sigma}_p = \sqrt{\widehat{\mathbf{w}}^{*T} \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{w}}^*}$ bestimmt. Wiederholt man dieses Vorgehen 500 Mal, berechnet für jeden Wert von μ_* den Mittelwert der zugehörigen Risiken und plottet μ_* gegen dieses mittlere Risiko $\bar{\sigma}_p$, so erhält man eine Annäherung der *in-sample* Kurve im restringierten Fall. In den Abbildungen 5.3 - 5.5 sind die Ergebnisse der Simulationen dargestellt. Neben der wahren Minimum-Varianz-Kurve und der *in-sample* Kurve sind außerdem das 5%- und das 95%-Quantil der empirischen Minimum-Varianz-Kurve eingezeichnet, um deren Streuung zu veranschaulichen. Für die Quantile werden 50000 empirische Minimum-Varianz-Kurven generiert, entweder wie in Satz 4.3.2 oder numerisch mit den empirischen Parametern von $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -verteilten Stichproben, und anschließend für jeden Wert von μ_* die jeweils 5% niedrigsten und höchsten Werte für $\widehat{\sigma}_p$ ermittelt.

Aus den Abbildungen geht hervor, dass unabhängig von m , n und der Leerverkaufsbeschränkung die *in-sample* Kurve im Mittel links von der wahren Minimum-Varianz-Kurve liegt. Dies resultiert daraus, dass ein Asset, dessen Risiko zu optimistisch geschätzt wird, durch das Minimum-Varianz-Problem ein zu hohes Portfoliogewicht zugewiesen bekommt. Dadurch ist das *in-sample* Risiko häufig niedriger als das tatsächliche Risiko, mit dem eine erwartete Zielrendite verbunden ist. Weiterhin fällt auf, dass die geschätzte Kurve im Bereich des globalen varianzminimalen Portfolios die wahre Minimum-Varianz-Kurve relativ gut approximiert, die Differenz zwischen den beiden Kurven jedoch umso größer wird, je weiter man sich vom varianzminimalen Portfolio entfernt. Dabei nimmt auch die Streuung der empirischen Minimum-Varianz-Kurve stark zu. Mit zunehmendem n wird zwar die Approximation besser und die Streuung nimmt ab, jedoch eignet sich die empirische Minimum-Varianz-Kurve vor allem im Fall ohne Leerverkäufe auch für große n nur ungenügend für die Annäherung der wahren Kurve. Vergleicht man die Ergebnisse für $m = 14$ mit den Ergebnissen für den Fall $m = 25$, so erkennt man, dass eine Erhöhung der Anzahl der Assets im unrestringierten Fall zu einer niedrigeren Streuung der empirischen Minimum-Varianz-Kurve führt, die Approximation jedoch schlechter wird.

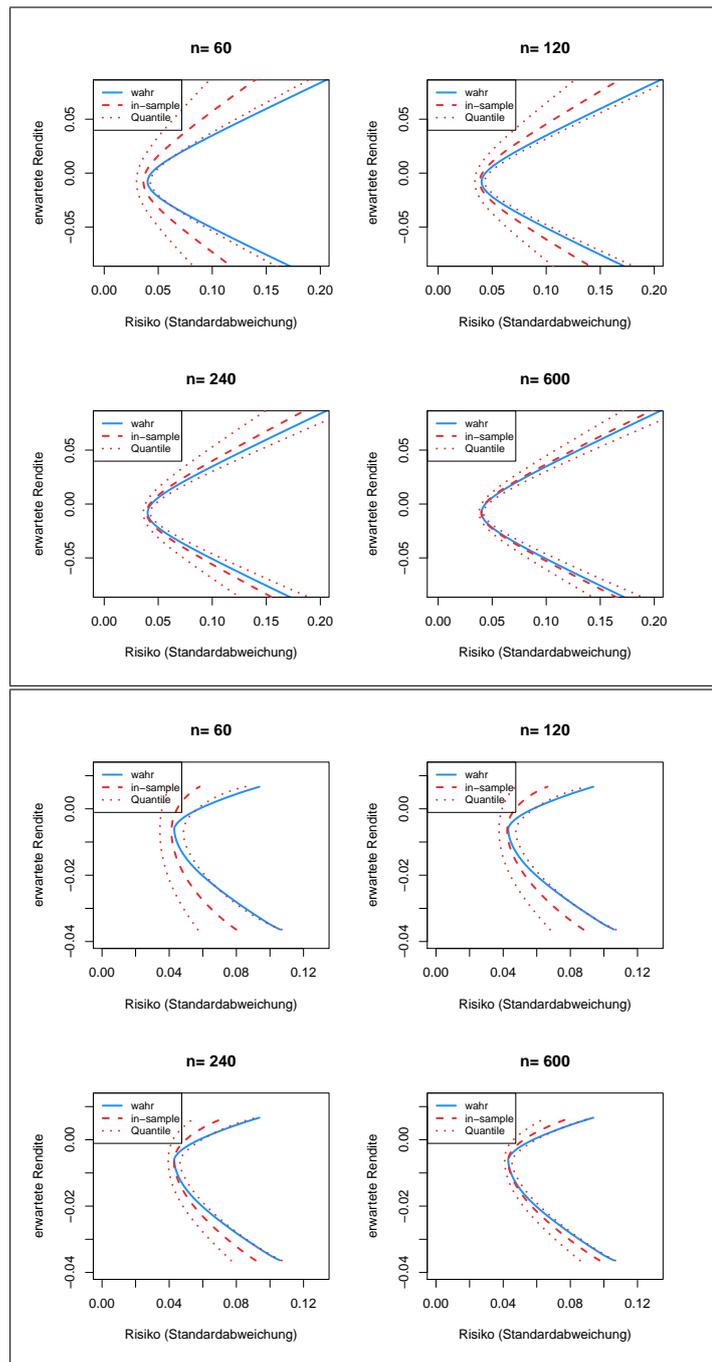


Abbildung 5.3.: *In-sample* Kurve für $m = 14$ Wertpapiere aus dem NASDAQ-100 für verschiedene n , oben mit Leerverkäufen, unten ohne Leerverkäufe. Die durchgezogene Linie stellt die wahre Minimum-Varianz-Kurve dar, bei der gestrichelten Linie handelt es sich um $\bar{\sigma}_p$. Die gepunkteten Linien sind die 5%- und 95%-Quantile der empirischen Minimum-Varianz-Kurve, die basierend auf 50000 Simulationen bestimmt wurden.

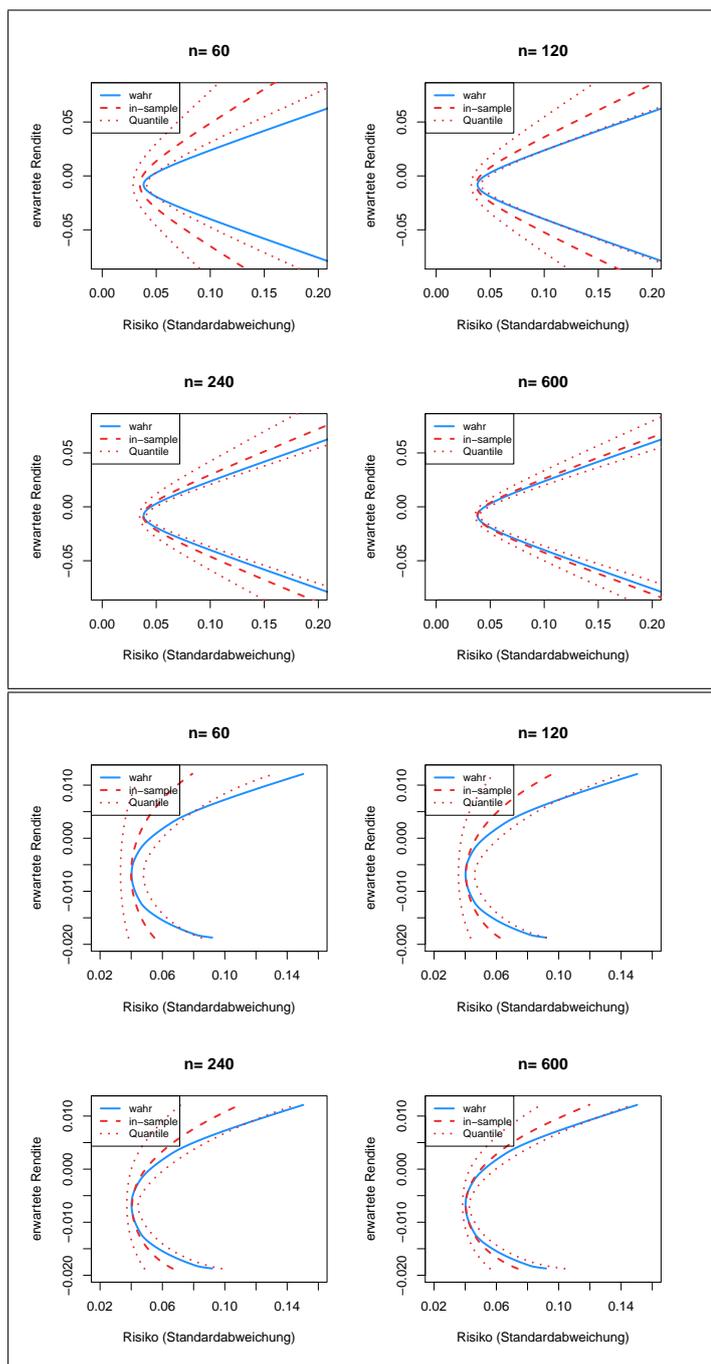


Abbildung 5.4.: *In-sample* Kurve für $m = 14$ Wertpapiere aus dem DAX für verschiedene n , oben mit Leerverkäufen, unten ohne Leerverkäufe.

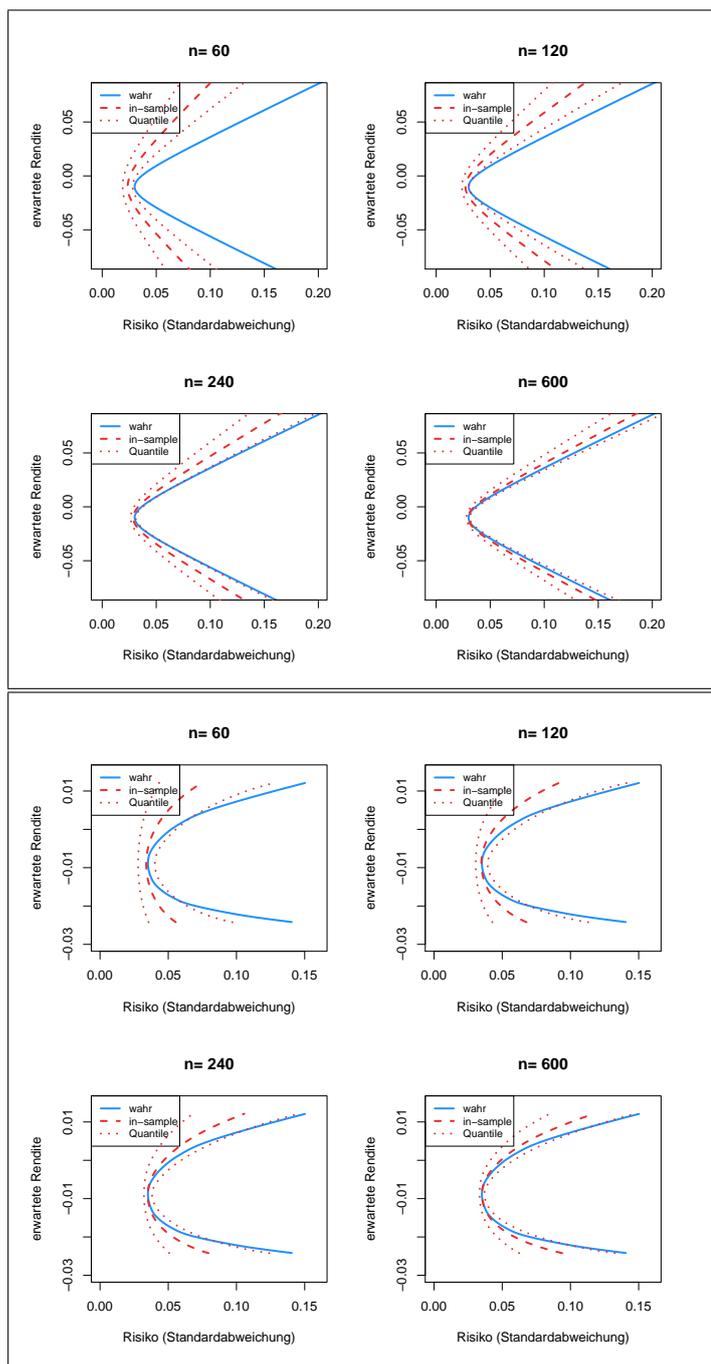


Abbildung 5.5.: *In-sample* Kurve für $m = 25$ Wertpapiere aus dem DAX für verschiedene n , oben mit Leerverkäufen, unten ohne Leerverkäufe.

Da die *in-sample* Kurve die wahre Minimum-Varianz-Kurve nur ungenügend approximiert, haben Kan und Smith (2008) für den unrestringierten Fall einen verbesserten Schätzer der Kurve definiert. Der Erwartungswert dieses Schätzers wird in (4.20) bestimmt. Dass dieser tatsächlich besser als die *in-sample* Kurve abschneidet, bestätigen die Abbildungen 5.6 - 5.8. Die wahre Minimum-Varianz-Kurve lässt sich mit dem verbesserten Schätzer im Mittel zwar viel besser annähern, jedoch muss dafür eine höhere Streuung als die der empirischen Minimum-Varianz-Kurve in Kauf genommen werden.

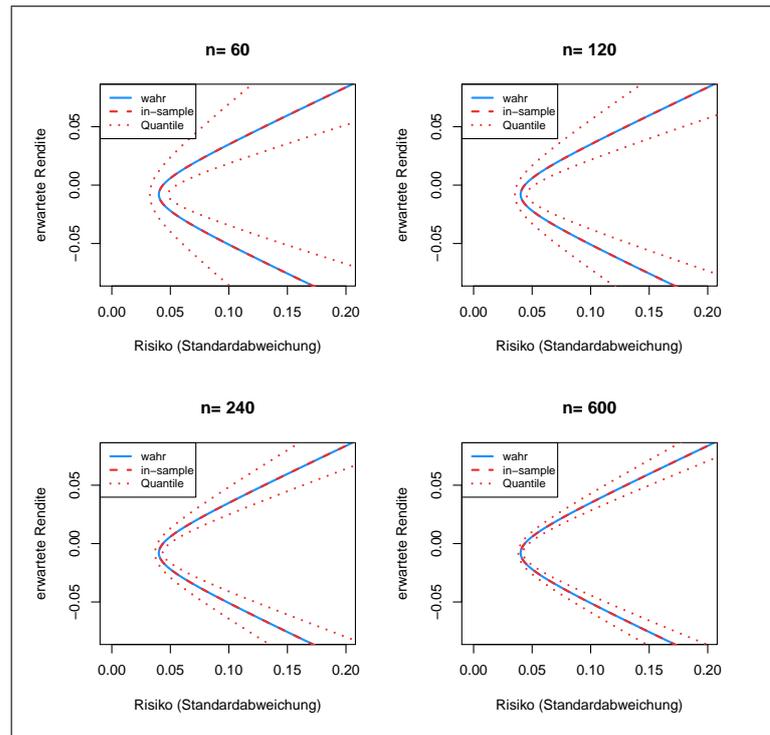


Abbildung 5.6.: Verbesserte *in-sample* Kurve für $m = 14$ Wertpapiere aus dem NASDAQ-100 für verschiedene n . Die durchgezogene Linie stellt die wahre Minimum-Varianz-Kurve dar, bei der gestrichelten Linie handelt es sich um die Wurzel des Erwartungswerts der verbesserten *in-sample* Kurve, $\sqrt{E[\hat{\sigma}_{pa}^2]}$. Die gepunkteten Linien sind die 5%- und 95%-Quantile der verbesserten Minimum-Varianz-Kurve $\hat{\sigma}_{pa}$, die basierend auf 50000 Simulationen bestimmt wurden.

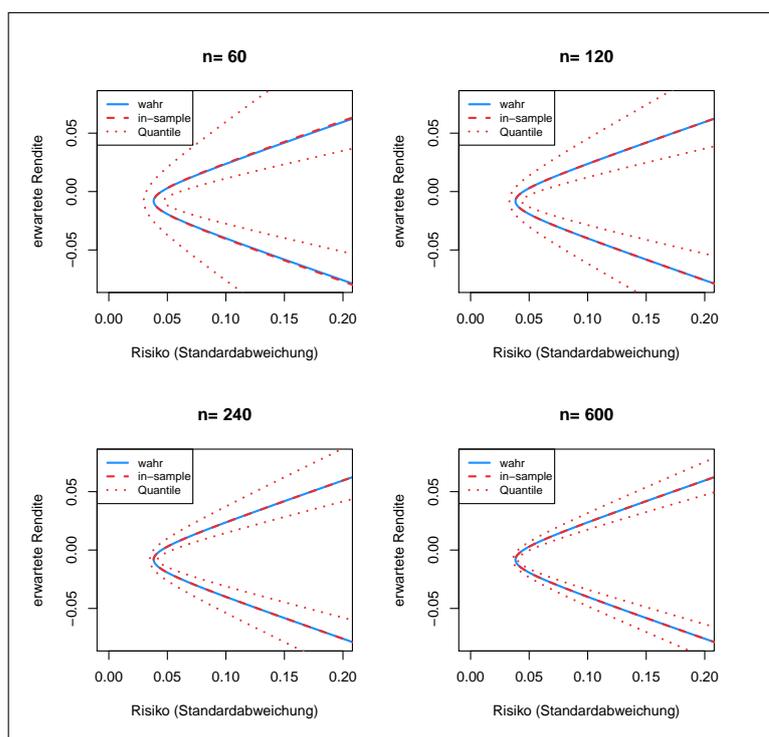


Abbildung 5.7.: Verbesserte *in-sample* Kurve für $m = 14$ Wertpapiere aus dem DAX für verschiedene n .

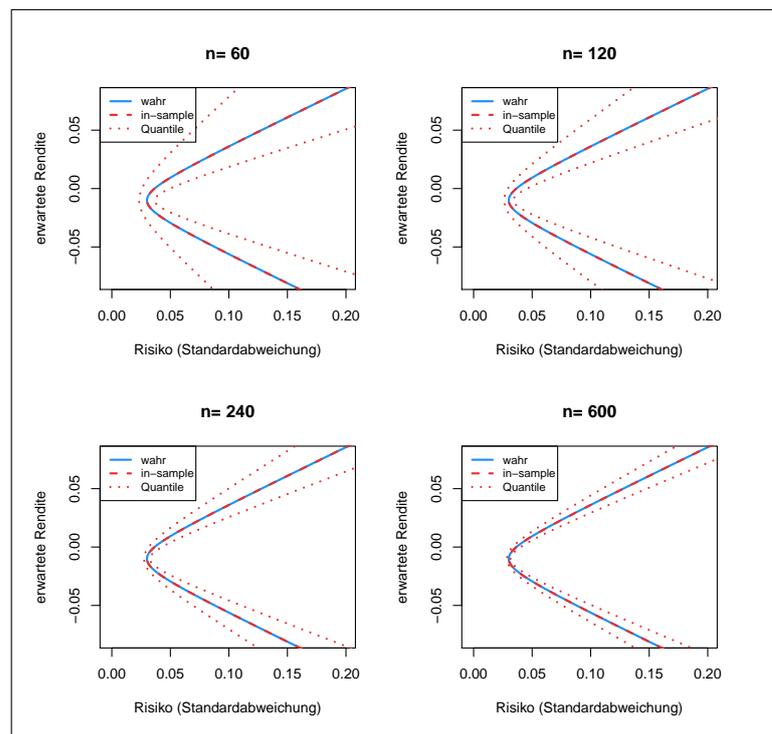


Abbildung 5.8.: Verbesserte *in-sample* Kurve für $m = 25$ Wertpapiere aus dem DAX für verschiedene n .

Mit der *out-of-sample* erwarteten Rendite und dem *out-of-sample* Risiko lässt sich die Performance eines empirischen Minimum-Varianz-Portfolios unter tatsächlichen Bedingungen beurteilen. Für den Fall, dass Leerverkäufe zulässig sind, werden $\underline{\mu}_p$ und $\underline{\sigma}_p = \sqrt{E[\tilde{\sigma}_p^2]}$ wie in (4.22) und (4.23) für verschiedene Werte für μ_* bestimmt und in ein (μ, σ) -Diagramm eingetragen. Ohne Leerverkäufe werden mit dem empirischen Mittelwert und der empirischen Kovarianzmatrix einer Stichprobe $\{\mathbf{r}_1, \dots, \mathbf{r}_n\}$ mit $\mathbf{r}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \dots, n$, die Gewichte der empirischen Minimum-Varianz-Portfolios $\widehat{\mathbf{w}}^*$ für verschiedene Werte für μ_* numerisch ermittelt und mit ihnen die jeweilige *out-of-sample* erwartete Rendite $\underline{\mu}_p = \widehat{\mathbf{w}}^{*T} \boldsymbol{\mu}$ und das jeweilige *out-of-sample* Risiko $\underline{\sigma}_p = \sqrt{\widehat{\mathbf{w}}^{*T} \boldsymbol{\Sigma} \widehat{\mathbf{w}}^*}$ bestimmt. Wiederholt man dieses Vorgehen 500 Mal und bildet den Mittelwert der *out-of-sample* Parameter für die jeweils gleichen Werte von μ_* , erhält man eine Annäherung der mittleren *out-of-sample* Kurve mit Leerverkaufsbeschränkung. Welche Form die *out-of-sample* Kurve annimmt und wie sie im Vergleich zur *in-sample* Kurve abschneidet, ist in den Abbildungen 5.9 - 5.11 zu sehen.

Die *out-of-sample* Kurve liegt in allen Fällen rechts von der wahren Minimum-Varianz-Kurve und damit auch rechts von der *in-sample* Kurve. Das bedeutet, dass empirische Minimum-Varianz-Portfolios im Mittel riskanter sind als die Portfolios, die man mit den wahren Parametern erhält. Außerdem sind sie riskanter als von der *in-sample* Kurve vorhergesagt. Die erwartete Rendite eines geschätzt effizienten Portfolios fällt in Wirklichkeit im Schnitt geringer aus als die erwartete Zielrendite μ_* . Schließlich gilt auch für die *out-of-sample* Kurve, dass sie im Bereich des globalen varianzminimalen Portfolios die wahre Minimum-Varianz-Kurve relativ gut approximiert, ihr Abstand jedoch umso größer wird, je weiter man sich vom varianzminimalen Portfolio entfernt. Mit steigendem n wird diese Differenz jedoch immer geringer.

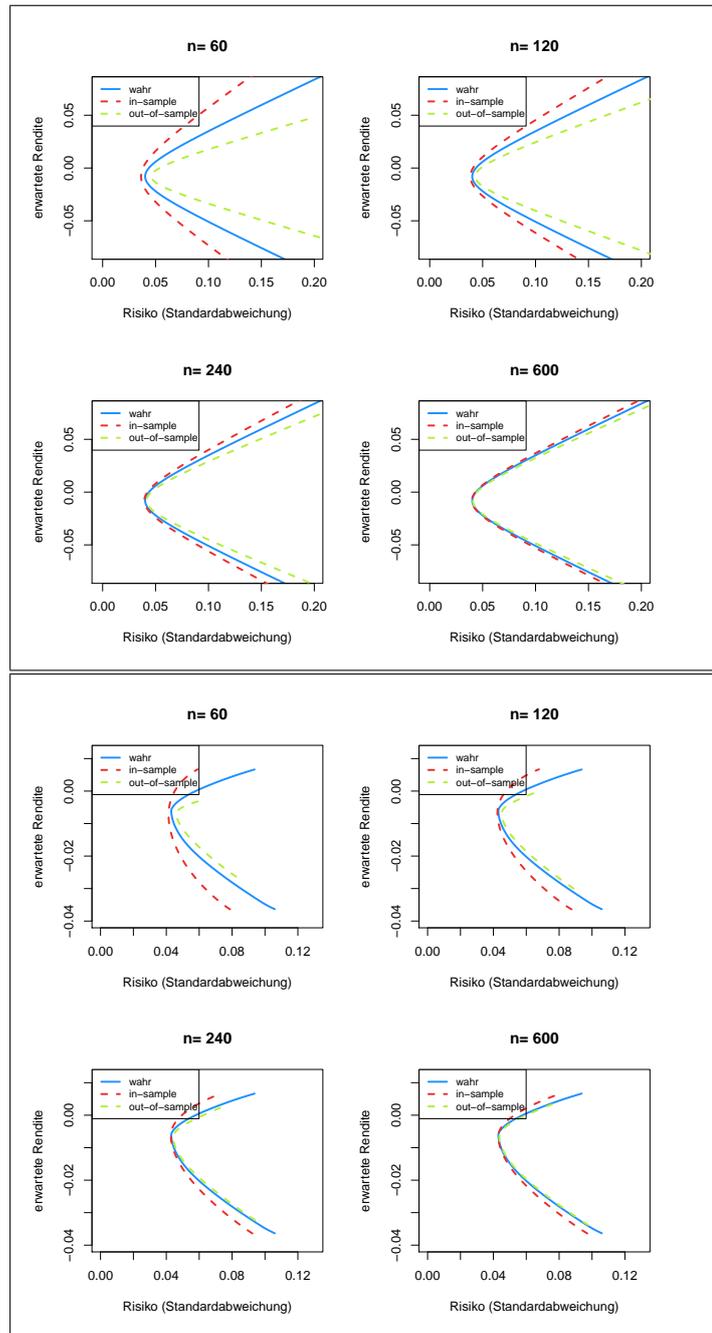


Abbildung 5.9.: *In-sample* Kurve und *out-of-sample* Kurve für $m = 14$ Wertpapiere aus dem NASDAQ-100 für verschiedene n , oben mit Leerverkäufen, unten ohne Leerverkäufe. Die durchgezogene Linie stellt die wahre Minimum-Varianz-Kurve dar, bei der roten gestrichelten Linie handelt es sich um $\bar{\sigma}_p$ und bei der grünen gestrichelten Linie um $\underline{\sigma}_p$.

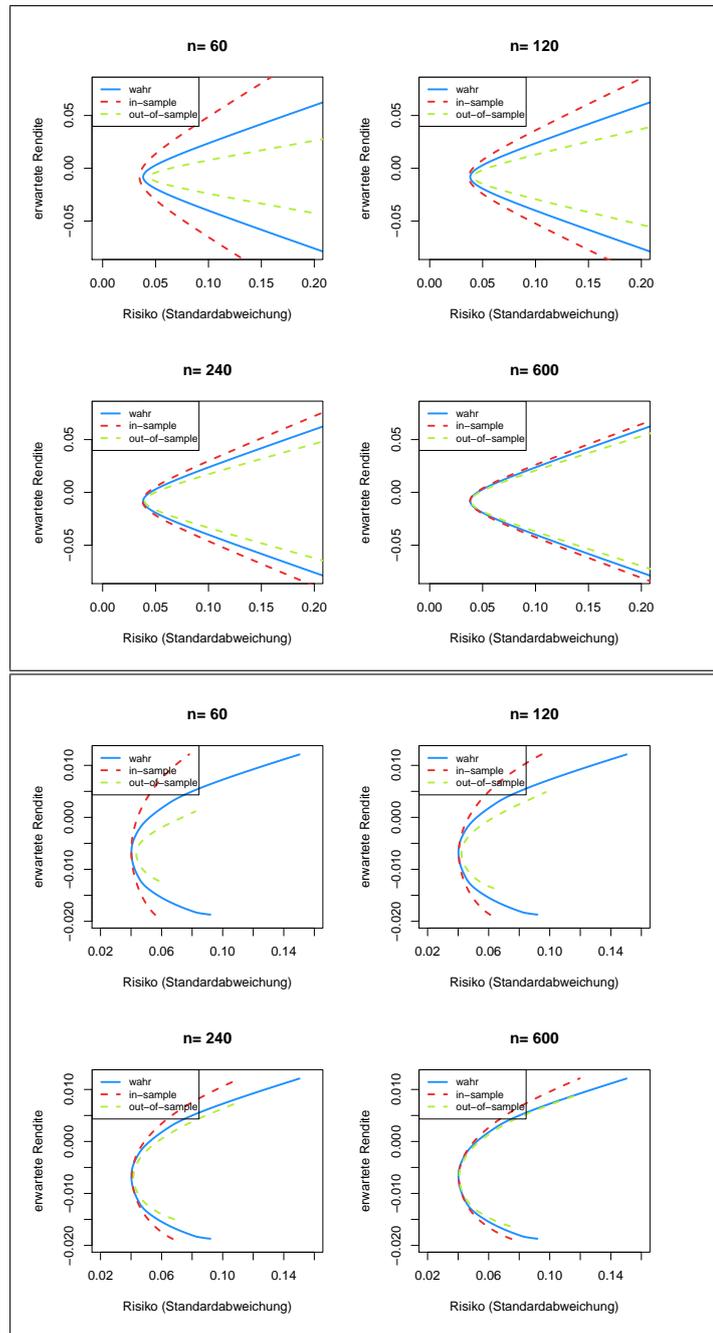


Abbildung 5.10.: *In-sample* Kurve und *out-of-sample* Kurve für $m = 14$ Wertpapiere aus dem DAX für verschiedene n , oben mit Leerverkäufen, unten ohne Leerverkäufe.

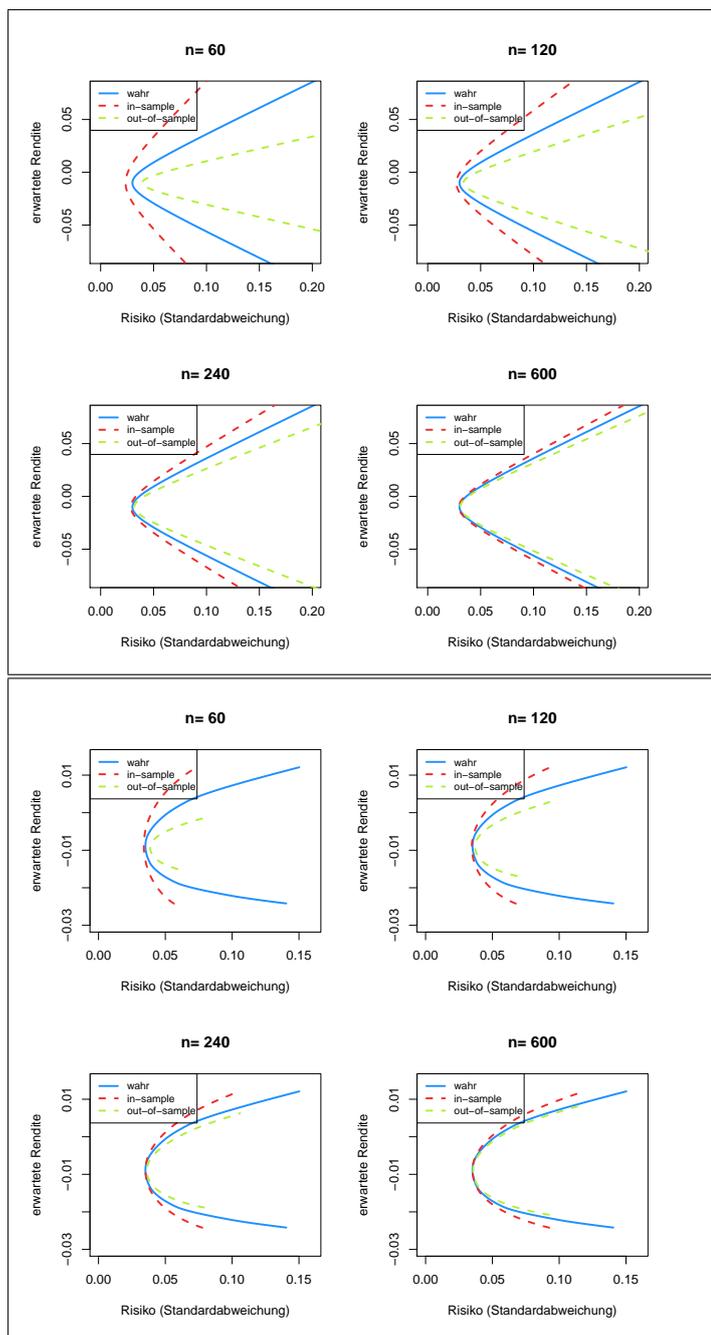


Abbildung 5.11.: *In-sample* Kurve und *out-of-sample* Kurve für $m = 25$ Wertpapiere aus dem DAX für verschiedene n , oben mit Leerverkäufen, unten ohne Leerverkäufe.

NASDAQ-100 ($m = 14$)														
	Amazon	Angen	Apple	Cisco	Comcast	Dell	Ebay	Intel	Microsoft	Oracle	Qualcomm	Starbucks	Vodafone	Yahoo!
μ	-0,02050	-0,00495	-0,03640	-0,00361	-0,00680	0,00670	-0,00855	-0,00420	-0,00351	-0,00857	-0,01077	-0,01334	-0,00630	-0,00646
	0,01450	0,00158	0,00402	0,00395	0,00083	0,00434	0,00641	0,00461	0,00282	0,00354	0,00369	0,00308	0,00121	0,00366
	0,00158	0,00478	0,00066	0,00097	0,00126	0,00149	0,00202	0,00128	0,00082	0,00090	0,00184	0,00118	0,00062	0,00017
	0,00402	0,00066	0,01127	0,00307	0,00137	0,00417	0,00434	0,00436	0,00280	0,00339	0,00298	0,00314	0,00189	0,00257
	0,00395	0,00097	0,00307	0,00686	0,00258	0,00412	0,00421	0,00386	0,00276	0,00360	0,00245	0,00347	0,00173	0,00308
	0,00083	0,00126	0,00137	0,00258	0,00547	0,00323	0,00273	0,00209	0,00144	0,00226	0,00172	0,00177	0,00191	0,00210
	0,00434	0,00149	0,00417	0,00412	0,00323	0,00883	0,00467	0,00423	0,00321	0,00365	0,00328	0,00318	0,00210	0,00365
Σ	0,00641	0,00202	0,00434	0,00421	0,00273	0,00467	0,01092	0,00401	0,00284	0,00310	0,00426	0,00422	0,00454	0,00454
	0,00461	0,00128	0,00436	0,00386	0,00209	0,00423	0,00401	0,00642	0,00225	0,00333	0,00210	0,00279	0,00134	0,00274
	0,00282	0,00082	0,00280	0,00276	0,00144	0,00321	0,00284	0,00225	0,00427	0,00234	0,00210	0,00203	0,00144	0,00112
	0,00354	0,00090	0,00339	0,00360	0,00226	0,00365	0,00310	0,00333	0,00234	0,00542	0,00228	0,00228	0,00180	0,00195
	0,00369	0,00184	0,00298	0,00245	0,00172	0,00328	0,00426	0,00210	0,00132	0,00225	0,00652	0,00258	0,00101	0,00253
	0,00308	0,00118	0,00314	0,00347	0,00177	0,00318	0,00422	0,00279	0,00203	0,00228	0,00258	0,00774	0,00060	0,00285
	0,00121	0,00062	0,00189	0,00173	0,00191	0,00210	0,00240	0,00134	0,00144	0,00180	0,00101	0,00060	0,00387	0,00124
	0,00366	0,00017	0,00257	0,00308	0,00210	0,00365	0,00454	0,00274	0,00112	0,00195	0,00253	0,00285	0,00124	0,01035
DAX ($m = 14$)														
	Allianz	BASF	Bayer	Bettersdorf	BMW	Commerzbank	Daimler	Deutsche Bank	E.ON	Henkel	Merck	RWE	Siemens	Telekom
μ	-0,00418	-0,01875	-0,01355	-0,00777	-0,00936	0,01213	-0,00617	-0,00110	-0,01646	-0,01106	-0,01182	-0,00541	-0,00786	0,00003
	0,01124	0,00478	0,00446	0,00134	0,00447	0,00924	0,00616	0,00797	0,00484	0,00286	0,00230	0,00384	0,00564	0,00232
	0,00478	0,00863	0,00293	0,00083	0,00317	0,00518	0,00442	0,00380	0,00208	0,00155	0,00128	0,00256	0,00425	0,00151
	0,00446	0,00293	0,00619	0,00161	0,00297	0,00512	0,00350	0,00386	0,00255	0,00229	0,00249	0,00284	0,00303	0,00193
	0,00134	0,00083	0,00161	0,00343	0,00157	0,00163	0,00193	0,00166	0,00054	0,00134	0,00081	0,00107	0,00098	0,00032
	0,00447	0,00317	0,00297	0,00157	0,00693	0,00586	0,00618	0,00493	0,00165	0,00270	0,00114	0,00221	0,00396	0,00118
	0,00924	0,00518	0,00512	0,00163	0,00586	0,02270	0,00813	0,01159	0,00561	0,00370	0,00264	0,00529	0,00675	0,00284
Σ	0,00616	0,00442	0,00350	0,00193	0,00618	0,00618	0,00975	0,00582	0,00436	0,00324	0,00166	0,00351	0,00504	0,00150
	0,00797	0,00380	0,00386	0,00166	0,00493	0,01159	0,00582	0,01200	0,00448	0,00321	0,00296	0,00425	0,00557	0,00191
	0,00484	0,00208	0,00255	0,00054	0,00165	0,00561	0,00436	0,00448	0,01584	0,00219	0,00233	0,00420	0,00348	0,00234
	0,00286	0,00155	0,00229	0,00134	0,00270	0,00370	0,00324	0,00321	0,00219	0,00403	0,00133	0,00185	0,00275	0,00105
	0,00230	0,00128	0,00249	0,00081	0,00114	0,00264	0,00166	0,00296	0,00233	0,00541	0,00211	0,00211	0,00169	0,00103
	0,00384	0,00256	0,00284	0,00107	0,00221	0,00529	0,00351	0,00425	0,00420	0,00185	0,00211	0,00256	0,00256	0,00174
	0,00564	0,00425	0,00303	0,00098	0,00396	0,00675	0,00504	0,00557	0,00348	0,00275	0,00169	0,00256	0,00636	0,00207
	0,00232	0,00151	0,00193	0,00032	0,00118	0,00284	0,00150	0,00191	0,00234	0,00105	0,00103	0,00174	0,00207	0,00347

Tabelle 5.3.: Daten für die Simulationen.

6. Fazit und Ausblick

In Kapitel 2 wurde die Portfoliotheorie von Markowitz vorgestellt, deren Kern das Mean-Variance-Optimierungsproblem ist. Es wurden Bedingungen hergeleitet, unter denen sich verschiedene Formulierungen des Problems ineinander überführen lassen. Durch Lösen des Problems erhält man Portfolios, deren Risiko-Rendite-Kombinationen auf der Effizienzkurve bzw. auf der Minimum-Varianz-Kurve liegen. Um in der Praxis (annähernd) effiziente Portfolios konstruieren zu können, muss die Effizienzkurve näherungsweise bestimmt werden, wofür μ und Σ geeignet geschätzt werden müssen. Dafür wurden in Kapitel 3 verschiedene Methoden vorgestellt. In der Simulationsstudie in Kapitel 5 wurden die empirischen Schätzer, die *resampling*-Methode sowie die Bayes-Schätzer miteinander verglichen. Während große Unterschiede zwischen den geschätzten Kurven im unrestringierten Fall zu beobachten waren, verliefen die Kurven im restringierten Fall fast gleich. Mit Leerverkäufen erhielt man für hohe Werte für μ_* mit den Bayes-Schätzern die beste Annäherung der wahren Effizienzkurve. Für niedrige erwartete Zielrenditen schnitten jedoch in beiden Fällen alle drei Methoden nur ungenügend ab. Da die empirischen Schätzer in der Praxis dennoch von großer Bedeutung sind, beschäftigte sich Kapitel 4 mit der Verteilung der empirischen Minimum-Varianz-Kurve. Allerdings bezogen sich die Untersuchungen dabei lediglich auf den unrestringierten Fall, da sich die Lösung des Minimum-Varianz-Problems nur dann explizit mithilfe der Effizienzmengenkennkonstanten darstellen lässt. Es wurden insbesondere die Verteilungen der *in-sample* und der *out-of-sample* Kurve ermittelt. Die Simulationsstudie in Kapitel 5 bestätigte, dass das *in-sample* Risiko der empirischen Minimum-Varianz-Portfolios das Risiko der wahren Minimum-Varianz-Portfolios im Durchschnitt unterschätzt, während das mittlere *out-of-sample* Risiko im Vergleich zum Risiko der wahren Portfolios zu hoch ist.

In dieser Arbeit wurde den Renditen meist eine Normalverteilung mit Parametern μ und Σ unterstellt. Diese Annahme ist zwar gebräuchlich, berücksichtigt aber nicht, dass die Verteilung der Renditen in Wirklichkeit schwerere *tails* als die Normalverteilung besitzt. Aus diesem Grund wäre es interessant zu untersuchen, wie andere Modellannahmen sich auf die Ergebnisse dieser Arbeit auswirken würden. So könnte man zum Beispiel annehmen, dass die Renditen einer studentschen *t*-Verteilung oder einer Lévy-Verteilung folgen. Eine weitere Möglichkeit besteht darin, die Renditen als GARCH-Zeitreihe zu modellieren.

Bisher ist eine analytische Untersuchung der Effizienzkurve nur dann möglich, wenn Leerverkäufe zugelassen sind. Da Leerverkäufe in der EU nach der EU-Leerverkaufsverordnung (EU-Kommission, Verordnung (EU) Nr. 236/2012 vom 14. März 2012) nur beschränkt möglich und insbesondere für kleinere Investoren in der Regel mit einem erhöhten Risiko verbunden sind, sollten sich zukünftige Forschungen intensiver auf den restringierten Fall beziehen, um eine umfassende Analyse der Effizienzkurve unter Leerverkaufsbeschränkungen zu ermöglichen.

Literaturverzeichnis

- [1] Technical report.
- [2] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Courier Dover Publications, 1965.
- [3] J. Berk and P. DeMarzo. *Grundlagen der Finanzwirtschaft: Analyse, Entscheidung und Umsetzung*. Pearson, 2011.
- [4] M. J. Best and R. R. Grauer. On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *The Review of Financial Studies*, 4:315–342, 1991.
- [5] M. E. Bock, G. G. Judge, and T. A. Yancey. A simple form for the inverse moments of non-central χ^2 and F random variables and certain confluent hypergeometric functions. *Journal of Econometrics*, 25:217–234, 1984.
- [6] T. Cool. Proper definitions for Risk and Uncertainty. Working paper, EconWPA, General Economics and Teaching 9902002, 1993.
- [7] A. P. Dawid. Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, 68:265–274, 1981.
- [8] J. M. Dickey. Matricvariate Generalizations of the Multivariate t Distribution and the Inverted Multivariate t Distribution. *The Annals of Mathematical Statistics*, 38:511–518, 1967.
- [9] P. S. Dwyer. Some Applications of Matrix Derivatives in Multivariate Analysis. *Journal of the American Statistical Association*, 62:607 – 625, 1967.
- [10] E. J. Elton and M. J. Gruber. Estimating the dependence structure of share prices - Implications for portfolio selection. *The Journal of Finance*, 28:1203–1232, 1973.
- [11] E. J. Elton and M. J. Gruber. Modern Portfolio Theory, 1950 to Date. Working paper, New York University, Stern School of Business, Finance Department, 1997.
- [12] L. Fahrmeir, A. Hamerle, and G. Tutz. *Multivariate statistische Verfahren*. de Gruyter, 1996.
- [13] T. S. Ferguson. *A course in large sample theory*. Chapman & Hall, 1996.
- [14] R. Fletcher. *Practical methods of optimization*. Wiley, 1991.

-
- [15] G. M. Frankfurter, H. E. Phillips, and J. P. Seagle. Performance of the Sharpe portfolio selection model: A comparison. *Journal of Financial and Quantitative Analysis*, 6:191–204, 1976.
- [16] Y. Fujikoshi, V. V. Ulyanov, and R. Shimizu. *Multivariate Statistics: High Dimensional and Large-Sample Approximations*. Wiley, 2010.
- [17] C. Geiger and C. Kanzow. *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer, 2002.
- [18] L. R. Haff. An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, 9:531–544, 1979.
- [19] W. Hausmann, K. Diener, and J. Käsler. *Derivate, Arbitrage und Portfolio-Selection*. Vieweg, 2002.
- [20] W. Härdle and L. Simar. *Applied Multivariate Statistical Analysis*. Springer, 2003.
- [21] J. D. Jobson. Confidence regions for the mean-variance efficient set: An alternative approach to estimation risk. *Review of Quantitative Finance and Accounting*, 1:235–257, 1991.
- [22] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Volume 2*. Wiley, 1995.
- [23] P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, 3rd edition, 2007.
- [24] R. Kan and D. R. Smith. The Distribution of the Sample Minimum-Variance Frontier. Working paper, 2007.
- [25] R. Kan and D. R. Smith. The Distribution of the Sample Minimum-Variance Frontier. *Management Science*, 54:1364–1380, 2008.
- [26] R. Kan, X. Wang, and G. Hillier. Computationally efficient recursions for top-order invariant polynomials with applications. Working paper, The Institute for Fiscal Studies, Department of Economics, UCL, 2007.
- [27] S. Karlin and H. M. Taylor. *A Second Course in Stochastic Processes, Band 2*. Academic Press, 1981.
- [28] F. H. Knight. *Risk, Uncertainty and Profit*. Houghton Mifflin Co., Boston, 1921.
- [29] M. Koecher. *Lineare Algebra Und Analytische Geometrie*. Springer, 2003.
- [30] J. Kremer. *Einführung in die Diskrete Finanzmathematik*. Springer, 2006-a.
- [31] J. Kremer. *Portfoliotheorie, Risikomanagement und die Bewertung von Derivaten*. Springer, 2006-b.

-
- [32] K. Krishnamoorthy. *Handbook of Statistical Distributions with Applications*. Chapman & Hall, 2006.
- [33] T.L. Lai, H. Xing, and Z. Chen. Mean-variance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics*, 5:798–823, 2011.
- [34] O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621, 2003.
- [35] W. Luh, J. Müller, S. Ponnusamy, and P. Vasundhra. Polynomial series expansions for confluent and gaussian hypergeometric functions. *Mathematics of Computation*, 74:1937–1952, 2005.
- [36] B. Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36:394–419, 1963.
- [37] H. Markowitz. Portfolio Selection. *The Journal of Finance*, 7:77–91, 1952.
- [38] R. C. Merton. An Analytic Derivation of the Efficient Portfolio Frontier. *The Journal of Financial and Quantitative Analysis*, 7:1851–1872, 1972.
- [39] A. Meucci. *Risk and Asset Allocation*. Springer, 2009.
- [40] R. O. Michaud and R. Michaud. *Efficient Asset Management*. Harvard Business School Press, Boston, 1989.
- [41] R. O. Michaud and R. Michaud. The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal*, 45:31–42, 1989.
- [42] R. O. Michaud and R. Michaud. Portfolio optimization by means of resampled efficient frontiers. Patent Application, http://www.patentlens.net/patentlens/patent/WO_1999_050773_A1/en/, 1999.
- [43] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, 1982.
- [44] K.M. Ramachandran and C.P. Tsokos. *Mathematical Statistics with Applications*. Elsevier, 2009.
- [45] R. Remmert and G. Schumacher. *Funktionentheorie 2*. Springer, 2007.
- [46] C. P. Robert. *The Bayesian Choice*. Springer, 2nd edition, 2007.
- [47] S. A. Ross. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13:341–360, 1976.
- [48] F.J. Samaniego. *A Comparison of the Bayesian and Frequentist Approaches to Estimation*. Springer, 2010.

- [49] T. Sawa. Finite-Sample Properties of the k -Class Estimators. *Econometrica*, 40:653–680, 1972.
- [50] W. F. Sharpe. A simplified model for portfolio analysis. *Management Science*, 9:277–293, 1963.
- [51] W. F. Sharpe. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19:425–442, 1964.
- [52] M. C. Steinbach. Markowitz Revisited: Mean-Variance Models in Financial Portfolio Analysis. *SIAM Review*, 43:31–85, 2001.
- [53] F. G. Tricomi. Konfluente hypergeometrische Funktionen. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, 6:257–274, 1955.
- [54] J. Trölb. *Angewandte Mathematik mit Mathcad. Lehr- und Arbeitsbuch: Band 2*. Springer, 2008.
- [55] R. S. Tsay. *Analysis of Financial Time Series*. Wiley, 2010.

A. Capital Asset Pricing Model

Aufbauend auf die Portfoliotheorie von Markowitz hat William Sharpe 1964 das Capital Asset Pricing Modell (CAPM) entwickelt. Dabei stellte er fest, dass sich das Gesamtrisiko eines Portfolios in zwei Komponenten aufteilen lässt, das *systematische* und das *unsystematische Risiko*, wobei sich nur letzteres durch Diversifikation reduzieren lässt. Beim systematischen Risiko, auch Marktrisiko genannt, handelt es sich um die Sensitivität eines Assets in Bezug auf allgemeine Marktbewegungen. Wie stark und in welche Richtung ein Asset auf die Marktbewegungen reagiert, wird durch das *Beta* des Assets ausgedrückt, das weiter unten definiert wird. Das unsystematische Risiko eines Assets stammt von spezifischen Einflüssen auf dessen Rendite, die unabhängig vom Markt sind. Da das unsystematische Risiko mittels Diversifikation eliminiert werden kann, sollte nach Sharpe ein Investor nur für die Aufnahme eines höheren systematischen Risikos durch eine höhere Risikoprämie entlohnt werden. Dazu setzte Sharpe in seinem Modell die Risikoprämie in direkte Beziehung zum Beta eines Assets. Dabei machte er zunächst folgende Annahmen:

- Alle Investoren kennen μ und Σ und sind sich insbesondere über sie einig.
- Jeder Investor strebt ein (μ, σ) -effizientes Portfolio nach Markowitz unter individuellen Risikopräferenzen an.
- Alle Investoren investieren während desselben Zeithorizonts.
- Es gibt eine sichere Anlage. Diese liefert für jeden Investor dieselbe Rendite.
- Es gibt keine Transaktionskosten.
- Jedes Asset ist beliebig teilbar.

Beim CAPM wird davon ausgegangen, dass sich der Markt in einem Gleichgewicht befindet, alle Assets also jeweils den Preis besitzen, der Angebot und Nachfrage ausgleicht. Nimmt man nämlich an, dass der Kapitalmarkt zunächst beliebig strukturiert ist und jeder Investor ein beliebiges Portfolio hält und nun auf einmal die Voraussetzungen für das CAPM eintreten, wird jeder Investor sein Portfolio derart umstrukturieren, dass es effizient ist. Er wird also einen Anteil in das Marktportfolio und den Rest in die sichere Anlage investieren. Der riskante Anteil des Portfolios weist dann die gleiche Struktur wie das Marktportfolio auf. Durch die dadurch entstehende Veränderung von Angebot und Nachfrage passen sich die Preise der Assets am Markt so an, dass Angebot und Nachfrage wieder im Gleichgewicht sind. Durch die Preisänderungen ändert sich natürlich auch das Marktportfolio, und zwar solange, bis jeder Investor mit seinem Portfolio zufrieden ist. Da die (σ, μ) -Kombination eines beliebigen Assets stets unterhalb der Kapitalmarktklinie liegt, sind die einzelnen Assets

ineffizient. Diese Überlegungen können bei Sharpe (1964) sowie bei zahlreichen weiteren Autoren, wie z.B. Hausmann, Diener und Käsler (2002) nachgelesen werden.

Die zentrale Aussage des CAPM lautet nun: Im Gleichgewichtszustand eines Kapitalmarktes, der obige Annahmen erfüllt, gilt für die erwartete Rendite μ_i jedes riskanten Assets i des Marktes

$$\mu_i - r_f = \frac{\mu_M - r_f}{\sigma_M} \cdot \frac{\sigma_{i,M}}{\sigma_M} = (\mu_M - r_f) \beta_i \quad \text{mit} \quad \beta_i = \frac{\sigma_{i,M}}{\sigma_M^2}, \quad (\text{A.1})$$

wobei $\sigma_{i,M}$ die Kovarianz der Rendite von Asset i und des Marktportfolios M ist, σ_M^2 die Varianz des Marktportfolios und β_i das Beta von Asset i . Wie oben bereits erwähnt, steht β_i für das systematische Risiko von Asset i . Es lässt sich im Gleichgewicht also nur dann eine höhere Risikoprämie erzielen, wenn dafür ein höheres systematisches Risiko aufgenommen wird. Für den Beweis von (A.1) siehe z.B. Kremer (2006-a).

B. Bayes'sche Entscheidungstheorie

Der Inhalt dieses Abschnitts orientiert sich an Robert (2007).

Wird ein zufälliges Phänomen von einem Parameter θ beeinflusst, lassen sich einerseits anhand von Beobachtungen mithilfe statistischer Methoden Rückschlüsse auf θ ziehen, andererseits ermöglicht die Bildung von Wahrscheinlichkeiten, die von θ abhängen, Aussagen über das zukünftige Verhalten zu treffen. Dieser inverse Zusammenhang von Wahrscheinlichkeiten wird insbesondere in der Likelihoodfunktion deutlich, da sie im Prinzip nichts anderes als die Dichtefunktion ist, jedoch als Funktion in θ anstatt in der Beobachtung x . Ist also $f(x | \theta)$ die Dichte einer Zufallsvariablen X bedingt auf θ , so ist bei unbekanntem Parameter θ und gegebener Beobachtung x die Likelihoodfunktion

$$l(\theta | x) = f(x | \theta).$$

Eine allgemeine Beschreibung dieser Inversion von Wahrscheinlichkeiten liefert das Bayes-Theorem: Für zwei Ereignisse A und B mit $P(B) \neq 0$ gilt

$$P(A | B) = \frac{P(B | A)P(A)}{P(B | A)P(A) + P(B | A^c)P(A^c)} = \frac{P(B | A)P(A)}{P(B)}.$$

Mit der stetigen Version des Theorems für zwei Zufallsvariablen X und Y mit bedingter Dichte $f_{X|Y}$ und Marginaldichte g_Y lässt sich die bedingte Dichte von Y gegeben X bestimmen durch

$$g_{Y|X}(y | x) = \frac{f_{X|Y}(x | y)g_Y(y)}{\int f_{X|Y}(x | y)g_Y(y)dy}.$$

In einem Bayes'schen Modell wird der unbekannte Parameter θ als Zufallsvariable betrachtet, deren Unsicherheit durch eine Verteilung π auf dem Parameterraum Θ modelliert wird. π wird als *A-priori-Verteilung* von θ bezeichnet und auf der Grundlage von Vorwissen und Annahmen über den Parameter gebildet. Die Verteilung von θ gegeben die Beobachtung x heißt *A-posteriori-Verteilung*. Sie lässt sich mithilfe des Bayes-Theorems aus der Likelihoodfunktion und der A-priori-Verteilung bestimmen:

$$\pi(\theta | x) = \frac{f(x | \theta)\pi(\theta)}{\int f(x | \theta)\pi(\theta)d\theta}.$$

Beim Schätzen eines Parameters entscheidet man sich letztendlich mithilfe einer geeigneten Vorschrift für einen Wert, den man als Annäherung für den unbekanntem wahren Wert des Parameters annimmt. Im Folgenden bezeichne \mathcal{D} die Menge aller möglichen solcher Entscheidungen, man nennt \mathcal{D} deswegen auch *Entscheidungsraum*. Meist gilt $\mathcal{D} = \Theta$. Die Entscheidung wird auf Grundlage von Beobachtungen $x \sim f(x | \theta)$ getroffen, die aus dem

Beobachtungsraum \mathcal{X} stammen. Die Regel δ , die allen Beobachtungen eine Entscheidung d zuordnet, ist also eine Funktion aus dem Beobachtungsraum \mathcal{X} in den Entscheidungsraum \mathcal{D} . Man bezeichnet δ auch als *Schätzer* und $\delta(x) = d$ als Schätzwert. Um eine Entscheidung für einen Schätzwert zu bewerten, verwendet man *Verlustfunktionen*. Jede Funktion L von der Form

$$L : \Theta \times \mathcal{D} \rightarrow [0, +\infty)$$

ist eine Verlustfunktion. Sie misst den Fehler $L(\theta, d)$, der entsteht, wenn man für den wahren Parameter θ den Wert d annimmt. Man verwendet häufig die *quadratische Verlustfunktion*

$$L(\theta, d) = (\theta - d)^2.$$

Anstelle des Verlusts, der mit einer konkreten Entscheidung d verbunden ist, lässt sich auch der mittlere Verlust aller Entscheidungen, die man mit δ treffen kann, betrachten, also

$$\begin{aligned} R(\theta, \delta) &= E_{\theta} [L(\theta, \delta(x))] \\ &= \int_{\mathcal{X}} L(\theta, \delta(x)) f(x | \theta) dx. \end{aligned}$$

Diese Funktion heißt *frequentistisches Risiko*. Da θ unbekannt bzw. nach Bayes'schen Ansichten vielmehr zufällig ist, integriert man in einem Bayes'schen Modell über Θ statt über \mathcal{X} . Dadurch erhält man den *A-posteriori erwarteten Verlust*

$$\begin{aligned} \rho(\pi, d | x) &= E^{\pi} [L(\theta, d) | x] \\ &= \int_{\Theta} L(\theta, d) \pi(\theta | x) d\theta. \end{aligned}$$

Gegeben die Beobachtung x ist der durchschnittliche Fehler, der durch die Entscheidung d entsteht, $\rho(\pi, d | x)$. Bildet man beim frequentistischen Risiko den Durchschnitt über die Werte von θ gemäß ihrer A-priori-Verteilung, erhält man das *integrierte Risiko*

$$\begin{aligned} r(\pi, \delta) &= E^{\pi} [R(\theta, \delta)] \\ &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) f(x | \theta) dx \pi(\theta) d\theta. \end{aligned}$$

Mit diesen Definitionen können nun Bayes-Schätzer definiert werden:

Definition B.0.1 (Bayes-Schätzer, Bayes-Risiko) *Ein Schätzer δ^{π} , der das integrierte Risiko $r(\pi, \delta)$ minimiert, heißt Bayes-Schätzer. Der Wert $r(\pi) = r(\pi, \delta^{\pi})$ heißt dann Bayes-Risiko.*

Der Wert von δ^{π} hängt also insbesondere von der A-priori-Verteilung π und der Verlustfunktion L ab.

Satz B.0.2 *Der Bayes-Schätzer δ^{π} , den man mit der A-priori-Verteilung π und der quadratischen Verlustfunktion L erhält, ist der bedingte Erwartungswert*

$$\delta^{\pi}(x) = E^{\pi} [\theta | x].$$

C. Verteilungen

In diesem Abschnitt werden einige Verteilungen definiert und Eigenschaften bewiesen, die in dieser Arbeit benötigt werden.

C.1. Multivariate Normalverteilung

Satz C.1.1 Seien $\mathbf{X}_1, \dots, \mathbf{X}_n$ unabhängig und identisch verteilte Zufallsvariablen mit $\mathbf{X}_i \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \dots, n$. Dann sind der empirische Mittelwert $\bar{\mathbf{X}} = 1/n \sum_{i=1}^n \mathbf{X}_i$ und die empirische Kovarianzmatrix $\hat{\boldsymbol{\Sigma}} = 1/n \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ die Maximum-Likelihood-Schätzer für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$.

Der Beweis ist bei Dwyer (1967) zu finden.

Beweis. Die Likelihood-Funktion von $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ist

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; x) = \frac{1}{(2\pi)^{\frac{nm}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu})}$$

mit $x \in \mathbb{R}^n$ und die Log-Likelihood-Funktion ist

$$\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; x) = \text{const} - \frac{n}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n (x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu}).$$

Da nach Dwyer (1967)

$$\frac{\partial}{\partial \boldsymbol{\mu}} (x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu}) = -2\boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu})$$

gilt, folgt

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; x) = \boldsymbol{\Sigma}^{-1} \sum_{i=1}^n (x_i - \boldsymbol{\mu}).$$

Diese Ableitung ist für $\boldsymbol{\mu} = \bar{x}$ gleich null, sodass $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$ der Maximum-Likelihood-Schätzer für $\boldsymbol{\mu}$ ist.

Weiterhin gilt nach Dwyer (1967)

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} \ln |\boldsymbol{\Sigma}| = \boldsymbol{\Sigma}^{-1}$$

und

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} (x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu}) = -\boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu}) (x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1},$$

sodass

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\Sigma}} \ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; x) &= -\frac{n}{2} \boldsymbol{\Sigma}^{-1} + \frac{1}{2} \boldsymbol{\Sigma}^{-1} \sum_{i=1}^n (x_i - \boldsymbol{\mu})(x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} \\ &= \frac{1}{2} \boldsymbol{\Sigma}^{-1} \left(-n + \sum_{i=1}^n (x_i - \boldsymbol{\mu})(x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} \right)\end{aligned}$$

ist. Mit $\boldsymbol{\mu} = \bar{x}$ ist diese Ableitung gleich null, wenn $\boldsymbol{\Sigma} = 1/n \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$ ist. Somit ist $\hat{\boldsymbol{\Sigma}} = 1/n \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ der Maximum-Likelihood-Schätzer von $\boldsymbol{\Sigma}$. \square

Satz C.1.2 *Seien $\mathbf{X}_1, \dots, \mathbf{X}_n$ unabhängige $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -verteilte Zufallsvariablen. Dann sind der empirische Mittelwert $\bar{\mathbf{X}} = 1/n \sum_{i=1}^n \mathbf{X}_i$ und die korrigierte empirische Kovarianzmatrix $\mathbf{S}^* = 1/(n-1) \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ unverzerrte Schätzer für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$.*

Beweis. Für den empirischen Mittelwert gilt

$$E \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \right] = \frac{1}{n} \sum_{i=1}^n E[\mathbf{X}_i] = \frac{1}{n} \cdot n \cdot \boldsymbol{\mu} = \boldsymbol{\mu}.$$

Also ist der empirische Mittelwert ein unverzerrter Schätzer für $\boldsymbol{\mu}$. Nach Muirhead (1982) ist $\bar{\mathbf{X}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$. Somit gilt für die korrigierte empirische Kovarianzmatrix

$$\begin{aligned}& E \left[\frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \right] \\ &= E \left[\frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu} + \boldsymbol{\mu} - \bar{\mathbf{X}})(\mathbf{X}_i - \boldsymbol{\mu} + \boldsymbol{\mu} - \bar{\mathbf{X}})^T \right] \\ &= E \left[\frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^T \right] + E \left[\frac{2}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu})(\boldsymbol{\mu} - \bar{\mathbf{X}})^T \right] \\ &\quad + E \left[\frac{n}{n-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})^T \right] \\ &= \frac{n}{n-1} \boldsymbol{\Sigma} + E \left[\frac{2}{n-1} \left(\sum_{i=1}^n \mathbf{X}_i - n\boldsymbol{\mu} \right) (\boldsymbol{\mu} - \bar{\mathbf{X}})^T \right] + \frac{n}{n-1} \frac{\boldsymbol{\Sigma}}{n} \\ &= \frac{n}{n-1} \boldsymbol{\Sigma} - E \left[\frac{2n}{n-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})^T \right] + \frac{n}{n-1} \frac{\boldsymbol{\Sigma}}{n} \\ &= \frac{n}{n-1} \boldsymbol{\Sigma} - \frac{2n}{n-1} \frac{\boldsymbol{\Sigma}}{n} + \frac{n}{n-1} \frac{\boldsymbol{\Sigma}}{n} = \boldsymbol{\Sigma}.\end{aligned}$$

Somit ist \mathbf{S}^* ein unverzerrter Schätzer für $\boldsymbol{\Sigma}$. \square

Satz C.1.3 *Sei $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ und seien \mathbf{A} und \mathbf{B} zwei Matrizen. Dann sind $\mathbf{A}\mathbf{X}$ und $\mathbf{B}\mathbf{X}$ genau dann unabhängig, wenn $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^T = \mathbf{0}$ ist; siehe Härdle und Simar (2003).*

Satz C.1.4 Für eine m -dimensionale Zufallsvariable \mathbf{X} mit $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_m)$ und einen konstanten Vektor $\mathbf{h} \in \mathbb{R}^m$ gilt

$$E[(\mathbf{X}^T \mathbf{X})(\mathbf{X}^T \mathbf{h})] = (m + 2 + \boldsymbol{\mu}^T \boldsymbol{\mu})(\boldsymbol{\mu}^T \mathbf{h}).$$

Beweis. Sei $\mathbf{X} = (X_1, \dots, X_m)^T$. Wegen $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_m)$ ist dann $X_i \sim \mathcal{N}(\mu_i, 1)$ für $i = 1, \dots, m$ und X_i und X_j sind für $i \neq j$ unabhängig. Sei weiterhin $\mathbf{h} = (h_1, \dots, h_m)^T$. Durch Ausmultiplizieren erhält man

$$(\mathbf{X}^T \mathbf{X})(\mathbf{X}^T \mathbf{h}) = \left(\sum_{i=1}^m X_i^2 \right) \left(\sum_{i=1}^m X_i h_i \right) = \sum_{i=1}^m h_i X_i^3 + \sum_{i=1}^m \left(X_i h_i \sum_{\substack{j=1 \\ j \neq i}}^m X_j^2 \right). \quad (\text{C.1})$$

Wendet man darauf den Erwartungswert an, erhält man mit der Unabhängigkeit der X_i 's

$$\begin{aligned} E[(\mathbf{X}^T \mathbf{X})(\mathbf{X}^T \mathbf{h})] &= E \left[\sum_{i=1}^m h_i X_i^3 + \sum_{i=1}^m \left(X_i h_i \sum_{\substack{j=1 \\ j \neq i}}^m X_j^2 \right) \right] \\ &= \sum_{i=1}^m h_i E[X_i^3] + \sum_{i=1}^m \left(h_i E[X_i] \sum_{\substack{j=1 \\ j \neq i}}^m E[X_j^2] \right) \\ &= \sum_{i=1}^m h_i (\mu_i^3 + 3\mu_i) + \sum_{i=1}^m \left(h_i \mu_i \sum_{\substack{j=1 \\ j \neq i}}^m (\mu_j^2 + 1) \right) \\ &= \sum_{i=1}^m h_i \mu_i^3 + 3 \sum_{i=1}^m h_i \mu_i + \sum_{i=1}^m h_i \mu_i \sum_{\substack{j=1 \\ j \neq i}}^m \mu_j^2 + (m-1) \sum_{i=1}^m h_i \mu_i \\ &= \sum_{i=1}^m h_i \mu_i^3 + \sum_{i=1}^m h_i \mu_i \sum_{\substack{j=1 \\ j \neq i}}^m \mu_j^2 + (m+2)(\boldsymbol{\mu}^T \mathbf{h}) \\ &= (\boldsymbol{\mu}^T \boldsymbol{\mu})(\boldsymbol{\mu}^T \mathbf{h}) + (m+2)(\boldsymbol{\mu}^T \mathbf{h}) \\ &= (m+2 + \boldsymbol{\mu}^T \boldsymbol{\mu})(\boldsymbol{\mu}^T \mathbf{h}), \end{aligned}$$

wobei im vorletzten Schritt die Gleichung in (C.1) mit $\boldsymbol{\mu}$ anstelle von \mathbf{X} ausgenutzt wurde. \square

C.2. χ^2 -Verteilung

Definition C.2.1 (χ^2 -Verteilung) Seien X_1, \dots, X_n unabhängige $\mathcal{N}(0, 1)$ -verteilte Zufallsvariablen. Dann ist

$$Z = \sum_{i=1}^n X_i^2$$

χ^2 -verteilt mit n Freiheitsgraden, also $Z \sim \chi_n^2$; siehe Krishnamoorthy (2006).

Satz C.2.2 Der Erwartungswert und die Varianz von $Z \sim \chi_n^2$ sind

$$E[Z] = n \quad \text{und} \quad \text{Var}(Z) = 2n;$$

siehe Krishnamoorthy (2006).

Satz C.2.3 Sei $X \sim \chi_n^2$. Dann ist

$$E[X^{-1}] = \frac{1}{n-2} \quad \text{und} \quad E[X^{-2}] = \frac{1}{(n-2)(n-4)}.$$

Beweis. Die Dichte einer χ^2 -verteilten Zufallsvariablen X mit n Freiheitsgraden ist

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad \text{für } x > 0.$$

Für eine Funktion $g(X)$ gilt $E[g(X)] = \int_0^\infty g(x)f(x)dx$. Somit ist

$$\begin{aligned} E[X^{-1}] &= \int_0^\infty \frac{1}{x} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx \\ &= \int_0^\infty \frac{1}{2(\frac{n}{2}-1)} \frac{1}{2^{\frac{n-2}{2}} \Gamma(\frac{n-2}{2})} x^{\frac{n-2}{2}-1} e^{-\frac{x}{2}} dx \\ &= \frac{1}{n-2} \int_0^\infty \frac{1}{2^{\frac{n-2}{2}} \Gamma(\frac{n-2}{2})} x^{\frac{n-2}{2}-1} e^{-\frac{x}{2}} dx \\ &= \frac{1}{n-2}, \end{aligned}$$

wobei im letzten Schritt ausgenutzt wurde, dass sich die Dichte einer χ_{n-2}^2 -verteilten Zufallsvariable zu 1 integriert.

Weiter ist

$$\begin{aligned}
 E[X^{-2}] &= \int_0^\infty \frac{1}{x^2} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx \\
 &= \int_0^\infty \frac{1}{2(\frac{n}{2}-1)} \frac{1}{2^{\frac{n-2}{2}} \Gamma(\frac{n-2}{2})} x^{\frac{n-4}{2}-1} e^{-\frac{x}{2}} dx \\
 &= \int_0^\infty \frac{1}{2(\frac{n}{2}-1)} \frac{1}{2(\frac{n-2}{2}-1)} \frac{1}{2^{\frac{n-4}{2}} \Gamma(\frac{n-4}{2})} x^{\frac{n-4}{2}-1} e^{-\frac{x}{2}} dx \\
 &= \frac{1}{(n-2)(n-4)} \int_0^\infty \frac{1}{2^{\frac{n-4}{2}} \Gamma(\frac{n-4}{2})} x^{\frac{n-4}{2}-1} e^{-\frac{x}{2}} dx \\
 &= \frac{1}{(n-2)(n-4)},
 \end{aligned}$$

wobei im letzten Schritt ausgenutzt wurde, dass sich die Dichte einer χ_{n-4}^2 -verteilten Zufallsvariable zu 1 integriert. \square

Satz C.2.4 Für unabhängige Zufallsvariablen X_1, \dots, X_n mit $X_i \sim \chi_{n_i}^2$, $i = 1, \dots, n$, gilt

$$\sum_{i=1}^n X_i \sim \chi_n^2 \quad \text{mit } n = \sum_{i=1}^n n_i;$$

siehe Krishnamoorthy (2006).

C.3. Nichtzentrale χ^2 -Verteilung

Definition C.3.1 (Nichtzentrale χ^2 -Verteilung) Seien X_1, \dots, X_n unabhängige normalverteilte Zufallsvariablen mit $X_i \sim \mathcal{N}(\mu_i, 1)$, $i = 1, \dots, n$, und sei $\delta = \sum_{i=1}^n \mu_i^2$. Dann folgt

$$Z = \sum_{i=1}^n X_i^2$$

einer nichtzentralen χ^2 -Verteilung mit n Freiheitsgraden und Nichtzentralitätsparameter δ , also $Z \sim \chi_n^2(\delta)$; siehe Krishnamoorthy (2006).

Insbesondere gilt dann für zwei Zufallsvariablen $X \sim \chi_n^2$ und $Y \sim \mathcal{N}(\mu, 1)$, dass

$$X + Y^2 \sim \chi_{n+1}^2(\mu)$$

ist.

Satz C.3.2 Der Erwartungswert und die Varianz von $Z \sim \chi_n^2(\delta)$ sind

$$E[Z] = n + \delta \quad \text{und} \quad \text{Var}(Z) = 2(n + 2\delta);$$

siehe Krishnamoorthy (2006).

C.4. Wishart-Verteilung

Definition C.4.1 (Wishart-Verteilung) Seien $\mathbf{X}_1, \dots, \mathbf{X}_n$ unabhängig und identisch verteilte Zufallsvariablen mit $\mathbf{X}_i \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}), i = 1, \dots, n$, und sei $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$ eine $(n \times p)$ -Matrix, deren Zeilen die Vektoren $\mathbf{X}_1, \dots, \mathbf{X}_n$ sind. Dann ist die $(p \times p)$ -Matrix

$$\mathbf{W} = \mathbf{X}^T \mathbf{X} = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T$$

Wishart-verteilt mit der Matrix $\mathbf{\Sigma}$ und n Freiheitsgraden, also $\mathbf{W} \sim \mathcal{W}_p(n, \mathbf{\Sigma})$; siehe Fahrmeir, Hamerle und Tutz (1996).

Für $p = 1$ und $\mathbf{X}_i \sim \mathcal{N}(0, 1), i = 1, \dots, n$, ist die Wishart-Verteilung mit der χ_n^2 -Verteilung identisch.

Definition C.4.2 (Inverse Wishart-Verteilung) Sei \mathbf{W} eine Matrix, deren Inverse Wishart-verteilt ist, also $\mathbf{W}^{-1} \sim \mathcal{W}_p(n, \mathbf{\Sigma}^{-1})$. Dann ist \mathbf{W} invers Wishart-verteilt, also

$$\mathbf{W} \sim \mathcal{IW}_p(n, \mathbf{\Sigma});$$

siehe Meucci (2009).

Satz C.4.3 Der Erwartungswert einer $\mathcal{IW}_p(n, \mathbf{\Sigma})$ -verteilten Matrix \mathbf{W} ist

$$E[\mathbf{W}] = \frac{1}{n-p-1} \mathbf{\Sigma};$$

siehe Fujikoshi, Ulyanov und Shimizu (2010).

Satz C.4.4 Eine Zufallsvariable X ist genau dann $\mathcal{W}_1(n, \sigma^2)$ -verteilt, wenn $1/\sigma^2 X$ χ_n^2 -verteilt ist.

Beweis. Sei $X \sim \mathcal{W}_1(n, \sigma^2)$. Dann gilt mit $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, wobei Y_i mit $i = 1, \dots, n$ unabhängig und identisch $\mathcal{N}(0, \sigma^2)$ -verteilt sind, dass

$$X = \mathbf{Y}^T \mathbf{Y}$$

ist. Dann ist

$$\frac{1}{\sigma^2} X = \frac{1}{\sigma^2} \mathbf{Y}^T \mathbf{Y} = \sum_{i=1}^n \left(\frac{Y_i}{\sigma} \right)^2$$

mit $Y_i/\sigma \sim \mathcal{N}(0, 1)$ für $i = 1, \dots, n$. Damit ist $1/\sigma^2 X \sim \chi_n^2$. □

C.5. F -Verteilung

Definition C.5.1 (F -Verteilung) Sei $X \sim \chi_m^2$ und $Y \sim \chi_n^2$ und seien X und Y unabhängig. Dann ist die Zufallsvariable

$$F = \frac{\frac{1}{m}X}{\frac{1}{n}Y}$$

F -verteilt mit m und n Freiheitsgraden, also $F \sim F_{m,n}$; siehe Krishnamoorthy (2006).

Satz C.5.2 Der Erwartungswert von $F \sim F_{m,n}$ ist

$$E[F] = \frac{n}{n-2};$$

siehe Krishnamoorthy (2006).

C.6. Nichtzentrale F -Verteilung

Definition C.6.1 (Nichtzentrale F -Verteilung) Sei $X \sim \chi_m^2(\delta)$ und $Y \sim \chi_n^2$ und seien X und Y unabhängig. Dann ist die Zufallsvariable

$$F = \frac{\frac{1}{m}X}{\frac{1}{n}Y}$$

nichtzentral F -verteilt mit m und n Freiheitsgraden und Nichtzentralitätsparameter δ , also $F \sim F_{m,n}(\delta)$; siehe Krishnamoorthy (2006).

C.7. Multivariate t -Verteilung

Definition C.7.1 (Multivariate t -Verteilung) Sei

$$\mathbf{t} = \left(\frac{v}{\nu}\right)^{-\frac{1}{2}} \mathbf{y}$$

mit $v \sim \chi_\nu^2$ und $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}^{-1})$, wobei v und \mathbf{y} unabhängig voneinander sind. Dann ist \mathbf{t} multivariat t -verteilt mit ν Freiheitsgraden; vgl. Dickey (1967).

C.8. Matrix- t -Verteilung

Definition C.8.1 (Matrix- t -Verteilung) Sei \mathbf{T} eine stochastische $(p \times q)$ -Matrix der Form

$$\mathbf{T} = \left(\mathbf{U}^{\frac{1}{2}T}\right)^{-1} \mathbf{Z},$$

wobei $\mathbf{U}^{\frac{1}{2}}\mathbf{U}^{\frac{1}{2}T} = \mathbf{U} \sim \mathcal{W}(m-q, \mathbf{P})$ mit der symmetrischen und positiv definiten Matrix \mathbf{P} und mit $m > p+q-1$ unabhängig von der Matrix \mathbf{Z} ist, deren Zeilenvektoren unabhängig

$\mathcal{N}(\mathbf{0}, \mathbf{Q})$ -verteilt sind, wobei \mathbf{Q} eine symmetrische und positiv definite Matrix ist. Dann folgt \mathbf{T} einer Matrix-t-Verteilung, also $\mathbf{T} \sim \mathcal{T}_\nu(\mathbf{P}, \mathbf{Q}, m)$. Die Anzahl der Freiheitsgrade ist $\nu = m - p - q + 1$. (Bei der Notation der Verteilung handelt es sich um eine Kombination aus der von Dickey (1967) und der von Kan und Smith (2008).)

D. Sätze und Lemmata anderer Autoren

Für manche Beweise aus Kapitel 4 werden Ergebnisse anderer Autoren verwendet. Die Sätze, in denen diese Ergebnisse festgehalten werden, können in diesem Abschnitt nachgelesen werden.

D.1. Theorem 3.2 von Haff (1979)

Satz D.1.1 Sei $\mathbf{S} = (s_{ij})_{i,j=1,\dots,p}$ invers Wishart-verteilt mit k Freiheitsgraden und der Matrix $\mathbf{\Sigma}^{-1} = (\sigma_{ij})_{i,j=1,\dots,p}$, also $\mathbf{S} \sim \mathcal{IW}(k, \mathbf{\Sigma}^{-1})$ und sei $k - p - 3 > 0$. Dann gilt

$$(i) E[s_{ij}] = \sigma_{ij}/(k - p - 1),$$

$$(ii) \text{Cov}(s_{ij}, s_{kl}) = \left(\frac{2}{k-p-1}\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{kj}\right)/((k-p)(k-p-1)(k-p-3)).$$

D.2. Theoreme und Korollar von Muirhead (1982)

Satz D.2.1 (Theorem 3.2.11) Sei $\mathbf{A} \sim \mathcal{W}_m(n, \mathbf{\Sigma})$ und \mathbf{M} eine $(k \times m)$ -Matrix mit Rang k . Dann ist

$$(\mathbf{M}\mathbf{A}^{-1}\mathbf{M}^T)^{-1} \sim \mathcal{W}_k(n - m + k, (\mathbf{M}\mathbf{\Sigma}^{-1}\mathbf{M}^T)^{-1}).$$

Satz D.2.2 (Theorem 3.2.10 und Korollar 3.2.6) Sei $\mathbf{A} \sim \mathcal{W}_m(n, \mathbf{\Sigma})$, wobei sich \mathbf{A} und $\mathbf{\Sigma}$ folgendermaßen in Blöcke einteilen lassen:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix},$$

wobei \mathbf{A}_{11} und $\mathbf{\Sigma}_{11}$ $(k \times k)$ -Matrizen sind. Dann ist $\mathbf{A}_{11} \sim \mathcal{W}_k(n, \mathbf{\Sigma}_{11})$. Sei weiter $\mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$ und $\mathbf{\Sigma}_{11.2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$. Dann gilt

$$(i) \mathbf{A}_{11.2} \sim \mathcal{W}_k(n - m + k, \mathbf{\Sigma}_{11.2}) \text{ ist unabhängig von } \mathbf{A}_{12} \text{ und } \mathbf{A}_{22},$$

$$(ii) \text{ die bedingte Verteilung von } \mathbf{A}_{12} \text{ gegeben } \mathbf{A}_{22} \text{ ist } \mathcal{N}(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{A}_{22}, \mathbf{\Sigma}_{11.2} \otimes \mathbf{A}_{22}),$$

$$(iii) \mathbf{A}_{22} \sim \mathcal{W}_{m-k}(n, \mathbf{\Sigma}_{22}).$$

In (ii) bezeichnet \otimes das Kronecker-Produkt, wobei für eine $(m \times n)$ -Matrix \mathbf{B} mit Einträgen b_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$, und eine $(p \times q)$ -Matrix \mathbf{C} das Kronecker-Produkt definiert ist als

$$\mathbf{D} = \mathbf{B} \otimes \mathbf{C} = \begin{pmatrix} b_{11}\mathbf{C} & \cdots & b_{1n}\mathbf{C} \\ \vdots & \ddots & \vdots \\ b_{m1}\mathbf{C} & \cdots & b_{mn}\mathbf{C} \end{pmatrix} \in \mathbb{R}^{mp \times nq}.$$

D.3. Lemma 1 von Sawa (1972)

Lemma D.3.1 *Sei X_1 eine fast überall positive Zufallsvariable und X_2 eine beliebige Zufallsvariable. Angenommen, es existiere eine gemeinsame momenterzeugende Funktion von X_1 und X_2 :*

$$\varphi(\theta_1, \theta_2) = E[\exp(\theta_1 X_1 + \theta_2 X_2)]$$

mit $\theta_1 < \epsilon$ und $|\theta_2| < \epsilon$, wobei ϵ eine beliebige positive Konstante sei. Dann ist das α -te Moment ($\alpha > 0$) von X_2/X_1 gegeben durch

$$\frac{1}{\Gamma(\alpha)} \int_{-\infty}^0 (-\theta_1)^{\alpha-1} \frac{\partial^\alpha \varphi(\theta_1, \theta_2)}{\partial \theta_2^\alpha} \Big|_{\theta_2=0} d\theta_1,$$

vorausgesetzt, es existiert oder es ist unendlich.

E. Sonstiges

E.1. Suffizienz

Definition E.1.1 (Suffizienz) Sei $\{X_1, \dots, X_n\}$ eine Stichprobe aus einer Wahrscheinlichkeitsverteilung mit unbekanntem Parameter θ . Dann ist die Statistik $T = g(X_1, \dots, X_n)$ suffizient für θ , wenn die bedingte Dichte bzw. Verteilungsfunktion von $\{X_1, \dots, X_n\}$ gegeben $T = t$ für keinen Wert von t von θ abhängt; siehe Ramachandran und Tsokos (2009).

Satz E.1.2 (Faktorisierungssatz von Neyman-Fisher) Sei T eine Statistik der Stichprobe $\{X_1, \dots, X_n\}$. Dann ist T genau dann suffizient für θ , wenn sich die von θ abhängige Dichte (bzw. Verteilungsfunktion) $f(x_1, \dots, x_n; \theta)$ schreiben lässt als

$$f(x_1, \dots, x_n; \theta) = g(t, \theta)h(x_1, \dots, x_n), \quad \text{für alle } x_1, \dots, x_n,$$

wobei $g(t, \theta)$ eine Funktion nur in t und θ und $h(x_1, \dots, x_n)$ eine Funktion nur in x_1, \dots, x_n ist; siehe Ramachandran und Tsokos (2009).

Satz E.1.3 Sei $\mathbf{X}_1, \dots, \mathbf{X}_n$ eine Stichprobe einer m -dimensionalen $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -verteilten Zufallsvariablen mit unbekanntem Parameter $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$. Dann sind der empirische Mittelwert $\bar{\mathbf{X}} = 1/n \sum_{i=1}^n \mathbf{X}_i$ und die empirische Kovarianzmatrix $\mathbf{S} = 1/n \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ suffiziente Statistiken für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$.

Beweis. Die gemeinsame Dichte von $\mathbf{X}_1, \dots, \mathbf{X}_n$ ist

$$\begin{aligned} f(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{(2\pi)^{\frac{m}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu})} \\ &= \frac{1}{(2\pi)^{\frac{nm}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu})} \end{aligned}$$

mit $x_i \in \mathbb{R}^m, i = 1, \dots, n$. Aufgrund der Symmetrie von $\boldsymbol{\Sigma}^{-1}$ lässt sich der Exponent folgendermaßen umformen:

$$\begin{aligned} -\frac{1}{2} \sum_{i=1}^n (x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu}) &= -\frac{1}{2} \sum_{i=1}^n \text{tr} \left(\boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu}) (x_i - \boldsymbol{\mu})^T \right) \\ &= -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^n (x_i - \boldsymbol{\mu}) (x_i - \boldsymbol{\mu})^T \right) \\ &= -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \boldsymbol{\mu}) (x_i - \bar{x} + \bar{x} - \boldsymbol{\mu})^T \right) \end{aligned}$$

mit $\bar{x} = 1/n \sum_{i=1}^n x_i$. Wegen

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})(\bar{x} - \boldsymbol{\mu})^T &= \left(\sum_{i=1}^n x_i - n\bar{x} \right) (\bar{x} - \boldsymbol{\mu})^T \\ &= \left(\sum_{i=1}^n x_i - \sum_{i=1}^n x_i \right) (\bar{x} - \boldsymbol{\mu})^T = \mathbf{0}_{m \times m} \end{aligned}$$

gilt weiter

$$\begin{aligned} &-\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \boldsymbol{\mu})(x_i - \bar{x} + \bar{x} - \boldsymbol{\mu})^T \right) \\ &= -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \left(\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T + \sum_{i=1}^n (\bar{x} - \boldsymbol{\mu})(\bar{x} - \boldsymbol{\mu})^T \right) \right) \\ &= -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \left(\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T + n(\bar{x} - \boldsymbol{\mu})(\bar{x} - \boldsymbol{\mu})^T \right) \right). \end{aligned}$$

Die gemeinsame Dichte von $\mathbf{X}_1, \dots, \mathbf{X}_n$ lässt sich somit schreiben als

$$f(x_1, \dots, x_n) = g(\bar{x}, s; \boldsymbol{\mu}, \boldsymbol{\Sigma}) h(x_1, \dots, x_n)$$

mit

$$g(\bar{x}, s; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{n}{2}}} e^{-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}(n \cdot s + n(\bar{x} - \boldsymbol{\mu})(\bar{x} - \boldsymbol{\mu})^T))}$$

und

$$h(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{nm}{2}}},$$

wobei \bar{x} und s der empirische Mittelwert und die empirische Kovarianzmatrix für die Realisationen x_1, \dots, x_n seien. Nach dem Faktorisierungssatz von Neyman-Fisher sind somit $\bar{\mathbf{X}}$ und \mathbf{S} suffiziente Statistiken für $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$. \square

E.2. Optimierung

Die folgenden Definitionen und Sätze sind bei Geiger und Kanzow (2002) zu finden.

Definition E.2.1 (Restringierte und konvexe Optimierungsprobleme) *Das Problem*

$$\min f(x) \quad \text{unter} \quad g_i(x) \leq 0, \quad h_j(x) = 0, \quad i = 1, \dots, m, \quad j = 1, \dots, p, \quad (\text{E.1})$$

wobei $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, m$, und $h_j : \mathbb{R}^n \rightarrow \mathbb{R}$, $j = 1, \dots, p$, stetig differenzierbar seien, ist ein restringiertes Optimierungsproblem. *Das Problem*

$$\min f(x) \quad \text{unter} \quad g_i(x) \leq 0, \quad b_j^T x = \beta_j, \quad i = 1, \dots, m, \quad j = 1, \dots, p, \quad (\text{E.2})$$

wobei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ und $g_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, m$, stetig differenzierbare konvexe Funktionen, $b_j \in \mathbb{R}^n$ gegebene Vektoren und $\beta_j \in \mathbb{R}$ gegebene Skalare für $j = 1, \dots, p$ seien, ist ein konvexes Optimierungsproblem. Ein konvexes Optimierungsproblem ist also ein restringiertes Optimierungsproblem mit konvexer Zielfunktion und konvexen Ungleichungsrestriktionen sowie linearen Gleichungsnebenbedingungen.

Definition E.2.2 (Lagrangefunktion) Sei $\lambda = (\lambda_1, \dots, \lambda_m)^T \in \mathbb{R}^m$ und $\mu = (\mu_1, \dots, \mu_p)^T \in \mathbb{R}^p$. Die durch

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_j(x)$$

definierte Abbildung $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ heißt Lagrangefunktion des restringierten Optimierungsproblems (E.1).

Da die Nebenbedingung $h_j(x) = 0$ äquivalent zu $-h_j(x) = 0$ ist, lässt sich die Lagrangefunktion eines Problems nur mit Gleichungsrestriktionen auch schreiben als

$$L(x, \mu) = f(x) - \sum_{j=1}^p \mu_j h_j(x).$$

Definition E.2.3 (KKT-Bedingungen, KKT-Punkt) Betrachte das Optimierungsproblem (E.1) mit stetig differenzierbaren Funktionen f, g und h . Die Bedingungen

$$\begin{aligned} \nabla_x L(x, \lambda, \gamma) &= 0, \\ h(x) &= 0, \\ \lambda &\geq 0, \quad g(x) \leq 0, \quad \lambda^T g(x) = 0 \end{aligned}$$

heißen Karush-Kuhn-Tucker-Bedingungen (KKT-Bedingungen) des Optimierungsproblems (E.1), wobei

$$\nabla_x L(x, \lambda, \mu) = \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{j=1}^p \mu_j \nabla h_j(x)$$

den Gradienten der Lagrangefunktion L bzgl. x bezeichnet. Jeder Vektor $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$, der den KKT-Bedingungen genügt, heißt Karush-Kuhn-Tucker-Punkt (KKT-Punkt) des Optimierungsproblems (E.1), die Komponenten von λ^* und μ^* werden auch Lagrangemultiplikatoren genannt. (Häufig werden auch λ^* und μ^* selbst als Lagrangemultiplikatoren bezeichnet.)

Die KKT-Bedingungen sind die notwendigen Optimalitätsbedingungen für eine Lösung eines nichtlinearen Optimierungsproblems.

Satz E.2.4 Betrachte das konvexe Optimierungsproblem (E.2), wobei alle Nebenbedingungen durch lineare Funktionen beschrieben seien. Dann ist $x^* \in \mathbb{R}^n$ genau dann ein (lokales=globales) Minimum von (E.2), wenn es Lagrangemultiplikatoren $\lambda^* \in \mathbb{R}^m$ und $\mu^* \in \mathbb{R}^p$ gibt, sodass (x^*, λ^*, μ^*) ein KKT-Punkt von (E.2) ist.

E.3. Konfluente hypergeometrische Funktion und Gammafunktion

Definition E.3.1 (Konfluente hypergeometrische Funktion) *Die konfluente hypergeometrische Funktion ist definiert als*

$${}_1F_1(a, b; c) = \sum_{\nu=0}^{\infty} \frac{(a)_{\nu} c^{\nu}}{(b)_{\nu} \nu!}$$

mit $a, b, c \in \mathbb{C}$, $b \neq 0, -1, -2, \dots$, und $(a)_{\nu} := a(a+1) \cdots (a+\nu-1)$, $(a)_0 := 1$; siehe Luh, Müller, Ponnusamy und Vasundhara (2005).

Satz E.3.2 (Eigenschaften der Gammafunktion) *Die Gammafunktion $\Gamma : \mathbb{R}^+ \rightarrow \mathbb{R}$ besitzt folgende Eigenschaften:*

$$(i) \quad \Gamma(n) = (n-1)! \quad \forall n \in \mathbb{N}, \quad \Gamma(1) = 1$$

$$(ii) \quad \Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!}{n!4^n} \sqrt{\pi} \quad \forall n \in \mathbb{N};$$

vgl. Remmert und Schumacher (2007).

E.4. Zerlegung der Kovarianz

Satz E.4.1 *Für drei Zufallsvariablen X, Y und Z gilt*

$$\text{Cov}(X, Y) = E[\text{Cov}(X, Y|Z)] + \text{Cov}(E[X|Z], E[Y|Z]).$$

Beweis.

$$\begin{aligned} \text{Cov}(X, Y) &= E[XY] - E[X]E[Y] = E[E[XY|Z]] - E[E[X|Z]]E[E[Y|Z]] \\ &= E[E[XY|Z]] - E[E[X|Z]E[Y|Z]] + E[E[X|Z]E[Y|Z]] \\ &\quad - E[E[X|Z]]E[E[Y|Z]] \\ &= E[\text{Cov}(X, Y|Z)] + \text{Cov}(E[X|Z], E[Y|Z]) \end{aligned}$$

□

F. Notationen

m : Anzahl riskanter Assets

n : Anzahl vergangener Renditen eines Assets

r_{it} : Rendite von Asset i zum Zeitpunkt t , $\mathbf{r}_t = (r_{1t}, \dots, r_{mt})^T$

$\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)^T$, $\mu_i = E[r_{i,t+1}]$

$\boldsymbol{\Sigma} = (\sigma_{i,j})_{i,j=1,\dots,m}$, $\sigma_{i,j} = Cov(r_{i,t+1}, r_{j,t+1})$ für $i \neq j$, $\sigma_{ii} = \sigma_i^2 = Var(r_{i,t+1})$

$\mathbf{w} = (w_1, \dots, w_m)^T$, w_i : Portfoliogewicht von Asset i

$$\mu_{Pf} = \mathbf{w}^T \boldsymbol{\mu}$$

$$\sigma_{Pf}^2 = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$$

μ_* : erwartete Zielrendite

λ : Risikoaversionsparameter

$$\gamma = 1/\lambda$$

$\tilde{\lambda}$: Budgetmultiplikator

$\tilde{\gamma}$: Renditemultiplikator

$$\bar{\mathbf{r}} = 1/n \sum_{t=1}^n \mathbf{r}_t = \hat{\boldsymbol{\mu}}_{emp}, \quad \text{empirischer Mittelwert}$$

$$\mathbf{S} = 1/n \sum_{t=1}^n (\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^T = \hat{\boldsymbol{\Sigma}}_{emp}, \quad \text{empirische Kovarianzmatrix}$$

$$A = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \mu_g / \sigma_g^2$$

$$B = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \psi^2 + \mu_g^2 / \sigma_g^2$$

$$C = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = 1/\sigma_g^2$$

$$D = BC - A^2$$

$$\psi^2 = B - A^2/C$$

$$\mu_g = A/C$$

$$\sigma_g^2 = 1/C$$

$$\mathbf{M} = (\boldsymbol{\mu}, \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}, \mathbf{1}) = \begin{pmatrix} B & A \\ A & C \end{pmatrix}$$

$$\tilde{\mathbf{M}} = (\hat{\boldsymbol{\mu}}, \mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\mu}}, \mathbf{1})$$

$$\widehat{\mathbf{M}} = (\widehat{\boldsymbol{\mu}}, \mathbf{1})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1}) = \begin{pmatrix} \widehat{B} & \widehat{A} \\ \widehat{A} & \widehat{C} \end{pmatrix}$$

$$\mathbf{M}^{-1} = \begin{pmatrix} 1/\psi^2 & -\mu_g/\psi^2 \\ -\mu_g/\psi^2 & \sigma_g^2 + \mu_g^2/\psi^2 \end{pmatrix}$$

$$\mathbf{w}^* = \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}, \mathbf{1}) \mathbf{M}^{-1} (\mu_*, 1)^T, \quad \text{Lösung des Mean-Variance-Optimierungsproblems}$$

$$\widehat{\mathbf{w}}^* = \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1}) \widehat{\mathbf{M}}^{-1} (\mu_*, 1)^T, \quad \text{Lösung des Problems mit } \widehat{\boldsymbol{\Sigma}} \text{ und } \widehat{\boldsymbol{\mu}}$$

$$\widetilde{\mathbf{w}} = \boldsymbol{\Sigma}^{-1} (\widehat{\boldsymbol{\mu}}, \mathbf{1}) \widetilde{\mathbf{M}}^{-1} (\mu_*, 1)^T, \quad \text{Lösung des Problems mit } \boldsymbol{\Sigma} \text{ und } \widehat{\boldsymbol{\mu}}$$

$$y = \sqrt{n} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \widehat{\boldsymbol{\mu}} / \sqrt{C} \sim \mathcal{N}(\mu_y, 1)$$

$$\mathbf{z} = \sqrt{n} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \widehat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}_z, \mathbf{I}_{m-1})$$

$$\mu_y = \sqrt{n} \mu_g / \sigma_g$$

$$\boldsymbol{\mu}_z = \sqrt{n} \mathbf{P}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}$$

$$\mathbf{P} \in \mathbb{R}^{m \times (m-1)}, \quad \text{orthonormale Spalten, orthogonal zu } \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1}$$

$$u = \mathbf{z}^T \mathbf{z} \sim \chi_{m-1}^2(n\psi^2)$$

$$v = u / \widehat{\psi}^2 \sim \chi_{n-m+1}^2, \quad v = n \widehat{\sigma}_p^2 / \check{\sigma}_p^2 \sim \chi_{n-m+1}^2$$

$$s = u/v \sim (m-1)/(n-m+1) F_{m-1, n-m+1}(n\psi^2)$$

$$q = \widehat{\sigma}_g^2 n / \sigma_g^2 \sim \chi_{n-m}^2$$

$$x = (y + \widehat{\psi} y_1 - \sqrt{n} \mu_g / \sigma_g) / (1 + \widehat{\psi}^2)^{\frac{1}{2}} \sim \mathcal{N}(0, 1)$$

$$\delta = (\mu_* - \mu_g) / \sigma_g$$

$$\widetilde{y} = \sqrt{n} \mu_* / \sigma_g - y \sim \mathcal{N}(\sqrt{n} \delta, 1)$$

$$h = 1 + n \delta^2 = E[\widetilde{y}^2]$$

$$\sigma_p^2 = (C \mu_*^2 - 2A \mu_* + B) / (BC - A^2) = (\mu_*, 1) \mathbf{M}^{-1} (\mu_*, 1)^T = \sigma_g^2 + (\mu_* - \mu_g)^2 / \psi^2$$

$$= \sigma_g^2 (1 + \delta^2 / \psi^2), \quad \text{wahre Minimum-Varianz-Kurve}$$

$$\widehat{\sigma}_p^2 = \widehat{\mathbf{w}}^{*T} \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{w}}^* = (\widehat{C} \mu_*^2 - 2\widehat{A} \mu_* + \widehat{B}) / (\widehat{D}) = (\mu_*, 1) \widehat{\mathbf{M}}^{-1} (\mu_*, 1)^T$$

$$= \widehat{\sigma}_g^2 + (\mu_* - \widehat{\mu}_g)^2 / \widehat{\psi}^2, \quad \text{in-sample Varianz, empirische Minimum-Varianz-Kurve}$$

$$\check{\sigma}_p^2 = E[\widehat{\sigma}_p^2], \quad \text{in-sample Kurve}$$

$$\widehat{\sigma}_{pa}^2 = n \widehat{\sigma}_g^2 / (n-m) + \max(1 / \widehat{\psi}_a^2 ((\mu_* - \widehat{\mu}_g)^2 - \widehat{\sigma}_g^2 (1 + \widehat{\psi}^2) / (n-m)), -n \widehat{\sigma}_g^2 / (n-m)),$$

$$\text{verbesserter Schätzer der Minimum-Varianz-Kurve}$$

$$\widetilde{\mu}_p = \widehat{\mathbf{w}}^{*T} \boldsymbol{\mu}, \quad \text{out-of-sample erwartete Rendite}$$

$$\widetilde{\sigma}_p^2 = \widehat{\mathbf{w}}^{*T} \boldsymbol{\Sigma} \widehat{\mathbf{w}}^*, \quad \text{out-of-sample Varianz}$$

$$\underline{\mu}_p = E[\widetilde{\mu}_p]$$

$$\underline{\sigma}_p^2 = E[\widetilde{\sigma}_p^2], \quad \text{bildet zusammen mit } \underline{\mu}_p \text{ die out-of-sample Kurve}$$

$$\mathbf{X} = \boldsymbol{\Sigma}^{-\frac{1}{2}} (\widehat{\boldsymbol{\mu}}, \mathbf{1}), \quad \mathbf{X}^T \mathbf{X} = \widetilde{\mathbf{M}}$$

$$\mathbf{W} = n\boldsymbol{\Sigma}^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-\frac{1}{2}} \sim \mathcal{W}_m(n-1, \mathbf{I}_m)$$

$\mathbf{Q} \in \mathbb{R}^{m \times (m-2)}$, orthonormale Spalten, orthogonal zu \mathbf{X}

$$\check{\mu}_p = \widetilde{\mathbf{w}}^T \boldsymbol{\mu}$$

$$\check{\sigma}_p^2 = \widetilde{\mathbf{w}}^T \boldsymbol{\Sigma} \widetilde{\mathbf{w}} = \sigma_g^2(1 + \widetilde{y}^2/u) = (\mu_*, 1) \widetilde{\mathbf{M}}^{-1}(\mu_*, 1)^T$$

$$\phi = n\psi^2/2 \int_0^1 e^{\frac{n\psi^2(y-1)}{2}} y^{\frac{m-3}{2}} dy = E[\mathbf{z}^T \boldsymbol{\mu}_z / (\mathbf{z}^T \mathbf{z})]$$

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}_{m-2}, \mathbf{I}_{m-2})$$

$$\widetilde{r} \sim \chi_{n-m+2}^2$$

$$\mathbf{Y} = \mathbf{Q}^T \mathbf{W}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mu_*, 1)^T = \frac{\check{\sigma}_p \mathbf{x}}{\widetilde{r}^{\frac{1}{2}}}$$

$$\widetilde{d} = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{Q} \mathbf{Q}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} (\mathbf{I}_m - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu}$$

$$\widetilde{x} = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{Q} \mathbf{x} / \widetilde{d}^{\frac{1}{2}} \sim \mathcal{N}(0, 1)$$

$$\widetilde{q} = \mathbf{x}^T \mathbf{x} - \widetilde{x}^2 \sim \chi_{m-3}^2$$

$$\widetilde{z} = \mathbf{z}^T \boldsymbol{\mu}_z / (\boldsymbol{\mu}_z^T \boldsymbol{\mu}_z)^{\frac{1}{2}} = \mathbf{z}^T \boldsymbol{\mu}_z / (\sqrt{n}\psi) \sim \mathcal{N}(\sqrt{n}\psi, 1)$$

$$\widetilde{u} \sim \chi_{m-2}^2, \quad u = \widetilde{z}^2 + \widetilde{u} \sim \chi_{m-1}^2(n\psi^2)$$

$$\widetilde{t} = \widetilde{x} / \widetilde{r}^{\frac{1}{2}}$$

Danksagung

Mein herzlichster Dank gilt Herrn Prof. Dr. Hajo Holzmann für seine umfangreiche Betreuung während meiner Masterarbeit. Besonderer Dank geht auch an Florian Schwaiger für seine wertvollen Anregungen. Bedanken möchte ich mich außerdem bei Daniela Thorn und Robert Gerstberger, die meine Arbeit Korrektur gelesen haben.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich diese Masterarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Stellen meiner Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht. Diese Arbeit hat in dieser oder einer ähnlichen Form noch nicht im Rahmen einer anderen Prüfung vorgelegen.

Marburg, 25. Januar 2013

Unterschrift _____