

Schätzung von hochdimensionalen Kovarianzmatrizen in Faktormodellen

Masterarbeit

vorgelegt von
Anna Maria Leister

angefertigt am
Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg

betreut von
Prof. Dr. Holzmann

Marburg, 17. Januar 2013

Inhaltsverzeichnis

Notation	IV
Einleitung	1
1. Faktormodelle	3
2. Kovarianzmatrixschätzung in Faktormodellen	8
2.1. Schätzung der Komponenten des Faktormodells	8
2.2. Annahmen und Notation	9
2.3. Unabhängige Faktoren	15
2.4. Stationäre Faktoren	21
3. Auswirkungen der Kovarianzmatrixschätzung auf Anwendungen in der Finanzwirtschaft	26
3.1. Portfoliotheorie	26
3.1.1. Allgemeiner Markowitzfall	27
3.1.2. Das globale varianzminimale Portfolio	33
3.2. Risikomanagement	35
4. Simulation	38
4.1. Kalibrierung	38
4.2. Simulation des Faktormodells	39
4.3. Auswertung	41
5. Schätzung der Autokovarianzmatrix im Faktormodell	47
Zusammenfassung und Ausblick	55
A. Hilfssätze und Beweise	IV
A.1. Kovarianzmatrizen	IV
A.2. Bedingte Erwartungen	IX
A.3. Exponentielle Ungleichungen	X
B. Lineare Algebra	XII
B.1. Spureigenschaften	XII
B.2. Normen	XII
B.3. Matrixtheorie	XIV

B.4. Eigenwerte	XV
B.5. Hutmatrix	XV
C. Herleitung der Konvergenzraten für Abschnitt 2.3	XVI
D. Herleitung der Konvergenzraten für Abschnitt 2.4	XL
Literaturverzeichnis	XLVI

Notation

$(\mathbf{A})_{ij}$	Element in Zeile i , Spalte j der Matrix $\mathbf{A}$.
$E(X)$	Erwartungswert der Zufallsvariable X .
$E(X \mathcal{C})$	Bedingte Erwartung der Zufallsvariable X gegeben $\mathcal{C}$.
O, o	Landausymbole (deterministisch).
O_P, o_P	Landausymbole (stochastisch).
$\Gamma(\alpha, \beta)$	Gammaverteilung mit Parametern α und β .
$\bar{\mathbf{x}}$	Arithmetisches Mittel der Einträge von $\mathbf{x} = (x_1, \dots, x_n)$.
$\circ$	Hadamard Produkt - elementweise Multiplikation von Matrizen.
$\text{cov}(X, Y)$	Kovarianz der Zufallsvariablen X und Y .
$\det(\mathbf{A})$	Determinante der Matrix $\mathbf{A}$.
$\text{diag}(\cdot)$	Für eine Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$: $\text{diag}(\mathbf{A}) := \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & a_{nn} \end{pmatrix}$. Für $a_1, \dots, a_n \in \mathbb{R}$: $\text{diag}(a_1, \dots, a_n) = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & a_n \end{pmatrix}$.
$\lambda_j(\mathbf{A})$	j -größter Eigenwert der Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$.
$\mathbb{R}$	Körper der reellen Zahlen.
$\mathbb{Z}$	Menge der ganzen Zahlen.
$\mathcal{N}(\mu, \sigma^2)$	Normalverteilung mit Erwartungswert μ und Varianz σ^2 .
$\mathbf{1}$	Einsvektor $(1, \dots, 1)^T$.
$\mathbf{A} \geq \mathbf{B}$	Für symmetrisch positiv semidefinite Matrizen $\mathbf{A}$ und $\mathbf{B}$: $\mathbf{A} - \mathbf{B}$ ist positiv semidefinit.
$\mathbf{I}_p$	Einheitsmatrix der Dimension $p \times p$.
$\text{rg}(\mathbf{A})$	Rang der Matrix $\mathbf{A}$.
$\xrightarrow{P}$	Stochastische Konvergenz.
$\text{tr}(\mathbf{A})$	Spur der Matrix $\mathbf{A}$.
$\text{var}(X)$	Varianz der Zufallsvariable X .
$\text{mean}(a_1, \dots, a_n)$	Mittelwert von $a_1, \dots, a_n$.
$\text{sd}(a_1, \dots, a_n)$	Standardabweichung von $a_1, \dots, a_n$.

Einleitung

In der heutigen Zeit sind in vielen Anwendungen große Datenmengen verfügbar, die für die Gewinnung von Informationen und das Verständnis von Prozessen genutzt werden können. Besonders in naturwissenschaftlichen Disziplinen kann durch verbesserte Verfahren eine immer größer werdende Anzahl an Daten gemessen und beobachtet werden. Beispielsweise ist in der Signalverarbeitung, durch die Digitalisierung, das Datenaufkommen für Musik, Videos und Bilder stark gewachsen, in der Klimatologie werden hochdimensionale Daten genutzt um neue Vorhersage- und Klimamodelle zu entwickeln, in der Molekularbiologie sind immer detailliertere genetische Informationen für die Forschung verfügbar. Um mit diesen Datenmengen umgehen zu können werden zuverlässige Methoden aus der multivariaten Statistik benötigt, denn häufig bedeutet eine große Datenmenge nicht nur die Verfügbarkeit von Beobachtungen über einen langen Zeitraum hinweg, sondern darüber hinaus eine große Anzahl von Messungen und Daten für jeden Beobachtungszeitpunkt.

In dieser Arbeit wollen wir eine Anwendung aus der Finanzwelt betrachten, in der heutzutage oft hochdimensionale Portfolios, also eine große Anzahl an Investitionsobjekten und deren Verhalten untereinander betrachtet werden müssen. Besonders für das Risikomanagement ist es wichtig, zuverlässige Aussagen über die einzelnen Positionen und über ihren Beitrag zum Gesamtrisiko treffen zu können, um Vorhersagen für das Portfoliorisiko zu ermöglichen. Zu diesem Zweck werden die Kovarianzmatrizen der Portfolios betrachtet. Da diese jedoch nicht beobachtbar sind, müssen sie aus den vorhandenen Daten geschätzt werden.

Bei der Wahl der Schätzmethode ist es einerseits wichtig den Rechenaufwand zu betrachten, andererseits muss die Güte der Schätzung berücksichtigt werden. Häufig wird für die Schätzung die Stichprobenkovarianzmatrix genutzt. Allerdings tritt, sobald die Dimension der Matrix in der gleichen Größenordnung vorliegt wie die Stichprobengröße, das Problem der Singularität auf. Da es aber insbesondere für die Portfolioselektion wichtig ist, eine gute Schätzung für die Inverse der Kovarianzmatrix zu haben, geht man häufig zu anderen Schätzmethoden über.

Weiterhin ist es für die Anwendung in der Finanzwirtschaft wichtig eine Schätzung zu erhalten, die positiv semidefinit ist, da andernfalls Portfolios mit negativen Varianzen bestimmt werden könnten. Den Hintergrund werden wir in Kapitel 3 erläutern.

Um dem Problem der großen Dimensionen entgegenzutreten, werden häufig Strukturannahmen getroffen. Ein beliebtes Beispiel hierfür sind Faktormodelle, die vorwiegend in der Ökonomie und der Finanzwissenschaft genutzt werden. Wir werden diesen Ansatz aufgreifen und die Rendite eines hochdimensionalen Portfolios durch eine möglichst kleine Anzahl von Faktoren erklären. Je weniger gemeinsame Faktoren für die Erklärung der Rendite nötig sind, desto größer ist der Effekt der Dimensionsreduktion durch das

Faktormodell.

Dieser Ansatz wurde intensiv von Fama und French (1993) untersucht und sie fanden heraus, dass die Rendite eines Portfolios bestehend aus Aktien und Anleihen schon durch ein Faktormodell mit fünf Faktoren gut erklärt werden kann.

Wir werden die Entstehung der Faktormodelle in Kapitel 1 motivieren, verschiedene Anpassungsmöglichkeiten erläutern und die Struktur dann für die Schätzung der Kovarianzmatrix nutzen. Der Vorteil der Faktormodelle liegt darin, dass nicht die Kovarianzmatrix des Portfolios direkt geschätzt wird, sondern die Komponenten des Faktormodells, aus denen anschließend die Kovarianzmatrix rekonstruiert werden kann.

Nach der Einführung der Faktormodelle in Kapitel 1, nutzen wir diese in Kapitel 2 um die Kovarianzmatrix eines hochdimensionalen Portfolios zu schätzen. Wir wenden dazu den Ansatz aus Fan et al. (2007) an und vergleichen die Schätzung der Kovarianzmatrix aus dem Faktormodell mit der Stichprobenkovarianzmatrix unter der Annahme, dass die Dimension des Portfolios und die Anzahl der Faktoren mit der Stichprobengröße wachsen können. Als Vergleichswert betrachten wir die Konvergenzraten der beiden Schätzer unter verschiedenen Normen. Außerdem analysieren wir unterschiedliche Annahmen an die Faktoren des Faktormodells. Im einfachsten Fall gehen wir von unabhängig identisch verteilten Faktoren aus. Da in der Praxis aber häufig makroökonomische oder branchenspezifische Daten als Faktoren genutzt werden, gehen wir zu der Annahme über, dass die Faktoren einer stationären Zeitreihe entstammen.

Es stellt sich heraus, dass die Auswirkung dieser veränderten Annahme auf das asymptotische Verhalten der Kovarianzmatrixschätzer nur gering ist. In beiden Ansätzen zeigen wir, dass der Vorteil durch die Schätzung der Kovarianzmatrix mithilfe des Faktormodells gegenüber der Stichprobenkovarianzmatrix nur gering ist. Dahingegen kann die Inverse der Kovarianzmatrix mithilfe des Faktormodells deutlich besser geschätzt werden als durch Invertieren der Stichprobenkovarianzmatrix.

In Kapitel 3 untersuchen wir, wie sich diese Resultate auf Anwendungen in der Finanzwirtschaft auswirken. Dazu betrachten wir Ansätze der Portfolioselektion für die Konstruktion optimaler Portfolios und die Berechnung des Risikos bei einem vorgegebenen Portfolio. In beiden Anwendungsgebieten wird die Kovarianzmatrix verwendet und wir analysieren den Einfluss der Schätzung der Kovarianzmatrix in beiden Bereichen. Es stellt sich heraus, dass besonders die Portfolioselektion von der Verwendung von Faktormodellen profitiert, da dabei die Schätzung der inversen Kovarianzmatrix eingeht. Im Risikomanagement wird hingegen die Kovarianzmatrix direkt angewendet, daher ist der Nutzen des Faktormodells nur gering.

In Kapitel 4 führen wir mithilfe des Dreifaktormodells für Aktienrenditen aus Fama und French (1993) eine Simulation unserer Untersuchungen mithilfe eines realen Datensatzes durch. Dazu nutzen wir die Software R (<http://www.R-project.org/>).

In Kapitel 5 untersuchen wir, ob sich der Faktormodellansatz unter der Annahme von stationären Faktoren auch für die Schätzung der Autokovarianzmatrix zu einem Zeitabstand h eignet. Die Resultate in diesem Abschnitt vergleichen wir dann für $h = 0$ mit den Ergebnissen aus Kapitel 2 für die Kovarianzmatrix.

1. Faktormodelle

Das Konzept der Faktormodelle wurde in Ross (1976) als neuartiger Ansatz für die Bewertung von Kapitalanlagen eingeführt. Insbesondere sollen sie eine Alternative zu dem von Ross kritisierten *Capital Asset Pricing Model* (CAPM) darstellen.

Das CAPM, basierend auf dem Mean-Variance Ansatz von Markowitz (1952), ist ein Modell zur Bewertung einzelner Kapitalanlagen unter den Annahmen eines Marktgleichgewichts, normalverteilter Wertpapierrenditen und quadratischer Nutzenfunktionen von Investoren. Um die erwartete Rendite einer Kapitalanlage mithilfe des CAPMs zu bestimmen, muss ein Marktportfolio bekannt sein, anhand dessen die Kapitalanlage gemessen werden kann. Die erwartete Rendite wird als Linearkombination der risikolosen Rendite (R_F) und der erwarteten Überschussrendite des Marktportfolios bestimmt. Letztere wird mit dem sogenannten Betafaktor (β_{iM}) gewichtet, der die Schwankung der Kapitalanlage im Vergleich zum Marktportfolio beschreibt:

$$E(R_i) = R_F + \beta_{iM}(E(R_M) - R_F)$$
$$\beta_{iM} = \frac{\text{cov}(R_i, R_M)}{\text{var}(R_M)}.$$

Dabei bezeichne R_i die Rendite der Anlage i für $i \in \{1, \dots, p, M\}$, wobei der Index M für das Marktportfolio steht.

Das von Ross neu eingeführte Konzept basiert auf einem Arbitrageargument und trägt daher den Namen *Arbitrage Pricing Theory* (APT). Ein besonderer Vorteil gegenüber dem CAPM ist, dass die Annahme des Marktgleichgewichts fallen gelassen werden kann. Auch die Bestimmung eines Marktportfolios ist nicht mehr nötig, da die Rendite der Kapitalanlagen durch gemeinsame Marktfaktoren approximiert wird.

Der Ausgangspunkt für die APT ist die Darstellung der Rendite einer Kapitalanlage R_i als Linearkombination der Faktorausprägungen (f_j):

$$R_i = a_i + \mathbf{b}_i \mathbf{f} + \varepsilon_i,$$

mit einem Intercept a_i , einem Störterm ε_i und anlagespezifischen Faktorladungen

$$\mathbf{b}_i = (b_{i1}, \dots, b_{iK}), \quad i = 1, \dots, p,$$

sowie dem Vektor der Faktorausprägungen

$$\mathbf{f} = (f_1, \dots, f_K)^T.$$

Dabei bezeichne p die Anzahl der Kapitalanlagen und K die Anzahl der gemeinsamen Faktoren. Für den Störterm müssen die Bedingungen

$$\begin{aligned} E(\varepsilon_i|\mathbf{f}) &= 0 \text{ und} \\ \text{var}(\varepsilon_i) &= \sigma_i^2 = \sigma^2 < \infty \end{aligned}$$

gelten. Unter der Annahme der Arbitragefreiheit entwickelte Ross (1976, 1977) aus dieser Darstellung der Renditen einen approximativen linearen Zusammenhang zwischen der erwarteten Überschussrendite einer Kapitalanlage und den Risikoprämien $\tilde{f}_j$ der gemeinsamen Faktoren:

$$(E(R_i) - R_F) \approx \mathbf{b}_i \tilde{\mathbf{f}}.$$

Basierend auf der APT wurden verschiedene Arten von Faktormodellen für unterschiedliche Anwendungen entwickelt.

In dieser Arbeit nutzen wir Faktormodelle, um die Überschussrenditen eines Portfolios durch gemeinsame Marktfaktoren zu erklären. Dazu betrachten wir für $i = 1, \dots, p$ die Überschussrendite Y_i einer Kapitalanlage i und bezeichnen die Dimension unseres Portfolios, also die Anzahl der Kapitalanlagen, mit p . Als Faktoren in unserem Modell dienen uns K Überschussrenditen von gemeinsamen Marktfaktoren. Diese sollen die Überschussrenditen der p Kapitalanlagen erklären. Dabei ist die Wirkungsweise der Faktoren anlagespezifisch und wird durch Faktorladungen b_{ij} , $i = 1, \dots, p$, $j = 1, \dots, K$ bestimmt. Das Faktormodell für unser Portfolio besitzt dann die folgende Form

$$\mathbf{y} = \mathbf{B}\mathbf{f} + \boldsymbol{\varepsilon}. \quad (1.0.1)$$

Dabei ist

$$\begin{aligned} \mathbf{y} &= (Y_1, \dots, Y_p)^T \in \mathbb{R}^p \text{ der Vektor der Überschussrenditen unseres Portfolios,} \\ \mathbf{b}_i &= (b_{i1}, \dots, b_{iK}) \in \mathbb{R}^K \text{ der Vektor der Faktorladungen für Kapitalanlage } i, \\ \mathbf{B} &= (\mathbf{b}_1, \dots, \mathbf{b}_p)^T \in \mathbb{R}^{p \times K} \text{ die Matrix aller Faktorladungen des Portfolios,} \\ \mathbf{f} &= (f_1, \dots, f_K)^T \in \mathbb{R}^K \text{ der Vektor der Faktoren und} \\ \boldsymbol{\varepsilon} &= (\varepsilon_1, \dots, \varepsilon_p)^T \in \mathbb{R}^p \text{ der Vektor der kapitalanlagespezifischen Fehlerterme.} \end{aligned}$$

Für diesen gilt

$$\begin{aligned} E(\boldsymbol{\varepsilon}|\mathbf{f}) &= \mathbf{0} \text{ und} \\ \text{cov}(\boldsymbol{\varepsilon}|\mathbf{f}) &= \boldsymbol{\Sigma}_0. \end{aligned} \quad (1.0.2)$$

Weiterhin sind die Faktoren und die Fehlerterme unkorreliert, das heißt

$$\text{cov}(\boldsymbol{\varepsilon}, \mathbf{f}) = \mathbf{0}.$$

Dies ist eine zentrale Bedingung für diese Arbeit, da sie es ermöglicht die Kovarianzmatrix des Portfolios mithilfe der Darstellung (A.2.4) in folgende Komponenten zu zerlegen

$$\begin{aligned}\Sigma &:= \text{cov}(\mathbf{y}) = \text{cov}(\mathbf{B}\mathbf{f} + \boldsymbol{\varepsilon}) = \text{cov}(\mathbf{B}\mathbf{f}) + \text{cov}(\boldsymbol{\varepsilon}) + 2\mathbf{B}\text{cov}(\mathbf{f}, \boldsymbol{\varepsilon}) \\ &= \mathbf{B}\text{cov}(\mathbf{f})\mathbf{B}^T + E(\text{cov}(\boldsymbol{\varepsilon}|\mathbf{f})) + \text{cov}(E(\boldsymbol{\varepsilon}|\mathbf{f})) = \mathbf{B}\text{cov}(\mathbf{f})\mathbf{B}^T + \Sigma_0.\end{aligned}\quad (1.0.3)$$

In dieser Darstellung wird der Vorteil der Dimensionsreduktion deutlich, solange die Anzahl der Faktoren im Vergleich zur Dimension des Portfolios klein ist. Wegen der Symmetrie der Kovarianzmatrix sind bei direkter Schätzung $\frac{p(p+1)}{2}$ Parameter zu schätzen. Bei der Betrachtung des Faktormodells müssen die Einträge der $p \times K$ Matrix der Faktorladungen, die symmetrische $K \times K$ Kovarianzmatrix der Faktoren und die Einträge der Fehlerkovarianzmatrix geschätzt werden. Häufig wird diese als diagonal (vgl. Fan et al., 2007) oder dünn besetzte Matrix (vgl. Fan et al., 2011) angenommen, daher reduziert sich die Anzahl der zu schätzenden Parameter deutlich.

Für die datenbasierte Anpassung eines Faktormodells benötigen wir n beobachtete Daten des Portfolios, $\mathbf{y}_1, \dots, \mathbf{y}_n$, die durch das Faktormodell generiert werden sollen. Häufig unterscheidet man drei verschiedene Typen von Faktormodellen. In Tsay (2010) und Zivot und Wang (2006) werden sie als makroökonomische, fundamentale und statistische Faktormodelle beschrieben.

In makroökonomischen Faktormodellen werden neben den Portfoliodaten auch die Faktorausprägungen beobachtet. Dabei kann es sich beispielsweise um Aktienindizes, Zinsen, Renditen oder andere makroökonomische Größen handeln. Die Modellanpassung erfolgt dann meist durch eine Kleinst-Quadrat-Schätzung für die Matrix der Faktorladungen. Die Schätzung kann für jede der p Kapitalanlagen einzeln oder als multivariate Regression für das gesamte Portfolio gleichzeitig durchgeführt werden. Da wir diesen Ansatz in den folgenden Kapiteln verfolgen werden, verweisen wir für die Schätzung auf Kapitel 2.

In fundamentalen Faktormodellen werden beobachtbare, kapitalanlagespezifische Fundamentalwerte genutzt, um das Modell anzupassen. Dabei kann es sich beispielsweise um Bilanzwerte, Wachstumswerte oder Branchenklassifikationen handeln. Innerhalb der fundamentalen Faktormodelle gibt es zwei verschiedene Ansätze.

Im BARRA Ansatz (vgl. Burmeister et al., 1994) werden die gewählten Fundamentalwerte genutzt, um zeitinvariante Faktorladungen zu konstruieren. Die Faktoren werden dann durch Regression geschätzt.

Im Fama–French Ansatz wird für jeden Fundamentalwert ein Hedgeportfolio gebildet. Die Renditen dieser Hedgeportfolios werden als Faktoren im Faktormodell genutzt, darauf basierend werden analog zum makroökonomischen Faktormodell die Faktorladungen durch Regression geschätzt. In Fama und French (1993) wird ein Modell beschrieben, das zwei Fundamentalwerte nutzt: die Unternehmensgröße (Börsenkurs = Anzahl Aktien * Aktienkurs) und das Verhältnis von Bilanzkurs zu Börsenkurs. Die Faktoren, die aus diesen Fundamentalwerten gebildet werden, nutzen wir für die Simulation in Kapitel 4.

In statistischen Faktormodellen wird die Modellanpassung ohne zusätzlich beobachtete

Daten durchgeführt. Dazu wendet man häufig eine Faktoranalyse oder eine Hauptkomponentenanalyse an.

Bei der Faktoranalyse gilt die Annahme, dass die Beobachtungen $\mathbf{y}_1, \dots, \mathbf{y}_n$ einer orthogonalen Faktorstruktur folgen. Dies bedeutet, dass für $i = 1, \dots, n$ ein Faktormodell folgender Form gilt

$$\begin{aligned}\mathbf{y}_i &= \mathbf{B}\mathbf{f}_i + \boldsymbol{\varepsilon}_i \\ \text{cov}(\mathbf{f}_i, \boldsymbol{\varepsilon}_i) &= \mathbf{0} \\ E(\mathbf{f}_i) = E(\boldsymbol{\varepsilon}_i) &= \mathbf{0} \\ \text{var}(\mathbf{f}_i) &= \mathbf{I}_K \\ \text{var}(\boldsymbol{\varepsilon}_i) &= \text{diag}(\sigma_1^2, \dots, \sigma_p^2).\end{aligned}$$

Diese Darstellung kann durch Transformation des Faktormodells (1.0.1) immer erreicht werden, indem die Faktoren und die Matrix der Faktorladungen wie folgt transformiert werden

$$\begin{aligned}\tilde{\mathbf{f}} &= \text{cov}(\mathbf{f})^{-\frac{1}{2}} \mathbf{f} \\ \tilde{\mathbf{B}} &= \mathbf{B} \text{cov}(\mathbf{f})^{\frac{1}{2}}.\end{aligned}$$

Die Faktoren und Faktorladungen sind nicht eindeutig bestimmt, da mit jeder beliebigen orthogonalen Matrix $\mathbf{M}$ Transformationen vorgenommen werden können, sodass das Faktormodell

$$\mathbf{y}_i = \mathbf{BMM}^T \mathbf{f}_i + \boldsymbol{\varepsilon}_i$$

gilt. Ausgehend von der orthogonalen Faktorstruktur werden in der Faktoranalyse in einem ersten Schritt Maximum Likelihood Schätzungen für die Faktorladungen und die Fehlerkovarianzmatrix $\mathbf{D}$ unter der Annahme unabhängig identisch normalverteilter Überschussrenditen durchgeführt. Basierend auf diesen Schätzungen können dann die Faktoren durch das verallgemeinerte Kleinste-Quadrate Verfahren bestimmt werden. Da die Darstellung wie oben beschrieben nicht eindeutig ist, kann in einem dritten Schritt eine Transformation mithilfe einer orthogonalen Matrix vorgenommen werden, um Faktoren zu erhalten die sinnvoll interpretierbar sind.

Für die Hauptkomponentenanalyse betrachten wir den Vektor der Faktorüberschussrenditen $\mathbf{y}$ mit der Kovarianzmatrix $\boldsymbol{\Sigma}$. Das Ziel der Hauptkomponentenanalyse ist es, die Varianz durch möglichst wenige Komponenten zu erklären. Dazu betrachten wir folgendes Verfahren.

Für $i = 1, \dots, p$ betrachten wir standardisierte Gewichtsvektoren $\boldsymbol{\omega}_i \in \mathbb{R}^p$ mit $\boldsymbol{\omega}_i^T \boldsymbol{\omega}_i = 1$. Die Hauptkomponenten x_i sind Linearkombinationen der Portfolioüberschussrenditen mit diesen Gewichtsvektoren:

$$x_i = \boldsymbol{\omega}_i^T \mathbf{y}.$$

Insbesondere gilt für die Hauptkomponenten

$$\text{var}(x_i) = \boldsymbol{\omega}_i^T \boldsymbol{\Sigma} \boldsymbol{\omega}_i, \quad i = 1, \dots, p \text{ und}$$

$$\text{cov}(x_i, x_j) = \boldsymbol{\omega}_i^T \boldsymbol{\Sigma} \boldsymbol{\omega}_j, \quad i, j = 1, \dots, p.$$

Die Gewichte müssen dabei so bestimmt werden, dass die Hauptkomponenten folgende Bedingungen erfüllen:

1. Die erste Hauptkomponente $x_1^* = \boldsymbol{\omega}_1^T \mathbf{y}$ wird so bestimmt, dass $x_1^* = \max_{x_1} \text{var}(x_1)$ unter der Bedingung $\boldsymbol{\omega}_1^T \boldsymbol{\omega}_1 = 1$ erfüllt ist.
2. Die zweite Hauptkomponente $x_2^* = \boldsymbol{\omega}_2^T \mathbf{y}$ wird so bestimmt, dass $x_2^* = \max_{x_2} \text{var}(x_2)$ unter den Bedingungen $\boldsymbol{\omega}_2^T \boldsymbol{\omega}_2 = 1$ und $\text{cov}(x_1, x_2) = 0$ erfüllt ist.
3. Analog wird die i -te Hauptkomponente $x_i^* = \boldsymbol{\omega}_i^T \mathbf{y}$ so bestimmt, dass $x_i^* = \max_{x_i} \text{var}(x_i)$ unter den Bedingungen $\boldsymbol{\omega}_i^T \boldsymbol{\omega}_i = 1$ und $\text{cov}(x_i, x_j) = 0$ für $j = 1, \dots, i-1$ erfüllt ist.

Da die Kovarianzmatrix positiv semidefinit ist, können die Hauptkomponenten durch Spektralzerlegung der Kovarianzmatrix berechnet werden. Wir betrachten die Eigenwerte $\lambda_1, \dots, \lambda_p$ der Kovarianzmatrix, sodass $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ gilt. Die zugehörigen normierten Eigenvektoren bezeichnen wir mit $\mathbf{v}_1, \dots, \mathbf{v}_p$. Sie dienen uns bei der Berechnung der Hauptkomponenten als Gewichtsvektoren. Dann gilt

$$\begin{aligned} x_i &= \mathbf{v}_i^T \mathbf{y}, \\ \text{var}(x_i) &= \mathbf{v}_i^T \boldsymbol{\Sigma} \mathbf{v}_i = \lambda_i \end{aligned}$$

und wegen der Orthogonalität der Eigenwerte von symmetrischen Matrizen

$$\text{cov}(x_i, x_j) = \mathbf{v}_i^T \boldsymbol{\Sigma} \mathbf{v}_j = 0 \quad \text{für } i \neq j.$$

Durch diesen Ansatz sind also alle Bedingungen für die Bestimmung der Hauptkomponenten erfüllt.

In unserem Faktormodell nutzen wir die Hauptkomponentenanalyse, indem wir das oben beschriebene Verfahren auf die empirische Kovarianzmatrix der $\mathbf{y}_1, \dots, \mathbf{y}_n$ anwenden. Die ersten K Hauptkomponenten dienen uns dann als Faktoren. Die Faktorladungen können wir basierend auf diesen geschätzten Faktoren wieder durch Regression berechnen.

Ein Vorteil der angenommenen Faktorstruktur ist, dass die mitunter hochdimensionalen Probleme durch ein niedrigdimensionales Faktormodell gelöst werden können. Diesen Ansatz nutzen auch Fan et al. (2007) für die Schätzung von Kovarianzmatrizen. Wir wollen ihren Ansatz im folgenden Kapitel darstellen.

2. Kovarianzmatrixschätzung in Faktormodellen

In diesem Kapitel nutzen wir die Theorie der Faktormodelle aus Kapitel 1, um die Kovarianzmatrix eines Portfolios zu schätzen. Dafür wenden wir den Ansatz von Fan et al. (2007) und Fan et al. (2011) an und leiten das asymptotische Verhalten der Schätzer unter unterschiedlichen Annahmen an die beobachtete Stichprobe her. Insbesondere betrachten wir das Szenario, in dem die Dimension des Portfolios p mit wachsender Stichprobengröße n gegen unendlich geht und auch die Anzahl der beobachteten Faktoren K wachsen kann.

Unter verschiedenen Voraussetzungen an die Faktoren der beobachteten Stichprobe vergleichen wir das asymptotische Verhalten der aus dem Faktormodell hervorgehenden Schätzung der Kovarianzmatrix nach (1.0.3) mit der Stichprobenkovarianzmatrix.

2.1. Schätzung der Komponenten des Faktormodells

Für die Schätzung der Kovarianzmatrix betrachten wir eine Stichprobe vom Umfang n mit Beobachtungen der Portfolioüberschussrenditen und der Faktoren

$$(\mathbf{y}_1, \mathbf{f}_1), \dots, (\mathbf{y}_n, \mathbf{f}_n).$$

Für diese Stichprobe formulieren wir das Faktormodell (1.0.1) mit folgender Notation

$$\begin{aligned}\mathbf{X} &= (\mathbf{f}_1, \dots, \mathbf{f}_n) \in \mathbb{R}^{K \times n} \\ \mathbf{Y} &= (\mathbf{y}_1, \dots, \mathbf{y}_n) \in \mathbb{R}^{p \times n} \\ \mathbf{E} &= (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n) \in \mathbb{R}^{p \times n}.\end{aligned}$$

Das Faktormodell (1.0.1) gilt dann in der Form

$$\mathbf{Y} = \mathbf{B}_n \mathbf{X} + \mathbf{E}, \tag{2.1.1}$$

dabei handelt es sich, wie oben schon angedeutet, um Matrizen:

$$\begin{pmatrix} Y_{11} & \dots & Y_{1n} \\ \vdots & \ddots & \vdots \\ Y_{p1} & \dots & Y_{pn} \end{pmatrix} = \begin{pmatrix} b_{11} & \dots & b_{1K} \\ \vdots & \ddots & \vdots \\ b_{p1} & \dots & b_{pK} \end{pmatrix} \begin{pmatrix} f_{11} & \dots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{K1} & \dots & f_{Kn} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} & \dots & \varepsilon_{1n} \\ \vdots & \ddots & \vdots \\ \varepsilon_{p1} & \dots & \varepsilon_{pn} \end{pmatrix}.$$

Wir betrachten die Anzahl der Faktoren K und die Dimension des Portfolios p in Abhängigkeit von n . Um diese Abhängigkeit bei der Matrix der Faktorladungen $\mathbf{B}$ zu verdeutlichen, nutzen wir den Index n und setzen $b_{ij} := b_{n,ij}$.

Wir bestimmen mithilfe der Stichprobe einen Schätzer für die Kovarianzmatrix Σ_n , indem wir in (1.0.3) die Schätzer der einzelnen Matrizen einsetzen.

$$\hat{\Sigma}_n = \hat{\mathbf{B}}_n \hat{\text{cov}}(\mathbf{f}) \hat{\mathbf{B}}_n^T + \hat{\Sigma}_0. \quad (2.1.2)$$

Der Kovarianzmatrixschätzer setzt sich zusammen aus der Stichprobenkovarianzmatrix der Faktoren in der Darstellung von Lemma A.1.1,

$$\hat{\text{cov}}(\mathbf{f}) = \frac{1}{n-1} \mathbf{X} \mathbf{X}^T - \frac{1}{n(n-1)} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T, \quad (2.1.3)$$

der empirischen Fehlervarianz

$$\hat{\Sigma}_0 = \text{diag}\left(\frac{1}{n} \hat{\mathbf{E}} \hat{\mathbf{E}}^T\right),$$

wobei

$$\hat{\mathbf{E}} = \mathbf{Y} - \hat{\mathbf{B}}_n \mathbf{X} = \mathbf{Y} (\mathbf{I}_n - \mathbf{H})$$

die Matrix der Residuen ist und dem Kleinste Quadrate Schätzer für die Matrix der Faktorladungen

$$\hat{\mathbf{B}}_n = \mathbf{Y} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1}, \quad (2.1.4)$$

der sich durch Minimieren des Ausdrucks $\|\mathbf{Y} - \mathbf{B}_n \mathbf{X}\|^2$ berechnen lässt. Dabei ist $\mathbf{H}$ die symmetrisch positiv definite $n \times n$ Hutmatrix aus dem linearen Modell

$$\mathbf{H} = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X}. \quad (2.1.5)$$

Eine weitere Möglichkeit die Kovarianzmatrix zu schätzen, ist die Stichprobenkovarianzmatrix. Sie bietet die Möglichkeit einer Schätzung ohne Anwendung des Faktormodells. Wir nutzen dazu die Darstellung aus Lemma A.1.1

$$\hat{\Sigma}_{sam} = \frac{1}{n-1} \mathbf{Y} \mathbf{Y}^T - \frac{1}{n(n-1)} \mathbf{Y} \mathbf{1} \mathbf{1}^T \mathbf{Y}^T. \quad (2.1.6)$$

Der Nachteil der Stichprobenkovarianzmatrix ist, dass sie für das hochdimensionale Szenario mit $p > n$ singulär ist.

2.2. Annahmen und Notation

In den verschiedenen Szenarien, die wir in diesem Kapitel betrachten, sollen einige gemeinsame Annahmen gelten, die wir nun diskutieren.

Annahme 1 Bei gegebenen Faktoren $\mathbf{f}$ gilt für den Erwartungswert und die Kovarianzmatrix der Fehlerterme die Aussage (1.0.2). Die Fehler sind unkorreliert und ihre Kovarianzmatrix hat Diagonalstruktur

$$\text{cov}(\boldsymbol{\varepsilon}|\mathbf{f}) = \boldsymbol{\Sigma}_0 = \text{diag}(\sigma_1^2, \dots, \sigma_p^2).$$

Diese Annahme ist in Faktormodellen üblich, wenngleich sie eine starke Einschränkung an die Fehler darstellt.

Annahme 2 Die vierten Momente der Fehler und Faktoren sind gleichmäßig beschränkt: $\max_{1 \leq i \leq K} E(f_i^4) = O(1)$ und $\max_{1 \leq i \leq p} E(\varepsilon_i^4) = O(1)$.

Die Annahme an die vierten Momente der Fehler und Faktoren wird genutzt, um die Konvergenzraten der jeweiligen Stichprobenkovarianzmatrizen herzuleiten. Für die Überschussrenditen des Portfolios gilt diese Annahme nicht automatisch, da dafür auch Annahmen an die Faktorladungen nötig wären.

Mit der Annahme an die vierten Momente der Faktoren gilt mit der Cauchy–Schwarz–Ungleichung für $i = 1, \dots, K$ und eine Konstante $C > 0$:

$$E(f_i^2) \leq [E(f_i^4)]^{\frac{1}{2}} \leq \left[\max_{1 \leq i \leq K} E(f_i^4) \right]^{\frac{1}{2}} \leq C^{\frac{1}{2}}.$$

Demnach sind auch die zweiten Momente der Faktoren gleichmäßig beschränkt.

Dieses Resultat folgt analog auch für die zweiten Momente der Fehler.

Annahme 3 Es existiert ein $c_1 > 0$, sodass für den kleinsten Eigenwert der Kovarianzmatrix der Faktoren $\lambda_K(\text{cov}(\mathbf{f})) \geq c_1$ für alle n gilt.

Annahme 4 Es existiert ein $c_2 > 0$, sodass für den kleinsten Eigenwert der Kovarianzmatrix der Fehler $\lambda_p(\boldsymbol{\Sigma}_0) \geq c_2$ für alle n gilt.

Diese Annahmen werden benötigt, um die Invertierbarkeit der beiden Kovarianzmatrizen sicher zu stellen und die Konvergenzraten der inversen Kovarianzmatrix des Portfolios herzuleiten.

Annahme 5 Es gilt $E(\|\mathbf{y}\|^2) = O(p)$.

Obwohl wir die Überschussrenditen der Portfoliopositionen mithilfe der Faktoren und der Faktorladungen durch das Faktormodell erklären, treffen wir diese Annahme um damit die Ordnung der Frobeniusnorm für die Matrix der Faktorladungen herzuleiten. Dazu nutzen wir die Annahme 3, dann gilt mit (B.4.1)

$$c_1 \mathbf{I}_K \leq \text{cov}(\mathbf{f}). \quad (2.2.1)$$

Mit der Faktormodelldarstellung der Kovarianzmatrix nach (1.0.3) folgt dann:

$$\boldsymbol{\Sigma}_n = \mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{B}_n^T + \boldsymbol{\Sigma}_0 \geq \mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{B}_n^T \geq \mathbf{B}_n c_1 \mathbf{I}_K \mathbf{B}_n^T = c_1 \mathbf{B}_n \mathbf{B}_n^T$$

und mit

$$\text{tr}\{\mathbf{\Sigma}_n\} = \sum_{i=1}^p \text{var}(Y_i) = \sum_{i=1}^p E(Y_i^2) - E(Y_i)^2 \leq \sum_{i=1}^p E(Y_i^2) = E(\|\mathbf{y}\|^2) \quad (2.2.2)$$

folgt mithilfe von $E(\|\mathbf{y}\|^2) = O(p)$ daher

$$\|\mathbf{B}_n\|^2 = \text{tr}\{\mathbf{B}_n \mathbf{B}_n^T\} \stackrel{(B.1.9)}{\leq} \frac{1}{c_1} \text{tr}\{\mathbf{\Sigma}_n\} \leq \frac{1}{c_1} E(\|\mathbf{y}\|^2) = O(p)$$

und damit insbesondere

$$\|\mathbf{B}_n\| = O(p^{\frac{1}{2}}). \quad (2.2.3)$$

Umgekehrt ist es nicht möglich, die Ordnung der Frobeniusnorm für die Matrix der Faktorladungen als Annahme zu formulieren und daraus mithilfe der übrigen getroffenen Annahmen die Aussage von Annahme 5 herzuleiten. Ein Grund dafür ist, dass für die Umkehrung die Aussage (2.2.1) durch die analoge Aussage für den größten Eigenwert $\text{cov}(\mathbf{f}) \leq \lambda_1(\text{cov}(\mathbf{f}))\mathbf{I}_K$ ersetzt werden müsste. Dazu wäre eine zusätzliche Annahme an die Spektralnorm der Kovarianzmatrix der Faktoren nötig. Da zusätzlich die Abschätzung aus (2.2.2) nicht umgekehrt werden kann, könnte damit ohnehin nur eine Aussage für die zentrierten Momente der Überschussrenditen getroffen werden.

Alternativ ist es möglich, folgende direkte Abschätzung vorzunehmen

$$E(\|\mathbf{y}\|^2) = E(\|\mathbf{B}_n \mathbf{f} + \varepsilon\|^2) \leq E(\|\mathbf{B}_n \mathbf{f}\|^2) + E(\|\varepsilon\|^2),$$

da der zweite Summand nach Annahme 2 von der Größenordnung $O(p)$ ist, schätzen wir nur den ersten Summanden mithilfe der Cauchy–Schwarz-Ungleichung und Annahme 2 ab.

$$\begin{aligned} E(\|\mathbf{y}\|^2) &\leq E\left(\sum_{i=1}^p \left|\sum_{j=1}^K b_{ij} f_j\right|^2\right) \leq E\left(\sum_{i=1}^p \left(\sum_{j=1}^K |b_{ij} f_j|\right)^2\right) = E\left(\sum_{i=1}^p \sum_{j=1}^K \sum_{l=1}^K |b_{ij}| |b_{il}| E(|f_j f_l|)\right) \\ &\leq \sum_{i=1}^p \sum_{j=1}^K \sum_{l=1}^K |b_{ij}| |b_{il}| \sqrt{E(|f_j|^2) E(|f_l|^2)} \leq C \sum_{i=1}^p \left(\sum_{j=1}^K |b_{ij}|\right)^2 = C \sum_{i=1}^p \|\mathbf{b}_i\|_1^2. \end{aligned}$$

Durch eine zusätzliche Annahme $\max_{1 \leq i \leq p} \|\mathbf{b}_i\|_1^2 = O(1)$ könnte dann mit (B.2.7) die Ordnung der Frobeniusnorm der Matrix der Faktorladungen und Annahme 5 hergeleitet werden.

Annahme 6 Die Verteilung der Faktoren $\mathbf{f}$ ist stetig und die Anzahl der Faktoren ist kleiner als die Dimension und die Stichprobengröße: $K < p$ und $K < n$.

Diese Annahme wird benötigt, damit die Matrix $\mathbf{X}\mathbf{X}^T$ für $n \geq K$ fast sicher invertierbar ist. Die Annahme an die Anzahl der Faktoren ist sinnvoll, da das Faktormodell genutzt werden soll, um die Dimension des betrachteten Problems zu reduzieren.

Annahme 7 Die Faktoren $f_1, \dots, f_K$ sind über n fest gewählt. Es existiert eine symmetrisch positiv definite Matrix $\mathbf{A}$, deren kleinster Eigenwert von Null weg beschränkt ist, für die gilt

$$\frac{1}{p} \mathbf{B}_n^T \mathbf{B}_n \longrightarrow \mathbf{A}, \quad n \rightarrow \infty.$$

Diese Annahme benötigen wir für die Herleitung der Konvergenzrate der inversen Kovarianzmatrix des Portfolios.

Für den Vergleich der beiden Kovarianzmatrixschätzer (2.1.2) und (2.1.6) nutzen wir die Frobeniusnorm $\|\cdot\|$ und die Spektralnorm $\|\cdot\|_S$. Für die Definition und Eigenschaften dieser Normen verweisen wir auf den Anhang B. Zusätzlich definieren wir eine weitere Norm, unter der die Faktorstruktur besser berücksichtigt wird.

Definition 2.2.1 Sei Σ_n eine Folge von positiv definiten Kovarianzmatrizen der Dimension $p_n \times p_n$. Für eine Matrix $\mathbf{A}$ der Dimension $p_n \times p_n$ definieren wir folgende Norm:

$$\|\mathbf{A}\|_{\Sigma_n} = \frac{1}{\sqrt{p_n}} \|\Sigma_n^{-\frac{1}{2}} \mathbf{A} \Sigma_n^{-\frac{1}{2}}\|. \quad (2.2.4)$$

Die Definition dieser Norm impliziert folgende Verlustfunktion für symmetrische $p \times p$ Matrizen

$$\begin{aligned} L(\hat{\Sigma}, \Sigma) &= \sqrt{p} \|\hat{\Sigma} - \Sigma\|_{\Sigma} = \frac{\sqrt{p}}{\sqrt{p}} \|\Sigma^{-\frac{1}{2}} (\hat{\Sigma} - \Sigma) \Sigma^{-\frac{1}{2}}\| \\ &= \|\Sigma^{-\frac{1}{2}} \hat{\Sigma} \Sigma^{-\frac{1}{2}} - \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}}\| = [\text{tr}\{[\Sigma^{-\frac{1}{2}} \hat{\Sigma} \Sigma^{-\frac{1}{2}} - \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}}]^2\}]^{\frac{1}{2}} \\ &= [\text{tr}\{\Sigma^{-\frac{1}{2}} \hat{\Sigma} \Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} \hat{\Sigma} \Sigma^{-\frac{1}{2}} - 2 \Sigma^{-\frac{1}{2}} \hat{\Sigma} \Sigma^{-\frac{1}{2}} + \mathbf{I}_p\}]^{\frac{1}{2}} \\ &= [\text{tr}\{(\hat{\Sigma} \Sigma^{-1})^2 - 2 \hat{\Sigma} \Sigma^{-1} + \mathbf{I}_p\}]^{\frac{1}{2}} \\ &= [\text{tr}\{[\hat{\Sigma} \Sigma^{-1} - \mathbf{I}_p]^2\}]^{\frac{1}{2}}. \end{aligned}$$

Diese ist angelehnt an die Verlustfunktion, die James und Stein (1961) definieren

$$L(\hat{\Sigma}, \Sigma) = \text{tr}\{\hat{\Sigma} \Sigma^{-1}\} - \log(\det(\hat{\Sigma} \Sigma^{-1})) - p.$$

Insbesondere gilt für eine Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ folgende Beziehung zwischen der Norm $\|\cdot\|_{\Sigma}$ und der Spektralnorm:

$$\begin{aligned} \|\mathbf{A}\|_{\Sigma}^2 &= \frac{1}{n} \|\Sigma^{-\frac{1}{2}} \mathbf{A} \Sigma^{-\frac{1}{2}}\|^2 = \frac{1}{n} \text{tr}\{\Sigma^{-\frac{1}{2}} \mathbf{A} \Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} \mathbf{A}^T \Sigma^{-\frac{1}{2}}\} = \frac{1}{n} \text{tr}\{\Sigma^{-1} \mathbf{A} \Sigma^{-1} \mathbf{A}^T\} \\ &\stackrel{(B.2.3)}{\leq} \frac{1}{n} \|\Sigma^{-1} \mathbf{A}\| \|\Sigma^{-1} \mathbf{A}^T\| \stackrel{(B.2.5)}{\leq} \frac{1}{n} \|\Sigma^{-1}\|_S \|\mathbf{A}\| \|\Sigma^{-1}\|_S \|\mathbf{A}^T\| \\ &= \frac{1}{n} \|\Sigma^{-1}\|_S^2 \|\mathbf{A}\|^2. \end{aligned} \quad (2.2.5)$$

Aus den oben getroffenen Annahmen können wir direkt die Ordnung der Spektralnormen einiger Matrizen in unserem Faktormodell bestimmen. Für die Inverse der Kovarianzmatrix der Faktoren gilt nach Annahme 3

$$\|\text{cov}(\mathbf{f})^{-1}\|_S = O(1). \quad (2.2.6)$$

Außerdem gilt nach Annahme 4 für die Inverse der Fehlerkovarianzmatrix

$$\|\boldsymbol{\Sigma}_0^{-1}\|_S = O(1). \quad (2.2.7)$$

Daraus können wir auch die Spektralnorm der Inversen der Kovarianzmatrix $\boldsymbol{\Sigma}_n$ bestimmen. Mit der Darstellung des Faktormodells (1.0.3) gilt

$$\boldsymbol{\Sigma}_n = \mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{B}_n^T + \boldsymbol{\Sigma}_0 \geq \boldsymbol{\Sigma}_0,$$

also gilt insbesondere

$$\boldsymbol{\Sigma}_n^{-1} = (\mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{B}_n^T + \boldsymbol{\Sigma}_0)^{-1} \leq \boldsymbol{\Sigma}_0^{-1}.$$

Wegen (B.4.2) folgt dann auch

$$\|\boldsymbol{\Sigma}_n^{-1}\|_S \leq \|\boldsymbol{\Sigma}_0^{-1}\|_S = O(1). \quad (2.2.8)$$

Bevor wir mit den Abschätzungen der Schätzfehler unter unterschiedlichen Annahmen an die Stichprobe beginnen, führen wir Notationen für die Schätzfehler der einzelnen Komponenten des Faktormodells ein. Der Schätzfehler für die $p \times K$ Matrix der Faktorladungen $\mathbf{B}_n$ lässt sich schreiben als

$$\mathbf{C}_n := \mathbf{E} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} = \hat{\mathbf{B}}_n - \mathbf{B}_n. \quad (2.2.9)$$

Der Schätzfehler der $K \times K$ Kovarianzmatrix der Faktoren besitzt die Darstellung

$$\mathbf{D}_n := \frac{1}{n-1} \mathbf{X} \mathbf{X}^T - \frac{1}{n(n-1)} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T - \text{cov}(\mathbf{f}) = \hat{\text{cov}}(\mathbf{f}) - \text{cov}(\mathbf{f}). \quad (2.2.10)$$

Der Schätzfehler der $p \times p$ Kovarianzmatrix der Fehlerterme besitzt unter Berücksichtigung des Faktormodells folgende Struktur

$$\mathbf{F}_n := \mathbf{I}_p \circ \frac{1}{n} \mathbf{E} (\mathbf{I}_n - \mathbf{H}) \mathbf{E}^T - \boldsymbol{\Sigma}_0 = \hat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0. \quad (2.2.11)$$

Wird die Faktorstruktur nicht berücksichtigt, so nutzen wir die Stichprobenkovarianzmatrix nach Lemma A.1.1 für die Definition des Schätzfehlers der Kovarianzmatrix der Fehlerterme

$$\mathbf{G}_n := \frac{1}{n-1} \mathbf{E} \mathbf{E}^T - \frac{1}{n(n-1)} \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{E}^T - \boldsymbol{\Sigma}_0. \quad (2.2.12)$$

Mithilfe dieser Notationen formulieren wir folgendes Lemma

Lemma 2.2.2 Die Darstellung des Kovarianzmatrixschätzers aus dem Faktormodell (2.1.2) lässt sich in die folgenden Summanden zerlegen:

$$\hat{\Sigma}_n = \Sigma_n + \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + [\mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{B}_n^T + \mathbf{B}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T] + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{F}_n. \quad (2.2.13)$$

Für die Stichprobenkovarianzmatrix nach (2.1.6) erhalten wir analog folgende Darstellung:

$$\begin{aligned} \hat{\Sigma}_{sam} &= \Sigma_n + \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + \mathbf{G}_n + \frac{1}{n-1} (\mathbf{B}_n \mathbf{X} \mathbf{E}^T + \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T) \\ &\quad - \frac{1}{n(n-1)} (\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T + \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T). \end{aligned} \quad (2.2.14)$$

Beweis. Durch Einsetzen der Faktormodelldarstellung für die Kovarianzmatrix (2.1.2) und (2.2.11), sowie (2.2.9) gilt

$$\begin{aligned} \hat{\Sigma}_n &= \hat{\mathbf{B}}_n \hat{\text{cov}}(\mathbf{f}) \hat{\mathbf{B}}_n^T + \hat{\Sigma}_0 = (\mathbf{B}_n + \mathbf{C}_n) \hat{\text{cov}}(\mathbf{f}) (\mathbf{B}_n + \mathbf{C}_n)^T + \Sigma_0 + \mathbf{F}_n \\ &= \mathbf{B}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{B}_n^T + \mathbf{B}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{B}_n^T + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \Sigma_0 + \mathbf{F}_n \\ &\stackrel{(2.2.10)}{=} \mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{B}_n^T + \Sigma_0 + \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + \mathbf{B}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{B}_n^T \\ &\quad + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{F}_n \\ &\stackrel{(1.0.3)}{=} \Sigma_n + \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + [\mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{B}_n^T + \mathbf{B}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T] + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{F}_n. \end{aligned}$$

Für die Stichprobenkovarianzmatrix gilt in der Darstellung von (2.1.6)

$$\begin{aligned} \hat{\Sigma}_{sam} &= \frac{1}{n-1} \mathbf{Y} \mathbf{Y}^T - \frac{1}{n(n-1)} \mathbf{Y} \mathbf{1} \mathbf{1}^T \mathbf{Y}^T \\ &\stackrel{(2.1.1)}{=} \frac{1}{n-1} (\mathbf{B}_n \mathbf{X} + \mathbf{E}) (\mathbf{X}^T \mathbf{B}_n^T + \mathbf{E}^T) - \frac{1}{n(n-1)} (\mathbf{B}_n \mathbf{X} + \mathbf{E}) \mathbf{1} \mathbf{1}^T (\mathbf{X}^T \mathbf{B}_n^T + \mathbf{E}^T) \\ &= \frac{1}{n-1} (\mathbf{B}_n \mathbf{X} \mathbf{X}^T \mathbf{B}_n^T + \mathbf{B}_n \mathbf{X} \mathbf{E}^T + \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T + \mathbf{E} \mathbf{E}^T) \\ &\quad - \frac{1}{n(n-1)} (\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T + \mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T + \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T + \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{E}^T) \\ &= \frac{1}{n-1} \mathbf{B}_n \mathbf{X} \mathbf{X}^T \mathbf{B}_n^T - \frac{1}{n(n-1)} \mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T + \frac{1}{n-1} (\mathbf{B}_n \mathbf{X} \mathbf{E}^T + \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T) \\ &\quad + \frac{1}{n-1} \mathbf{E} \mathbf{E}^T - \frac{1}{n(n-1)} \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{E}^T - \frac{1}{n(n-1)} (\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T + \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T). \end{aligned}$$

Einsetzen von (2.2.10) und (2.2.12) liefert

$$\hat{\Sigma}_{sam} = \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + \mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{B}_n^T + \Sigma_0 + \frac{1}{n-1} (\mathbf{B}_n \mathbf{X} \mathbf{E}^T + \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T) + \mathbf{G}_n$$

$$\begin{aligned}
 & - \frac{1}{n(n-1)} (\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T + \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T) \\
 & \stackrel{(1.0.3)}{=} \mathbf{\Sigma}_n + \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + \mathbf{G}_n + \frac{1}{n-1} (\mathbf{B}_n \mathbf{X} \mathbf{E}^T + \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T) \\
 & - \frac{1}{n(n-1)} (\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T + \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T).
 \end{aligned}$$

□

Diese Darstellungen sind besonders nützlich für die Behandlung der Schätzfehler, da sie die Kovarianzmatrix $\mathbf{\Sigma}_n$ enthalten. Es genügt also die Konvergenzraten der jeweils vier anderen Summanden zu bestimmen und die jeweils dominierende Rate zu identifizieren, um den Schätzfehler zu quantifizieren.

Hier wird deutlich, dass wir auch für die Herleitung der Konvergenzraten für die Stichprobenkovarianzmatrix die Faktorstruktur anwenden, da dadurch die Beweise vereinfacht werden.

Für die Betrachtung der Inversen der Kovarianzmatrix im Faktormodell nutzen wir die Sherman–Morrison–Woodbury Formel nach (B.3.1). Diese wenden wir auf (1.0.3) und (2.1.2) an:

$$\mathbf{\Sigma}_n^{-1} = \mathbf{\Sigma}_0^{-1} - \mathbf{\Sigma}_0^{-1} \mathbf{B}_n [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \mathbf{\Sigma}_0^{-1} \mathbf{B}_n]^{-1} \mathbf{B}_n^T \mathbf{\Sigma}_0^{-1}, \quad (2.2.15)$$

$$\hat{\mathbf{\Sigma}}_n^{-1} = \hat{\mathbf{\Sigma}}_0^{-1} - \hat{\mathbf{\Sigma}}_0^{-1} \hat{\mathbf{B}}_n [\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\mathbf{\Sigma}}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\mathbf{\Sigma}}_0^{-1}. \quad (2.2.16)$$

Bei der Behandlung des Schätzfehlers subtrahieren wir diese beiden Darstellungen voneinander und ordnen um, um das Problem für die Bestimmung der Konvergenzrate wieder in kleinere Teilprobleme zu zerlegen. Die Details dazu folgen an gegebener Stelle im Beweis.

2.3. Unabhängige Faktoren

Wir beginnen nun mit der Herleitung der Konvergenzraten für die Kovarianzmatrixschätzer unter den verschiedenen Normen. In diesem Abschnitt betrachten wir den einfachsten Fall der Schätzung und treffen zusätzlich zu den Annahmen 1 bis 7 folgende Annahme

Annahme 8 Die betrachtete Stichprobe $(\mathbf{y}_1, \mathbf{f}_1), \dots, (\mathbf{y}_n, \mathbf{f}_n)$ ist unabhängig identisch verteilt.

Wir wollen nun das asymptotische Verhalten des Kovarianzmatrixschätzers aus dem Faktormodell und der Stichprobenkovarianzmatrix vergleichen, der Ansatz basiert auf der Arbeit von Fan et al. (2007). Dazu nutzen wir die Darstellungen (2.2.13) und (2.2.14). Die Beweise werden mithilfe von Abschätzungen der einzelnen Komponenten dieser Darstellungen hergeleitet, die wir ausführlich in Anhang C behandeln werden.

Zunächst betrachten wir die Konvergenz der Schätzer (2.1.2) und (2.1.6) unter der Frobeniusnorm.

Satz 2.3.1 Konvergenzraten unter der Frobeniusnorm

Unter den Annahmen 1 - 6 und 8 gelten für die Kovarianzmatrixschätzung im Faktormodell und die Stichprobenkovarianzmatrix folgende Konvergenzraten:

$$E(\|\hat{\Sigma}_n - \Sigma_n\|^2) = O(n^{-1}p^2K^2) \quad (2.3.1)$$

und

$$E(\|\hat{\Sigma}_{sam} - \Sigma_n\|^2) = O(n^{-1}p^2K^2). \quad (2.3.2)$$

Beweis. Für den Beweis nutzen wir die Darstellung des Schätzers aus dem Faktormodell aus (2.2.13) und wenden dann die Dreiecksungleichung an, um die Resultate der Lemmata C.0.3, C.0.6, C.0.4 und C.0.9 nutzen zu können. Damit erhalten wir

$$\begin{aligned} E(\|\hat{\Sigma}_n - \Sigma_n\|^2) &= E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + [\mathbf{B}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{B}_n^T] + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{F}_n\|^2) \\ &\leq 4(E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|^2) + E(\|\mathbf{B}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{B}_n^T\|^2) + E(\|\mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T\|^2) \\ &\quad + E(\|\mathbf{F}_n\|^2)) \\ &= O(n^{-1}p^2K^2) + O(n^{-1}p^2K) + O(n^{-2}p^2K) + O(n^{-1}p) + O(n^{-2}pK^2) \\ &= O(n^{-1}p^2K^2), \end{aligned}$$

da die Rate des ersten Summanden dominiert.

Die Argumentation für die Stichprobenkovarianzmatrix erfolgt analog. Mithilfe der Darstellung (2.2.14) und der Dreiecksungleichung erhalten wir

$$\begin{aligned} E(\|\hat{\Sigma}_{sam} - \Sigma_n\|^2) &= E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + \mathbf{G}_n + \frac{1}{n-1}(\mathbf{B}_n \mathbf{X} \mathbf{E}^T + \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T) \\ &\quad - \frac{1}{n(n-1)}(\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T + \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T)\|^2) \\ &\leq 4(E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|^2) + E(\|\mathbf{G}_n\|^2) + \frac{1}{(n-1)^2}E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|^2) \\ &\quad + \frac{1}{n^2(n-1)^2}E(\|\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T\|^2)). \end{aligned}$$

Aus den Konvergenzraten der Lemmata C.0.3, C.0.10, C.0.7 und C.0.8 folgt

$$\begin{aligned} E(\|\hat{\Sigma}_{sam} - \Sigma_n\|^2) &= O(n^{-1}p^2K^2) + O(n^{-1}p^2) + O(n^{-1}p^2K) + O(n^{-1}p^2K) \\ &= O(n^{-1}p^2K^2), \end{aligned}$$

da auch hier die Rate des ersten Summanden dominiert. $\square$

Wir bemerken, dass bei beiden Schätzern die Rate dominiert, die aus der Schätzung der Stichprobenkovarianzmatrix der Faktoren resultiert und die Dimension des Portfolios durch die Matrix der Faktorladungen eingeht. In der Frobeniusnorm ist daher kein Vorteil durch die Schätzung der Kovarianzmatrix mithilfe des Faktormodells erkennbar

und beide Schätzer sind nur konsistent, falls die Dimension deutlich kleiner als die Stichprobengröße ist ($p^2 < n$). In unserem hochdimensionalen Szenario ist die Betrachtung der Frobeniusnorm daher eher ungeeignet.

Wir stellen weiterhin fest, dass wir für die Stichprobenkovarianzmatrix eine schlechtere Rate erhalten, als im allgemeinen Beweis von Lemma A.1.2. Die Ursache hierfür ist, dass wir für die Wertpapierüberschussrenditen Y_i , $i = 1, \dots, p$ keine Annahme über gleichmäßig beschränkte vierte Momente treffen. Eine solche Annahme würde in Annahme 5 stärkere Einschränkungen an die Matrix der Faktorladungen erfordern.

Wir untersuchen nun das Konvergenzverhalten der beiden Kovarianzmatrixschätzer unter der in Definition 2.2.1 eingeführten Norm. Unsere Ergebnisse über die Stichprobenkovarianzmatrix weichen von denen in Fan et al. (2007) ab, da wir durch die Anwendung der Spektralnorm wie in Fan et al. (2011) eine verbesserte Abschätzung einiger Konvergenzraten erreichen.

Satz 2.3.2 Konvergenzraten unter der Norm $\|\cdot\|_{\Sigma_n}$

Unter den Annahmen 1 - 6 und 8 gelten für die Schätzung der Kovarianzmatrix aus dem Faktormodell und die Stichprobenkovarianzmatrix folgende Aussagen:

$$E(\|\hat{\Sigma}_n - \Sigma_n\|_{\Sigma_n}^2) = O(n^{-1}K) + O(n^{-2}pK) \quad (2.3.3)$$

und

$$E(\|\hat{\Sigma}_{sam} - \Sigma_n\|_{\Sigma_n}^2) = O(n^{-1}p). \quad (2.3.4)$$

Beweis. Wir betrachten die Konvergenzrate des ersten Ausdrucks. Wegen der Darstellung des Kovarianzmatrixschätzers nach (2.2.13) erhalten wir mithilfe der Dreiecksungleichung und den Resultaten der Lemmata C.0.3, C.0.6, C.0.4 und C.0.9 folgende Abschätzung

$$\begin{aligned} & E(\|\hat{\Sigma}_n - \Sigma_n\|_{\Sigma_n}^2) \\ &= E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + [\mathbf{B}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{B}_n^T] + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{F}_n\|_{\Sigma_n}^2) \\ &\leq 4(E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|_{\Sigma_n}^2) + E(\|\mathbf{B}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{B}_n^T\|_{\Sigma_n}^2) \\ &\quad + E(\|\mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T\|_{\Sigma_n}^2) + E(\|\mathbf{F}_n\|_{\Sigma_n}^2)) \\ &= O(n^{-1}p^{-1}K^2) + O(n^{-1}K) + O(n^{-2}pK) + O(n^{-1}) + O(n^{-2}K^2). \end{aligned}$$

Da nach Annahme 6 die Dimension größer ist als die Anzahl der Faktoren ($p \geq K$), dominieren die Raten des zweiten und des dritten Summanden und wir erhalten als Resultat

$$E(\|\hat{\Sigma}_n - \Sigma_n\|_{\Sigma_n}^2) = O(n^{-1}K) + O(n^{-2}pK).$$

Für die Abschätzungen der beiden dominierenden Summanden betrachten wir nicht die Raten der einzelnen Faktoren, sondern das Produkt der Komponenten, da sich der Ausdruck dadurch jeweils vereinfacht. Die Details über die Herkunft der Raten sind in Lemma

C.0.6 bzw. Lemma C.0.4 nachzulesen.

Für die Stichprobenkovarianzmatrix gilt nach (2.2.14) und mit der Dreiecksungleichung die Darstellung

$$\begin{aligned}
 E(\|\hat{\Sigma}_{sam} - \Sigma_n\|_{\Sigma_n}^2) &= E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T + \mathbf{G}_n + \frac{1}{n-1}(\mathbf{B}_n \mathbf{X} \mathbf{E}^T + \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T) \\
 &\quad - \frac{1}{n(n-1)}(\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T + \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T)\|_{\Sigma_n}^2) \\
 &\leq 4(E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|_{\Sigma_n}^2) + E(\|\mathbf{G}_n\|_{\Sigma_n}^2) + \frac{1}{(n-1)^2} E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T + \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T\|_{\Sigma_n}^2) \\
 &\quad + \frac{1}{n^2(n-1)^2} E(\|\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T + \mathbf{E} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T\|_{\Sigma_n}^2)).
 \end{aligned}$$

Aus den Lemmata C.0.3 und C.0.10 und Korollaren C.0.7 und C.0.8 folgt

$$\begin{aligned}
 E(\|\hat{\Sigma}_{sam} - \Sigma_n\|_{\Sigma_n}^2) &= O(n^{-1}p^{-1}K^2) + O(n^{-1}p) + O(n^{-1}K) + O(n^{-1}K) \\
 &= O(n^{-1}p),
 \end{aligned}$$

da nach Annahme 6 gilt, dass $K \leq p$ und daher die Rate des zweiten Summanden dominiert.

Diese finale Rate stammt aus der Konvergenzrate des Schätzfehlers der Fehlerkovarianzmatrix und lässt sich mit Annahme 2 über die gleichmäßig beschränkten vierten Momente der Fehler herleiten (vgl. Lemma A.1.1). $\square$

Bemerkung 2.3.3 Für den Vergleich betrachten wir die Raten der stochastischen Konvergenz, die aus der Konvergenz im quadratischen Mittel im Satz folgen. Dazu analysieren wir die Anzahl der Faktoren und die Dimension in Abhängigkeit von der Stichprobengröße. Wir setzen $K = O(n^{\alpha_1})$ und $p = O(n^\alpha)$ für $\alpha, \alpha_1 \geq 0$. Dann definieren wir $\beta = \min\{1 - \alpha_1, 2 - \alpha - \alpha_1\}$, sodass wir mit (2.3.3) schreiben können

$$\|\hat{\Sigma}_n - \Sigma_n\|_{\Sigma_n} = O_P(n^{-\beta/2})$$

und $\beta_1 = 1 - \alpha$ ist. Daher können wir mithilfe von (2.3.4) die Darstellung

$$\|\hat{\Sigma}_{sam} - \Sigma_n\|_{\Sigma_n} = O_P(n^{-\beta_1/2})$$

wählen. Da nach Annahme 6 gilt, dass $K \leq p$, ist auch $\alpha_1 \leq \alpha$.

Solange $\alpha \leq 1$ gilt, die Dimension also höchstens in der gleichen Größenordnung wächst wie die Stichprobengröße, erhalten wir $\beta = 1 - \alpha_1$. Nach Annahme gilt immer $\alpha_1 \leq \alpha$, daher hat der Schätzer aus dem Faktormodell unter diesen Bedingungen eine bessere Konvergenzrate als die Stichprobenkovarianzmatrix.

Für Werte $\alpha > 1$, wenn also die Dimension schneller wächst als die Stichprobengröße, erhalten wir $\beta = 2 - \alpha - \alpha_1$. Dann hat der Schätzer aus dem Faktormodell immer dann eine bessere Konvergenzrate als die Stichprobenkovarianzmatrix, wenn $\alpha_1 < 1$ gilt. Die Konvergenzrate der Stichprobenkovarianzmatrix ist also nur dann besser als die Rate des Schätzers aus dem Faktormodell, wenn $\alpha > 1$ und $\alpha_1 > 1$ gelten. In dieser Situation ist aber $\beta_1 < 0$, das heißt der Schätzer ist nicht mehr konsistent.

Zuletzt vergleichen wir die Konvergenzraten der Inversen der Kovarianzmatrixschätzer.

Satz 2.3.4 Konvergenzraten der Inversen unter der Frobeniusnorm

Unter den Annahmen 1 - 8 gilt für die Inverse des Kovarianzmatrixschätzers aus dem Faktormodell folgende Konvergenzrate

$$\|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| = O_P(n^{-\frac{1}{2}}pK^{\frac{3}{2}}) \quad (2.3.5)$$

und falls $p^{\frac{3}{2}}K = o(n^{\frac{1}{2}})$, so gilt für die Inverse der Stichprobenkovarianzmatrix die Konvergenzrate

$$\|\hat{\Sigma}_{sam}^{-1} - \Sigma_n^{-1}\| = O_P(n^{-\frac{1}{2}}p^2K). \quad (2.3.6)$$

Beweis. Für den Beweis von (2.3.5) betrachten wir die Darstellungen der Inversen nach (2.2.15) und (2.2.16)

$$\begin{aligned} \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| &= \|\hat{\Sigma}_0^{-1} - \hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1} \\ &\quad - \Sigma_0^{-1} + \Sigma_0^{-1}\mathbf{B}_n[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\Sigma_0^{-1}\mathbf{B}_n]^{-1}\mathbf{B}_n^T\Sigma_0^{-1}\|. \end{aligned}$$

Mithilfe von Nulladditionen teilen wir diese Gleichung in sechs Terme auf, die wir in Lemma C.0.12 bis Lemma C.0.15 abschätzen.

$$\begin{aligned} \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| &= \|\hat{\Sigma}_0^{-1} - \Sigma_0^{-1} - \hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1} \\ &\quad + \Sigma_0^{-1}\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1} \\ &\quad + \Sigma_0^{-1}\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\Sigma_0^{-1} - \Sigma_0^{-1}\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\Sigma_0^{-1} \\ &\quad + \Sigma_0^{-1}\mathbf{B}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\Sigma_0^{-1} - \Sigma_0^{-1}\mathbf{B}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\Sigma_0^{-1} \\ &\quad + \Sigma_0^{-1}\mathbf{B}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\Sigma_0^{-1} - \Sigma_0^{-1}\mathbf{B}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\Sigma_0^{-1} \\ &\quad + \Sigma_0^{-1}\mathbf{B}_n[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\Sigma_0^{-1}\mathbf{B}_n]^{-1}\mathbf{B}_n^T\Sigma_0^{-1}\| \\ &= \|\hat{\Sigma}_0^{-1} - \Sigma_0^{-1} - (\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1} \\ &\quad - \Sigma_0^{-1}\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}) \\ &\quad - \Sigma_0^{-1}(\hat{\mathbf{B}}_n - \mathbf{B}_n)[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\Sigma_0^{-1} \\ &\quad - \Sigma_0^{-1}\mathbf{B}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}(\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T)\Sigma_0^{-1} \\ &\quad - \Sigma_0^{-1}\mathbf{B}_n\{[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\Sigma_0^{-1}\mathbf{B}_n]^{-1}\}\mathbf{B}_n^T\Sigma_0^{-1}\| \\ &\leq \|\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}\| + \|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\| \\ &\quad + \|\Sigma_0^{-1}\hat{\mathbf{B}}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\| \\ &\quad + \|\Sigma_0^{-1}(\hat{\mathbf{B}}_n - \mathbf{B}_n)[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\Sigma_0^{-1}\| \\ &\quad + \|\Sigma_0^{-1}\mathbf{B}_n[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}(\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T)\Sigma_0^{-1}\| \\ &\quad + \|\Sigma_0^{-1}\mathbf{B}_n\{[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\Sigma_0^{-1}\mathbf{B}_n]^{-1}\}\mathbf{B}_n^T\Sigma_0^{-1}\|. \end{aligned} \quad (2.3.7)$$

Aus den Lemmata C.0.12, C.0.13, C.0.14 und C.0.15 erhalten wir für diese Summanden die Konvergenzraten

$$\begin{aligned}\|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| &= (O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(n^{-1}p^{\frac{1}{2}}K)) + (O_P(n^{-\frac{1}{2}}p) + O_P(n^{-1}pK)) \\ &\quad + (O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(n^{-1}p^{\frac{1}{2}}K)) + O_P(n^{-\frac{1}{2}}pK^{\frac{3}{2}}) + O_P(n^{-\frac{1}{2}}pK^{\frac{3}{2}}) \\ &\quad + O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}K^{\frac{1}{2}}) \\ &= O_P(n^{-\frac{1}{2}}pK^{\frac{3}{2}}).\end{aligned}$$

Die finale Rate stammt im Wesentlichen aus der Schätzung der Matrix der Faktorladungen und der Kovarianzmatrix der Faktoren, vgl. Lemma C.0.14.

Für den Beweis der zweiten Aussage (2.3.6) nutzen wir eine andere Beweistechnik und leiten die Rate aus dem Resultat von Satz 2.3.1 her. Wir betrachten (B.3.3) mit $\mathbf{A}_1 := \Sigma_n$ und $\mathbf{A}_2 := \hat{\Sigma}_{sam} - \Sigma_n$. Dann gilt $\mathbf{A}_1 + \mathbf{A}_2 = \hat{\Sigma}_{sam}$ und für $\|\Sigma_n^{-1}\| \|\hat{\Sigma}_{sam} - \Sigma_n\| < 1$ folgt die Ungleichung

$$\|\hat{\Sigma}_{sam}^{-1} - \Sigma_n^{-1}\| \leq \frac{\|\Sigma_n^{-1}\|^2 \|\hat{\Sigma}_{sam} - \Sigma_n\|}{1 - \|\Sigma_n^{-1}\| \|\hat{\Sigma}_{sam} - \Sigma_n\|}.$$

Mit (B.2.3) folgt weiter

$$\|\hat{\Sigma}_{sam}^{-1} - \Sigma_n^{-1}\| \leq \frac{\|\Sigma_n^{-1}\|^2 \|\hat{\Sigma}_{sam} - \Sigma_n\|}{1 - \|\Sigma_n^{-1}\| \|\hat{\Sigma}_{sam} - \Sigma_n\|}. \quad (2.3.8)$$

Mithilfe der Darstellung (2.1.2) gilt

$$\Sigma_n = \Sigma_0 + \mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{B}_n^T \geq \Sigma_0,$$

daraus folgt

$$\Sigma_n^{-1} = [\Sigma_0 + \mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{B}_n^T]^{-1} \leq \Sigma_0^{-1}.$$

Mit (B.1.9) und Annahme 4 folgt

$$\text{tr}\{\Sigma_n^{-1}\} = \text{tr}\{[\Sigma_0 + \mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{B}_n^T]^{-1}\} \leq \text{tr}\{\Sigma_0^{-1}\} = O(p),$$

also insbesondere

$$\|\Sigma_n^{-1}\| = \text{tr}\{\Sigma_n^{-1} \Sigma_n^{-1}\}^{\frac{1}{2}} \leq \text{tr}\{\Sigma_0^{-2}\}^{\frac{1}{2}} = O(p^{\frac{1}{2}}). \quad (2.3.9)$$

Aus diesem Resultat und der Aussage von Satz 2.3.1 folgt für den Nenner der rechten Seite in (2.3.8):

$$\|\Sigma_n^{-1}\| \|\hat{\Sigma}_{sam} - \Sigma_n\| = O_P(n^{-\frac{1}{2}}p^{\frac{3}{2}}K)$$

und mit der Annahme $p^{\frac{3}{2}}K = o(n^{\frac{1}{2}})$ konvergiert der Nenner stochastisch gegen 1. Für den Zähler gilt

$$\|\Sigma_n^{-1}\|^2 \|\hat{\Sigma}_{sam} - \Sigma_n\| = O_P(n^{-\frac{1}{2}}p^2K)$$

und damit folgt die Aussage (2.3.6). $\square$

Wir bemerken, dass die Schätzung aus dem Faktormodell bei der Betrachtung der inversen Matrizen einen deutlichen Vorteil gegenüber der Stichprobenkovarianzmatrix hat. Solange wir eine kleine Anzahl von Faktoren im Vergleich zur Dimension betrachten, erhalten wir in beiden Fällen konsistente Schätzungen. In Kapitel 3 werden wir sehen, dass dieses Ergebnis sich auch auf Anwendungsfälle innerhalb der Portfolios Selektion auswirkt.

2.4. Stationäre Faktoren

Die in Abschnitt 2.3 getroffene Annahme, dass die Faktoren unabhängig identisch verteilt sind, stellt eine starke Vereinfachung dar. Da für die Faktoren häufig Zeitreihen makro-ökonomischer Größen genutzt werden, muss diese Annahme fallen gelassen werden, damit die Abhängigkeitsstruktur der Zeitreihe abgebildet werden kann. Diesen Ansatz verfolgen Fan et al. (2011). Wir wenden ihn nun an, um einen Vergleich zu den Resultaten aus Abschnitt 2.3 anzustellen. Trotz der Abhängigkeitsstruktur in den Faktoren betrachten wir in dieser Arbeit nur die Schätzung der unbedingten Kovarianzmatrix.

Für die Herleitung von Konvergenzraten unter der Betrachtung von Zeitreihen sind weitere Annahmen nötig.

Annahme 9 Die Zeitreihe der Faktoren $\{\mathbf{f}_t\}_{t \geq 1}$ ist stationär und ergodisch. Die Fehler $(\epsilon_1, \dots, \epsilon_n)$ sind unabhängig identisch verteilt. Die Zeitreihen der Faktoren $\{\mathbf{f}_t\}_{t \geq 1}$ und der Fehler $\{\epsilon_t\}_{t \geq 1}$ sind unabhängig.

Diese Annahmen sind üblich um Aussagen über die gesamte Zeitreihe, unabhängig von dem betrachteten Zeitpunkt, machen zu können.

Wir betrachten nun die σ -Algebra der Vergangenheit und die von den Ereignissen nach n erzeugte σ -Algebra,

$$\begin{aligned}\mathcal{F}_{-\infty}^0 &:= \sigma((\mathbf{f}_t, \epsilon_t) : -\infty \leq t \leq 0) \\ \mathcal{F}_n^\infty &:= \sigma((\mathbf{f}_t, \epsilon_t) : n \leq t \leq \infty).\end{aligned}$$

Dazu betrachten wir den α -Mischungskoeffizienten

$$\alpha(n) := \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty} |P(A)P(B) - P(AB)|.$$

Zusätzlich treffen wir folgende Annahme

Annahme 10 Der α -Mischungskoeffizient zum Abstand t klingt exponentiell ab: Es existieren Konstanten $r_1, C > 0$, sodass für alle $t \in \mathbb{Z}^+$ gilt

$$\alpha(t) \leq \exp(-Ct^{r_1}).$$

Weiterhin müssen wir folgende Annahme an die Verteilung der Faktoren treffen

Annahme 11 Es existieren Konstanten $r_2, b_1 > 0$ mit $3r_2^{-1} + r_1^{-1} > 1$, sodass für alle $s > 0$ und alle $i \in \{1, \dots, K\}$ gilt

$$P(|f_{it}| > s) \leq \exp(-(s/b_1)^{r_2}).$$

Diese Annahmen benötigen wir für die Betrachtung des Schätzfehlers bei der Stichprobenkovarianzmatrix der stationären Faktoren, da der Ansatz aus Lemma C.0.2 auf Annahme 8 basiert, die wir nun nicht mehr berücksichtigen.

Die Resultate für die Faktoren aus Lemma C.0.5 basieren nur auf Annahme 2, daher müssen wir für den Vergleich zwischen unabhängigen und stationären Faktoren nur eine neue Rate für $\|\mathbf{D}_n\|$ als Ersatz für Lemma C.0.2 herleiten. Der Beweis dafür basiert wegen der Abhängigkeitsstruktur auf einer anderen Beweistechnik. Wir nutzen die Annahme an das Tailverhalten der Verteilung der Faktoren, daher ändert sich die Rate und wir erhalten im Vergleich zu dem Fall mit unabhängigen Faktoren einen zusätzlichen Faktor $\log n$.

Satz 2.4.1 *Unter den Annahmen 1 - 6 und 9 - 11 gelten für die Schätzung der Kovarianzmatrix im Faktormodell und die Stichprobenkovarianzmatrix folgende Konvergenzraten*

$$\begin{aligned} E(\|\hat{\Sigma}_n - \Sigma_n\|^2) &= O(n^{-1}[\log n]p^2K^2), \\ E(\|\hat{\Sigma}_{sam} - \Sigma_n\|^2) &= O(n^{-1}[\log n]p^2K^2). \end{aligned}$$

Beweis. Wir nutzen analog zum Beweis von Satz 2.3.1 die Darstellung (2.2.13) und Korollar D.0.17, Bemerkung D.0.18 und die Lemmata C.0.4, C.0.6 und C.0.9.

$$\begin{aligned} E(\|\hat{\Sigma}_n - \Sigma_n\|^2) &\leq 4(E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|^2) + 2E(\|\mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{C}_n^T\|^2) + E(\|\mathbf{C}_n \text{cov}(\mathbf{f}) \mathbf{C}_n^T\|^2) \\ &\quad + E(\|\mathbf{F}_n\|^2)) \\ &= O(n^{-1}[\log n]p^2K^2) + O(n^{-1}p^2K) + O(n^{-2}p^2K) + O(n^{-1}p) \\ &\quad + O(n^{-2}pK^2). \end{aligned}$$

Dabei dominiert wie in Satz 2.3.1 die Rate des ersten Summanden, hier mit zusätzlichem Faktor $\log n$, daher erhalten wir das Resultat

$$E(\|\hat{\Sigma}_n - \Sigma_n\|^2) = O(n^{-1}[\log n]p^2K^2).$$

Analog erhalten wir mit (2.2.14), Korollar D.0.17, Bemerkung D.0.18, Lemma C.0.10 und den Korollaren C.0.7 und C.0.8 folgende Abschätzung für die Stichprobenkovarianzmatrix:

$$\begin{aligned} E(\|\hat{\Sigma}_{sam} - \Sigma_n\|^2) &\leq 4(E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|^2) + E(\|\mathbf{G}_n\|^2) + \frac{1}{(n-1)^2}E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|^2) \\ &\quad + \frac{1}{n^2(n-1)^2}E(\|\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T\|^2)) \end{aligned}$$

$$= O(n^{-1}[\log n]p^2K^2) + O(n^{-1}p^2) + O(n^{-1}p^2K) + O(n^{-1}p^2K).$$

Auch hier dominiert die Rate des ersten Summanden und wir erhalten das Resultat

$$E(\|\hat{\Sigma}_{sam} - \Sigma_n\|^2) = O(n^{-1}[\log n]p^2K^2).$$

□

Da der Übergang zu den stationären Faktoren nur die Rate des Summanden verschlechtert, der schon in Abschnitt 2.3 die dominierende Rate lieferte, geht die Änderung hier direkt in die Aussage des Satzes ein. Da der Faktor $\log n$ nur sehr langsam wächst, macht sich der Übergang zu stationären Faktoren in den Konvergenzraten kaum bemerkbar. Dies werden wir auch in den Simulationen in Kapitel 4 sehen. Wir untersuchen nun, wie sich der zusätzliche Faktor $\log n$ auf das Konvergenzverhalten in der Norm $\|\cdot\|_{\Sigma_n}$ auswirkt.

Satz 2.4.2 *Unter den Annahmen 1 - 6 und 9 - 11 gelten für die Schätzung der Kovarianzmatrix aus dem Faktormodell und die Stichprobenkovarianzmatrix folgende Aussagen:*

$$E(\|\hat{\Sigma}_n - \Sigma_n\|_{\Sigma_n}^2) = O(n^{-1}[\log n]p^{-1}K^2) + O(n^{-1}K) + O(n^{-2}pK)$$

und

$$E(\|\hat{\Sigma}_{sam} - \Sigma_n\|_{\Sigma_n}^2) = O(n^{-1}[\log n]p^{-1}K^2) + O(n^{-1}p).$$

Beweis. Der Beweis erfolgt analog zum Beweis von Satz 2.3.2. Wir nutzen die Darstellung aus (2.2.13) und die Resultate aus Korollar D.0.17, Bemerkung D.0.18 und den Lemmata C.0.4, C.0.6 und C.0.9.

$$\begin{aligned} E(\|\hat{\Sigma}_n - \Sigma_n\|_{\Sigma_n}^2) &\leq 4(E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|_{\Sigma_n}^2) + 2E(\|\mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{C}_n^T\|_{\Sigma_n}^2) + E(\|\mathbf{C}_n \text{cov}(\mathbf{f}) \mathbf{C}_n^T\|_{\Sigma_n}^2) \\ &\quad + E(\|\mathbf{F}_n\|_{\Sigma_n}^2)) \\ &= O(n^{-1}[\log n]p^{-1}K) + O(n^{-1}K) + O(n^{-2}pK) + O(n^{-1}) + O(n^{-2}K^2). \end{aligned}$$

Dabei dominieren die Raten der ersten drei Summanden und wir erhalten

$$E(\|\hat{\Sigma}_n - \Sigma_n\|_{\Sigma_n}^2) = O(n^{-1}[\log n]p^{-1}K) + O(n^{-1}K) + O(n^{-2}pK).$$

Mit der Darstellung (2.2.14), Korollar D.0.17, Bemerkung D.0.18, Lemma C.0.10 und den Korollaren C.0.7 und C.0.8 erhalten wir für die Stichprobenkovarianzmatrix

$$\begin{aligned} E(\|\hat{\Sigma}_{sam} - \Sigma_n\|_{\Sigma_n}^2) &\leq 4(E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|_{\Sigma_n}^2) + E(\|\mathbf{G}_n\|_{\Sigma_n}^2) + \frac{1}{(n-1)^2} E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|_{\Sigma_n}^2) \\ &\quad + \frac{1}{n^2(n-1)^2} E(\|\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{E}^T\|_{\Sigma_n}^2)) \\ &= O(n^{-1}[\log n]p^{-1}K) + O(n^{-1}p) + O(n^{-1}K) + O(n^{-1}K). \end{aligned}$$

Dabei dominieren die Raten der ersten beiden Summanden und wir erhalten

$$E(\|\hat{\Sigma}_{sam} - \Sigma_n\|_{\Sigma_n}^2) = O(n^{-1}[\log n]p^{-1}K) + O(n^{-1}p).$$

□

Da der Faktor $\log n$ nur sehr langsam wächst, wir aber wachsende Portfoliodimensionen p betrachten, kann in den Raten beider Schätzer der erste Summand vernachlässigt werden. Wir erhalten daher die gleichen Konvergenzraten wie in Satz 2.3.2. Dort hatten wir gezeigt, dass für Dimensionen und Stichprobengrößen, für die die Schätzungen konsistent sind, die Schätzung mithilfe des Faktormodells besser ist als die Stichprobenkovarianzmatrix (vgl. Bemerkung 2.3.3).

Für die inverse Kovarianzmatrix gilt folgender Satz:

Satz 2.4.3 *Unter den Annahmen 1 - 7 und 9 - 11 gilt für die Inverse des Kovarianzmatrixschätzers aus dem Faktormodell*

$$\|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| = O_P(n^{-\frac{1}{2}}pK^{\frac{3}{2}}) + O_P(n^{-\frac{1}{2}}[\log n]^{\frac{1}{2}}p^{-1}K)$$

und falls $p^{\frac{3}{2}}K = o(n^{\frac{1}{2}})$, folgt

$$\|\hat{\Sigma}_{sam}^{-1} - \Sigma_n^{-1}\| = O_P(n^{-\frac{1}{2}}[\log n]^{\frac{1}{2}}p^2K).$$

Beweis. Wir führen den Beweis analog zu dem Beweis von Satz 2.3.4 und nutzen für die Inverse der Kovarianzmatrix im Faktormodell die Darstellung nach (2.3.7). Mit Lemma D.0.20, Bemerkung D.0.21 und den Lemmata C.0.12 bis C.0.14 erhalten wir die Abschätzung

$$\begin{aligned} \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| &\leq \|\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}\| + \|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\mathbf{B}}_n[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\| \\ &\quad + \|\Sigma_0^{-1}\hat{\mathbf{B}}_n[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\| \\ &\quad + \|\Sigma_0^{-1}(\hat{\mathbf{B}}_n - \mathbf{B}_n)[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}\hat{\mathbf{B}}_n^T\Sigma_0^{-1}\| \\ &\quad + \|\Sigma_0^{-1}\mathbf{B}_n[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1}(\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T)\Sigma_0^{-1}\| \\ &\quad + \|\Sigma_0^{-1}\mathbf{B}_n\{\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\Sigma}_0^{-1}\hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\Sigma_0^{-1}\mathbf{B}_n]^{-1}\}\mathbf{B}_n^T\Sigma_0^{-1}\| \\ &= (O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(n^{-1}p^{\frac{1}{2}}K)) + (O_P(n^{-\frac{1}{2}}p) + O_P(n^{-1}pK)) \\ &\quad + (O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(n^{-1}p^{\frac{1}{2}}K)) + O_P(n^{-\frac{1}{2}}pK^{\frac{3}{2}}) + O_P(n^{-\frac{1}{2}}pK^{\frac{3}{2}}) \\ &\quad + (O_P(n^{-\frac{1}{2}}[\log n]^{\frac{1}{2}}p^{-1}K) + O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}K^{\frac{1}{2}})). \end{aligned}$$

Hier dominiert die Rate des gleichen Summanden wie im Fall mit unabhängigen Faktoren, dazu kommt die Rate des Summanden, der wegen der stationären Faktoren einen zusätzlichen Faktor $[\log n]^{\frac{1}{2}}$ erhalten hat:

$$\|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| = O_P(n^{-\frac{1}{2}}pK^{\frac{3}{2}}) + O_P(n^{-\frac{1}{2}}[\log n]^{\frac{1}{2}}p^{-1}K).$$

Für die Stichprobenkovarianzmatrix erhalten wir das Resultat für die Inverse völlig analog zum Beweis von Satz 2.3.4. Hierbei ist lediglich zu beachten, dass sich die Rate der Stichprobenkovarianzmatrix der Faktoren in der Frobeniusnorm bei stationären Faktoren von der bei unabhängigen Faktoren unterscheidet. Wir erhalten

$$\|\hat{\Sigma}_{sam}^{-1} - \Sigma_n^{-1}\| = O_P(n^{-\frac{1}{2}}[\log n]^{\frac{1}{2}}p^2K).$$

□

In der Rate der Inversen der Kovarianzmatrix aus dem Faktormodell dominiert der zweite Summand, da $[\log n]^{\frac{1}{2}}$ im Vergleich zur Dimension p nur sehr langsam wächst. Wir erhalten daher die gleiche Rate wie unter unabhängig identisch verteilten Faktoren. Für die Inverse der Stichprobenkovarianzmatrix verschlechtert sich die Rate im Vergleich zu den unabhängig identisch verteilten Faktoren um den Faktor $[\log n]^{\frac{1}{2}}$. Weiterhin ist hier der Vorteil der Schätzung mithilfe des Faktormodells gegenüber der Stichprobenkovarianzmatrix sehr deutlich.

Aus den Resultaten dieses Abschnitts wird deutlich, dass der Übergang von unabhängigen zu stationären Faktoren im Faktormodell nur geringe Auswirkungen auf die Asymptotik der betrachteten Schätzer der Kovarianzmatrix hat. Dieses Ergebnis wird auch in der Simulation in Kapitel 4 deutlich, denn oft sind die Unterschiede zwischen den Simulationsergebnissen der beiden betrachteten Fälle kaum sichtbar.

3. Auswirkungen der Kovarianzmatrixschätzung auf Anwendungen in der Finanzwirtschaft

Zwei wichtige Bereiche der Finanzwirtschaft, in denen Kovarianzmatrizen eine entscheidende Rolle spielen, sind das Risikomanagement und die Portfoliotheorie. Im Risikomanagement werden Kovarianzmatrizen dazu benötigt, das Risiko eines Portfolios zu bestimmen, Risikovorhersagen zu machen oder das Risiko unter verschiedenen Krisenszenarien abschätzen zu können. In der Portfoliotheorie werden Kovarianzmatrizen für das Treffen von optimalen Investitionsentscheidungen und die Bestimmung der Portfoliodiversifikation unter verschiedenen Bedingungen benötigt.

In beiden Anwendungen basieren viele Rechnungen auf der Varianz eines linearen Portfolios. Betrachten wir Portfoliogewichte $\boldsymbol{\xi} \in \mathbb{R}^p$ und die Wertpapierüberschussrenditen $\mathbf{y} = (Y_1, \dots, Y_p)^T$, so berechnet sich die Überschussrendite Y_P eines linearen Portfolios durch

$$Y_P = \boldsymbol{\xi}^T \mathbf{y}.$$

Die Varianz dieses Portfolios lässt sich dann mithilfe der Kovarianzmatrix der Wertpapierüberschussrenditen schreiben als

$$\text{var}(Y_P) = \boldsymbol{\xi}^T \boldsymbol{\Sigma} \boldsymbol{\xi}.$$

Aus dieser Darstellung wird deutlich, dass die Kovarianzmatrix bzw. deren Schätzung notwendigerweise positiv semidefinit sein muss, da sonst negative Portfoliovarianzen auftreten könnten.

In den folgenden Abschnitten gehen wir näher auf die Rolle der Kovarianzmatrix in den beiden genannten Bereichen ein und untersuchen die Auswirkungen der Schätzungen der Kovarianzmatrix aus Kapitel 2 auf die betrachteten Anwendungen.

3.1. Portfoliotheorie

Eine wichtige Aufgabe der Portfoliotheorie ist die Entscheidung über die Verteilung von Kapital mit dem Ziel optimale Portfolios zu konstruieren. Optimale Portfolios können mithilfe diverser Zielsetzungen beschrieben werden, die meist die Betrachtung von Risiko und Rendite unter verschiedenen Nebenbedingungen umfassen. Die Konstruktion von optimalen Portfolios geht auf Markowitz (1952) zurück, dessen allgemeiner Ansatz auf dem Ausgleich von Rendite und Risiko basiert. Dabei werden die Wertpapierstückzahlen des $\mu - \sigma$ -optimalen Portfolios so bestimmt, dass unter einer gegebenen Renditevorgabe

die Varianz des Portfolios minimiert wird. In diesem Fall wird also ein höheres Risiko akzeptiert, solange mit dem höheren Risiko auch eine höhere Rendite einhergeht. Diesem Ansatz gegenüber steht die Konstruktion des globalen varianzminimalen Portfolios. Dabei geht es nur darum die Wertpapiergewichte so zu bestimmen, dass die Varianz des Portfolios ungeachtet der Portfoliorendite minimal wird.

Bei beiden Ansätzen ist es denkbar, weitere Nebenbedingungen zu beachten. So wäre es sinnvoll optimale Portfolios unter verschiedenen Handelsbeschränkungen wie beispielsweise Leerverkaufsverboten zu betrachten oder unternehmensspezifische Einschränkungen wie die Beschränkung der Wertpapieranteile, die in bestimmte Branchen investiert werden dürfen, zu berücksichtigen. Wir betrachten hier aber nur den einfachsten Fall, in dem Wertpapiergewichte beliebige Werte annehmen dürfen, solange die Summation zu Eins gewährleistet ist.

In den folgenden Abschnitten beschreiben wir die Konstruktion der optimalen Portfolios in den beiden oben beschriebenen Fällen und untersuchen dann, welche Auswirkungen die Schätzungen der Kovarianzmatrix auf die Varianz der konstruierten Portfolios haben.

3.1.1. Allgemeiner Markowitzfall

Das $\mu - \sigma$ -optimale Portfolio nach Markowitz ist definiert als Lösung $\xi_n \in \mathbb{R}^p$ des folgenden Minimierungsproblems: Wir suchen die varianzminimierenden Portfoliogewichte, sodass die Gewichte sich zu Eins summieren und das Portfolio die vorgegebene erwartete Portfoliorendite γ_n erzielt. Formal schreiben wir das Minimierungsproblem als

$$\min_{\xi_n} \xi_n^T \Sigma_n \xi_n \quad (3.1.1)$$

unter den Bedingungen

$$\begin{aligned} \xi_n^T \mathbf{1} &= 1 \\ \xi_n^T \mu_n &= \gamma_n. \end{aligned}$$

Dabei bezeichnet $\mu_n = E(\mathbf{y})$ die erwartete Portfoliorendite. Mit dem Lagrangeansatz bestimmen wir das optimale Portfolio wie folgt

$$L(\xi_n, \lambda_1, \lambda_2) = \xi_n^T \Sigma_n \xi_n + \lambda_1(1 - \xi_n^T \mathbf{1}) + \lambda_2(\gamma_n - \xi_n^T \mu_n).$$

Die partiellen Ableitungen liefern folgendes Gleichungssystem

$$\frac{\partial L}{\partial \xi_n^T} = \Sigma_n \xi_n - \lambda_1 \mathbf{1} - \lambda_2 \mu_n = 0 \quad (3.1.2)$$

$$\frac{\partial L}{\partial \lambda_1} = 1 - \xi_n^T \mathbf{1} = 0 \quad (3.1.3)$$

$$\frac{\partial L}{\partial \lambda_2} = \gamma_n - \xi_n^T \mu_n = 0. \quad (3.1.4)$$

Aus (3.1.2) folgt

$$\Sigma_n \xi_n = \lambda_1 \mathbf{1} + \lambda_2 \mu_n \text{ und insbesondere } \xi_n = \lambda_1 \Sigma_n^{-1} \mathbf{1} + \lambda_2 \Sigma_n^{-1} \mu_n. \quad (3.1.5)$$

Eingesetzt in (3.1.3) und (3.1.4) muss nun das folgende Gleichungssystem gelöst werden:

$$\begin{aligned} 1 &= \lambda_1 \mathbf{1}^T \Sigma_n^{-1} \mathbf{1} + \lambda_2 \mu_n^T \Sigma_n^{-1} \mathbf{1} \\ \gamma_n &= \lambda_1 \mathbf{1}^T \Sigma_n^{-1} \mu_n + \lambda_2 \mu_n^T \Sigma_n^{-1} \mu_n. \end{aligned}$$

Wir nutzen die folgende Notation:

$$\begin{aligned} \varphi_n &:= \mathbf{1}^T \Sigma_n^{-1} \mathbf{1} \\ \psi_n &:= \mathbf{1}^T \Sigma_n^{-1} \mu_n \\ \phi_n &:= \mu_n^T \Sigma_n^{-1} \mu_n, \end{aligned}$$

das heißt in Matrixschreibweise gilt nun

$$\begin{aligned} \begin{pmatrix} \varphi_n & \psi_n \\ \psi_n & \phi_n \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} &= \begin{pmatrix} 1 \\ \gamma_n \end{pmatrix} \\ \Rightarrow \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} &= \begin{pmatrix} \varphi_n & \psi_n \\ \psi_n & \phi_n \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \gamma_n \end{pmatrix} = \frac{1}{\varphi_n \phi_n - \psi_n^2} \begin{pmatrix} \phi_n & -\psi_n \\ -\psi_n & \varphi_n \end{pmatrix} \begin{pmatrix} 1 \\ \gamma_n \end{pmatrix} \\ \Rightarrow \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} &= \frac{1}{\varphi_n \phi_n - \psi_n^2} \begin{pmatrix} \phi_n - \psi_n \gamma_n \\ \varphi_n \gamma_n - \psi_n \end{pmatrix}. \end{aligned} \quad (3.1.6)$$

Eingesetzt in (3.1.5) erhalten wir die optimalen Wertpapiergewichte

$$\begin{aligned} \xi_n &= \lambda_1 \Sigma_n^{-1} \mathbf{1} + \lambda_2 \Sigma_n^{-1} \mu_n \\ &= \frac{\phi_n - \gamma_n \psi_n}{\varphi_n \phi_n - \psi_n^2} \Sigma_n^{-1} \mathbf{1} + \frac{\varphi_n \gamma_n - \psi_n}{\varphi_n \phi_n - \psi_n^2} \Sigma_n^{-1} \mu_n. \end{aligned}$$

Die Varianz des optimalen Portfolios lässt sich nun darstellen als

$$\begin{aligned} \xi_n^T \Sigma_n \xi_n &= (\lambda_2 \mu_n^T \Sigma_n^{-1} + \lambda_1 \mathbf{1}^T \Sigma_n^{-1}) \Sigma_n (\lambda_1 \Sigma_n^{-1} \mathbf{1} + \lambda_2 \Sigma_n^{-1} \mu_n) \\ &= (\lambda_2 \mu_n^T \Sigma_n^{-1} + \lambda_1 \mathbf{1}^T \Sigma_n^{-1}) (\lambda_1 \Sigma_n \Sigma_n^{-1} \mathbf{1} + \lambda_2 \Sigma_n \Sigma_n^{-1} \mu_n) \\ &= 2\lambda_1 \lambda_2 \mu_n^T \Sigma_n^{-1} \mathbf{1} + \lambda_2^2 \mu_n^T \Sigma_n^{-1} \mu_n + \lambda_1^2 \mathbf{1}^T \Sigma_n^{-1} \mathbf{1} \\ &= 2\lambda_1 \lambda_2 \psi_n + \lambda_2^2 \phi_n + \lambda_1^2 \varphi_n \\ &\stackrel{(3.1.6)}{=} 2\psi_n \frac{(\phi_n - \gamma_n \psi_n)(\varphi_n \gamma_n - \psi_n)}{(\varphi_n \phi_n - \psi_n^2)(\varphi_n \phi_n - \psi_n^2)} + \phi_n \frac{[\varphi_n \gamma_n - \psi_n]^2}{(\varphi_n \phi_n - \psi_n^2)(\varphi_n \phi_n - \psi_n^2)} \\ &\quad + \varphi_n \frac{[\phi_n - \psi_n \gamma_n]^2}{(\varphi_n \phi_n - \psi_n^2)(\varphi_n \phi_n - \psi_n^2)} \end{aligned}$$

$$\begin{aligned}
&= \frac{2\psi_n\phi_n\varphi_n\gamma_n - 2\psi_n^2\phi_n + \phi_n\psi_n^2 + \varphi_n\phi_n^2 - 2\psi_n^2\gamma_n^2\varphi_n + \phi_n\varphi_n^2\gamma_n^2 + \varphi_n\psi_n^2\gamma_n^2}{(\varphi_n\phi_n - \psi_n^2)(\varphi_n\phi_n - \psi_n^2)} \\
&+ \frac{-2\phi_n\varphi_n\gamma_n\psi_n + 2\psi_n^3\gamma_n - 2\varphi_n\phi_n\psi_n\gamma_n}{(\varphi_n\phi_n - \psi_n^2)(\varphi_n\phi_n - \psi_n^2)} \\
&= \frac{2\psi_n\phi_n\varphi_n\gamma_n + \phi_n(\varphi_n\phi_n - \psi_n^2) + \varphi_n\gamma_n^2(\varphi_n\phi_n - \psi_n^2)}{(\varphi_n\phi_n - \psi_n^2)(\varphi_n\phi_n - \psi_n^2)} \\
&+ \frac{-2\phi_n\varphi_n\gamma_n\psi_n - 2\psi_n\gamma_n(\varphi_n\phi_n - \psi_n^2)}{(\varphi_n\phi_n - \psi_n^2)(\varphi_n\phi_n - \psi_n^2)} \\
&= \frac{\varphi_n\gamma_n^2 + \phi_n - 2\psi_n\gamma_n}{\varphi_n\phi_n - \psi_n^2}. \tag{3.1.7}
\end{aligned}$$

Um die Auswirkungen der Schätzung der Kovarianzmatrix auf die Varianz des Portfolios zu betrachten, setzen wir in die Definitionen von φ_n , ψ_n und ϕ_n die Schätzer (2.1.2) und (2.1.6) ein. Außerdem schätzen wir $\boldsymbol{\mu}_n$ mithilfe der Stichprobe durch: $\hat{\boldsymbol{\mu}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$. Dann gilt folgendes Resultat:

Satz 3.1.1 *Falls $\varphi_n\phi_n - \psi_n^2$ von Null weg beschränkt ist und die Terme $\frac{\varphi_n}{(\varphi_n\phi_n - \psi_n^2)}$, $\frac{\psi_n}{(\varphi_n\phi_n - \psi_n^2)}$, $\frac{\phi_n}{(\varphi_n\phi_n - \psi_n^2)}$ und γ_n beschränkt sind, folgt unter den Annahmen 1-8 für die Portfoliovarianz im allgemeinen Markowitzfall*

$$\hat{\boldsymbol{\xi}}_n^T \hat{\boldsymbol{\Sigma}}_n \hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n = O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}}) \tag{3.1.8}$$

und

$$\hat{\boldsymbol{\xi}}_n^T \hat{\boldsymbol{\Sigma}}_{sam} \hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n = O_P(n^{-\frac{1}{2}} p^3 K). \tag{3.1.9}$$

Beweis. Für den Beweis nutzen wir die Varianzdarstellung nach (3.1.7). Wir beginnen mit (3.1.8) und schätzen die Konvergenzen der einzelnen Bestandteile von (3.1.7) zunächst einzeln ab, beginnend mit φ_n :

$$|\hat{\varphi}_n - \varphi_n| = |\mathbf{1}^T \hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1} - \mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} \mathbf{1}| = |\mathbf{1}^T (\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}) \mathbf{1}|,$$

da es sich hierbei um ein Skalar handelt, können wir von diesem Ausdruck die Spur betrachten und mit den Rechenregeln der Spur und (B.2.3) wie folgt umformen:

$$\begin{aligned}
|\hat{\varphi}_n - \varphi_n| &= \text{tr}\{|\mathbf{1}^T (\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}) \mathbf{1}|\} = |\text{tr}\{(\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}) \mathbf{1} \mathbf{1}^T\}| \\
&\leq \|\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}\| \|\mathbf{1} \mathbf{1}^T\| = p \|\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}\|.
\end{aligned}$$

Nach Satz 2.3.4 gilt $\|\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}\| = O_P(n^{-\frac{1}{2}} p K^{\frac{3}{2}})$, daher folgt hier

$$p \|\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}\| = O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}}).$$

Nach Annahme 5 gilt $b_n = E(\|\mathbf{y}\|^2) = E([Y_1]^2 + \dots + [Y_p]^2) = O(p)$. Daher gilt mit der Jensen Ungleichung

$$\|\boldsymbol{\mu}_n\|^2 = \|E(\mathbf{y})\|^2 = [E(Y_1)]^2 + \dots + [E(Y_p)]^2 \leq E([Y_1]^2) + \dots + E([Y_p]^2) = O(p),$$

also $\|\boldsymbol{\mu}_n\| = O(p^{\frac{1}{2}})$.

Außerdem gilt wegen der Bedingung $E(\|\mathbf{y}\|^2) = O(p)$ aus Annahme 5:

$$\begin{aligned} E(\|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\|^2) &= E(\|\frac{1}{n}\mathbf{Y}\mathbf{1} - E(\mathbf{y})\|^2) = E\left(\left\|\frac{1}{n}\sum_{i=1}^n \mathbf{y}_i - E(\mathbf{y})\right\|^2\right) \\ &= E\left(\sum_{j=1}^p \left[\frac{1}{n}\sum_{i=1}^n (Y_{ji} - E(Y_j))\right]^2\right) = \sum_{j=1}^p \frac{1}{n^2} E\left(\left[\sum_{i=1}^n (Y_{ji} - E(Y_j))\right]^2\right), \end{aligned}$$

da die $Y_{ji} - E(Y_j)$ für $j = 1, \dots, p$ unabhängig identisch verteilte, zentrierte Zufallsvariablen sind, folgt

$$\begin{aligned} E(\|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\|^2) &= \sum_{j=1}^p \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n (Y_{ji} - E(Y_j))\right) = \sum_{j=1}^p \frac{1}{n} \text{var}(Y_{j1} - E(Y_j)) \\ &= \frac{1}{n} \sum_{j=1}^p E([Y_{j1} - E(Y_j)]^2) = \frac{1}{n} \sum_{j=1}^p E(Y_{j1}^2) - [E(Y_j)]^2 \leq \frac{1}{n} \sum_{j=1}^p E(Y_{j1}^2) = O(n^{-1}p). \end{aligned}$$

Also insbesondere

$$\|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\| = O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}). \quad (3.1.10)$$

Weiter gilt

$$\begin{aligned} |\hat{\psi}_n - \psi_n| &= |\mathbf{1}^T \hat{\boldsymbol{\Sigma}}_n^{-1} \hat{\boldsymbol{\mu}}_n - \mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n| \\ &= |\mathbf{1}^T \hat{\boldsymbol{\Sigma}}_n^{-1} \hat{\boldsymbol{\mu}}_n - \mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} \hat{\boldsymbol{\mu}}_n + \mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} \hat{\boldsymbol{\mu}}_n - \mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n| \\ &= |\mathbf{1}^T (\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}) \hat{\boldsymbol{\mu}}_n + \mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n)| \end{aligned}$$

und mit der Dreiecksungleichung teilen wir diesen Ausdruck in zwei Summanden auf und schreiben die Skalare mithilfe der Spur um:

$$\begin{aligned} |\hat{\psi}_n - \psi_n| &\leq |\mathbf{1}^T (\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}) \hat{\boldsymbol{\mu}}_n| + |\mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n)| \\ &= \text{tr}\{|\mathbf{1}^T (\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}) \hat{\boldsymbol{\mu}}_n|\} + \text{tr}\{|\mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n)|\} \\ &= |\text{tr}\{(\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}) \hat{\boldsymbol{\mu}}_n \mathbf{1}^T\}| + |\text{tr}\{\boldsymbol{\Sigma}_n^{-1} (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n) \mathbf{1}^T\}|. \end{aligned}$$

Mithilfe von (B.2.3) und der Dreiecksungleichung schätzen wir diese Terme ab und ergänzen durch eine Nulladdition

$$|\hat{\psi}_n - \psi_n| \leq \|\mathbf{1}^T\| \|\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n\| + \|\mathbf{1}^T\| \|\boldsymbol{\Sigma}_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\|$$

$$\begin{aligned}
&= \|\mathbf{1}^T\| \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n + \boldsymbol{\mu}_n - \boldsymbol{\mu}_n\| + \|\mathbf{1}^T\| \|\Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\| \\
&\leq \|\mathbf{1}^T\| \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| (\|\boldsymbol{\mu}_n\| + \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\|) + \|\mathbf{1}^T\| \|\Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\| \\
&= p^{\frac{1}{2}} \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| (O(p^{\frac{1}{2}}) + O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}})) + p^{\frac{1}{2}} O(p^{\frac{1}{2}}) O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}}) \\
&= O(p) \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| + O_P(n^{-\frac{1}{2}} p^{\frac{3}{2}}).
\end{aligned}$$

Dabei dominiert die Rate des ersten Summanden und mit (2.3.5) folgt

$$|\hat{\psi}_n - \psi_n| \leq O(p) \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| = O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}}).$$

Analog gilt

$$\begin{aligned}
|\hat{\phi}_n - \phi_n| &= |\hat{\boldsymbol{\mu}}_n^T \hat{\Sigma}_n^{-1} \hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n^T \Sigma_n^{-1} \boldsymbol{\mu}_n| \\
&= |\hat{\boldsymbol{\mu}}_n^T \hat{\Sigma}_n^{-1} \hat{\boldsymbol{\mu}}_n - \hat{\boldsymbol{\mu}}_n^T \Sigma_n^{-1} \hat{\boldsymbol{\mu}}_n + \hat{\boldsymbol{\mu}}_n^T \Sigma_n^{-1} \hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n^T \Sigma_n^{-1} \hat{\boldsymbol{\mu}}_n + \boldsymbol{\mu}_n^T \Sigma_n^{-1} \hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n^T \Sigma_n^{-1} \boldsymbol{\mu}_n| \\
&= |\hat{\boldsymbol{\mu}}_n^T (\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}) \hat{\boldsymbol{\mu}}_n + (\hat{\boldsymbol{\mu}}_n^T - \boldsymbol{\mu}_n^T) \Sigma_n^{-1} \hat{\boldsymbol{\mu}}_n + \boldsymbol{\mu}_n^T \Sigma_n^{-1} (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n)| \\
&\leq |\hat{\boldsymbol{\mu}}_n^T (\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}) \hat{\boldsymbol{\mu}}_n| + |(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n)^T \Sigma_n^{-1} \hat{\boldsymbol{\mu}}_n| + |\boldsymbol{\mu}_n^T \Sigma_n^{-1} (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n)| \\
&= |\text{tr}\{(\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}) \hat{\boldsymbol{\mu}}_n \hat{\boldsymbol{\mu}}_n^T\}| + |\text{tr}\{\Sigma_n^{-1} \hat{\boldsymbol{\mu}}_n (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n)^T\}| + |\text{tr}\{\Sigma_n^{-1} (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n) \boldsymbol{\mu}_n^T\}|.
\end{aligned}$$

Wir schätzen durch (B.2.3) ab und ergänzen wie oben geeignet durch Nulladditionen

$$\begin{aligned}
|\hat{\phi}_n - \phi_n| &\leq \|\hat{\boldsymbol{\mu}}_n^T\| \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n\| + \|\hat{\boldsymbol{\mu}}_n^T - \boldsymbol{\mu}_n^T\| \|\Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n\| + \|\boldsymbol{\mu}_n^T\| \|\Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\| \\
&= \|\hat{\boldsymbol{\mu}}_n^T - \boldsymbol{\mu}_n^T + \boldsymbol{\mu}_n^T\| \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n + \boldsymbol{\mu}_n\| \\
&\quad + \|\hat{\boldsymbol{\mu}}_n^T - \boldsymbol{\mu}_n^T\| \|\Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n + \boldsymbol{\mu}_n\| + \|\boldsymbol{\mu}_n^T\| \|\Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\|.
\end{aligned}$$

Dann folgt mit der Dreiecksungleichung

$$\begin{aligned}
|\hat{\phi}_n - \phi_n| &\leq (\|\hat{\boldsymbol{\mu}}_n^T - \boldsymbol{\mu}_n^T\| + \|\boldsymbol{\mu}_n^T\|) \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| (\|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\| + \|\boldsymbol{\mu}_n\|) \\
&\quad + \|\hat{\boldsymbol{\mu}}_n^T - \boldsymbol{\mu}_n^T\| \|\Sigma_n^{-1}\| (\|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\| + \|\boldsymbol{\mu}_n\|) + \|\boldsymbol{\mu}_n^T\| \|\Sigma_n^{-1}\| \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\| \\
&= O(p^{\frac{1}{2}}) \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| O(p^{\frac{1}{2}}) + O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}}) O(p^{\frac{1}{2}}) O(p^{\frac{1}{2}}) + O(p^{\frac{1}{2}}) O(p^{\frac{1}{2}}) O_P(p^{\frac{1}{2}} n^{-\frac{1}{2}}) \\
&= O(p) \|\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}\| + O_P(p^{\frac{3}{2}} n^{-\frac{1}{2}}) + O_P(p^{\frac{3}{2}} n^{-\frac{1}{2}}).
\end{aligned}$$

Dabei dominiert die Rate des ersten Summanden und es folgt mit (2.3.5)

$$|\hat{\phi}_n - \phi_n| = O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}}).$$

Nach Voraussetzung ist $\varphi_n \phi_n - \psi_n^2$ von Null weg beschränkt und $\frac{\varphi_n}{(\varphi_n \phi_n - \psi_n^2)}$, $\frac{\psi_n}{(\varphi_n \phi_n - \psi_n^2)}$, $\frac{\phi_n}{(\varphi_n \phi_n - \psi_n^2)}$ sowie γ_n sind beschränkt. Es folgt

$$|\hat{\boldsymbol{\xi}}_n^T \hat{\Sigma}_n \hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_n^T \Sigma_n \boldsymbol{\xi}_n|$$

$$\stackrel{(3.1.7)}{=} \left| \frac{\hat{\varphi}_n^2 \gamma_n^2 + \hat{\phi}_n - 2\hat{\psi}_n \gamma_n}{\hat{\varphi}_n \hat{\phi}_n - \hat{\psi}_n^2} - \frac{\varphi_n \gamma_n^2 + \phi_n - 2\psi_n \gamma_n}{\varphi_n \phi_n - \psi_n^2} \right| = O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}}).$$

Für den Beweis von (3.1.9) können wir die Abschätzungen aus dem ersten Teil des Beweises direkt übernehmen. Dort hatten wir gezeigt, dass $|\hat{\phi}_n - \phi_n| \leq p \|\hat{\Sigma}_{sam}^{-1} - \Sigma_n^{-1}\|$ gilt. Daher gilt mit der Aussage (2.3.6) aus Satz 2.3.4:

$$|\hat{\varphi}_n - \varphi_n| = O_P(n^{-\frac{1}{2}} p^3 K).$$

Analog zum ersten Teil des Beweises gelten

$$|\hat{\psi}_n - \psi_n| = O_P(n^{-\frac{1}{2}} p^3 K)$$

und

$$|\hat{\phi}_n - \phi_n| = O_P(n^{-\frac{1}{2}} p^3 K).$$

Es folgt für die Varianz des Portfolios

$$\begin{aligned} & |\hat{\xi}_n^T \hat{\Sigma}_{sam} \hat{\xi}_n - \xi_n^T \Sigma_n \xi_n| \\ &= \left| \frac{\hat{\varphi}_n^2 \gamma_n^2 + \hat{\phi}_n - 2\hat{\psi}_n \gamma_n}{\hat{\varphi}_n \hat{\phi}_n - \hat{\psi}_n^2} - \frac{\varphi_n \gamma_n^2 + \phi_n - 2\psi_n \gamma_n}{\varphi_n \phi_n - \psi_n^2} \right| = O_P(n^{-\frac{1}{2}} p^3 K). \end{aligned}$$

□

Da in diesen Beweis direkt die Konvergenzrate der Inversen der Kovarianzmatrix aus Satz 2.3.4 eingeht, ist hier wieder deutlich der Vorteil der Schätzung aus dem Faktormodell zu erkennen. Die Dimension des Portfolios geht durch einen zusätzlichen Faktor p ein. Für die Betrachtung der stationären Faktoren muss die veränderte Rate für die Inverse der Kovarianzmatrix aus Satz 2.4.3 eingesetzt werden. Für den Beweis ist nur eine kleine Modifikation notwendig.

Korollar 3.1.2 *Falls $\varphi_n \phi_n - \psi_n^2$ von Null weg beschränkt ist und die Terme $\frac{\varphi_n}{(\varphi_n \phi_n - \psi_n^2)}$, $\frac{\psi_n}{(\varphi_n \phi_n - \psi_n^2)}$, $\frac{\phi_n}{(\varphi_n \phi_n - \psi_n^2)}$ und γ_n beschränkt sind, so folgt unter den Annahmen 1-7 und 9 - 11 für die Portfoliovarianz im allgemeinen Markowitzfall*

$$\hat{\xi}_n^T \hat{\Sigma}_{sam} \hat{\xi}_n - \xi_n^T \Sigma_n \xi_n = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} K) + O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}})$$

und

$$\hat{\xi}_n^T \hat{\Sigma}_{sam} \hat{\xi}_n - \xi_n^T \Sigma_n \xi_n = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} p^3 K).$$

Beweis. Wir müssen zeigen, dass für $\|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\|$ die gleiche Konvergenzrate gilt, wie im unabhängig identisch verteilten Fall. Für $j = 1, \dots, p$ sind die $Y_{jt} - E(Y_j)$ stationär und zentriert, daher folgt analog zum Beweis von (3.1.10) mit Annahme 5

$$\begin{aligned} E(\|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\|^2) &= \sum_{j=1}^p \frac{1}{n^2} E\left(\left[\sum_{t=1}^n (Y_{jt} - E(Y_j))\right]^2\right), \\ &= \sum_{j=1}^p \frac{1}{n^2} \text{var}\left(\sum_{t=1}^n (Y_{jt} - E(Y_j))\right) = \sum_{j=1}^p \frac{1}{n} \text{var}(Y_{j1} - E(Y_j)) \\ &= \frac{1}{n} \sum_{j=1}^p E\left([Y_{j1} - E(Y_j)]^2\right) = \frac{1}{n} \sum_{j=1}^p E(Y_{j1}^2) - [E(Y_j)]^2 \leq \frac{1}{n} \sum_{j=1}^p E(Y_{j1}^2) = O(n^{-1}p). \end{aligned}$$

Also insbesondere

$$\|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n\| = O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}).$$

Die Behauptung folgt dann analog zum Beweis von Satz 3.1.1 mit den Raten aus Satz 2.4.3. $\square$

Da $[\log n]^{\frac{1}{2}}$ im Vergleich zu p^2 nur langsam wächst, dominiert bei der Schätzung der Portfoliovarianz mithilfe des Faktormodells die zweite Rate und wir erhalten die gleiche Rate wie bei unabhängig identisch verteilten Faktoren.

3.1.2. Das globale varianzminimale Portfolio

Beim globalen varianzminimalen Portfolio betrachten wir das Minimierungsproblem (3.1.1) ohne Renditevorgabe

$$\min_{\boldsymbol{\xi}_{ng}} \boldsymbol{\xi}_{ng}^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_{ng}$$

unter der Bedingung

$$\boldsymbol{\xi}_{ng}^T \mathbf{1} = 1.$$

Mit dem Lagrangeansatz erhalten wir dann die optimalen Wertpapiergewichte in folgender Darstellung

$$L(\boldsymbol{\xi}_{ng}, \lambda) = \boldsymbol{\xi}_{ng}^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_{ng} + \lambda(1 - \boldsymbol{\xi}_{ng}^T \mathbf{1}).$$

Mit den partiellen Ableitungen ergibt sich folgendes Gleichungssystem:

$$\frac{\partial L}{\partial \boldsymbol{\xi}_{ng}^T} = \boldsymbol{\Sigma}_n \boldsymbol{\xi}_{ng} - \lambda \mathbf{1} = 0 \tag{3.1.11}$$

$$\frac{\partial L}{\partial \lambda} = 1 - \boldsymbol{\xi}_{ng}^T \mathbf{1} = 0. \quad (3.1.12)$$

Die erste Gleichung (3.1.11) umgestellt nach $\boldsymbol{\xi}_{ng}$ ergibt

$$\boldsymbol{\xi}_{ng} = \lambda \boldsymbol{\Sigma}_n^{-1} \mathbf{1}$$

und das Einsetzen in (3.1.12) liefert

$$\begin{aligned} 1 &= \lambda \mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} \mathbf{1} = \lambda \varphi_n \\ \Rightarrow \lambda &= \frac{1}{\varphi_n}. \end{aligned}$$

Die optimalen Wertpapiergewichte lassen sich demnach darstellen als

$$\boldsymbol{\xi}_{ng} = \frac{1}{\varphi_n} \boldsymbol{\Sigma}_n^{-1} \mathbf{1}$$

und die Varianz beim globalen varianzminimalen Portfolio berechnet sich durch

$$\boldsymbol{\xi}_{ng}^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_{ng} = \frac{1}{\varphi_n^2} \mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Sigma}_n \boldsymbol{\Sigma}_n^{-1} \mathbf{1} = \frac{1}{\varphi_n^2} \mathbf{1}^T \boldsymbol{\Sigma}_n^{-1} \mathbf{1} = \frac{1}{\varphi_n}. \quad (3.1.13)$$

Im Folgenden betrachten wir die Auswirkungen der Schätzung der Kovarianzmatrix (2.1.2) und (2.1.6) auf die Varianz des globalen varianzminimalen Portfolios.

Satz 3.1.3 *Sind alle φ_n von Null weg beschränkt, so gilt unter den Annahmen 1-8 für die Varianz des globalen varianzminimalen Portfolios unter dem Kovarianzmatrixschätzer des Faktormodells*

$$\hat{\boldsymbol{\xi}}_{ng}^T \hat{\boldsymbol{\Sigma}}_n \hat{\boldsymbol{\xi}}_{ng} - \boldsymbol{\xi}_{ng}^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_{ng} = O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}}) \quad (3.1.14)$$

und unter der Stichprobenkovarianzmatrix

$$\hat{\boldsymbol{\xi}}_{ng}^T \hat{\boldsymbol{\Sigma}}_{sam} \hat{\boldsymbol{\xi}}_{ng} - \boldsymbol{\xi}_{ng}^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_{ng} = O_P(n^{-\frac{1}{2}} p^3 K). \quad (3.1.15)$$

Beweis. Wir betrachten zunächst die erste Aussage (3.1.14):

Im Beweis von Satz 3.1.1 hatten wir bereits gezeigt, dass

$$|\hat{\varphi}_n - \varphi_n| \leq p \|\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}\|$$

gilt. Nach (2.3.5) gilt $\|\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}\| = O_P(n^{-\frac{1}{2}} p K^{\frac{3}{2}})$ und daher folgt hier

$$p \|\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}_n^{-1}\| = O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}}).$$

Da nach Voraussetzung alle φ_n von Null weg beschränkt sind, gilt mit (3.1.13):

$$|\hat{\xi}_{ng}^T \hat{\Sigma}_n \hat{\xi}_{ng} - \xi_{ng}^T \Sigma_n \xi_{ng}| = \left| \frac{1}{\hat{\varphi}_n} - \frac{1}{\varphi_n} \right| = O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}}).$$

Das Resultat (3.1.15) folgt analog: Wegen $|\hat{\varphi}_n - \varphi_n| \leq p \|\hat{\Sigma}_{sam}^{-1} - \Sigma_n^{-1}\|$ und der Aussage von (2.3.6), $\|\hat{\Sigma}_{sam}^{-1} - \Sigma_n^{-1}\| = O_P(n^{-\frac{1}{2}} p^2 K)$, folgt mit (3.1.13):

$$|\hat{\xi}_{ng}^T \hat{\Sigma}_{sam} \hat{\xi}_{ng} - \xi_{ng}^T \Sigma_n \xi_{ng}| = \left| \frac{1}{\hat{\varphi}_n} - \frac{1}{\varphi_n} \right| = O_P(n^{-\frac{1}{2}} p^3 K).$$

□

Auch für die Varianzschätzung beim globalen varianzminimalen Portfolio stellt der Faktormodellansatz eine Verbesserung gegenüber der Schätzung über die Stichprobenkovarianzmatrix dar, da auch hier die Konvergenzrate der Inversen aus Satz 2.3.4 zusätzlich zur Dimension p eingeht. Für den Fall der stationären Faktoren erhalten wir die Konvergenzraten direkt durch Einsetzen der veränderten Raten aus Satz 2.4.3:

Korollar 3.1.4 *Sind alle φ_n von Null weg beschränkt, so gilt unter den Annahmen 1 - 7 und 9 - 11*

$$\hat{\xi}_{ng}^T \hat{\Sigma}_n \hat{\xi}_{ng} - \xi_{ng}^T \Sigma_n \xi_{ng} = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} K) + O_P(n^{-\frac{1}{2}} p^2 K^{\frac{3}{2}})$$

und

$$\hat{\xi}_{ng}^T \hat{\Sigma}_{sam} \hat{\xi}_{ng} - \xi_{ng}^T \Sigma_n \xi_{ng} = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} p^3 K).$$

Beweis. Der Beweis folgt direkt durch Einsetzen der Konvergenzraten aus Satz 2.4.3 in den Beweis von Satz 3.1.3. □

Aus allen Resultaten in diesem Abschnitt über Portfolioselektion erkennen wir, dass die Schätzung mithilfe des Faktormodells bessere Ergebnisse liefert als die Schätzung durch die Stichprobenkovarianzmatrix. Dies ist auf die Nutzung der Inversen der Kovarianzmatrix zurückzuführen. Allerdings ist es wichtig zu beachten, dass bei beiden Schätzungen die Portfoliogröße und die Anzahl der Faktoren nur langsam wachsen darf, da die Schätzer sonst nicht konsistent sind und die Varianz der Portfolios unter Umständen nur sehr schlecht geschätzt wird.

3.2. Risikomanagement

Im Risikomanagement betrachten wir, im Gegensatz zur Portfolioselektion, die Varianz eines Portfolios mit fest vorgegebenen Wertpapiergewichten $\xi_n \in \mathbb{R}^p$. Die Gewichte summieren sich zu Eins auf, das heißt $\xi_n^T \mathbf{1} = 1$. Um diese Bedingung zu erfüllen, ohne extreme Shortpositionen eingehen zu müssen, nehmen wir an, dass die Komponenten des Gewichtsvektors gleichmäßig beschränkt sind, $\xi_n = O(1)\mathbf{1}$.

Der Einfluss der Schätzung der Kovarianzmatrix mithilfe von (2.1.2) und (2.1.6) auf das Risikomanagement kann wie folgt beschrieben werden:

Satz 3.2.1 *Unter den Annahmen 1 - 6 und 8 gelten für die Varianz des vorgegebenen Portfolios folgende Aussagen*

$$\boldsymbol{\xi}_n^T \hat{\boldsymbol{\Sigma}}_n \boldsymbol{\xi}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n = O_P(n^{-\frac{1}{2}} p^2 K) \quad (3.2.1)$$

und

$$\boldsymbol{\xi}_n^T \hat{\boldsymbol{\Sigma}}_{sam} \boldsymbol{\xi}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n = O_P(n^{-\frac{1}{2}} p^2 K). \quad (3.2.2)$$

Falls keine Shortpositionen im Portfolio $\boldsymbol{\xi}_n$ zugelassen sind ($\boldsymbol{\xi}_n \geq \mathbf{0}$, komponentenweise), gelten

$$\boldsymbol{\xi}_n^T \hat{\boldsymbol{\Sigma}}_n \boldsymbol{\xi}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n = O_P(n^{-\frac{1}{2}} p K) \quad (3.2.3)$$

und

$$\boldsymbol{\xi}_n^T \hat{\boldsymbol{\Sigma}}_{sam} \boldsymbol{\xi}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n = O_P(n^{-\frac{1}{2}} p K). \quad (3.2.4)$$

Beweis. Wir beweisen zunächst den Fall mit zugelassenen Shortpositionen (3.2.1). Wegen $\boldsymbol{\xi}_n = O(1)\mathbf{1}$ gilt mit (B.2.3)

$$\begin{aligned} |\boldsymbol{\xi}_n^T \hat{\boldsymbol{\Sigma}}_n \boldsymbol{\xi}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n| &= |\boldsymbol{\xi}_n^T (\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n) \boldsymbol{\xi}_n| = |\text{tr}\{(\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n) \boldsymbol{\xi}_n \boldsymbol{\xi}_n^T\}| \\ &\leq \|\boldsymbol{\xi}_n^T\| \|\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n\| \|\boldsymbol{\xi}_n\| = O(1) \|\mathbf{1}^T\| \|\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n\| O(1) \|\mathbf{1}\| \\ &= O(p^{\frac{1}{2}}) \|\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n\| O(p^{\frac{1}{2}}) = O(p) \|\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n\|. \end{aligned}$$

Mit der Aussage von Satz 2.3.1 folgt dann

$$|\boldsymbol{\xi}_n^T \hat{\boldsymbol{\Sigma}}_n \boldsymbol{\xi}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n| = O_P(n^{-\frac{1}{2}} p^2 K).$$

Analog gilt für die Stichprobenkovarianzmatrix nach dem Beweis von (2.3.2)

$$|\boldsymbol{\xi}_n^T \hat{\boldsymbol{\Sigma}}_{sam} \boldsymbol{\xi}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n| \leq O(p) \|\hat{\boldsymbol{\Sigma}}_{sam} - \boldsymbol{\Sigma}_n\|,$$

also

$$\boldsymbol{\xi}_n^T \hat{\boldsymbol{\Sigma}}_{sam} \boldsymbol{\xi}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n = O_P(n^{-\frac{1}{2}} p^2 K).$$

Für den Fall ohne Shortpositionen liegen alle Wertpapiergewichte im Intervall $[0, 1]$, also gilt insbesondere die Ungleichung

$$\|\boldsymbol{\xi}_n\| = \sqrt{\xi_1^2 + \cdots + \xi_p^2} \leq \sqrt{\xi_1 + \cdots + \xi_p} = \sqrt{1} = 1$$

und daher gilt für (3.2.3) mithilfe von (B.2.3)

$$|\boldsymbol{\xi}_n^T \hat{\boldsymbol{\Sigma}}_n \boldsymbol{\xi}_n - \boldsymbol{\xi}_n^T \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n| = |\boldsymbol{\xi}_n^T (\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n) \boldsymbol{\xi}_n| = |\text{tr}\{(\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n) \boldsymbol{\xi}_n \boldsymbol{\xi}_n^T\}|$$

$$\leq \|\xi_n^T\| \|\hat{\Sigma}_n - \Sigma_n\| \|\xi_n\| = \|\hat{\Sigma}_n - \Sigma_n\|.$$

Es folgt daher nach Satz 2.3.1

$$\xi_n^T \hat{\Sigma}_n \xi_n - \xi_n^T \Sigma_n \xi_n = O_P(n^{-\frac{1}{2}} p K).$$

Analog gilt für die Stichprobenkovarianzmatrix

$$|\xi_n^T \hat{\Sigma}_{sam} \xi_n - \xi_n^T \Sigma_n \xi_n| \leq \|\hat{\Sigma}_{sam} - \Sigma_n\|,$$

also

$$\xi_n^T \hat{\Sigma}_{sam} \xi_n - \xi_n^T \Sigma_n \xi_n = O_P(n^{-\frac{1}{2}} p K).$$

□

Da die Aussage mithilfe des Resultats aus Satz 2.3.1 bewiesen wird, in dem wir die gleichen Konvergenzraten für beide Schätzer gezeigt haben, erhalten wir auch für die Portfoliovarianz bei gegebenen Portfoliogewichten unter beiden Schätzmethoden die gleichen Konvergenzraten. Beim Übergang zu stationären Faktoren ändert sich die Rate nur durch den zusätzlichen Faktor $[\log n]^{\frac{1}{2}}$.

Korollar 3.2.2 *Unter den Annahmen 1 - 6 und 9 - 11 gelten für die Varianz des vorgegebenen Portfolios folgende Aussagen*

$$\xi_n^T \hat{\Sigma}_n \xi_n - \xi_n^T \Sigma_n \xi_n = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} p^2 K)$$

und

$$\xi_n^T \hat{\Sigma}_{sam} \xi_n - \xi_n^T \Sigma_n \xi_n = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} p^2 K).$$

Falls keine Shortpositionen im Portfolio ξ_n zugelassen sind ($\xi_n \geq \mathbf{0}$, komponentenweise), gelten

$$\xi_n^T \hat{\Sigma}_n \xi_n - \xi_n^T \Sigma_n \xi_n = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} p K)$$

und

$$\xi_n^T \hat{\Sigma}_{sam} \xi_n - \xi_n^T \Sigma_n \xi_n = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} p K).$$

Beweis. Der Beweis folgt mit den Resultaten aus Satz 2.4.1 analog zum Beweis von Satz 3.2.1. □

Wir erkennen aus den Resultaten dieses Abschnitts, dass die Varianzschätzung im Risikomanagement nicht von der Anpassung eines Faktormodells profitiert, da nicht die Inverse der Kovarianzmatrix sondern die Kovarianz selbst für die Berechnung genutzt wird. Insbesondere im Fall mit erlaubten Leerverkäufen muss in hohem Maße darauf geachtet werden, dass die Dimension des Portfolios und die Anzahl der Faktoren nicht zu schnell mit der Stichprobengröße wachsen, da hier ein zusätzlicher Faktor p in die Rate eingeht.

4. Simulation

In diesem Kapitel führen wir eine Simulation für die Schätzung der Kovarianzmatrix in Faktormodellen durch. In einer ersten Kalibrierungsphase wenden wir das bekannte Dreifaktormodell aus Fama und French (1993) auf einen realen Datensatz an, um die Parameter für die Verteilungen zu ermitteln, aus denen wir dann in der Simulationsphase das Faktormodell simulieren. Basierend auf dem simulierten Faktormodell führen wir die in Abschnitt 2.1 erläuterten Schätzungen durch und betrachten deren Güte in den Normen, für die wir in den Abschnitten 2.3 und 2.4 Konvergenzraten hergeleitet haben. Zusätzlich simulieren wir die Ergebnisse aus den Anwendungen in Kapitel 3.

4.1. Kalibrierung

In Fama und French (1992) werden verschiedene unternehmensspezifische Größen auf ihren Einfluss auf Wertpapierrenditen untersucht, um auf deren Basis Faktoren für ein fundamentales Faktormodell zu konstruieren. Das Resultat ist, dass durch die Unternehmensgröße und das Verhältnis von Bilanzkurs zu Börsenkurs das Verhalten der Renditen gut erklärt werden kann. Darauf basierend führen Fama und French (1993) das Dreifaktormodell für Aktien ein und erläutern die Konstruktion der Faktoren. Der erste Faktor ist die Überschussrendite der Aktien im Vergleich zur sicheren Anlage. Die anderen beiden Faktoren werden jeweils aus den Werten für die Unternehmensgröße und dem Verhältnis von Bilanzkurs zu Börsenkurs bestimmt. Wir nutzen dieses Modell, um die Verteilungsparameter unserer Simulation zu bestimmen. Die Daten sind verfügbar unter http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (15.11.2012). Für die Kalibrierung passen wir ein lineares Modell an, in dem die Renditen von 30 Industriepportfolios die abhängigen Variablen und die Fama/French Faktoren die unabhängigen Variablen sind

$$\begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix} = b_{i1} \begin{pmatrix} f_{11} \\ \vdots \\ f_{1n} \end{pmatrix} + b_{i2} \begin{pmatrix} f_{21} \\ \vdots \\ f_{2n} \end{pmatrix} + b_{i3} \begin{pmatrix} f_{31} \\ \vdots \\ f_{3n} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in} \end{pmatrix}, \quad i = 1, \dots, 30.$$

Wir nutzen dazu verschiedene Stichprobengrößen, da wir auch die Simulation für verschiedene Werte von n durchführen. Die Datenbank umfasst aktuell Tagesdaten für den Zeitraum zwischen dem 01.07.1926 und dem 31.07.2012 und wir nutzen die n aktuellsten Datenpunkte für $n \in \{756, 1250, 5000, 10000\}$. Dieses Verfahren liefert uns die in Tabelle 4.1 aufgelisteten Zeiträume.

n	Zeitraum
756	03.08.2009-31.07.2012
1250	16.08.2007-31.07.2012
5000	28.09.1992-31.07.2012
10000	13.12.1972-31.07.2012

Tabelle 4.1.: Zeiträume für verschiedene Stichprobengrößen

Aus dem linearen Modell erhalten wir 30 Vektoren der Regressionskoeffizienten, von denen wir den Mittelwert $\hat{\boldsymbol{\mu}}_b$ und die Stichprobenkovarianzmatrix $\text{cov}(\mathbf{b})$ bilden. Für die Residuen des Modells bestimmen wir die Standardabweichungen $\hat{\sigma}_1, \dots, \hat{\sigma}_{30}$. Für den Fall der unabhängigen Faktoren nutzen wir die Daten für die Faktoren aus der Datenbasis und schätzen deren Mittelwert $\hat{\boldsymbol{\mu}}_f$ und die Stichprobenkovarianzmatrix $\text{cov}(\mathbf{f})$. Für den Fall der stationären Faktoren nutzen wir den Ansatz aus Fan et al. (2011) und passen ein Vektorautoregressives (VAR-)Modell der Ordnung 1 an die beobachteten Daten an:

$$\mathbf{f}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \quad t \in \{1, \dots, n\}$$

mit

$$\boldsymbol{\eta}_t \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{\Sigma}_\eta).$$

Daraus erhalten wir Schätzungen für die Modellparameter $\hat{\boldsymbol{\mu}} \in \mathbb{R}^3$, $\hat{\boldsymbol{\Phi}} \in \mathbb{R}^{3 \times 3}$ und $\hat{\boldsymbol{\Sigma}}_\eta \in \mathbb{R}^{3 \times 3}$. Diese Schätzungen nutzen wir nun, um das Faktormodell zu simulieren.

4.2. Simulation des Faktormodells

Die Simulation für die unabhängigen Faktoren verläuft nach folgenden Schritten:

1. Wir erzeugen n Werte für die drei Faktoren, $(f_1, f_2, f_3)^T \sim \mathcal{N}_3(\hat{\boldsymbol{\mu}}_f, \text{cov}(\mathbf{f}))$ und bilden daraus die Matrix $\mathbf{X} \in \mathbb{R}^{3 \times n}$.
2. Wir setzen die Portfoliodimension auf den Startwert $p := 16$.
3. Wir generieren p Werte für die Faktorladungen, $(b_{i1}, b_{i2}, b_{i3}) \sim \mathcal{N}_3(\hat{\boldsymbol{\mu}}_b, \text{cov}(\mathbf{b}))$ und bilden daraus die Matrix $\mathbf{B}_n \in \mathbb{R}^{p \times 3}$.
4. Die Stichprobe für die Standardabweichungen der Fehler soll aus einer Gamma-verteilung generiert werden, für die wir vorher die Parameter bestimmen müssen. Um die Annahme 4 zu erfüllen, ermitteln wir zunächst einen Threshold $\tau := \min_{1 \leq i \leq 30} \hat{\sigma}_i$.

Um die Parameter passend zu den Daten zu wählen, nutzen wir das folgende Verfahren aus Fan et al. (2007): Wir approximieren den Erwartungswert und das zweite Moment, bedingt auf Werte über dem Threshold durch

$$E(\sigma | \sigma > \tau) = \frac{\alpha\beta - \frac{\tau}{2}q}{1 - q},$$

$$E(\sigma^2 | \sigma > \tau) = \frac{\alpha\beta^2 + \alpha^2\beta^2 - \frac{\tau^2}{2}q}{1 - q},$$

$$q := P(\sigma < \tau).$$

Die Startwerte für die Parameter berechnen wir mithilfe der Momente der Gamma-Verteilung. Für $X \sim \Gamma(\alpha, \beta)$ gilt $E(X) = \alpha\beta$ und $\text{var}(X) = \alpha\beta^2$. Wir wählen daher die Startwerte wie folgt:

$$\alpha_0\beta_0 = \text{mean}(\hat{\sigma}_1, \dots, \hat{\sigma}_{30})$$

$$\alpha_0^{\frac{1}{2}}\beta_0 = \text{sd}(\hat{\sigma}_1, \dots, \hat{\sigma}_{30}).$$

Mithilfe dieser Startwerte bestimmen wir die Parameter rekursiv, sodass gilt

$$E(\sigma | \sigma > \tau) = \text{mean}(\hat{\sigma}_1, \dots, \hat{\sigma}_{30})$$

und

$$E(\sigma^2 | \sigma > \tau) = \text{var}(\hat{\sigma}_1, \dots, \hat{\sigma}_{30}) + [\text{mean}(\hat{\sigma}_1, \dots, \hat{\sigma}_{30})]^2.$$

Die Lösung für diese beiden Gleichungen lautet

$$\alpha = -\frac{(-\tau \cdot q - 2m + 2mq)^2}{\tau^2 q^2 + 4\tau \cdot qm - 4\tau \cdot q^2 m + 4m^2 - 8m^2 q + 4m^2 q^2 - 2\tau^2 q - 4n + 4nq}$$

$$\beta = \frac{1}{2} \frac{\tau^2 q^2 + 4\tau \cdot qm - 4\tau \cdot q^2 m + 4m^2 - 8m^2 q + 4m^2 q^2 - 2\tau^2 q - 4n + 4nq}{-\tau \cdot p - 2m + 2mq},$$

mit

$$m := \text{mean}(\hat{\sigma}_1, \dots, \hat{\sigma}_{30}),$$

$$n := \text{var}(\hat{\sigma}_1, \dots, \hat{\sigma}_{30}) + (\text{mean}(\hat{\sigma}_1, \dots, \hat{\sigma}_{30}))^2.$$

Nach der Berechnung dieser Parameter simulieren wir für $i = 1, \dots, p$ Standardabweichungen für die Fehler: $\sigma_i \sim \Gamma(\alpha, \beta)$ und bilden daraus die Kovarianzmatrix der Fehler $\mathbf{\Sigma}_0 = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$.

5. Wie generieren n Werte für die Fehler, $(\varepsilon_1, \dots, \varepsilon_n) \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_0)$ und bilden daraus die Matrix $\mathbf{E} \in \mathbb{R}^{p \times n}$.
6. Wir berechnen die Werte für die Wertpapierüberschussrenditen $\mathbf{Y} = \mathbf{B}_n \mathbf{X} + \mathbf{E}$.
7. Wir berechnen $\mathbf{\Sigma}_n$ und die Schätzer $\hat{\mathbf{\Sigma}}_n, \hat{\mathbf{\Sigma}}_{sam}$ nach Abschnitt 2.1.
8. Solange $p < 1000$ ist, erhöhen wir p um 20 und führe folgende Schritte durch

- a) wir generieren 20 Werte für die Faktorladungen, $(b'_{i1}, b'_{i2}, b'_{i3}) \sim \mathcal{N}_3(\hat{\boldsymbol{\mu}}, \text{côv}(\mathbf{b}))$ und bilden die Matrix $\mathbf{B}'_n \in \mathbb{R}^{20 \times 3}$. Diese fügen wir mit der Matrix aus Schritt 3 zusammen: $\mathbf{B}_n = (\mathbf{B}_n, \mathbf{B}'_n)^T$.
- b) Wir generieren für $i = 1, \dots, 20$ Werte für Standardabweichungen der Fehler, $\sigma'_i \sim \Gamma(\alpha, \beta)$ und bilde die Matrix $\boldsymbol{\Sigma}'_0 = \text{diag}((\sigma'_1)^2, \dots, (\sigma'_{20})^2)$. Diese fügen wir mit der Matrix aus Schritt 4 zusammen: $\boldsymbol{\Sigma}_0 = \text{diag}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}'_0)$.
- c) Wir generieren n Werte für die Fehler, $(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n) \sim \mathcal{N}_{20}(\mathbf{0}, \boldsymbol{\Sigma}'_0)$ und bilden daraus die Matrix $\mathbf{E}' \in \mathbb{R}^{20 \times n}$. Diese fügen wir mit der Matrix aus Schritt 5 zusammen: $\mathbf{E} = (\mathbf{E}, \mathbf{E}')^T$.
- d) Wir führen die Schritte 6 und 7 durch.

Für die Simulation des Faktormodells mit stationären Faktoren muss nur Schritt 1 ersetzt werden. Anstelle der Simulation aus einer Normalverteilung simulieren wir hier aus dem VAR-Modell: Wir generieren n Werte für die Faktoren: $(f_1, f_2, f_3)^T \sim AR(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Sigma}}_\eta)$ und bilden daraus die Matrix $\mathbf{X} \in \mathbb{R}^{3 \times n}$.

4.3. Auswertung

Insgesamt führen wir die Simulation für jeden Wert von n 500 mal durch. Aus den simulierten Kovarianzmatrizen und den Schätzungen berechnen wir zunächst die Schätzfehler und betrachten diese in der Frobeniusnorm, der Norm $\|\cdot\|_{\mathbf{\Sigma}_n}$ und zusätzlich die inverse Matrix in der Frobeniusnorm, um die Ergebnisse aus den Sätzen 2.3.1 - 2.4.3 zu visualisieren. Dazu bilden wir den Mittelwert und die Standardabweichung über die 500 Simulationen und stellen diese in Abhängigkeit von der Dimension p dar.

Um einen direkten Vergleich zwischen den unterschiedlichen Annahmen an die Faktoren zu erhalten, stellen wir jeweils die Schätzfehler der Schätzungen aus dem Faktormodell und der Stichprobenkovarianzmatrix für unabhängig identisch verteilte (rot) und stationäre Faktoren (grün) in einer Abbildung dar.

Weiterhin nutzen wir die Simulation, um die Auswirkungen der Schätzung auf die in Kapitel 3 beschriebenen Anwendung aus der Portfolioselektion und dem Risikomanagement zu visualisieren. Dazu berechnen wir jeweils die Varianz des Portfolios mithilfe der Wertpapiergewichte des optimalen Portfolios nach Markowitz, des globalen varianz-minimalen Portfolios und des gleichgewichteten Portfolios und betrachten die mittleren quadratischen Abweichungen über die 500 Simulationen. Für den allgemeinen Markowitzfall unterstellen wir eine erwartete Rendite von 10%.

Die Resultate, deren Berechnungen die Inverse der Kovarianzmatrix beinhalten, bilden wir nur für Dimensionen kleiner als die Stichprobengröße ab, da der Schätzfehler der Inversen der Stichprobenkovarianzmatrix andernfalls extrem hohe Werte liefert.

Aus den Abbildungen 4.1(a) - 4.4(a) können wir erkennen, dass die Schätzfehler beider Schätzer sich in der Frobeniusnorm für wachsende Dimension p nahezu gleich verhalten. Auch der Unterschied zwischen stationären und unabhängig identisch verteilten Faktoren, der in der Theorie durch den Faktor $\log n$ beschrieben wird, ist vernachlässigbar.

Die Abbildungen 4.1(b) - 4.4(b) bestätigen das Resultat aus Satz 2.3.2 bzw. Satz 2.4.2, dass unter der Norm $\|\cdot\|_{\Sigma_n}$ die Schätzung aus dem Faktormodell besser ist als die Stichprobenkovarianzmatrix. Der Unterschied zwischen stationären und unabhängig identisch verteilten Faktoren ist in den Simulationsergebnissen nicht erkennbar. Wir sehen, dass für jedes feste n die Schätzfehler über wachsende p fast auf konstantem Niveau bleiben. Diese Beobachtung ist konsistent zu Bemerkung 2.3.3, da die dominierende Rate in dem betrachteten Fall unabhängig von p ist.

In den Abbildungen 4.1(c) - 4.4(c) ist sehr deutlich zu erkennen, dass die Schätzung der Inversen der Kovarianzmatrix stark von der Nutzung des Faktormodells profitiert. Diese Beobachtung deckt sich mit dem Resultat aus Satz 2.3.4 bzw. Satz 2.4.3. Auch hier ist der Unterschied zwischen stationären und unabhängig identisch verteilten Faktoren in der Grafik nicht erkennbar.

Aus den Abbildungen 4.1(d) - 4.4(d) erkennen wir, dass sich die Ergebnisse für die verschiedenen Anwendungen stark unterscheiden.

Bei der Varianz des optimalen Portfolios nach Markowitz ist ein deutlicher Unterschied zwischen den Schätzfehlern der Stichprobenkovarianzmatrix und der Schätzung aus dem Faktormodell erkennbar. Bei größeren Werten von n ist sichtbar, dass beim Schätzer aus dem Faktormodell die Schätzfehler im Fall der unabhängig identisch verteilten Faktoren kleiner sind, als bei den stationären Faktoren. Insbesondere bemerken wir, dass die Höhe der Schätzfehler sehr datensensibel ist. In Tabelle 4.1 ist erkennbar, dass für die Stichprobengröße 1250 die Daten der Subprimekrise ein starkes Gewicht haben, während der Zeitraum zur Stichprobengröße 756 die besonders turbulenten Jahre 2007 und 2008 nicht umfasst.

Bei den anderen Anwendungsfällen, dem globalen varianzminimalen Portfolio und dem gleichgewichteten Portfolio im Risikomanagement sind zwischen den unterschiedlichen Schätzverfahren und auch zwischen den verschiedenen Annahmen an die Faktoren keine deutlichen Unterschiede bei den Schätzfehlern erkennbar.

Bei der Betrachtung der Ergebnisse ist zu beachten, dass die Anzahl der Faktoren K immer den Wert 3 annimmt und n in den Abbildungen 4.1- 4.4 jeweils fest gewählt ist. Wir beobachten, dass durch die Vergrößerung der Stichprobe die Schätzungen exakter werden.

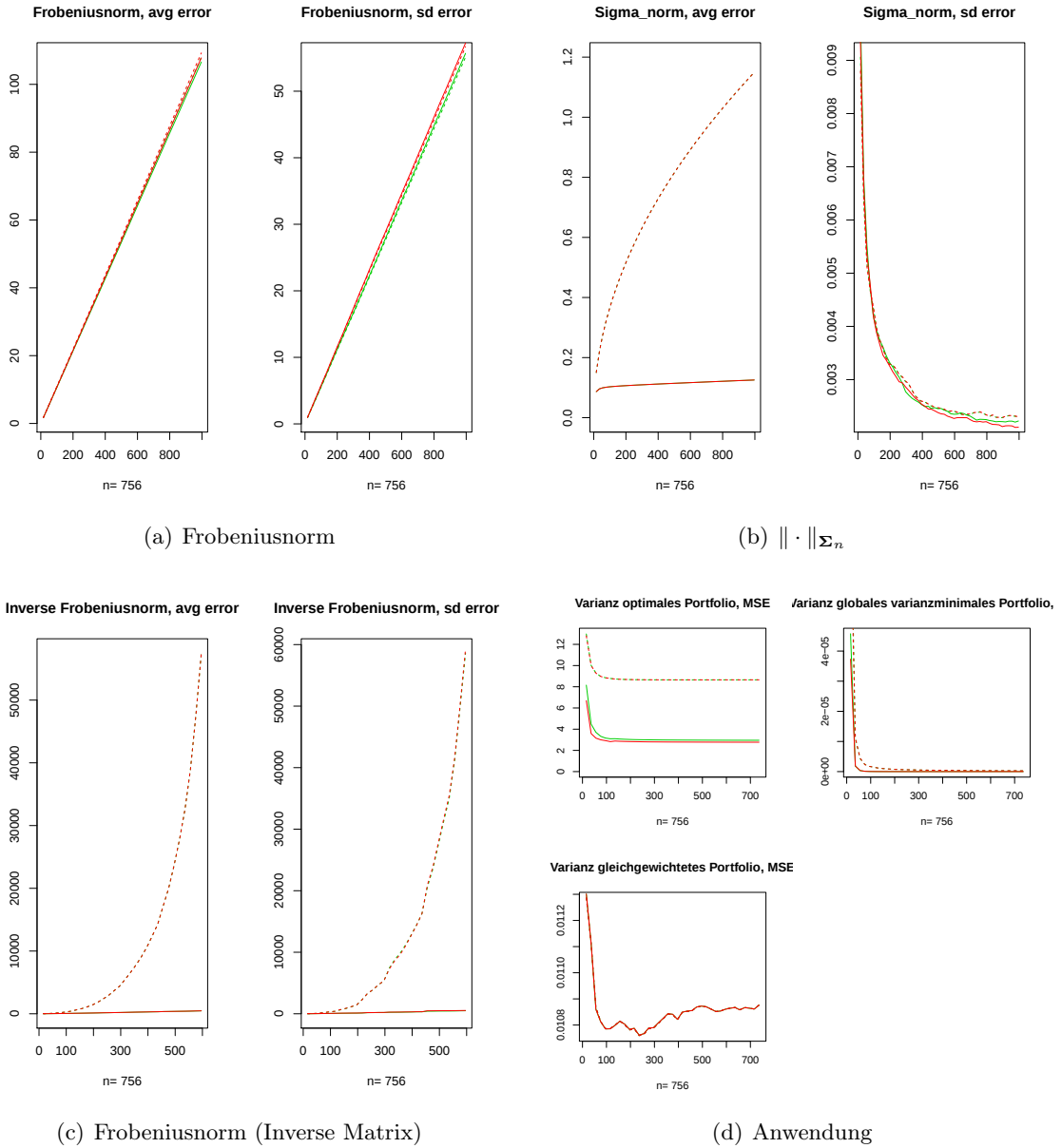


Abbildung 4.1.: Stichprobengröße 756. 4.1(a) - 4.1(c): Mittelwerte und Standardabweichungen der Schätzfehler über 500 Simulationen, 4.1(d): Mittlere quadratische Abweichung der Schätzfehler über 500 Simulationen. Die Ergebnisse unter der Annahme unabhängig identisch verteilter Faktoren sind jeweils in rot eingezeichnet, die Ergebnisse bei stationären Faktoren in grün. Dabei sind die durchgezogenen Linien die Schätzungen aus dem Faktormodell und die gestrichelten Linien die Ergebnisse der Stichprobenkovarianzmatrix.

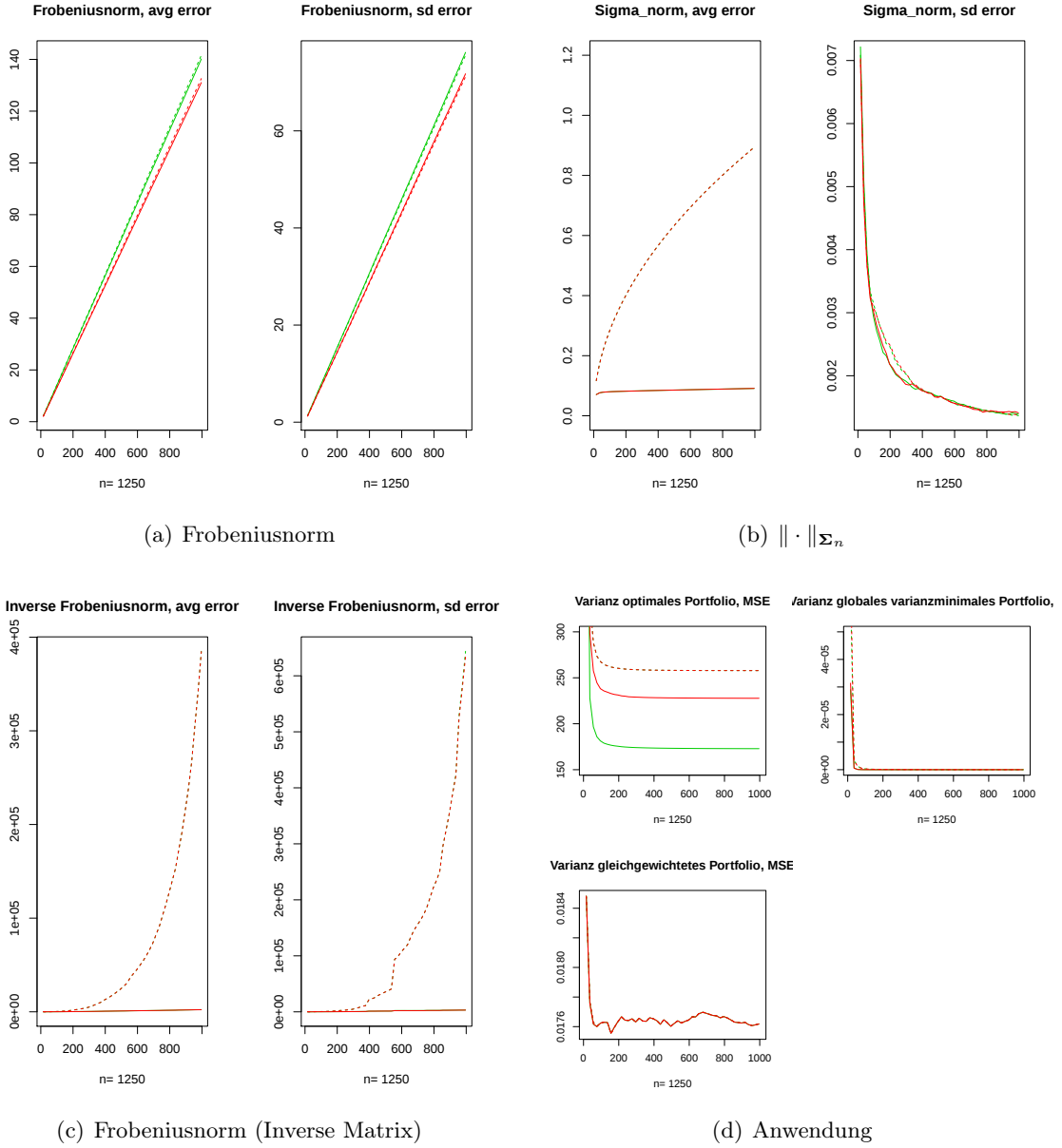


Abbildung 4.2.: Stichprobengröße 1250. 4.2(a) - 4.2(c): Mittelwerte und Standardabweichungen der Schätzfehler über 500 Simulationen, 4.2(d): Mittlere quadratische Abweichung der Schätzfehler über 500 Simulationen. Die Ergebnisse unter der Annahme unabhängig identisch verteilter Faktoren sind jeweils in rot eingezeichnet, die Ergebnisse bei stationären Faktoren in grün. Dabei sind die durchgezogenen Linien die Schätzungen aus dem Faktormodell und die gestrichelten Linien die Ergebnisse der Stichprobenkovarianzmatrix.

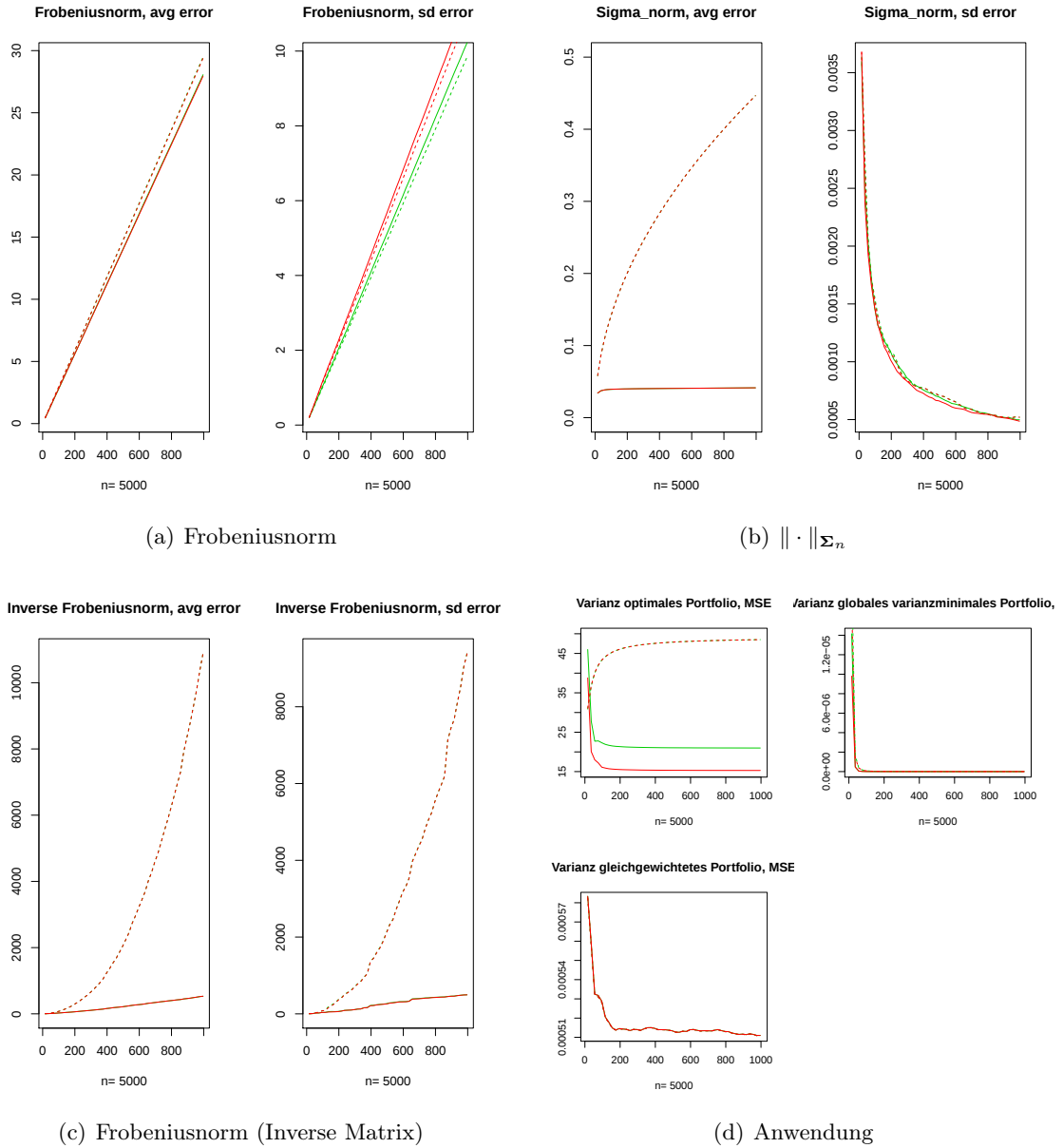


Abbildung 4.3.: Stichprobengröße 5000. 4.3(a) - 4.3(c): Mittelwerte und Standardabweichungen der Schätzfehler über 500 Simulationen, 4.3(d): Mittlere quadratische Abweichung der Schätzfehler über 500 Simulationen. Die Ergebnisse unter der Annahme unabhängig identisch verteilter Faktoren sind jeweils in rot eingezeichnet, die Ergebnisse bei stationären Faktoren in grün. Dabei sind die durchgezogenen Linien die Schätzungen aus dem Faktormodell und die gestrichelten Linien die Ergebnisse der Stichprobenkovarianzmatrix.

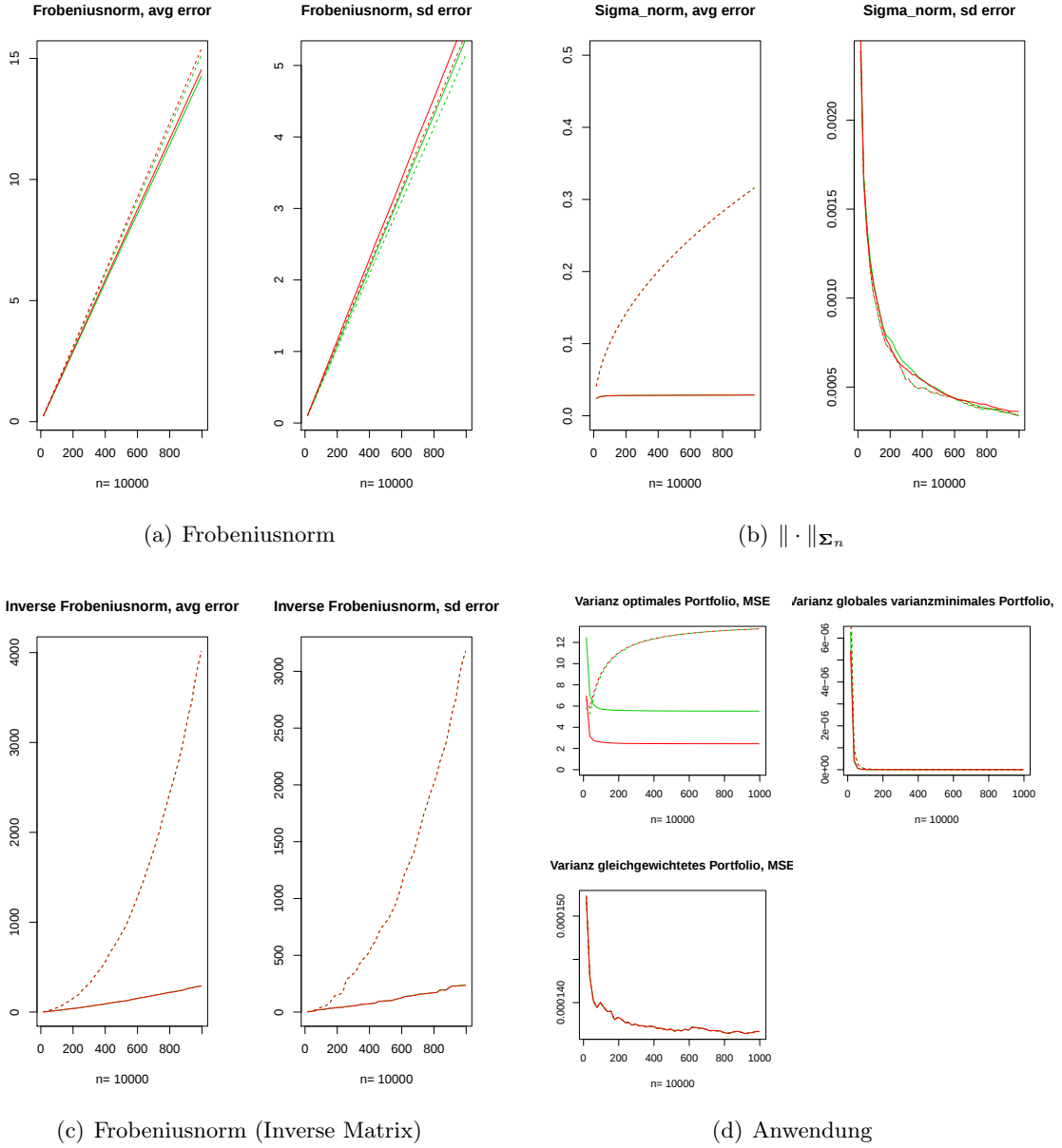


Abbildung 4.4.: Stichprobengröße 10000. 4.4(a) - 4.4(c): Mittelwerte und Standardabweichungen der Schätzfehler über 500 Simulationen, 4.4(d): Mittlere quadratische Abweichung der Schätzfehler über 500 Simulationen. Die Ergebnisse unter der Annahme unabhängig identisch verteilter Faktoren sind jeweils in rot eingezeichnet, die Ergebnisse bei stationären Faktoren in grün. Dabei sind die durchgezogenen Linien die Schätzungen aus dem Faktormodell und die gestrichelten Linien die Ergebnisse der Stichprobenkovarianzmatrix.

5. Schätzung der Autokovarianzmatrix im Faktormodell

Bislang haben wir uns mit der Schätzung der Kovarianzmatrix mithilfe von Faktormodellen beschäftigt. Häufig interessiert man sich auch für den Zusammenhang der Beobachtungen zu verschiedenen Zeitpunkten, der für einen Zeitabstand h durch die Kovarianz $\text{cov}(\mathbf{y}_{t+h}, \mathbf{y}_t)$ beschrieben wird. Typischerweise ist die Autokorrelation bei Finanzzeitreihen nicht signifikant ungleich Null, unter den 30 Industriepportfolios, die wir für die Simulation in Kapitel 4 genutzt haben zeigt sich aber, dass die Autokorrelation für einige Sektoren (z.B. Unterhaltungsindustrie, Gesundheits- und Arzneimittelsektor, Dienstleistungssektor und Geschäftsausstattungsindustrie) für kleine Zeitabstände signifikant ist. Für einen Vektor $\mathbf{y} \in \mathbb{R}^p$ wird die $p \times p$ Autokovarianzmatrix berechnet durch

$$\Sigma_n(h) := \text{cov}(\mathbf{y}_{t+h}, \mathbf{y}_t) = E((\mathbf{y}_{t+h} - E(\mathbf{y}_{t+h}))(\mathbf{y}_t - E(\mathbf{y}_t))^T).$$

Diese beinhaltet die Zusammenhänge zwischen allen Komponenten des Vektors über den Zeitraum h .

Für stationäre Zeitreihen hängt die Autokovarianz nicht von dem betrachteten Zeitpunkt $t \in \{1, \dots, n\}$, sondern nur von dem Zeitabstand h ab.

Die empirische Schätzung der Autokovarianzmatrix wird durch

$$\begin{aligned} \hat{\Sigma}_{sam}(h) &= \frac{1}{n} \sum_{t=1}^{n-h} (\mathbf{y}_{t+h} - \bar{\mathbf{y}}_n)(\mathbf{y}_t - \bar{\mathbf{y}}_n)^T, \\ \bar{\mathbf{y}}_n &= \frac{1}{n} \sum_{t=1}^n \mathbf{y}_t \end{aligned} \tag{5.0.1}$$

berechnet (vgl. Brockwell und Davis, 2009, S. 407ff.). Da es nur sinnvoll ist, die Autokovarianzmatrix zu schätzen, wenn $\frac{h}{n} \rightarrow 0$ für $n \rightarrow \infty$, treffen wir folgende Annahme:

Annahme 12 Für den Zeitabstand h gilt $h = o(n)$.

Wir wollen nun untersuchen, wie Faktormodelle für die Schätzung der Autokovarianzmatrix genutzt werden können. Dazu wenden wir das Faktormodell mit stationären Faktoren aus Abschnitt 2.4 an, die zentralen Annahmen dafür wiederholen wir an dieser Stelle noch einmal.

Annahme 9 Die Zeitreihe der Faktoren $\{\mathbf{f}_t\}_{t \geq 1}$ ist stationär und ergodisch. Die Fehler $(\epsilon_1, \dots, \epsilon_n)$ sind unabhängig identisch verteilt. Die Zeitreihen der Faktoren $\{\mathbf{f}_t\}_{t \geq 1}$ und

der Fehler $\{\varepsilon_t\}_{t \geq 1}$ sind unabhängig.

Diese Annahme nutzen wir in Lemma 5.0.1, um die Darstellung der Autokovarianzmatrix im Faktormodell herzuleiten.

Annahme 10 Der α -Mischungskoeffizient zum Abstand t klingt exponentiell ab: Es existieren Konstanten $r_1, C > 0$, sodass für alle $t \in \mathbb{Z}^+$ gilt

$$\alpha(t) \leq \exp(-Ct^{r_1}).$$

Annahme 11 Es existieren Konstanten $r_2, b_1 > 0$ mit $3r_2^{-1} + r_1^{-1} > 1$, sodass für alle $s > 0$ und alle $i \in \{1, \dots, K\}$ gilt

$$P(|f_{it}| > s) \leq \exp(-(s/b_1)^{r_2}).$$

Diese Annahmen benötigen wir, um in Lemma 5.0.2 die Konvergenzrate der empirischen Autokovarianz herzuleiten. Hierfür wenden wir wegen der Abhängigkeit der Faktoren das Resultat aus Merlevède et al. (2009) an.

Lemma 5.0.1 *Unter den Annahmen 1 und 9 besitzt die Autokovarianzmatrix im Faktormodell folgende Form*

$$\Sigma_n(h) = \mathbf{B}_n \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{B}_n^T.$$

Beweis. Mit der Darstellung des Faktormodells (1.0.1) erhalten wir mithilfe von Annahme 1

$$\begin{aligned} \Sigma_n(h) &= \text{cov}(\mathbf{y}_{t+h}, \mathbf{y}_t) = E((\mathbf{y}_{t+h} - E(\mathbf{y}_{t+h}))(\mathbf{y}_t - E(\mathbf{y}_t))^T) \\ &= E((\mathbf{B}_n \mathbf{f}_{t+h} + \varepsilon_{t+h} - \mathbf{B}_n E(\mathbf{f}_{t+h}))(\mathbf{B}_n \mathbf{f}_t + \varepsilon_t - \mathbf{B}_n E(\mathbf{f}_t))^T) \\ &= E((\mathbf{B}_n(\mathbf{f}_{t+h} - E(\mathbf{f}_{t+h})) + \varepsilon_{t+h})(\mathbf{B}_n(\mathbf{f}_t - E(\mathbf{f}_t)) + \varepsilon_t)^T) \\ &= E(\mathbf{B}_n(\mathbf{f}_{t+h} - E(\mathbf{f}_{t+h}))(\mathbf{f}_t - E(\mathbf{f}_t))^T \mathbf{B}_n^T + \mathbf{B}_n(\mathbf{f}_{t+h} - E(\mathbf{f}_{t+h}))\varepsilon_t^T \\ &\quad + \varepsilon_{t+h}(\mathbf{f}_t - E(\mathbf{f}_t))^T \mathbf{B}_n^T + \varepsilon_{t+h}\varepsilon_t^T) \\ &= \mathbf{B}_n E((\mathbf{f}_{t+h} - E(\mathbf{f}_{t+h}))(\mathbf{f}_t - E(\mathbf{f}_t))^T) \mathbf{B}_n^T + \mathbf{B}_n E((\mathbf{f}_{t+h} - E(\mathbf{f}_{t+h}))\varepsilon_t^T) \\ &\quad + E(\varepsilon_{t+h}(\mathbf{f}_t - E(\mathbf{f}_t))^T) \mathbf{B}_n^T + E(\varepsilon_{t+h}\varepsilon_t^T). \end{aligned}$$

Nach Annahme 9 sind die Fehler und die Faktoren unabhängig, sowie die Fehler über die Zeit unabhängig. Daher folgt mit Annahme 1

$$\Sigma_n(h) = \mathbf{B}_n E((\mathbf{f}_{t+h} - E(\mathbf{f}_{t+h}))(\mathbf{f}_t - E(\mathbf{f}_t))^T) \mathbf{B}_n^T = \mathbf{B}_n \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{B}_n^T.$$

□

Um die Autokovarianzmatrix mithilfe von n Beobachtungen zu schätzen, setzen wir in die Darstellung von Lemma 5.0.1 analog zur Kovarianzmatrixschätzung in Abschnitt 2.1 die Kleinste Quadrate Schätzung (2.1.4) für die Matrix der Faktorladungen und die empirische Autokovarianzmatrix der Faktoren analog zu (5.0.1) ein:

$$\hat{\Sigma}_n(h) = \hat{\mathbf{B}}_n \hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) \hat{\mathbf{B}}_n^T. \quad (5.0.2)$$

Im Vergleich zur Schätzung der Kovarianzmatrix (2.1.2) wird deutlich, dass hier weniger Parameter zu schätzen sind, da die Fehler über die Zeit unabhängig sind.

Unser Ziel ist die Analyse des asymptotischen Verhaltens des aus (5.0.2) resultierenden Schätzfehlers. Diesen betrachten wir unter Ausnutzung von (2.2.9) in folgender Form

$$\begin{aligned}\hat{\Sigma}_n(h) - \Sigma_n(h) &= \hat{\mathbf{B}}_n \hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) \hat{\mathbf{B}}_n^T - \mathbf{B}_n \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{B}_n^T \\ &= (\mathbf{B}_n + \mathbf{C}_n) \hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) (\mathbf{B}_n + \mathbf{C}_n)^T - \mathbf{B}_n \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{B}_n^T \\ &= \mathbf{B}_n (\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) - \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t)) \mathbf{B}_n^T + \mathbf{B}_n \hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{C}_n^T \\ &\quad + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{B}_n^T + \mathbf{C}_n \hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{C}_n^T.\end{aligned}\quad (5.0.3)$$

Wir werden den Schätzfehler später wieder in der Frobeniusnorm betrachten und benötigen dafür Abschätzungen für $\|\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t)\|$ und $\|\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) - \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t)\|$. Für die Matrix der Faktorladungen und deren Schätzfehler nutzen wir die Raten aus (2.2.3) und Lemma D.0.19.

Lemma 5.0.2 *Für die empirische Autokovarianzmatrix der Faktoren gilt unter den Annahmen 2 und 12*

$$E(\|\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t)\|^2) = O(n^{-1}(n-h)K^2).$$

Beweis. Wir nutzen die Darstellung der empirischen Autokovarianzmatrix nach (5.0.1)

$$\begin{aligned}\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) &= \frac{1}{n} \sum_{t=1}^{n-h} (\mathbf{f}_{t+h} - \bar{\mathbf{f}}_n)(\mathbf{f}_t - \bar{\mathbf{f}}_n)^T = \frac{1}{n} \sum_{t=1}^{n-h} (\mathbf{f}_{t+h} \mathbf{f}_t^T - \mathbf{f}_{t+h} \bar{\mathbf{f}}_n^T - \bar{\mathbf{f}}_n \mathbf{f}_t^T + \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T) \\ &= \frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T - \frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \frac{1}{n} \mathbf{1}^T \mathbf{X}^T - \frac{1}{n} \sum_{t=1}^{n-h} \frac{1}{n} \mathbf{X} \mathbf{1} \mathbf{f}_t^T + \frac{(n-h)}{n} \frac{1}{n^2} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T\end{aligned}$$

und betrachten die Konvergenz dieser vier Summanden getrennt voneinander.

1. $\frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T$
Mithilfe der Cauchy-Schwarz-Ungleichung erhalten wir

$$\begin{aligned}E\left(\left\|\frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T\right\|^2\right) &= \frac{1}{n^2} E\left(\sum_{i=1}^K \sum_{j=1}^K \left(\sum_{t=1}^{n-h} f_{t+h,i} f_{t,j}\right)^2\right) \\ &= \frac{1}{n^2} \sum_{i=1}^K \sum_{j=1}^K E\left(\left(\sum_{t=1}^{n-h} f_{t+h,i} f_{t,j}\right)^2\right) \\ &\leq \frac{n-h}{n^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{t=1}^{n-h} E(f_{t+h,i}^2 f_{t,j}^2) \leq \frac{n-h}{n^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{t=1}^{n-h} \sqrt{E(f_{t+h,i}^4) E(f_{t,j}^4)}.\end{aligned}$$

Da die vierten Momente der Faktoren nach Annahme 2 gleichmäßig beschränkt sind, folgt

$$E\left(\left\|\frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T\right\|^2\right) = O(n^{-2}(n-h)^2 K^2). \quad (5.0.4)$$

2. $\frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{1}^T \mathbf{X}^T$
Wir setzen

$$\mathbf{1}_{h+1,n}^T = (0 \quad \dots \quad 0 \quad 1 \quad \dots \quad 1)^T. \quad (5.0.5)$$

$h \quad h+1$

Vorbereitend betrachten wir zunächst

$$\|\mathbf{1}_{h+1,n} \mathbf{1}^T\|^2 = \text{tr}\{\mathbf{1}_{h+1,n} \mathbf{1}^T \mathbf{1} \mathbf{1}_{h+1,n}^T\} = \text{tr}\{\mathbf{1}_{h+1,n}^T \mathbf{1}_{h+1,n} \mathbf{1}^T \mathbf{1}\} = (n-h)n. \quad (5.0.6)$$

Dann gilt mit (B.2.4) und der Rate aus (C.0.19)

$$\begin{aligned} E\left(\left\|\frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \frac{1}{n} \mathbf{1}^T \mathbf{X}^T\right\|^2\right) &= E\left(\left\|\frac{1}{n^2} \mathbf{X} \mathbf{1}_{h+1,n} \mathbf{1}^T \mathbf{X}^T\right\|^2\right) \\ &\leq \frac{1}{n^4} E(\|\mathbf{X}^T \mathbf{X}\|^2) \|\mathbf{1}_{h+1,n} \mathbf{1}^T\|^2 = \frac{1}{n^4} O(n^2 K^2) (n-h)n = O(n^{-1}(n-h)K^2). \end{aligned} \quad (5.0.7)$$

3. $\frac{1}{n} \sum_{t=1}^{n-h} \frac{1}{n} \mathbf{X} \mathbf{1}^T \mathbf{f}_t^T$

Analog zu (5.0.5) setzen wir $\mathbf{1}_{1,n-h} = (1, \dots, 1, 0, \dots, 0)^T$. Dann gilt mit der Rate aus (C.0.19) und analoger Rechnung zu (5.0.6)

$$\begin{aligned} E\left(\left\|\frac{1}{n} \sum_{t=1}^{n-h} \frac{1}{n} \mathbf{X} \mathbf{1}^T \mathbf{f}_t^T\right\|^2\right) &= E\left(\left\|\frac{1}{n^2} \mathbf{X} \mathbf{1} \mathbf{1}_{1,n-h}^T \mathbf{X}^T\right\|^2\right) \leq \frac{1}{n^4} E(\|\mathbf{X}^T \mathbf{X}\|^2) \|\mathbf{1} \mathbf{1}_{1,n-h}^T\|^2 \\ &= \frac{1}{n^4} O(n^2 K^2) n(n-h) = O(n^{-1}(n-h)K^2). \end{aligned} \quad (5.0.8)$$

4. $\frac{(n-h)}{n} \frac{1}{n^2} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T$

Mit der Rate aus (C.0.19) und analoger Rechnung zu (5.0.6) folgt

$$\begin{aligned} E\left(\left\|\frac{(n-h)}{n^3} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T\right\|^2\right) &\leq \frac{(n-h)^2}{n^6} E(\|\mathbf{X}^T \mathbf{X}\|^2) \|\mathbf{1} \mathbf{1}^T\|^2 \\ &= \frac{(n-h)^2}{n^6} O(n^2 K^2) n^2 = O(n^{-2}(n-h)^2 K^2). \end{aligned} \quad (5.0.9)$$

Dann folgt mit der Dreiecksungleichung und den Abschätzungen (5.0.4) und (5.0.7) - (5.0.9)

$$\begin{aligned} E(\|\text{côv}(\mathbf{f}_{t+h}, \mathbf{f}_t)\|^2) &\leq E\left(\left\|\frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T\right\|^2\right) + E\left(\left\|\frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \frac{1}{n} \mathbf{1}^T \mathbf{X}^T\right\|^2\right) \\ &\quad + E\left(\left\|\frac{1}{n} \sum_{t=1}^{n-h} \frac{1}{n} \mathbf{X} \mathbf{1}^T \mathbf{f}_t^T\right\|^2\right) + E\left(\left\|\frac{1}{n} (n-h) \frac{1}{n^2} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T\right\|^2\right) \end{aligned}$$

$$\begin{aligned}
 &= O(n^{-2}(n-h)^2K^2) + O(n^{-1}(n-h)K^2) + O(n^{-1}(n-h)K^2) + O(n^{-2}(n-h)^2K^2) \\
 &= O(n^{-1}(n-h)K^2),
 \end{aligned}$$

wegen $n-h \leq n$. □

Wir bemerken, dass sich die Rate für $h = 0$ zu $O(K^2)$ reduziert. Dies entspricht der Rate für die empirische Kovarianzmatrix der Faktoren in (C.0.56).

Für das asymptotische Verhalten des Schätzfehlers betrachten wir folgendes Lemma

Lemma 5.0.3 *Unter den Annahmen 2 und 9 - 12 existiert ein $C > 0$, sodass der Schätzfehler der empirischen Autokovarianzmatrix der Faktoren für alle n und h mit $n > h$ gleichmäßig beschränkt ist:*

$$E(\|\text{côv}(\mathbf{f}_{t+h}, \mathbf{f}_t) - \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t)\|^2) \leq C \max\{([n-h]^{-1}[\log(n-h)]K^2), (n^{-1}hK^2)\}.$$

Beweis.

$$\begin{aligned}
 \text{côv}(\mathbf{f}_{t+h}, \mathbf{f}_t) - \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t) &= \frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T - \frac{1}{n} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \bar{\mathbf{f}}_n^T - \frac{1}{n} \sum_{t=1}^{n-h} \bar{\mathbf{f}}_n \mathbf{f}_t^T + \frac{1}{n} \sum_{t=1}^{n-h} \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T \\
 &\quad - E(\mathbf{f}_{t+h} \mathbf{f}_t^T) + E(\mathbf{f}_{t+h}) E(\mathbf{f}_t^T) \\
 &= \left(\frac{1}{n-h} - \frac{h}{n(n-h)} \right) \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T - \frac{1}{n} \left(\sum_{t=1}^T \mathbf{f}_t - \sum_{t=1}^h \mathbf{f}_t \right) \bar{\mathbf{f}}_n^T - \frac{1}{n} \bar{\mathbf{f}}_n \left(\sum_{t=1}^T \mathbf{f}_t^T - \sum_{t=n-h+1}^n \mathbf{f}_t^T \right) \\
 &\quad + \frac{(n-h)}{n} \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T - E(\mathbf{f}_{t+h} \mathbf{f}_t^T) + E(\mathbf{f}_{t+h}) E(\mathbf{f}_t^T) \\
 &= \frac{1}{n-h} \sum_{t=1}^{n-h} (\mathbf{f}_{t+h} \mathbf{f}_t^T - E(\mathbf{f}_{t+h} \mathbf{f}_t^T)) - \frac{h}{n(n-h)} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T - \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T + \frac{1}{n} \sum_{t=1}^h \mathbf{f}_t \bar{\mathbf{f}}_n^T - \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T \\
 &\quad + \frac{1}{n} \bar{\mathbf{f}}_n \sum_{t=n-h+1}^n \mathbf{f}_t^T + \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T - \frac{h}{n} \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T + E(\mathbf{f}_{t+h}) E(\mathbf{f}_t^T) \\
 &= \frac{1}{n-h} \sum_{t=1}^{n-h} (\mathbf{f}_{t+h} \mathbf{f}_t^T - E(\mathbf{f}_{t+h} \mathbf{f}_t^T)) - (\bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T - E(\mathbf{f}_{t+h}) E(\mathbf{f}_t^T)) - \frac{h}{n(n-h)} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T \\
 &\quad + \frac{1}{n} \sum_{t=1}^h \mathbf{f}_t \bar{\mathbf{f}}_n^T + \frac{1}{n} \bar{\mathbf{f}}_n^T \sum_{t=n-h+1}^n \mathbf{f}_t^T - \frac{h}{n} \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T.
 \end{aligned}$$

Nun schätzen wir die Normen der Summanden getrennt voneinander ab.

1. $\frac{1}{n-h} \sum_{t=1}^{n-h} (\mathbf{f}_{t+h} \mathbf{f}_t^T - E(\mathbf{f}_{t+h} \mathbf{f}_t^T)) - (\bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T - E(\mathbf{f}_{t+h}) E(\mathbf{f}_t^T))$
 Wegen der Stationarität der Faktoren und Annahme 12 können wir direkt das Ergebnis für die empirische Kovarianzmatrix aus Lemma D.0.16 auf die empirische Autokovarianzmatrix übertragen. Dazu benötigen wir die Annahme der exponentiell abfallenden Tails der Verteilung von den Faktoren und den exponentiell abklingenden Mischungskoeffizienten zum Zeitabstand t , um Satz A.3.2 (vgl. Merlevède

et al. (2009)) anwenden zu können. Wir erhalten die Rate

$$\begin{aligned} & \frac{1}{(n-h)^2} E(\|\sum_{t=1}^{n-h} (\mathbf{f}_{t+h} \mathbf{f}_t^T - E(\mathbf{f}_{t+h} \mathbf{f}_t^T)) - (\bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T - E(\mathbf{f}_{t+h}) E(\mathbf{f}_t^T))\|^2) \\ &= O([n-h]^{-1} [\log(n-h)] K^2). \end{aligned}$$

2. $\frac{h}{n(n-h)} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T$
Mithilfe der Cauchy-Schwarz-Ungleichung und Annahme 2 erhalten wir

$$\begin{aligned} E(\|\frac{h}{n(n-h)} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T\|^2) &= \frac{h^2}{(n-h)^2 n^2} \sum_{i=1}^K \sum_{j=1}^K E\left(\left[\sum_{t=1}^{n-h} f_{t+h,i} f_{t,j}\right]^2\right) \\ &\leq \frac{h^2(n-h)}{(n-h)^2 n^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{t=1}^{n-h} E(f_{t+h,i}^2 f_{t,j}^2) \leq \frac{h^2}{(n-h)n^2} \sum_{i=1}^K \sum_{j=1}^K \sum_{t=1}^{n-h} \sqrt{E(f_{t+h,i}^4) E(f_{t,j}^4)} \\ &= O(h^2 n^{-2} K^2). \end{aligned}$$

3. $\frac{1}{n} \sum_{t=1}^h \mathbf{f}_t \bar{\mathbf{f}}_n^T$
Wir definieren $\mathbb{1}_{1,h}$ analog zu (5.0.5). Dann erhalten wir durch analoge Rechnung zu (5.0.6) und mit der Rate aus (C.0.19) folgendes Resultat

$$\begin{aligned} E(\|\frac{1}{n} \sum_{t=1}^h \mathbf{f}_t \bar{\mathbf{f}}_n^T\|^2) &= E(\|\frac{1}{n^2} \mathbf{X} \mathbb{1}_{1,h} \mathbb{1}^T \mathbf{X}^T\|^2) \leq \frac{1}{n^4} E(\|\mathbf{X}^T \mathbf{X}\|^2) \|\mathbb{1}_{1,h} \mathbb{1}^T\|^2 \\ &= \frac{1}{n^4} O(n^2 K^2) n h = O(n^{-1} h K^2). \end{aligned}$$

4. $\frac{1}{n} \bar{\mathbf{f}}_n^T \sum_{t=n-h+1}^n \mathbf{f}_t^T$
Analog zu (5.0.5) definieren wir $\mathbb{1}_{n-h+1,n}$. Dann erhalten wir mit analoger Rechnung zu (5.0.6) und (C.0.19)

$$\begin{aligned} E(\|\frac{1}{n} \bar{\mathbf{f}}_n^T \sum_{t=n-h+1}^n \mathbf{f}_t^T\|^2) &= E(\|\frac{1}{n^2} \mathbf{X} \mathbb{1} \mathbb{1}_{n-h+1,n}^T \mathbf{X}^T\|^2) \leq \frac{1}{n^4} E(\|\mathbf{X}^T \mathbf{X}\|^2) \|\mathbb{1} \mathbb{1}_{n-h+1,n}^T\|^2 \\ &= \frac{1}{n^4} O(n^2 K^2) n h = O(n^{-1} h K^2). \end{aligned}$$

5. $\frac{h}{n} \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T$
Mit (C.0.19) erhalten wir die Rate

$$\begin{aligned} E(\|\frac{h}{n} \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T\|^2) &= E(\|\frac{h}{n^3} \mathbf{X} \mathbb{1} \mathbb{1}^T \mathbf{X}^T\|^2) = \frac{h^2}{n^6} E(\|\mathbf{X}^T \mathbf{X}\|^2) \|\mathbb{1} \mathbb{1}^T\|^2 \\ &= \frac{h^2}{n^6} O(n^2 K^2) n^2 = O(n^{-2} h^2 K^2). \end{aligned}$$

Insgesamt erhalten wir dann mithilfe der Dreiecksungleichung

$$\begin{aligned}
 & E(\|\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) - \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t)\|^2) \\
 & \leq 5(E(\|\frac{1}{n-h} \sum_{t=1}^{n-h} (\mathbf{f}_{t+h} \mathbf{f}_t^T - E(\mathbf{f}_{t+h} \mathbf{f}_t^T)) - (\bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T - E(\mathbf{f}_{t+h}) E(\mathbf{f}_t^T))\|^2) \\
 & \quad + E(\|\frac{h}{n(n-h)} \sum_{t=1}^{n-h} \mathbf{f}_{t+h} \mathbf{f}_t^T\|^2) + E(\|\frac{1}{n} \sum_{t=1}^h \mathbf{f}_t \bar{\mathbf{f}}_n^T\|^2) + E(\|\frac{1}{n} \bar{\mathbf{f}}_n^T \sum_{t=n-h+1}^n \mathbf{f}_t\|^2) \\
 & \quad + E(\|\frac{h}{n} \bar{\mathbf{f}}_n \bar{\mathbf{f}}_n^T\|^2)) \\
 & = O([n-h][\log(n-h)]K^2) + O(n^{-2}h^2K^2) + O(n^{-1}hK^2) + O(n^{-1}hK^2) \\
 & \quad + O(n^{-2}h^2K^2) \\
 & = O([n-h][\log(n-h)]K^2) + O(n^{-1}hK^2),
 \end{aligned}$$

wegen $h \leq n$ und damit folgt die Behauptung. $\square$

Auch hier erhalten wir für $h = 0$ die gleiche Konvergenzrate, wie in Lemma D.0.16 für den Schätzfehler der empirischen Kovarianzmatrix.

Wir können nun die Resultate der Lemmata 5.0.2 und 5.0.3 und die Darstellung (5.0.3) nutzen, um die Asymptotik für die Schätzung der Autokovarianzmatrix im Faktormodell herzuleiten.

Satz 5.0.4 *Unter den Annahmen 2, 6 und 9 - 12 existiert eine Konstante $C > 0$, sodass die Schätzung der Autokovarianzmatrix im Faktormodell für alle n und h mit $n > h$ gleichmäßig beschränkt ist:*

$$E(\|\hat{\Sigma}_n(h) - \Sigma_n(h)\|^2) = C \max\{([n-h]^{-1}[\log(n-h)]p^2K^2), (n^{-1}hp^2K^2), (n^{-2}[n-h]p^2K^3)\}.$$

Beweis. Mithilfe der Darstellung aus (5.0.3), der Dreiecksungleichung und (B.2.3) schreiben wir

$$\begin{aligned}
 E(\|\hat{\Sigma}_n(h) - \Sigma_n(h)\|^2) & \leq 4(E(\|\mathbf{B}_n(\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) - \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t))\mathbf{B}_n^T\|^2) \\
 & \quad + E(\|\mathbf{B}_n \hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{C}_n^T\|^2) + E(\|\mathbf{C}_n \hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{B}_n^T\|^2) + E(\|\mathbf{C}_n \hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{C}_n^T\|^2)) \\
 & \leq E(\|\mathbf{B}_n\|^2 \|(\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t) - \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t))\|^2 \|\mathbf{B}_n^T\|^2) + E(\|\mathbf{B}_n\|^2 \|\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t)\|^2 \|\mathbf{C}_n^T\|^2) \\
 & \quad + E(\|\mathbf{C}_n\|^2 \|\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t)\|^2 \|\mathbf{B}_n^T\|^2) + E(\|\mathbf{C}_n\|^2 \|\hat{\text{cov}}(\mathbf{f}_{t+h}, \mathbf{f}_t)\|^2 \|\mathbf{C}_n^T\|^2).
 \end{aligned}$$

Mit den Abschätzungen aus (2.2.3) und den Lemmata D.0.19, 5.0.2 und 5.0.3 folgt

$$\begin{aligned}
 E(\|\hat{\Sigma}_n(h) - \Sigma_n(h)\|^2) & = O([n-h]^{-1}[\log(n-h)]p^2K^2) + O(n^{-1}hp^2K^2) \\
 & \quad + O(n^{-2}[n-h]p^2K^3) + O(n^{-3}[n-h]p^2K^4) \\
 & = O([n-h]^{-1}[\log(n-h)]p^2K^2) + O(n^{-1}hp^2K^2) + O(n^{-2}[n-h]p^2K^3),
 \end{aligned}$$

wegen $K < n$. $\square$

Es fällt auf, dass wir für $h = 0$ im Vergleich mit Satz 2.4.1 zu einer abweichenden Konvergenzrate kommen. Der Grund dafür ist, dass wir in Lemma C.0.6 für die Abschätzung von $\|\mathbf{B}_n \text{cov}(\mathbf{f}) \mathbf{C}_n^T\|$ den Ausdruck als Produkt behandelt haben, hier aber die Komponenten von $\|\mathbf{B}_n \text{cov}(\mathbf{f}_{t+h}, \mathbf{f}_t) \mathbf{C}_n^T\|$ jeweils einzeln abgeschätzt haben. Dadurch verbleiben die Raten $O(n^{-1}[\log n]p^2K^2)$ und $O(n^{-1}p^2K^3)$, die finale Rate wird daher durch $\max\{\log n, K\}$ festgelegt.

Insgesamt ist festzuhalten, dass wir für die Herleitung der Asymptotik bei der Schätzung der Autokovarianzmatrix im Faktormodell analog zu dem Verfahren bei der Kovarianzmatrix vorgehen können. Die Ergebnisse, die wir in diesem Kapitel hergeleitet haben sind weitgehend konsistent zu denen in Abschnitt 2.4.

Zusammenfassung und Ausblick

Wir haben in dieser Arbeit untersucht, wie sich Faktormodelle für die Schätzung von hochdimensionalen Kovarianzmatrizen eignen. Dabei haben wir die Asymptotik der gängigen Stichprobenkovarianzmatrix mit der Asymptotik der Schätzung der Kovarianzmatrix aus dem Faktormodell unter verschiedenen Normen verglichen.

Es stellte sich heraus, dass die Dimensionsreduktion des Faktormodells keinen entscheidenden Vorteil für die Schätzung der Kovarianzmatrix bringt. Bei der Schätzung der inversen Kovarianzmatrix ist ein deutlicher Vorteil des Schätzers im Faktormodell gegenüber der Inversen der empirischen Kovarianzmatrix erkennbar.

Diese Resultate wirken sich direkt auf Anwendungen in der Finanzwirtschaft aus. Während in der Portfolioselektion der Vorteil der Schätzung der Inversen genutzt werden kann, wird im Risikomanagement eine Schätzung der Kovarianzmatrix benötigt. Für die von uns untersuchten Anwendungen im Risikomanagement bringt die Anwendung von Faktormodellen daher keinen entscheidenden Vorteil.

Unsere Untersuchungen bauten anfangs auf der stark vereinfachenden Annahme unabhängig identisch verteilter Faktoren im Faktormodell auf. Wir haben gezeigt, dass der Übergang zu stationären Faktoren, wie sie beispielsweise bei der Nutzung von Zeitreihen vorliegen, nur geringe Auswirkungen auf die Asymptotik der Kovarianzmatrixschätzung hat. Unter der Annahme von stationären Faktoren haben wir unsere Untersuchung erweitert und neben der Schätzung der Kovarianzmatrix im Faktormodell auch die Autokovarianzmatrix für einen Zeitabstand $h > 0$ betrachtet.

Eine weitere stark vereinfachende Annahme im Faktormodell ist die Unkorreliertheit der Fehlerterme. Die Lockerung dieser Annahme, wie sie schon in Chamberlain und Rothschild (1984) vorgeschlagen wurde, haben wir in dieser Arbeit nicht verfolgt. Fan et al. (2011) nutzen in diesem Zusammenhang approximative Faktormodelle, in denen sie die Annahme treffen, dass die Fehlerkovarianzmatrix im Faktormodell dünn besetzt ist. Für die Schätzung dieser dünn besetzten Matrix greifen sie auf ein Thresholdverfahren aus Cai und Liu (2011) zurück. Die zusätzliche Schätzung der Fehlerkovarianzmatrix hat insbesondere Auswirkungen auf die Konvergenzraten der Kovarianzmatrix des betrachteten Portfolios.

Die in dieser Arbeit gewonnenen Erkenntnisse über die verbesserten Schätzer durch Verwendung von Faktormodellen könnten daher mit einer weiteren Studie unter Berücksichtigung von korrelierten Fehlertermen verglichen werden.

A. Hilfssätze und Beweise

A.1. Kovarianzmatrizen

Lemma A.1.1 Sei $\mathbf{x} \in \mathbb{R}^p$ ein Zufallsvektor und $\mathbf{x}_1, \dots, \mathbf{x}_n$ eine unabhängig identisch verteilte Stichprobe von $\mathbf{x}$. Dann besitzt die Stichprobenkovarianzmatrix $\text{cov}(\mathbf{X})$ von $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{p \times n}$ die Darstellung

$$\text{cov}(\mathbf{X}) = \frac{1}{n-1} \mathbf{X} \mathbf{X}^T - \frac{1}{n(n-1)} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T. \quad (\text{A.1.1})$$

Beweis.

$$\begin{aligned} \text{cov}(\mathbf{X}) &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \\ &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^T - \mathbf{x}_i \bar{\mathbf{x}}^T - \bar{\mathbf{x}} \mathbf{x}_i^T + \bar{\mathbf{x}} \bar{\mathbf{x}}^T) \\ &= \frac{1}{n-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T - n \bar{\mathbf{x}} \bar{\mathbf{x}}^T - n \bar{\mathbf{x}} \bar{\mathbf{x}}^T + n \bar{\mathbf{x}} \bar{\mathbf{x}}^T \\ &= \frac{1}{n-1} (\mathbf{X} \mathbf{X}^T - n \bar{\mathbf{x}} \bar{\mathbf{x}}^T) \\ &= \frac{1}{n-1} (\mathbf{X} \mathbf{X}^T - n \frac{1}{n} \mathbf{X} \mathbf{1} \frac{1}{n} \mathbf{1}^T \mathbf{X}^T) = \frac{1}{n-1} (\mathbf{X} \mathbf{X}^T - \frac{1}{n} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T) \\ &= \frac{1}{n-1} \mathbf{X} \mathbf{X}^T - \frac{1}{n(n-1)} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T. \end{aligned}$$

□

Lemma A.1.2 Sei $\mathbf{x} = (x_1, \dots, x_p)^T$ ein Zufallsvektor mit $\max_{1 \leq i \leq p} E(x_i^4) \leq C$ für eine Konstante $C > 0$. Sei $\mathbf{x}_1, \dots, \mathbf{x}_n$ eine unabhängig identisch verteilte Stichprobe von $\mathbf{x}$ und $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{p \times n}$. Dann gilt für die Stichprobenkovarianzmatrix von $\mathbf{X}$ in der Darstellung (A.1.1) folgende Aussage

$$E(\|\text{cov}(\mathbf{X}) - \text{cov}(\mathbf{x})\|^2) = O(n^{-1}p^2).$$

Beweis. Wir bezeichnen die Komponenten der Stichprobenkovarianzmatrix mit $\text{cov}(x_i, x_j) := \frac{1}{n-1} \sum_{k=1}^n x_{ki} x_{kj} - \frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} \sum_{l=1}^n x_{lj}$ für $i, j = 1, \dots, p$. Dann gilt für die Konvergenz im quadratischen Mittel

$$E(\|\text{cov}(\mathbf{X}) - \text{cov}(\mathbf{x})\|^2) = E\left(\sum_{i=1}^p \sum_{j=1}^p [\text{cov}(x_i, x_j) - \text{cov}(x_i, x_j)]^2\right). \quad (\text{A.1.2})$$

In dieser Darstellung betrachten wir nun einen beliebigen Summanden für $i, j = 1, \dots, p$ und führen eine Abschätzung mithilfe der Cauchy–Schwarz-Ungleichung durch

$$\begin{aligned}
 & E([\text{côv}(x_i, x_j) - \text{cov}(x_i, x_j)]^2) \\
 &= E\left(\left[\frac{1}{n-1} \sum_{k=1}^n x_{ki} x_{kj} - E(x_i x_j) - \left(\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} \sum_{l=1}^n x_{lj} - E(x_i) E(x_j)\right)\right]^2\right) \\
 &\leq 2E\left(\left[\frac{1}{n-1} \sum_{k=1}^n x_{ki} x_{kj} - E(x_i x_j)\right]^2\right) + 2E\left(\left[\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} \sum_{l=1}^n x_{lj} - E(x_i) E(x_j)\right]^2\right), \\
 &=: 2\mathcal{A}_1 + 2\mathcal{A}_2.
 \end{aligned} \tag{A.1.3}$$

Wegen der gleichmäßig beschränkten vierten Momente sind mit der Cauchy–Schwarz-Ungleichung

$$E(x_i^2) \leq \sqrt{E(x_i^4)}$$

auch die zweiten Momente der Komponenten gleichmäßig beschränkt.

Wir betrachten nun $\mathcal{A}_1$. Mit $\frac{1}{n-1} = \frac{1}{n} + \frac{1}{n(n-1)}$ und der Cauchy–Schwarz-Ungleichung gilt

$$\begin{aligned}
 \mathcal{A}_1 &= E\left(\left[\frac{1}{n-1} \sum_{k=1}^n x_{ki} x_{kj} - E(x_i x_j)\right]^2\right) \\
 &= E\left(\left[\frac{1}{n} \sum_{k=1}^n (x_{ki} x_{kj} - E(x_i x_j)) + \frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} x_{kj}\right]^2\right) \\
 &\leq 2E\left(\left[\frac{1}{n} \sum_{k=1}^n (x_{ki} x_{kj} - E(x_i x_j))\right]^2\right) + 2E\left(\left[\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} x_{kj}\right]^2\right) \\
 &= \frac{2}{n^2} E\left(\left[\sum_{k=1}^n (x_{ki} x_{kj} - E(x_i x_j))\right]^2\right) + 2E\left(\left[\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} x_{kj}\right]^2\right).
 \end{aligned}$$

Die $(x_{ki} x_{kj} - E(x_i x_j))$ sind unabhängig identisch verteilte zentrierte Zufallsvariablen, daher folgt weiter

$$\begin{aligned}
 \mathcal{A}_1 &\leq \frac{2}{n^2} \text{var}\left(\sum_{k=1}^n (x_{ki} x_{kj} - E(x_i x_j))\right) + 2E\left(\left[\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} x_{kj}\right]^2\right) \\
 &= \frac{2}{n} \text{var}((x_{1i} x_{1j} - E(x_i x_j))) + 2E\left(\left[\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} x_{kj}\right]^2\right) \\
 &= \frac{2}{n} E([x_{1i} x_{1j} - E(x_i x_j)]^2) + 2E\left(\left[\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} x_{kj}\right]^2\right) \\
 &= \frac{2}{n} (E(x_{1i}^2 x_{1j}^2) - [E(x_i x_j)]^2) + 2E\left(\left[\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} x_{kj}\right]^2\right) \\
 &\leq \frac{2}{n} E(x_{1i}^2 x_{1j}^2) + 2E\left(\left[\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} x_{kj}\right]^2\right)
 \end{aligned}$$

und mit der Cauchy-Schwarz-Ungleichung schätzen wir ab durch

$$\begin{aligned}\mathcal{A}_1 &\leq \frac{2}{n} \sqrt{E(x_{1i}^4)E(x_{1j}^4)} + \frac{2n}{n^2(n-1)^2} \sum_{k=1}^n E(x_{ki}^2 x_{kj}^2) \\ &\leq \frac{2}{n} \sqrt{E(x_{1i}^4)E(x_{1j}^4)} + \frac{2}{n(n-1)^2} \sum_{k=1}^n \sqrt{E(x_{ki}^4)E(x_{kj}^4)}.\end{aligned}$$

Da die vierten Momente nach Voraussetzung gleichmäßig beschränkt sind, folgt

$$\mathcal{A}_1 = O(n^{-1}) + O(n^{-2}) = O(n^{-1}).$$

Nun betrachten wir $\mathcal{A}_2$ mithilfe der Cauchy-Schwarz-Ungleichung. Wegen $\frac{1}{n(n-1)} = \frac{1}{n^2} + \frac{1}{n^2(n-1)}$ gilt

$$\begin{aligned}\mathcal{A}_2 &= E\left(\left[\frac{1}{n(n-1)} \sum_{k=1}^n x_{ki} \sum_{l=1}^n x_{lj} - E(x_i)E(x_j)\right]^2\right) \\ &= E\left(\left[\frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n (x_{ki}x_{lj} - E(x_i)E(x_j)) + \frac{1}{n^2(n-1)} \sum_{k=1}^n \sum_{l=1}^n x_{ki}x_{lj}\right]^2\right) \\ &\leq 2E\left(\left[\frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n (x_{ki}x_{lj} - E(x_i)E(x_j))\right]^2\right) + 2E\left(\left[\frac{1}{n^2(n-1)} \sum_{k=1}^n \sum_{l=1}^n x_{ki}x_{lj}\right]^2\right) \\ &=: 2\mathcal{A}_3 + 2\mathcal{A}_4.\end{aligned}$$

Wir schätzen zunächst $\mathcal{A}_4$ durch die Cauchy-Schwarz-Ungleichung ab:

$$\begin{aligned}\mathcal{A}_4 &= \frac{1}{n^4(n-1)^2} E\left(\left[\sum_{k=1}^n \sum_{l=1}^n x_{ki}x_{lj}\right]^2\right) \leq \frac{n^2}{n^4(n-1)^2} \sum_{k=1}^n \sum_{l=1}^n E(x_{ki}^2 x_{lj}^2) \\ &= \frac{n^2}{n^4(n-1)^2} \left(\sum_{k=1}^n E(x_{ki}^2 x_{kj}^2) + \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n E(x_{ki}^2)E(x_{lj}^2) \right) \\ &\leq \frac{n^2}{n^4(n-1)^2} \left(\sum_{k=1}^n \sqrt{E(x_{ki}^4)E(x_{kj}^4)} + \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n E(x_{ki}^2)E(x_{lj}^2) \right).\end{aligned}$$

Wegen der gleichmäßig beschränkten zweiten und vierten Momente erhalten wir

$$\mathcal{A}_4 = O(n^{-2}).$$

Wir kommen nun zur Abschätzung von $\mathcal{A}_3$.

$$\begin{aligned}&E\left(\left[\frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n (x_{ki}x_{lj} - E(x_i)E(x_j))\right]^2\right) \\ &= \frac{1}{n^4} \sum_{k=1}^n \sum_{l=1}^n \sum_{m=1}^n \sum_{o=1}^n E((x_{ki}x_{lj} - E(x_i)E(x_j))(x_{mi}x_{oj} - E(x_i)E(x_j))).\end{aligned}\tag{A.1.4}$$

Jeden dieser n^4 Summanden können wir einer der folgenden sieben Darstellungen zuordnen, die wir nun getrennt voneinander mithilfe der zweiten und vierten Momente der x_i , $i = 1, \dots, p$ abschätzen.

1. Für $k \neq l \neq m \neq o$ gilt, da die Stichprobe unabhängig identisch verteilt ist

$$\begin{aligned}
 & \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n \sum_{\substack{o=1 \\ o \notin \{k,l,m\}}}^n E((x_{ki}x_{lj} - E(x_i)E(x_j))(x_{mi}x_{oj} - E(x_i)E(x_j))) \\
 &= \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n \sum_{\substack{o=1 \\ o \notin \{k,l,m\}}}^n (E(x_{ki}x_{lj}x_{mi}x_{oj}) - E(x_{ki}x_{lj})E(x_i)E(x_j) \\
 &\quad - E(x_i)E(x_j)E(x_{mi}x_{oj}) + E(x_i)^2E(x_j)^2) \\
 &= \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n \sum_{\substack{o=1 \\ o \notin \{k,l,m\}}}^n (E(x_{ki})E(x_{lj})E(x_{mi})E(x_{oj}) - E(x_{ki})E(x_{lj})E(x_i)E(x_j) \\
 &\quad - E(x_i)E(x_j)E(x_{mi})E(x_{oj}) + E(x_i)^2E(x_j)^2) = 0.
 \end{aligned}$$

Für die Abschätzungen der anderen Summanden schätzen wir die Darstellung aus (A.1.4) ab durch

$$\mathcal{A}_3 \leq \frac{1}{n^4} \sum_{k=1}^n \sum_{l=1}^n \sum_{m=1}^n \sum_{o=1}^n (E(x_{ki}x_{lj}x_{mi}x_{oj}) + E(x_i)^2E(x_j)^2).$$

2. Für $k = l = m = o$ gilt mit der Cauchy-Schwarz-Ungleichung und den gleichmäßig beschränkten zweiten und vierten Momenten

$$\frac{1}{n^4} \sum_{k=1}^n (E(x_{ki}^2x_{kj}^2) + E(x_i)^2E(x_j)^2) \leq \frac{1}{n^4} \sum_{k=1}^n (\sqrt{E(x_{ki}^4)E(x_{lj}^4)} + E(x_i)^2E(x_j)^2) = O(n^{-3}).$$

3. Da die Stichprobe unabhängig identisch verteilt ist, gilt für $k = m$ und $l = o$ mithilfe der Jensen-Ungleichung und der Annahme an die vierten Momente

$$\begin{aligned}
 & \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (E(x_{ki}^2x_{lj}^2) + E(x_i)^2E(x_j)^2) = \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (E(x_{ki}^2)E(x_{lj}^2) + E(x_i)^2E(x_j)^2) \\
 & \leq \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (E(x_{ki}^2)E(x_{lj}^2) + E(x_i)^2E(x_j)^2) = O(n^{-2}).
 \end{aligned}$$

4. Für $k = l$ und $m = o$ gilt analog mit der Cauchy-Schwarz-Ungleichung und der Annahme an die vierten Momente

$$\begin{aligned}
 & \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (E(x_{ki}x_{kj}x_{li}x_{lj}) + E(x_i)^2E(x_j)^2) = \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (E(x_{ki}x_{kj})E(x_{li}x_{lj}) + E(x_i)^2E(x_j)^2) \\
 & \leq \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (\sqrt{E(x_{ki}^2)E(x_{kj}^2)E(x_{li}^2)E(x_{lj}^2)} + E(x_i)^2E(x_j)^2) = O(n^{-2}).
 \end{aligned}$$

5. Für $k = m = l \neq o$ oder $k = m = o \neq l$ folgt wieder mit der Cauchy-Schwarz-Ungleichung und der Annahme an die vierten Momente

$$\frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (E(x_{ki}x_{kj}x_{ki}x_{lj}) + E(x_i)^2E(x_j)^2) = \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (E(x_{ki}^2x_{kj}^2)E(x_{li}) + E(x_i)^2E(x_j)^2)$$

$$\leq \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (\sqrt{E(x_{ki}^4)E(x_{kj}^2)E(x_{lj}^2)} + E(x_i^2)E(x_j^2)) = O(n^{-2}).$$

Wegen der gleichmäßig beschränkten zweiten und vierten Momente gilt diese Rechnung analog für die Fälle $l = o = k \neq m$ und $l = o = m \neq k$:

$$\frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n (E(x_{ki}x_{kj}x_{li}x_{kj}) + E(x_i)^2E(x_j)^2) = O(n^{-2}).$$

6. Für $k = m \neq o \neq l$ gilt mithilfe der Jensen-Ungleichung, der Annahme an die vierten Momente und da die Stichprobe unabhängig identisch verteilt ist

$$\begin{aligned} & \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n (E(x_{ki}^2x_{lj}x_{mj}) + E(x_i)^2E(x_j)^2) \\ &= \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n (E(x_{ki}^2)E(x_{lj})E(x_{mj}) + E(x_i)^2E(x_j)^2) \\ &\leq \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n (E(x_{ki}^2)E(x_j^2) + E(x_i^2)E(x_j^2)) = O(n^{-1}). \end{aligned}$$

Wegen der gleichmäßig beschränkten zweiten Momente gilt dieses Ergebnis auch für den Fall $o = l \neq m \neq k$:

$$\frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n (E(x_{ki}x_{li}x_{mj}^2) + E(x_i)^2E(x_j)^2) = O(n^{-1}).$$

7. Für $k = l \neq m \neq o$ oder $k = o \neq l \neq m$ gilt mit der Cauchy-Schwarz-Ungleichung und den beschränkten vierten Momenten

$$\begin{aligned} & \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n (E(x_{ki}x_{kj}x_{li}x_{mj}) + E(x_i)^2E(x_j)^2) \\ &= \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n (E(x_{ki}x_{kj})E(x_{li})E(x_{mj}) + E(x_i)^2E(x_j)^2) \\ &\leq \frac{1}{n^4} \sum_{k=1}^n \sum_{\substack{l=1 \\ l \neq k}}^n \sum_{\substack{m=1 \\ m \notin \{k,l\}}}^n (\sqrt{E(x_{ki}^2)E(x_{kj}^2)E(x_{li}^2)E(x_{mj}^2)} + E(x_i^2)E(x_j^2)) = O(n^{-1}). \end{aligned}$$

Insgesamt erhalten wir für $\mathcal{A}_3$ in (A.1.4) die Rate $O(n^{-1})$ und für (A.1.3) folgt mit den Abschätzungen $\mathcal{A}_1$, $\mathcal{A}_3$ und $\mathcal{A}_4$

$$E([\text{côv}(x_i, x_j) - \text{cov}(x_i, x_j)]^2) = O(n^{-1}).$$

Wegen der gleichmäßig beschränkten Momente sind in dieser Abschätzung die Schranken unabhängig von i und j . Daher gilt das Resultat gleichmäßig für alle Summanden in der Darstellung (A.1.2):

$$E(\|\text{côv}(\mathbf{X}) - \text{cov}(\mathbf{x})\|^2) = E\left(\sum_{i=1}^p \sum_{j=1}^p [\text{côv}(x_i, x_j) - \text{cov}(x_i, x_j)]^2\right) = O(n^{-1}p^2).$$

□

A.2. Bedingte Erwartungen

Lemma A.2.1 *Seien X und Y integrierbare Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und sei $\mathcal{C}$ eine σ -Algebra mit $\mathcal{C} \subseteq \mathcal{A}$. Weiterhin sei auch XY integrierbar. Wir bezeichnen die Bedingte Erwartung von X unter $\mathcal{C}$ als $E(X|\mathcal{C})$. Dann gelten folgende Eigenschaften (vgl. Bauer, 2002, S.120)*

$$E(E(X|\mathcal{C})) = E(X). \quad (\text{A.2.1})$$

$$E(aX + bY|\mathcal{C}) = aE(X|\mathcal{C}) + bE(Y|\mathcal{C}) \text{ fast sicher für } a, b \in \mathbb{R}. \quad (\text{A.2.2})$$

$$E(XY|\mathcal{C}) = XE(Y|\mathcal{C}) \text{ fast sicher, falls } X \text{ messbar bezüglich } \mathcal{C} \text{ ist.} \quad (\text{A.2.3})$$

Mithilfe dieser Eigenschaften kann die Varianz einer quadratintegrierbaren Zufallsvariable X geschrieben werden als

$$\text{var}(X) = E(\text{var}(X|\mathcal{C})) + \text{var}(E(X|\mathcal{C})). \quad (\text{A.2.4})$$

Beweis. Es gilt

$$\begin{aligned} \text{var}(X) &= E([X - E(X)]^2) \stackrel{(\text{A.2.1})}{=} E(E([X - E(X)]^2|\mathcal{C})) \\ &= E(E([X - E(X|\mathcal{C}) + E(X|\mathcal{C}) - E(X)]^2|\mathcal{C})) \\ &= E(E([X - E(X|\mathcal{C})]^2|\mathcal{C}) + 2E([X - E(X|\mathcal{C})][E(X|\mathcal{C}) - E(X)]|\mathcal{C}) \\ &\quad + E([E(X|\mathcal{C}) - E(X)]^2|\mathcal{C})) =: E(\mathcal{A}_1 + 2\mathcal{A}_2 + \mathcal{A}_3). \end{aligned}$$

Mit der Linearität der Bedingten Erwartung nach (A.2.2) gelten

$$\begin{aligned} E(\mathcal{A}_1) &= E(E([X - E(X|\mathcal{C})]^2|\mathcal{C})) = E(\text{var}(X|\mathcal{C})), \\ E(\mathcal{A}_2) &= E(E([X - E(X|\mathcal{C})][E(X|\mathcal{C}) - E(X)]|\mathcal{C})) \\ &\stackrel{(\text{A.2.3})}{=} E([E(X|\mathcal{C}) - E(X)]E(X - E(X|\mathcal{C})|\mathcal{C})) \\ &\stackrel{(\text{A.2.2})}{=} E([E(X|\mathcal{C}) - E(X)][E(X|\mathcal{C}) - E(E(X|\mathcal{C})|\mathcal{C})]) \\ &\stackrel{(\text{A.2.1})}{=} E([E(X|\mathcal{C}) - E(X)][E(X|\mathcal{C}) - E(X|\mathcal{C})]) = 0 \end{aligned}$$

und

$$E(\mathcal{A}_3) = E(E([E(X|\mathcal{C}) - E(X)]^2|\mathcal{C}))$$

$$\begin{aligned}
 & \stackrel{(A.2.1)}{=} E(E([E(X|\mathcal{C}) - E(E(X|\mathcal{C}))]^2|\mathcal{C})) \\
 & = E(\text{var}(E(X|\mathcal{C}))|\mathcal{C}) = \text{var}(E(X|\mathcal{C})).
 \end{aligned}$$

Damit folgt die Behauptung. $\square$

A.3. Exponentielle Ungleichungen

Lemma A.3.1 *Seien Z_1, Z_2 Zufallsvariablen, für die Konstanten $r_1, r_2 \in (0, 1)$ und $b_1, b_2 > 0$ existieren, sodass für alle $s > 0$ gilt*

$$P(|Z_i| > s) \leq \exp(1 - (s/b_i)^{r_i}), \quad i = 1, 2.$$

Dann gilt für ein r_3 und ein $b_3 > 0$

$$P(|Z_1 Z_2| > s) \leq \exp(1 - (s/b_3)^{r_3})$$

für alle $s > 0$.

Beweis. (Vgl. Fan et al., 2011, Lemma A.2) Wir setzen für alle $s > 0$

$$\begin{aligned}
 b &:= b_1 b_2, \\
 r &:= \frac{r_1 r_2}{r_1 + r_2} \quad \text{und} \\
 M &:= \left(\frac{s b_2^{r_2/r_1}}{b_1} \right)^{\frac{r_1}{r_1+r_2}} = s^{\frac{r_1}{r_1+r_2}} b_2^{\frac{r_2}{r_1+r_2}} b_1^{-\frac{r_1}{r_1+r_2}}.
 \end{aligned}$$

Wir zeigen, dass $P(|Z_1 Z_2| > s) \leq P(M|Z_1| > s) + P(|Z_2| > M)$.

Um den Satz der totalen Wahrscheinlichkeit anwenden zu können, nehmen wir die Zerlegung $|Z_2| > M$ und $|Z_2| \leq M$ vor. Wegen $\{|Z_1 Z_2| > s \cap |Z_2| > M\} \subseteq \{|Z_2| > M\}$ und $\{|Z_1 Z_2| > s \cap |Z_2| \leq M\} \subseteq \{M|Z_1| > s\}$ gilt

$$\begin{aligned}
 P(|Z_1 Z_2| > s) &= P(|Z_1 Z_2| > s, |Z_2| > M) + P(|Z_1 Z_2| > s, |Z_2| \leq M) \\
 &\leq P(|Z_2| > M) + P(M|Z_1| > s).
 \end{aligned} \tag{A.3.1}$$

Nun setzen wir die Voraussetzung des Lemmas ein und erhalten

$$\begin{aligned}
 P(|Z_1 Z_2| > s) &\leq \exp\left(1 - \left(\frac{M}{b_2}\right)^{r_2}\right) + \exp\left(1 - \left(\frac{s}{M b_1}\right)^{r_1}\right) \\
 &= \exp\left(1 - \frac{s^{\frac{r_1 r_2}{r_1+r_2}} b_2^{\frac{r_2}{r_1+r_2}} b_1^{-\frac{r_1 r_2}{r_1+r_2}}}{b_2^{r_2}}\right) + \exp\left(1 - \frac{s^{r_1}}{s^{\frac{r_1^2}{r_1+r_2}} b_2^{\frac{r_1 r_2}{r_1+r_2}} b_1^{-\frac{r_1^2}{r_1+r_2}} b_1^{r_1}}\right) \\
 &= \exp\left(1 - s^{\frac{r_1 r_2}{r_1+r_2}} b_2^{-\frac{r_1 r_2}{r_1+r_2}} b_1^{-\frac{r_1 r_2}{r_1+r_2}}\right) + \exp\left(1 - \frac{s^{\frac{r_1 r_2}{r_1+r_2}}}{b_2^{\frac{r_1 r_2}{r_1+r_2}} b_1^{\frac{r_1 r_2}{r_1+r_2}}}\right) \\
 &= 2 \exp\left(1 - \left(\frac{s}{b}\right)^r\right).
 \end{aligned} \tag{A.3.2}$$

Sei nun $0 < r_3 < r$ und $b_3 > \max\{(\frac{r_3}{r})^{\frac{1}{r}}b, (1 + \log 2)^{\frac{1}{r}}b\} \geq b$.

Wir setzen $F(s) := (\frac{s}{b})^r - (\frac{s}{b_3})^{r_3}$ und zeigen, dass $F(s)$ für $s > b_3$ monoton wachsend ist. Dazu betrachten wir die erste Ableitung von $F(s)$:

$$F'(s) = \frac{rs^{r-1}}{b^r} - \frac{r_3s^{r_3-1}}{b_3^{r_3}} = \frac{r}{s}(\frac{s}{b})^r - \frac{r_3}{s}(\frac{s}{b_3})^{r_3} \geq \frac{r}{s}(\frac{s}{b})^r - \frac{r}{s}(\frac{s}{b_3})^r \geq 0,$$

da $b_3 \geq b$ ist.

Daher folgt für $s > b_3$ auch $F(s) > F(b_3)$. Es gilt

$$\begin{aligned} F(b_3) &= (\frac{b_3}{b})^r - (\frac{b_3}{b_3})^{r_3} = (\frac{b_3}{b})^r - 1 > \max\{(\frac{(\frac{r_3}{r})^{\frac{1}{r}}b}{b})^r - 1, (\frac{(1 + \log 2)^{\frac{1}{r}}b}{b})^r - 1\} \\ &= \max\{\frac{r_3}{r}, 1 + \log 2\} - 1 = \log 2, \end{aligned}$$

wegen $r_3 < r$. Daher folgt $F(s) > F(b_3) > \log 2$.

1. Fall: $s > b_3$: Wir nutzen (A.3.1) und setzen $(\frac{s}{b})^r = F(s) + (\frac{s}{b_3})^{r_3}$ in (A.3.2) ein:

$$\begin{aligned} P(|Z_1 Z_2| > s) &\leq 2 \exp\left(1 - (\frac{s}{b})^r\right) = 2 \exp\left(1 - (F(s) + (\frac{s}{b_3})^{r_3})\right) \\ &\leq 2 \exp\left(1 - \log 2 - (\frac{s}{b_3})^{r_3}\right) = \exp\left(1 - (\frac{s}{b_3})^{r_3}\right). \end{aligned}$$

Damit haben wir für diesen Fall die Behauptung des Lemmas gezeigt.

2. Fall: $s \leq b_3$. Wir betrachten $\exp(1 - (\frac{s}{b_3})^{r_3})$.

Falls $s = b_3$: $\exp(1 - (\frac{s}{b_3})^{r_3}) = \exp(0) = 1$.

Falls $s < b_3$: $\exp(1 - (\frac{s}{b_3})^{r_3}) \in (\exp(0), \exp(1)) > 1$.

Wegen $P(|Z_1 Z_2| > s) \leq 1$ folgt daraus:

$$P(|Z_1 Z_2| > s) \leq \exp\left(1 - (\frac{s}{b_3})^{r_3}\right).$$

Also folgt die Behauptung für alle $s > 0$. □

Satz A.3.2 (Vgl. Merlevède et al., 2009, Theorem 1, Remark 1)

Seien $(X_j)_{j \in \mathbb{Z}}$ reelle, zentrierte Zufallsvariablen mit

$$\sup_{k > 0} P(|X_k| > t) \leq \exp(1 - t^{\gamma_1}), \text{ für } t > 0, \gamma_1 > 0,$$

die die starke Mischungseigenschaft erfüllen, sodass Konstanten $\gamma_2 > 0$ und $c > 0$ existieren mit

$$\alpha(n) \leq \exp(-cn^{\gamma_2}) \text{ für alle } n > 0.$$

Sei $\frac{1}{\gamma} := \frac{1}{\gamma_1} + \frac{1}{\gamma_2} < 1$, dann existieren für alle $n \geq 4$ positive Konstanten C_1, C_2, C_3, C_4, C_5 , die abhängig von c, γ_1 und γ_2 sind, sodass für alle $x > 0$

$$P\left(\sup_{j \leq n} |S_j| > x\right) \leq n \exp\left(-\frac{x^\gamma}{C_1}\right) + \exp\left(-\frac{x^2}{C_2(1 + nC_3)}\right) + \exp\left(-\frac{x^2}{C_4 n} \exp\left(\frac{x^{\gamma(1-\gamma)}}{C_5(\log x)^\gamma}\right)\right)$$

gilt, wobei $S_n := X_1 + \dots + X_n$.

Beweis. Für den Beweis siehe Merlevède et al. (2009). □

B. Lineare Algebra

B.1. Spureigenschaften

Lemma B.1.1 *Wir betrachten Matrizen $\mathbf{A}$ und $\mathbf{B}$ in den Dimensionen, sodass die folgenden Matrixoperationen jeweils definiert sind. Dann gelten folgende Spureigenschaften (vgl. Kollo und von Rosen (2005, Prop. 1.1.4 und 1.2.3), Bernstein (2009, Lemma 8.4.12)):*

$$\text{tr}\{\mathbf{AB}\} = \text{tr}\{\mathbf{BA}\} \quad (\text{B.1.1})$$

$$\text{tr}\{a\mathbf{A} + b\mathbf{B}\} = a \cdot \text{tr}\{\mathbf{A}\} + b \cdot \text{tr}\{\mathbf{B}\} \quad a, b \in \mathbb{R} \quad (\text{B.1.2})$$

$$\text{tr}\{\mathbf{A}\} = \text{rg}(\mathbf{A}), \text{ falls } \mathbf{A} \text{ idempotent} \quad (\text{B.1.3})$$

$$\text{tr}\{\mathbf{A}^T \mathbf{A}\} = 0 \iff \mathbf{A} = 0 \quad (\text{B.1.4})$$

$$\text{tr}\{\mathbf{A}^T\} = \text{tr}\{\mathbf{A}\} \quad (\text{B.1.5})$$

$$\text{tr}\{\mathbf{AA}^T\} = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2, \quad \mathbf{A} \in \mathbb{R}^{m \times n} \quad (\text{B.1.6})$$

$$\mathbf{x}^T \mathbf{Ax} = \text{tr}\{\mathbf{Axx}^T\} \quad (\text{B.1.7})$$

$$\text{tr}\{\mathbf{A}^k\} = \sum_{i=1}^n \lambda_i^k(\mathbf{A}) \text{ f\"ur } \mathbf{A} \in \mathbb{R}^{n \times n}. \quad (\text{B.1.8})$$

F\"ur symmetrische Matrizen $\mathbf{A}, \mathbf{B}, \mathbf{C}$, mit $\mathbf{A} \leq \mathbf{C}$ und $\mathbf{C} \geq 0$ gilt

$$\text{tr}\{\mathbf{AC}\} \leq \text{tr}\{\mathbf{BC}\}. \quad (\text{B.1.9})$$

B.2. Normen

Definition B.2.1 (Frobeniusnorm) F\"ur eine Matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{q \times p}$ ist die Frobeniusnorm definiert als

$$\|\mathbf{A}\| = [\text{tr}\{\mathbf{AA}^T\}]^{\frac{1}{2}} = \left[\sum_{i=1}^q \sum_{j=1}^p |a_{ij}|^2 \right]^{\frac{1}{2}}. \quad (\text{B.2.1})$$

Falls $\mathbf{A}$ eine symmetrische $p \times p$ Matrix ist, so gilt

$$\|\mathbf{A}\| = \left[\sum_{i=1}^p \lambda_i(\mathbf{A})^2 \right]^{\frac{1}{2}}. \quad (\text{B.2.2})$$

Diese Identit\"at folgt direkt durch Einsetzen von (B.1.8) in die Definition.

Eigenschaften der Frobeniusnorm

Lemma B.2.2 Für eine $q \times r$ Matrix $\mathbf{A}_1$ und eine $r \times q$ Matrix $\mathbf{A}_2$ gilt

$$|\operatorname{tr}\{\mathbf{A}_1\mathbf{A}_2\}| \leq \|\mathbf{A}_1\| \|\mathbf{A}_2\| \text{ und } \|\mathbf{A}_1\mathbf{A}_2\| \leq \|\mathbf{A}_1\| \|\mathbf{A}_2\|. \quad (\text{B.2.3})$$

Insbesondere gilt für eine $q \times r$ Matrix $\mathbf{A}_1$ und eine $r \times r$ Matrix $\mathbf{A}_2$

$$|\operatorname{tr}\{\mathbf{A}_1\mathbf{A}_2\mathbf{A}_1^T\}| \leq \|\mathbf{A}_1^T\mathbf{A}_1\| \|\mathbf{A}_2\| \text{ und } \|\mathbf{A}_1\mathbf{A}_2\mathbf{A}_1^T\| \leq \|\mathbf{A}_1^T\mathbf{A}_1\| \|\mathbf{A}_2\|. \quad (\text{B.2.4})$$

Beweis. Nach (B.1.1), (B.1.2), (B.1.4) und (B.1.6) können wir durch $\langle \mathbf{A}_1, \mathbf{A}_2 \rangle := \operatorname{tr}\{\mathbf{A}_1\mathbf{A}_2^T\}$ ein Skalarprodukt definieren, für das dann auch die Cauchy–Schwarz-Ungleichung gilt. Dann folgt mit (B.1.1)

$$\begin{aligned} |\operatorname{tr}\{\mathbf{A}_1\mathbf{A}_2\}| &\leq [\operatorname{tr}\{\mathbf{A}_1^T\mathbf{A}_1\}]^{\frac{1}{2}} [\operatorname{tr}\{\mathbf{A}_2^T\mathbf{A}_2\}]^{\frac{1}{2}} = [\operatorname{tr}\{\mathbf{A}_1\mathbf{A}_1^T\}]^{\frac{1}{2}} [\operatorname{tr}\{\mathbf{A}_2\mathbf{A}_2^T\}]^{\frac{1}{2}} \\ &= \|\mathbf{A}_1\| \|\mathbf{A}_2\|. \end{aligned}$$

Weiterhin gilt wegen der Submultiplikativität der Frobeniusnorm (Vgl. Bernstein, 2009, Prop. 9.3.5.)

$$\|\mathbf{A}_1\mathbf{A}_2\| \leq \|\mathbf{A}_1\| \|\mathbf{A}_2\|.$$

Insbesondere folgen damit auch

$$|\operatorname{tr}\{\mathbf{A}_1\mathbf{A}_2\mathbf{A}_1^T\}| = |\operatorname{tr}\{\mathbf{A}_1^T\mathbf{A}_1\mathbf{A}_2\}| \leq \|\mathbf{A}_1^T\mathbf{A}_1\| \|\mathbf{A}_2\|$$

und

$$\begin{aligned} \|\mathbf{A}_1\mathbf{A}_2\mathbf{A}_1^T\| &= [\operatorname{tr}\{\mathbf{A}_1\mathbf{A}_2\mathbf{A}_1^T\mathbf{A}_1\mathbf{A}_2^T\mathbf{A}_1^T\}]^{\frac{1}{2}} \stackrel{(\text{B.2.4})}{\leq} \|\mathbf{A}_2^T\mathbf{A}_1^T\mathbf{A}_1\mathbf{A}_2\|^{\frac{1}{2}} \|\mathbf{A}_1^T\mathbf{A}_1\| \\ &\stackrel{(\text{B.2.4})}{\leq} \|\mathbf{A}_2\mathbf{A}_2^T\|^{\frac{1}{2}} \|\mathbf{A}_1^T\mathbf{A}_1\|^{\frac{1}{2}} \|\mathbf{A}_1^T\mathbf{A}_1\|^{\frac{1}{2}} \stackrel{(\text{B.2.3})}{\leq} \|\mathbf{A}_2\|^{\frac{1}{2}} \|\mathbf{A}_2^T\|^{\frac{1}{2}} \|\mathbf{A}_1^T\mathbf{A}_1\| \\ &\stackrel{(\text{B.1.5})}{=} \|\mathbf{A}_2\| \|\mathbf{A}_1^T\mathbf{A}_1\|. \end{aligned}$$

□

Definition B.2.3 (Spektralnrm) Für Matrizen $\mathbf{A} \in \mathbb{R}^{m \times n}$ ist die Spektralnrm definiert als

$$\|\mathbf{A}\|_S := [\lambda_{\max}(\mathbf{A}^T\mathbf{A})]^{\frac{1}{2}}.$$

Für symmetrische Matrizen ergibt sich daraus

$$\|\mathbf{A}\|_S = |\lambda_{\max}(\mathbf{A})|.$$

Definition B.2.4 (p-Normen) Für Vektoren $\mathbf{x} \in \mathbb{R}^n$ ist die p -Norm von $\mathbf{x}$ für $1 \leq p \leq \infty$ definiert als

$$\|\mathbf{x}\|_p := \begin{cases} [\sum_{i=1}^n |x_i|^p]^{\frac{1}{p}} & 1 \leq p < \infty \\ \max_{1 \leq i \leq n} |x_i| & p = \infty. \end{cases}$$

Beziehung zwischen der Frobeniusnorm und der Spektralnorm

Zwischen der Frobeniusnorm und der Spektralnorm gilt für zwei Matrizen $\mathbf{A}_1 \in \mathbb{R}^{n \times m}$ und $\mathbf{A}_2 \in \mathbb{R}^{m \times l}$ folgende Beziehung:

$$\|\mathbf{A}_1 \mathbf{A}_2\|_S \leq \|\mathbf{A}_1 \mathbf{A}_2\| \leq \|\mathbf{A}_1\|_S \|\mathbf{A}_2\| \leq \|\mathbf{A}_1\| \|\mathbf{A}_2\|. \quad (\text{B.2.5})$$

(Vgl. Bernstein, 2009, Cor. 9.3.7.)

Für die Spur eines Matrixprodukts von $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{n \times n}$, wobei $\mathbf{A}_1$ symmetrisch und $\mathbf{A}_2$ positiv semidefinit sei, gilt

$$\text{tr}\{\mathbf{A}_1 \mathbf{A}_2\} \leq \lambda_{\max}(\mathbf{A}_1) \text{tr}\{\mathbf{A}_2\} = \|\mathbf{A}_1\|_S \text{tr}\{\mathbf{A}_2\}. \quad (\text{B.2.6})$$

(Vgl. Bernstein, 2009, Fact 5.12.7)

Eigenschaften der p -Normen

Für $1 \leq p \leq q \leq \infty$ gilt

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_q \leq \|\mathbf{x}\|_p \leq \|\mathbf{x}\|_1. \quad (\text{B.2.7})$$

(Vgl. Bernstein, 2009, S.544f.)

B.3. Matrixtheorie

Sherman-Morrison-Woodbury Formel

Für Matrizen $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, $\mathbf{C} \in \mathbb{R}^{m \times n}$ und $\mathbf{D} \in \mathbb{R}^{m \times m}$ gilt: Falls $\mathbf{A}$, $\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ und $\mathbf{D}$ nichtsingulär sind, dann ist $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ nichtsingulär und es gilt die Sherman-Morrison-Woodbury Formel

$$(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad (\text{B.3.1})$$

(Vgl. Bernstein, 2009, Cor. 2.8.8)

Gleichungen und Ungleichungen für Inverse Matrizen

Für $n \times n$ Matrizen $\mathbf{A}_1$ und $\mathbf{A}_2$, mit $\mathbf{A}_1 + \mathbf{A}_2$ nichtsingulär, gilt

$$\mathbf{A}_1(\mathbf{A}_1 + \mathbf{A}_2)^{-1}\mathbf{A}_2 = \mathbf{A}_2(\mathbf{A}_1 + \mathbf{A}_2)^{-1}\mathbf{A}_1 = \mathbf{A}_1 - \mathbf{A}_1(\mathbf{A}_1 + \mathbf{A}_2)^{-1}\mathbf{A}_1 = \mathbf{A}_2 - \mathbf{A}_2(\mathbf{A}_1 + \mathbf{A}_2)^{-1}\mathbf{A}_2. \quad (\text{B.3.2})$$

(Vgl. Bernstein, 2009, Fact 2.16.13)

Wir betrachten nun zwei Matrizen $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{n \times n}$. Falls $\mathbf{A}_1$ nichtsingulär ist und $\|\mathbf{A}_1^{-1}\mathbf{A}_2\| < 1$ gilt, dann ist $\mathbf{A}_1 + \mathbf{A}_2$ nichtsingulär und es gilt

$$\|(\mathbf{A}_1 + \mathbf{A}_2)^{-1} - \mathbf{A}_1^{-1}\| \leq \frac{\|\mathbf{A}_2\| \|\mathbf{A}_1^{-1}\|^2}{1 - \|\mathbf{A}_1^{-1}\mathbf{A}_2\|}. \quad (\text{B.3.3})$$

(Vgl. Golub und Van Loan, 1989, Theorem 2.3.4).

Insbesondere kann hierbei auch die Abschätzung durch die Spektralnorm nach (B.2.5) genutzt werden, sodass folgt

$$\|(\mathbf{A}_1 + \mathbf{A}_2)^{-1} - \mathbf{A}_1^{-1}\| \leq \frac{\|\mathbf{A}_2\| \|\mathbf{A}_1^{-1}\|_S^2}{1 - \|\mathbf{A}_1^{-1}\|_S \|\mathbf{A}_2\|}. \quad (\text{B.3.4})$$

B.4. Eigenwerte

Ungleichungen

Sei $\mathbf{A}$ eine $n \times n$ Matrix mit den Eigenwerten $\lambda_1(\mathbf{A}), \dots, \lambda_n(\mathbf{A})$. Falls $\mathbf{A}$ symmetrisch ist, gilt

$$\lambda_{\min}(\mathbf{A})\mathbf{I}_n \leq \mathbf{A} \leq \lambda_{\max}(\mathbf{A})\mathbf{I}_n. \quad (\text{B.4.1})$$

(Vgl. Bernstein, 2009, Cor. 8.4.2).

Für symmetrische $n \times n$ Matrizen $\mathbf{A}_1$ und $\mathbf{A}_2$ mit $\mathbf{A}_1 \leq \mathbf{A}_2$ gilt für $i = 1, \dots, n$:

$$\lambda_i(\mathbf{A}_1) \leq \lambda_i(\mathbf{A}_2). \quad (\text{B.4.2})$$

(Vgl. Bernstein, 2009, Theorem 8.4.9).

B.5. Hutmatrix

Für die Hutmatrix aus dem Linearen Modell gilt folgendes Lemma

Lemma B.5.1 *Die Hutmatrix $\mathbf{H}$ ist fast sicher idempotent mit*

$$\text{tr}\{\mathbf{H}^2\} = \text{tr}\{\mathbf{H}\} = K \quad (\text{B.5.1})$$

und es gilt

$$\|\mathbf{H}\|_S = O(1). \quad (\text{B.5.2})$$

Beweis. $\mathbf{H}$ ist idempotent, da

$$\mathbf{H} = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}$$

und

$$\mathbf{H}^2 = \mathbf{H}\mathbf{H} = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X} = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X} = \mathbf{H}$$

gilt. Für die Spur der Hutmatrix folgt mit (B.1.1)

$$\text{tr}\{\mathbf{H}\} = \text{tr}\{\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\} = \text{tr}\{\mathbf{X}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\} = \text{tr}\{\mathbf{I}_K\} = K.$$

Die Eigenwerte von symmetrischen idempotenten Matrizen nehmen nur die Werte 0 und 1 an (vgl. Kollo und von Rosen, 2005, Prop. 1.2.3). Daher gilt für die Hutmatrix die Aussage (B.5.2). $\square$

C. Herleitung der Konvergenzraten für Abschnitt 2.3

Wir beweisen nun die Konvergenzraten der einzelnen Komponenten des Faktormodells und der Zerlegungen (2.2.13) und (2.2.14), die für die Gesamtab-schätzungen in Abschnitt 2.3 nötig sind. Wir beginnen mit dem Schätzfehler der Kovarianzmatrixschätzung für die Faktoren.

Lemma C.0.2 *Für den Schätzfehler $\mathbf{D}_n$ der Stichprobenkovarianzmatrix der Faktoren $\mathbf{f}$ gilt unter den Annahmen 2, 6 und 8*

$$E(\|\mathbf{D}_n\|^2) = O(n^{-1}K^2).$$

Beweis. Nach Annahme 8 sind die $\mathbf{f}_1, \dots, \mathbf{f}_n$ unabhängig identisch verteilt und nach Annahme 2 besitzen die Komponenten von $\mathbf{f}$ gleichmäßig beschränkte vierte Momente. Dann folgt das Resultat durch Anwendung von Lemma A.1.1. $\square$

Mithilfe dieser Abschätzung betrachten wir nun den Term $\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T$ aus (2.2.13) und (2.2.14) in den beiden Normen (B.2.1) und (2.2.4).

Lemma C.0.3 *Unter den Annahmen 1 - 6 und 8 gelten folgende Aussagen:*

$$E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|^2) = O(n^{-1}p^2K^2) \tag{C.0.1}$$

und

$$E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|_{\Sigma_n}^2) = O(n^{-1}p^{-1}K^2). \tag{C.0.2}$$

Beweis. Für die Abschätzung in der Frobeniusnorm genügt es, die einzelnen Komponenten abzuschätzen.

Wir nutzen dazu (2.2.3), die Eigenschaft (B.2.3) und die Spureigenschaften aus Kapitel B.1:

$$\|\mathbf{B}_n^T \mathbf{B}_n\| = \|\mathbf{B}_n \mathbf{B}_n^T\| \leq \|\mathbf{B}_n\| \|\mathbf{B}_n^T\| = \|\mathbf{B}_n\|^2 = O(p). \tag{C.0.3}$$

Mit diesem Resultat und Lemma C.0.2 schätzen wir (C.0.1) ab durch

$$\begin{aligned} E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|^2) &= E(\text{tr}\{[\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T]^2\}) \stackrel{(B.2.4)}{\leq} E(\|\mathbf{B}_n^T \mathbf{B}_n\|^2 \|\mathbf{D}_n\|^2) \\ &= \|\mathbf{B}_n^T \mathbf{B}_n\|^2 E(\|\mathbf{D}_n\|^2) = O(p^2) O(n^{-1}K^2) = O(n^{-1}p^2K^2). \end{aligned}$$

Für die zweite Aussage des Lemmas nutzen wir eine Abschätzung für $\|\mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\|_S$ und die Abschätzung für den Schätzfehler der Kovarianzmatrix der Faktoren aus Lemma C.0.2.

Für die Inverse der Kovarianzmatrix nutzen wir die Darstellung aus (2.2.15):

$$\Sigma_n^{-1} = \Sigma_0^{-1} - \Sigma_0^{-1} \mathbf{B}_n [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1} \mathbf{B}_n^T \Sigma_0^{-1}.$$

Nach (B.3.2) gilt mit $\mathbf{A}_1 := \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n$ und $\mathbf{A}_2 := \text{cov}(\mathbf{f})^{-1}$:

$$\begin{aligned} \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n &= \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n - \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n]^{-1} \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n \\ &= \text{cov}(\mathbf{f})^{-1} - \text{cov}(\mathbf{f})^{-1} [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n]^{-1} \text{cov}(\mathbf{f})^{-1} \end{aligned}$$

und daher folgt mit der Dreiecksungleichung

$$\|\mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n\|_S \leq \|\text{cov}(\mathbf{f})^{-1}\|_S + \|\text{cov}(\mathbf{f})^{-1} [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n]^{-1} \text{cov}(\mathbf{f})^{-1}\|_S. \quad (\text{C.0.4})$$

Die Matrix $\text{cov}(\mathbf{f})^{-1}$ ist symmetrisch positiv definit und $\mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n$ ist symmetrisch positiv semidefinit. Somit gilt

$$\begin{aligned} \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n &= \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n + \text{cov}(\mathbf{f})^{-1} - \text{cov}(\mathbf{f})^{-1} \geq 0 \\ &\Rightarrow \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n + \text{cov}(\mathbf{f})^{-1} \geq \text{cov}(\mathbf{f})^{-1} \\ &\Rightarrow [\mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n + \text{cov}(\mathbf{f})^{-1}]^{-1} \leq \text{cov}(\mathbf{f}). \end{aligned}$$

Insbesondere folgt also

$$\text{cov}(\mathbf{f})^{-1} [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n]^{-1} \text{cov}(\mathbf{f})^{-1} \leq \text{cov}(\mathbf{f})^{-1} \text{cov}(\mathbf{f}) \text{cov}(\mathbf{f})^{-1} = \text{cov}(\mathbf{f})^{-1}$$

und mit (B.4.2)

$$\|\text{cov}(\mathbf{f})^{-1} [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n]^{-1} \text{cov}(\mathbf{f})^{-1}\|_S \leq \|\text{cov}(\mathbf{f})^{-1}\|_S.$$

Damit genügt es für die Abschätzung von (C.0.4), den Ausdruck $\|\text{cov}(\mathbf{f})^{-1}\|_S$ zu betrachten. Nach (2.2.6) gilt $\|\text{cov}(\mathbf{f})^{-1}\|_S = O(1)$ und damit folgt direkt

$$\|\mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n\|_S = O(1). \quad (\text{C.0.5})$$

Für (C.0.2) gilt damit die Abschätzung

$$\begin{aligned} E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|_{\boldsymbol{\Sigma}_n}^2) &= \frac{1}{p} E(\|\boldsymbol{\Sigma}_n^{-\frac{1}{2}} \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-\frac{1}{2}}\|^2) \\ &= \frac{1}{p} E(\text{tr}\{\boldsymbol{\Sigma}_n^{-\frac{1}{2}} \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-\frac{1}{2}} \boldsymbol{\Sigma}_n^{-\frac{1}{2}} \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-\frac{1}{2}}\}) \\ &= \frac{1}{p} E(\text{tr}\{\boldsymbol{\Sigma}_n^{-\frac{1}{2}} \boldsymbol{\Sigma}_n^{-\frac{1}{2}} \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\}) = \frac{1}{p} E(\text{tr}\{[\mathbf{D}_n \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n]^2\}) \\ &\stackrel{(\text{B.2.3})}{\leq} \frac{1}{p} E(\|\mathbf{D}_n \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n\|^2) \stackrel{(\text{B.2.5})}{\leq} \frac{1}{p} E(\|\mathbf{D}_n\|^2) \|\mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n\|_S^2. \end{aligned}$$

Dann folgt mit Lemma C.0.2 und (C.0.5)

$$E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|_{\boldsymbol{\Sigma}_n}^2) = \frac{1}{p} (O(n^{-1} K^2) O(1)) = O(n^{-1} p^{-1} K^2).$$

□

Lemma C.0.4 *Unter den Annahmen 1 - 6 und 8 gelten*

$$E(\|\mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T\|^2) = O(n^{-2} p^2 K) \quad (\text{C.0.6})$$

und

$$E(\|\mathbf{C}_n \hat{\text{cov}}(\mathbf{f}) \mathbf{C}_n^T\|_{\boldsymbol{\Sigma}_n}^2) = O(n^{-2} p K). \quad (\text{C.0.7})$$

Beweis. Für diesen Beweis ist es sinnvoll, statt der einzelnen Komponenten den abzuschätzenden Ausdruck als Produkt zu betrachten. Wegen der Darstellungen nach (2.2.9) und (2.1.3) vereinfacht sich die Darstellung und wir zeigen, dass für die beiden resultierenden Summanden die gleiche Konvergenzrate gilt. Wir nutzen die Cauchy-Schwarz-Ungleichung und schreiben

$$\begin{aligned}
& E(\text{tr}\{[\mathbf{C}_n \text{cov}(\mathbf{f}) \mathbf{C}_n^T]^2\}) \\
&= E(\text{tr}\{[\mathbf{E} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} (\frac{1}{n-1} \mathbf{X} \mathbf{X}^T - \frac{1}{n(n-1)} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T) (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{E}^T]^2\}) \\
&= E(\text{tr}\{[\frac{1}{n-1} \mathbf{E} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{E}^T \\
&\quad - \frac{1}{n(n-1)} \mathbf{E} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{E}^T]^2\}) \\
&= E(\text{tr}\{[\frac{1}{n-1} \mathbf{E} \mathbf{H} \mathbf{E}^T - \frac{1}{n(n-1)} \mathbf{E} \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T]^2\}) \\
&\leq 2E(\text{tr}\{\frac{1}{(n-1)^2} [\mathbf{E} \mathbf{H} \mathbf{E}^T]^2 + \frac{1}{n^2(n-1)^2} [\mathbf{E} \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T]^2\}) \\
&= \frac{2}{(n-1)^2} E(\|\mathbf{E} \mathbf{H} \mathbf{E}^T\|^2) + \frac{2}{n^2(n-1)^2} E(\|\mathbf{E} \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T\|^2) =: \frac{2}{(n-1)^2} \mathcal{A}_1 + \frac{2}{n^2(n-1)^2} \mathcal{A}_2.
\end{aligned}$$

Wir betrachten die Matrix von $\mathcal{A}_1$ und schreiben ihre Norm geeignet in vier Summanden um, um diese jeweils einzeln abzuschätzen:

$$\begin{aligned}
\mathbf{E} \mathbf{H} \mathbf{E}^T &= (\varepsilon_1, \dots, \varepsilon_n) \begin{pmatrix} H_{11} & \dots & H_{1n} \\ \vdots & \ddots & \vdots \\ H_{n1} & \dots & H_{nn} \end{pmatrix} \begin{pmatrix} \varepsilon_1^T \\ \vdots \\ \varepsilon_n^T \end{pmatrix} = (\varepsilon_1, \dots, \varepsilon_n) \begin{pmatrix} \sum_{j=1}^n H_{1j} \varepsilon_j^T \\ \vdots \\ \sum_{j=1}^n H_{nj} \varepsilon_j^T \end{pmatrix} \\
&= \sum_{i=1}^n \sum_{j=1}^n H_{ij} \varepsilon_i \varepsilon_j^T. \tag{C.0.8}
\end{aligned}$$

Für $\mathcal{A}_1$ schreiben wir durch Anwendung der iterierten Erwartung

$$\begin{aligned}
\mathcal{A}_1 &= E(\|\mathbf{E} \mathbf{H} \mathbf{E}^T\|^2) = E(\text{tr}\{[\mathbf{E} \mathbf{H} \mathbf{E}^T]^2\}) \stackrel{(C.0.8)}{=} E(\text{tr}\{[\sum_{i=1}^n \sum_{j=1}^n H_{ij} \varepsilon_i \varepsilon_j^T]^2\}) \\
&= E(\text{tr}\{\sum_{i=1}^n \sum_{j=1}^n H_{ij} \varepsilon_i \varepsilon_j^T \sum_{k=1}^n \sum_{l=1}^n H_{kl} \varepsilon_k \varepsilon_l^T\}) \\
&= E(E(\text{tr}\{\sum_{i=1}^n \sum_{j=1}^n H_{ij} \varepsilon_i \varepsilon_j^T \sum_{k=1}^n \sum_{l=1}^n H_{kl} \varepsilon_k \varepsilon_l^T\} | \mathbf{X})) \\
&= E(\text{tr}\{E(\sum_{i=1}^n \sum_{j=1}^n H_{ij} \varepsilon_i \varepsilon_j^T \sum_{k=1}^n \sum_{l=1}^n H_{kl} \varepsilon_k \varepsilon_l^T | \mathbf{X})\}).
\end{aligned}$$

Die Summanden dieser Darstellung können wir je nach Indices in fünf Mengen einteilen:

1. $i = j = k = l$: $\{E(H_{ii}^2 [\varepsilon_i \varepsilon_i^T]^2 | \mathbf{X}), i = 1, \dots, n\}$,
2. $i = j, k = l, i \neq k$: $\{E(H_{ii} H_{jj} \varepsilon_i \varepsilon_i^T \varepsilon_j \varepsilon_j^T | \mathbf{X}), i, k = 1, \dots, n, i \neq k\}$
3. $i = k, j = l, i \neq j$: $\{E(H_{ij}^2 [\varepsilon_i \varepsilon_j^T]^2 | \mathbf{X}), i, j = 1, \dots, n, i \neq j\}$

4. $i = l, j = k, i \neq j$: $\{E(H_{ij}H_{ji}\epsilon_i\epsilon_j^T|\mathbf{X}), i, j = 1, \dots, n, i \neq j\}$

5. $i = j = k \neq l, i = j = l \neq k, i = k = l \neq j$ und $j = k = l \neq i$:

$$\begin{aligned} & \{E(H_{ii}H_{il}\epsilon_i\epsilon_l^T|\mathbf{X}), i, l = 1, \dots, n, i \neq l\} \\ & \cup \{E(H_{ii}H_{ki}\epsilon_i\epsilon_k^T|\mathbf{X}), i, k = 1, \dots, n, i \neq k\} \\ & \cup \{E(H_{ij}H_{ii}\epsilon_i\epsilon_j^T|\mathbf{X}), i, j = 1, \dots, n, i \neq j\} \\ & \cup \{E(H_{ij}H_{jj}\epsilon_i\epsilon_j^T|\mathbf{X}), i, j = 1, \dots, n, i \neq j\}. \end{aligned}$$

Nach Annahme 1 gilt $E(\epsilon|\mathbf{f}) = 0$, daher fallen alle Terme in 5. weg und wir erhalten folgende Zerlegung:

$$\begin{aligned} \mathcal{A}_1 &= E(\text{tr}\{\sum_{i=1}^n H_{ii}^2 E([\epsilon_i\epsilon_i^T]^2|\mathbf{X})\}) + E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ii}H_{jj} E(\epsilon_i\epsilon_i^T \epsilon_j\epsilon_j^T|\mathbf{X})\}) \\ &+ E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 E([\epsilon_i\epsilon_j^T]^2|\mathbf{X})\}) + E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}H_{ji} E(\epsilon_i\epsilon_j^T \epsilon_j\epsilon_i^T|\mathbf{X})\}) \\ &=: \mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3 + \mathcal{B}_4. \end{aligned} \tag{C.0.9}$$

Diese vier Summanden schätzen wir nun nacheinander ab. Die Konvergenzraten folgen jeweils aus der Eigenschaft der Spur der Hutmatrix nach (B.5.1) und den gleichmäßig beschränkten vierten Momenten der Fehler.

Für den ersten Summanden betrachten mithilfe von (B.5.1) und der Symmetrie den folgenden Zusammenhang für die Hutmatrix:

$$K \stackrel{\text{(B.5.1)}}{=} \text{tr}\{\mathbf{H}^2\} \stackrel{\text{Symmetrie}}{=} \sum_{i=1}^n \sum_{j=1}^n H_{ij}^2 = \sum_{i=1}^n H_{ii}^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 \geq \sum_{i=1}^n H_{ii}^2. \tag{C.0.10}$$

Damit können wir $\mathcal{B}_1$ wie folgt abschätzen: Wegen der unabhängig identisch verteilten Fehlerterme gilt

$$\begin{aligned} \mathcal{B}_1 &= E(\text{tr}\{\sum_{i=1}^n H_{ii}^2 E([\epsilon_i\epsilon_i^T]^2|\mathbf{X})\}) = E(\text{tr}\{\sum_{i=1}^n H_{ii}^2 E([\epsilon\epsilon^T]^2|\mathbf{X})\}) \\ &\stackrel{\text{(C.0.10)}}{\leq} KE(E(\text{tr}\{[\epsilon\epsilon^T]^2|\mathbf{X})\}) = KE(\sum_{i=1}^p \sum_{j=1}^p \epsilon_i^2 \epsilon_j^2) \end{aligned}$$

und wegen der gleichmäßig beschränkten zweiten und vierten Momente der Fehler nach Annahme 2 folgt die Dominanz der Rate des zweiten Summanden in der folgenden Darstellung

$$\mathcal{B}_1 \leq K \sum_{i=1}^p E(\epsilon_i^4) + K \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p E(\epsilon_i^2) E(\epsilon_j^2) = O(p^2 K). \tag{C.0.11}$$

Für die Abschätzung des zweiten Summanden aus (C.0.9) benötigen wir folgenden Zusammenhang aus (B.5.1)

$$K^2 = [\text{tr}\{\mathbf{H}\}]^2 = [H_{11} + \dots + H_{nn}]^2 = \sum_{i=1}^n H_{ii}^2 + \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n H_{jj} H_{kk} \geq \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n H_{jj} H_{kk}. \quad (\text{C.0.12})$$

Für $\mathcal{B}_2$ folgt damit wegen der unabhängig identisch verteilten Fehlerterme

$$\begin{aligned} \mathcal{B}_2 &= E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ii} H_{jj} E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j^T | \mathbf{X})\}) = E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ii} H_{jj} E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T | \mathbf{X}) E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T | \mathbf{X})\}) \\ &\stackrel{(\text{C.0.12})}{\leq} K^2 E(\text{tr}\{E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T | \mathbf{X}) E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T | \mathbf{X})\}) = K^2 E(\text{tr}\{\boldsymbol{\Sigma}_0^2\}) \end{aligned}$$

und da die zweiten Momente der Fehler nach Annahme 2 gleichmäßig beschränkt sind, folgt

$$\mathcal{B}_2 \leq K^2 \sum_{i=1}^p [E(\varepsilon_i^2)]^2 = O(pK^2). \quad (\text{C.0.13})$$

Für die Abschätzung des dritten Summanden aus (C.0.9) betrachten wir die erneut die Rechnung aus (C.0.10)

$$\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 \leq \sum_{i=1}^n H_{ii}^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 = K. \quad (\text{C.0.14})$$

Sei $\boldsymbol{\eta} \in \mathbb{R}^p$ eine unabhängige Ausprägung von $\boldsymbol{\varepsilon}$. Dann folgt für $\mathcal{B}_3$ aufgrund der unabhängig identisch verteilten Fehlerterme

$$\begin{aligned} \mathcal{B}_3 &= E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 E([\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j^T]^2 | \mathbf{X})\}) = E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 E([\boldsymbol{\varepsilon} \boldsymbol{\eta}^T]^2 | \mathbf{X})\}) \\ &\stackrel{(\text{C.0.14})}{\leq} K E(E(\text{tr}\{[\boldsymbol{\varepsilon} \boldsymbol{\eta}^T]^2 | \mathbf{X})\}) = K E(\sum_{i=1}^p \sum_{j=1}^p \varepsilon_i \eta_i \varepsilon_j \eta_j) \end{aligned}$$

und mit den nach Annahme 2 gleichmäßig beschränkten zweiten Momenten folgt

$$\mathcal{B}_3 \leq K \sum_{i=1}^p \sum_{j=1}^p \sqrt{E(\varepsilon_i^2) E(\eta_i^2) E(\varepsilon_j^2) E(\eta_j^2)} = O(p^2 K). \quad (\text{C.0.15})$$

Für den letzten Summanden aus (C.0.9) benötigen wir folgende Abschätzung, die wir mithilfe der Symmetrie von $\mathbf{H}$ und (C.0.10) herleiten:

$$\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij} H_{ji} = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 \leq \sum_{i=1}^n H_{ii}^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 = K. \quad (\text{C.0.16})$$

Damit folgt für $\mathcal{B}_4$ aufgrund der unabhängig identisch verteilten Fehlerterme:

$$\mathcal{B}_4 = E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij} H_{ji} E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_i^T | \mathbf{X})\}) = E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij} H_{ji} E(\boldsymbol{\varepsilon} \boldsymbol{\eta}^T \boldsymbol{\eta} \boldsymbol{\varepsilon}^T | \mathbf{X})\})$$

$$\stackrel{(C.0.16)}{\leq} KE(E(\text{tr}\{(\boldsymbol{\varepsilon}\boldsymbol{\eta}^T\boldsymbol{\eta}\boldsymbol{\varepsilon}^T)|\mathbf{X})\}) = KE(\text{tr}\{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\boldsymbol{\eta}^T\boldsymbol{\eta}\}),$$

da es sich bei diesem Ausdruck um ein Skalar handelt, verzichten wir auf die Betrachtung der Spur und schreiben

$$\mathcal{B}_4 \leq KE(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\boldsymbol{\eta}^T\boldsymbol{\eta}) = KE([\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}]^2) = KE([\sum_{i=1}^p \varepsilon_i^2]^2) = K(\sum_{i=1}^p E(\varepsilon_i^4) + \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p E(\varepsilon_i^2)E(\varepsilon_j^2)).$$

Wegen der gleichmäßig beschränkten zweiten und vierten Momente nach Annahme 2 folgt dann die Dominanz der Rate des zweiten Summanden, also

$$\mathcal{B}_4 = O(p^2 K). \quad (C.0.17)$$

Insgesamt erhalten wir aus (C.0.11), (C.0.13), (C.0.15) und (C.0.17) folgende Darstellung, in der nach Annahme 6 ($p \geq K$) die Rate von $\mathcal{B}_1$ bzw. $\mathcal{B}_3$ oder $\mathcal{B}_4$ dominiert:

$$\mathcal{A}_1 = E(\|\mathbf{E}\mathbf{H}\mathbf{E}^T\|^2) = \mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3 + \mathcal{B}_4 = O(p^2 K) + O(pK^2) + O(p^2 K) + O(p^2 K) = O(p^2 K).$$

Dieses Resultat nutzen wir auch für die Abschätzung des zweiten Ausdrucks, denn da $\mathbf{H}$ idempotent ist, erhalten wir mit (B.2.4) wieder den Ausdruck $\mathcal{A}_1$:

$$\begin{aligned} \mathcal{A}_2 &= E(\|\mathbf{E}\mathbf{H}\mathbf{1}\mathbf{1}^T\mathbf{H}\mathbf{E}^T\|^2) \leq E(\|\mathbf{H}\mathbf{E}^T\mathbf{E}\mathbf{H}\|^2\|\mathbf{1}\mathbf{1}^T\|^2) = E(\|\mathbf{E}\mathbf{H}\mathbf{H}\mathbf{E}^T\|^2\|\mathbf{1}\mathbf{1}^T\|^2), \\ &= E(\|\mathbf{E}\mathbf{H}\mathbf{E}^T\|^2)n^2 = O(n^2 p^2 K). \end{aligned}$$

Damit folgt in der Gesamtabeschätzung für beide Summanden die gleiche Rate und insgesamt für (C.0.6):

$$\begin{aligned} E(\text{tr}\{[\mathbf{C}_n \hat{\text{cov}}(\mathbf{f})\mathbf{C}_n^T]^2\}) &= \frac{2}{(n-1)^2} \mathcal{A}_1 + \frac{2}{n^2(n-1)^2} \mathcal{A}_2 \\ &= \frac{2}{(n-1)^2} O(p^2 K) + \frac{1}{n^2(n-1)^2} O(n^2 p^2 K) = O(n^{-2} p^2 K). \end{aligned}$$

Nun beweisen wir die zweite Aussage des Lemmas. Mithilfe von (2.2.5), (2.2.8) und (C.0.6) folgt

$$\begin{aligned} E(\|\mathbf{C}_n \hat{\text{cov}}(\mathbf{f})\mathbf{C}_n^T\|_{\boldsymbol{\Sigma}_n}^2) &\leq \frac{1}{p} \|\boldsymbol{\Sigma}_n^{-1}\|_S^2 E(\|\mathbf{C}_n \hat{\text{cov}}(\mathbf{f})\mathbf{C}_n^T\|^2) \\ &= \frac{1}{p} O(1) O(n^{-2} p^2 K) = O(n^{-2} p K). \end{aligned}$$

□

Lemma C.0.5 Für die Faktoren $\mathbf{f}$ gelten unter der Annahme 2 folgende Aussagen

$$E(\|\mathbf{f}\|^4) = O(K^2) \quad (C.0.18)$$

und

$$E(\|\mathbf{X}^T\mathbf{X}\|^2) = O(n^2 K^2). \quad (C.0.19)$$

Beweis. Mit der Jensenungleichung gilt

$$E(\|\mathbf{f}\|^4) = E(\|\mathbf{f}\|^2)^2 = E\left(\left[\sum_{i=1}^K f_i^2\right]^2\right) \leq K \sum_{i=1}^K [E(f_i^2)]^2 \leq K \sum_{i=1}^K E(f_i^4)$$

und daraus folgt

$$E(\|\mathbf{f}\|^4) = O(K^2).$$

Weiter gilt mit der unabhängig identisch verteilten Faktoren

$$E(\|\mathbf{X}\|^2) = E(\text{tr}\{\mathbf{X}\mathbf{X}^T\}) = E(\text{tr}\{\sum_{i=1}^n \mathbf{f}_i \mathbf{f}_i^T\}) = n \sum_{i=1}^K E(f_i^2) = O(nK). \quad (\text{C.0.20})$$

Daraus folgt insbesondere

$$E(\|\mathbf{X}^T \mathbf{X}\|^2) E(\|\mathbf{X}\|^4) = O(n^2 K^2).$$

□

Lemma C.0.6 *Unter den Annahmen 1 - 6 und 8 gilt*

$$E(\|\mathbf{B}_n \text{côv}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \text{côv}(\mathbf{f}) \mathbf{B}_n^T\|^2) = O(n^{-1} p^2 K) \quad (\text{C.0.21})$$

und

$$E(\|\mathbf{B}_n \text{côv}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \text{côv}(\mathbf{f}) \mathbf{B}_n^T\|_{\Sigma_n}^2) = O(n^{-1} K). \quad (\text{C.0.22})$$

Beweis. Für diesen Beweis betrachten wir wieder den abzuschätzenden Ausdruck als Produkt. Zunächst vereinfachen wir ihn durch Anwendung der Cauchy–Schwarz-Ungleichung zu

$$\begin{aligned} E(\|\mathbf{B}_n \text{côv}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \text{côv}(\mathbf{f}) \mathbf{B}_n^T\|^2) &= E(\text{tr}\{[\mathbf{B}_n \text{côv}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \text{côv}(\mathbf{f}) \mathbf{B}_n^T]^2\}) \\ &\leq 2E(\text{tr}\{[\mathbf{B}_n \text{côv}(\mathbf{f}) \mathbf{C}_n^T]^2 + [\mathbf{C}_n \text{côv}(\mathbf{f}) \mathbf{B}_n^T]^2\}) \stackrel{(\text{B.1.5})}{=} 4E(\text{tr}\{[\mathbf{B}_n \text{côv}(\mathbf{f}) \mathbf{C}_n^T]^2\}). \end{aligned}$$

Mithilfe der Darstellungen (2.1.3) und (2.2.9) vereinfachen wir diesen Ausdruck weiter und teilen ihn mithilfe der Cauchy–Schwarz-Ungleichung in zwei Summanden auf, für die wir identische Konvergenzraten herleiten:

$$\begin{aligned} E(\text{tr}\{[\mathbf{B}_n \text{côv}(\mathbf{f}) \mathbf{C}_n^T]^2\}) &= E(\text{tr}\{[\mathbf{B}_n (\frac{1}{n-1} \mathbf{X} \mathbf{X}^T - \frac{1}{n(n-1)} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T) (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{E}^T]^2\}) \\ &= E(\text{tr}\{[\frac{1}{n-1} \mathbf{B}_n \mathbf{X} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{E}^T - \frac{1}{n(n-1)} \mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{E}^T]^2\}) \\ &= E(\text{tr}\{[\frac{1}{n-1} \mathbf{B}_n \mathbf{X} \mathbf{E}^T - \frac{1}{n(n-1)} \mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T]^2\}) \\ &\leq 2E(\text{tr}\{\frac{1}{(n-1)^2} [\mathbf{B}_n \mathbf{X} \mathbf{E}^T]^2 + \frac{1}{n^2(n-1)^2} [\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T]^2\}) \\ &= \frac{2}{(n-1)^2} E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|^2) + \frac{2}{n^2(n-1)^2} E(\|\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T\|^2) =: \frac{2}{(n-1)^2} \mathcal{A}_1 + \frac{2}{n^2(n-1)^2} \mathcal{A}_2. \end{aligned}$$

Für die Abschätzung von $\mathcal{A}_1$ betrachten wir zunächst

$$\begin{aligned}
 E(\mathbf{E}^T \mathbf{E} | \mathbf{X}) &= E \left(\begin{pmatrix} \epsilon_1^T \\ \vdots \\ \epsilon_n^T \end{pmatrix} (\epsilon_1, \dots, \epsilon_n) | \mathbf{X} \right) \\
 &= E \left(\begin{pmatrix} \sum_{i=1}^p \epsilon_{i1}^2 & \dots & \sum_{i=1}^p \epsilon_{i1} \epsilon_{in} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^p \epsilon_{in} \epsilon_{i1} & \dots & \sum_{i=1}^p \epsilon_{in}^2 \end{pmatrix} | \mathbf{X} \right) \\
 &= \begin{pmatrix} E(\epsilon_{11}^2 + \dots + \epsilon_{p1}^2 | \mathbf{X}) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & E(\epsilon_{1n}^2 + \dots + \epsilon_{pn}^2 | \mathbf{X}) \end{pmatrix} = \text{tr}\{\mathbf{\Sigma}_0\} \cdot \mathbf{I}_n, \quad (\text{C.0.23})
 \end{aligned}$$

da

$$\text{tr}\{\mathbf{\Sigma}_0\} = E(\epsilon_1^2 | \mathbf{f}) + \dots + E(\epsilon_p^2 | \mathbf{f}).$$

Für die Kovarianzmatrix der Fehler gilt mithilfe der Jensenungleichung

$$\begin{aligned}
 E([\text{tr}\{\mathbf{\Sigma}_0\}]^2) &= E([\sum_{i=1}^p E(\epsilon_i^2 | \mathbf{f})]^2) = E(\sum_{i=1}^p \sum_{j=1}^p E(\epsilon_i^2 | \mathbf{f}) E(\epsilon_j^2 | \mathbf{f})) \\
 &= \sum_{i=1}^p E([E(\epsilon_i^2 | \mathbf{f})]^2) + \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p E(E(\epsilon_i^2 | \mathbf{f})) E(E(\epsilon_j^2 | \mathbf{f})) \\
 &\leq \sum_{i=1}^p E(\epsilon_i^4) + \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p E(\epsilon_i^2) E(\epsilon_j^2).
 \end{aligned}$$

Mit Annahme 2 folgt

$$E([\text{tr}\{\mathbf{\Sigma}_0\}]^2) = O(p) + O(p^2) = O(p^2).$$

Dann folgt auch die stochastische Konvergenz mit der Rate

$$\text{tr}\{\mathbf{\Sigma}_0\} = O_P(p). \quad (\text{C.0.24})$$

Neben der Rate für die Fehlerkovarianzmatrix folgt die Konvergenzrate von $\mathcal{A}_1$ aus den Faktoren in der Darstellung (C.0.19) und der Matrix der Faktorladungen. Durch Anwendung der iterierten Erwartung zeigen wir folgende Abschätzung

$$\begin{aligned}
 \mathcal{A}_1 &= E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|^2) = E(\text{tr}\{\mathbf{B}_n \mathbf{X} \mathbf{E}^T \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T\}) = E(E(\text{tr}\{\mathbf{B}_n \mathbf{X} \mathbf{E}^T \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T\} | \mathbf{X})) \\
 &= E(\text{tr}\{\mathbf{B}_n \mathbf{X} E(\mathbf{E}^T \mathbf{E} | \mathbf{X}) \mathbf{X}^T \mathbf{B}_n^T\}) \stackrel{(\text{C.0.23})}{=} E(\text{tr}\{\mathbf{B}_n \mathbf{X} \text{tr}\{\mathbf{\Sigma}_0\} \mathbf{I}_n \mathbf{X}^T \mathbf{B}_n^T\}) \\
 &= \text{tr}\{\mathbf{\Sigma}_0\} E(\text{tr}\{\mathbf{B}_n \mathbf{X} \mathbf{X}^T \mathbf{B}_n^T\}) \stackrel{(\text{B.2.4})}{\leq} \text{tr}\{\mathbf{\Sigma}_0\} \|\mathbf{B}_n^T \mathbf{B}_n\| E(\|\mathbf{X} \mathbf{X}^T\|),
 \end{aligned}$$

wegen der Symmetrie von $\mathbf{X} \mathbf{X}^T$ können wir diesen Ausdruck umschreiben und mit der Cauchy-Schwarz-Ungleichung abschätzen

$$\mathcal{A}_1 \leq \text{tr}\{\mathbf{\Sigma}_0\} \|\mathbf{B}_n^T \mathbf{B}_n\| E(\|\mathbf{X}^T \mathbf{X}\|) \leq \text{tr}\{\mathbf{\Sigma}_0\} \|\mathbf{B}_n^T \mathbf{B}_n\| [E(\|\mathbf{X}^T \mathbf{X}\|^2)]^{\frac{1}{2}}.$$

Mit (C.0.24), (C.0.19) und (C.0.3) erhalten wir die Rate

$$\mathcal{A}_1 = O(p)O(p)O(nK) = O(p^2 nK). \quad (\text{C.0.25})$$

Für die Abschätzung von $\mathcal{A}_2$ benötigen wir Raten für die Fehlerkovarianzmatrix, die Hutmatrix, die Matrix der Faktorladungen und die Faktoren in Darstellung (C.0.19). Damit können wir nun $\mathcal{A}_2$ durch Anwendung der iterierten Erwartung schreiben als

$$\begin{aligned} \mathcal{A}_2 &= E(\|\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T\|^2) = E(\text{tr}\{\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T \mathbf{E} \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T\}), \\ &= E(E(\text{tr}\{\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T \mathbf{E} \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T\} | \mathbf{X})) \\ &= E(\text{tr}\{\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}(\mathbf{E}^T \mathbf{E} | \mathbf{X}) \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T\}) \\ &\stackrel{(\text{C.0.23})}{=} E(\text{tr}\{\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \text{tr}\{\boldsymbol{\Sigma}_0\} \mathbf{I}_n \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T\}) = \text{tr}\{\boldsymbol{\Sigma}_0\} E(\text{tr}\{\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T\}), \end{aligned}$$

da $\mathbf{H}$ idempotent ist und mit (B.2.4), (B.2.5) und der Cauchy–Schwarz-Ungleichung folgt

$$\begin{aligned} \mathcal{A}_2 &= \text{tr}\{\boldsymbol{\Sigma}_0\} E(\text{tr}\{\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T\}) \leq \text{tr}\{\boldsymbol{\Sigma}_0\} E(\|\mathbf{B}_n^T \mathbf{B}_n\| \|\mathbf{X}^T \mathbf{X}\| \|\mathbf{H}\|_S \|\mathbf{1} \mathbf{1}^T \mathbf{1} \mathbf{1}^T\|) \\ &\leq \text{tr}\{\boldsymbol{\Sigma}_0\} \|\mathbf{B}_n^T \mathbf{B}_n\| \|\mathbf{H}\|_S \|\mathbf{1} \mathbf{1}^T \mathbf{1} \mathbf{1}^T\| [E(\|\mathbf{X}^T \mathbf{X}\|^2)]^{\frac{1}{2}}. \end{aligned}$$

Mit (C.0.24), (C.0.3), (B.5.2) und (C.0.19) gilt dann

$$\mathcal{A}_2 = O(p)O(p)O(1)n^2O(nK) = O(n^3 p^2 K). \quad (\text{C.0.26})$$

Nun betrachten wir die Gesamtaberschätzung und setzen die Resultate aus (C.0.25) und (C.0.26) ein

$$\frac{2}{(n-1)^2} \mathcal{A}_1 + \frac{2}{n^2(n-1)^2} \mathcal{A}_2 = \frac{2}{(n-1)^2} O(np^2 K) + \frac{2}{n^2(n-1)^2} O(n^3 p^2 K) = O(n^{-1} p^2 K),$$

die Behauptung (C.0.21) folgt, da beide Summanden die gleiche Rate liefern.

Für den Ausdruck aus (C.0.22) gilt analog zum ersten Teil des Beweises die Aufteilung

$$\begin{aligned} E(\|\mathbf{B}_n \text{côv}(\mathbf{f}) \mathbf{C}_n^T + \mathbf{C}_n \text{côv}(\mathbf{f}) \mathbf{B}_n^T\|_{\boldsymbol{\Sigma}_n}^2) &\leq 4E(\|\mathbf{B}_n \text{côv}(\mathbf{f}) \mathbf{C}_n^T\|_{\boldsymbol{\Sigma}_n}^2) \\ &\leq 8 \left[\frac{1}{(n-1)^2} E(\|\mathbf{B}_n \mathbf{X} \mathbf{X}^T \mathbf{C}_n^T\|_{\boldsymbol{\Sigma}_n}^2) + \frac{1}{n^2(n-1)^2} E(\|\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{C}_n^T\|_{\boldsymbol{\Sigma}_n}^2) \right] \\ &=: \frac{8}{(n-1)^2} \mathcal{B}_1 + \frac{8}{n^2(n-1)^2} \mathcal{B}_2. \end{aligned} \quad (\text{C.0.27})$$

Für die Abschätzung von $\mathcal{B}_1$ betrachten wir zunächst

$$\begin{aligned} E(\mathbf{E}^T \boldsymbol{\Sigma}^{-1} \mathbf{E} | \mathbf{X}) &= E \left(\begin{pmatrix} \varepsilon_1^T \\ \vdots \\ \varepsilon_n^T \end{pmatrix} \begin{pmatrix} (\boldsymbol{\Sigma}_n^{-1})_{11} & \dots & (\boldsymbol{\Sigma}_n^{-1})_{1p} \\ \vdots & \ddots & \vdots \\ (\boldsymbol{\Sigma}_n^{-1})_{p1} & \dots & (\boldsymbol{\Sigma}_n^{-1})_{pp} \end{pmatrix} \begin{pmatrix} \varepsilon_1 & \dots & \varepsilon_n \end{pmatrix} | \mathbf{X} \right) \\ &= E \left(\begin{pmatrix} \sum_{i=1}^p \sum_{j=1}^p \varepsilon_{1i} (\boldsymbol{\Sigma}_n^{-1})_{ij} \varepsilon_{1j} & \dots & \sum_{i=1}^p \sum_{j=1}^p \varepsilon_{1i} (\boldsymbol{\Sigma}_n^{-1})_{ij} \varepsilon_{nj} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^p \sum_{j=1}^p \varepsilon_{ni} (\boldsymbol{\Sigma}_n^{-1})_{ij} \varepsilon_{1j} & \dots & \sum_{i=1}^p \sum_{j=1}^p \varepsilon_{ni} (\boldsymbol{\Sigma}_n^{-1})_{ij} \varepsilon_{nj} \end{pmatrix} | \mathbf{X} \right) \end{aligned}$$

$$\begin{aligned}
 &= \begin{pmatrix} \sum_{i=1}^p E(\varepsilon_{1i}^2 | \mathbf{X})(\Sigma_n^{-1})_{ii} & 0 \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sum_{i=1}^p E(\varepsilon_{ni}^2 | \mathbf{X})(\Sigma_n^{-1})_{ii} \end{pmatrix} = \text{tr}\{\Sigma_n^{-1} \Sigma_0\} \cdot \mathbf{I}_n, \text{ da } \quad (\text{C.0.28}) \\
 \text{tr}\{\Sigma_n^{-1} \Sigma_0\} &= \text{tr} \left\{ \begin{pmatrix} (\Sigma_n^{-1})_{11} & \dots & (\Sigma_n^{-1})_{1p} \\ \vdots & \ddots & \vdots \\ (\Sigma_n^{-1})_{1p} & \dots & (\Sigma_n^{-1})_{nn} \end{pmatrix} \begin{pmatrix} E(\varepsilon_1^2 | \mathbf{f}) & 0 \\ & \ddots \\ 0 & E(\varepsilon_p^2 | \mathbf{f}) \end{pmatrix} \right\} \\
 &= \text{tr} \left\{ \begin{pmatrix} (\Sigma_n^{-1})_{11} E(\varepsilon_1^2 | \mathbf{f}) & \dots & (\Sigma_n^{-1})_{1p} E(\varepsilon_p^2 | \mathbf{f}) \\ \vdots & \ddots & \vdots \\ (\Sigma_n^{-1})_{1p} E(\varepsilon_1^2 | \mathbf{f}) & \dots & (\Sigma_n^{-1})_{pp} E(\varepsilon_p^2 | \mathbf{f}) \end{pmatrix} \right\} = \sum_{i=1}^p E(\varepsilon_i^2 | \mathbf{f})(\Sigma_n^{-1})_{ii}.
 \end{aligned}$$

Durch Einsetzen von (2.2.9) vereinfacht sich die Darstellung von $\mathcal{B}_1$, wie die folgende Rechnung zeigt. Die Konvergenzrate leiten wir dann im Wesentlichen aus der Rate für die Inverse der Kovarianzmatrix des Portfolios und der Annahme an die vierten Momente der Faktoren her. Durch Einsetzen von (2.2.9) und Anwendung der iterierten Erwartung erhalten wir

$$\begin{aligned}
 \mathcal{B}_1 &= E(\|\mathbf{B}_n \mathbf{X} \mathbf{X}^T \mathbf{C}_n^T\|_{\Sigma_n}^2) = \frac{1}{p} E(\|\Sigma_n^{-\frac{1}{2}} \mathbf{B}_n \mathbf{X} \mathbf{X}^T \mathbf{C}_n^T \Sigma_n^{-\frac{1}{2}}\|^2), \\
 &= \frac{1}{p} E(\text{tr}\{\Sigma_n^{-\frac{1}{2}} \mathbf{B}_n \mathbf{X} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{E}^T \Sigma_n^{-1} \mathbf{E} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-\frac{1}{2}}\}) \\
 &= \frac{1}{p} E(\text{tr}\{\Sigma_n^{-\frac{1}{2}} \mathbf{B}_n \mathbf{X} \mathbf{E}^T \Sigma_n^{-1} \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-\frac{1}{2}}\}) = \frac{1}{p} E(\text{tr}\{\mathbf{X} \mathbf{E}^T \Sigma_n^{-1} \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\}), \\
 &= \frac{1}{p} E(E(\text{tr}\{\mathbf{X} \mathbf{E}^T \Sigma_n^{-1} \mathbf{E} \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\} | \mathbf{X})) = \frac{1}{p} E(\text{tr}\{\mathbf{X} \mathbf{E} (\mathbf{E}^T \Sigma_n^{-1} \mathbf{E} | \mathbf{X}) \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\}) \\
 &\stackrel{(\text{C.0.28})}{=} \frac{1}{p} E(\text{tr}\{\mathbf{X} \text{tr}\{\Sigma_n^{-1} \Sigma_0\} \mathbf{I}_n \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\}) = \frac{1}{p} \text{tr}\{\Sigma_n^{-1} \Sigma_0\} E(\text{tr}\{\mathbf{X} \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\}) \\
 &\stackrel{(\text{B.2.6})}{\leq} \frac{1}{p} \text{tr}\{\Sigma_n^{-1} \Sigma_0\} E(\text{tr}\{\mathbf{X} \mathbf{X}^T\}) \|\mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\|_S \stackrel{(\text{B.2.3})}{\leq} \frac{1}{p} \text{tr}\{\Sigma_n^{-1} \Sigma_0\} \|\mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\|_S E(\|\mathbf{X}\|^2) \quad (\text{C.0.29})
 \end{aligned}$$

Mit (2.3.9) und den Annahmen 1 und 2 folgt $\text{tr}\{\Sigma_0 \Sigma_n^{-1}\} = \text{tr}\{\Sigma_n^{-1} O(1)\} = O(p)$ und mit (C.0.20) und (C.0.5) folgt dann

$$\mathcal{B}_1 = \frac{1}{p} O(p) O(nK) O(1) = O(nK). \quad (\text{C.0.30})$$

Analog setzen wir bei $\mathcal{B}_2$ die Darstellungen (2.2.9) und (2.1.5) ein und wenden die iterierte Erwartung an. Dann folgt

$$\begin{aligned}
 \mathcal{B}_2 &= E(\|\mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{C}_n^T\|_{\Sigma_n}^2) = \frac{1}{p} E(\|\Sigma_n^{-\frac{1}{2}} \mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{C}_n^T \Sigma_n^{-\frac{1}{2}}\|^2), \\
 &= \frac{1}{p} E(\text{tr}\{\Sigma_n^{-\frac{1}{2}} \mathbf{B}_n \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{E}^T \Sigma_n^{-1} \mathbf{E} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-\frac{1}{2}}\}) \\
 &= \frac{1}{p} E(\text{tr}\{\mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T \Sigma_n^{-1} \mathbf{E} \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\}) \\
 &= \frac{1}{p} E(E(\text{tr}\{\mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{H} \mathbf{E}^T \Sigma_n^{-1} \mathbf{E} \mathbf{H} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\} | \mathbf{X}))
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{p} E(\text{tr}\{\mathbf{X}\mathbf{1}\mathbf{1}^T \mathbf{H} E(\mathbf{E}^T \boldsymbol{\Sigma}_n^{-1} \mathbf{E} | \mathbf{X}) \mathbf{H} \mathbf{1}\mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n\}) \\
&\stackrel{(C.0.28)}{=} \frac{1}{p} E(\text{tr}\{\mathbf{X}\mathbf{1}\mathbf{1}^T \mathbf{H} \text{tr}\{\boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Sigma}_0\} \mathbf{I}_n \mathbf{H} \mathbf{1}\mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n\}) \\
&= \frac{1}{p} \text{tr}\{\boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Sigma}_0\} E(\text{tr}\{\mathbf{X}\mathbf{1}\mathbf{1}^T \mathbf{H} \mathbf{H} \mathbf{1}\mathbf{1}^T \mathbf{X}^T \mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n\}).
\end{aligned}$$

Da $\mathbf{H}$ idempotent ist, erhalten wir durch Anwendung von (B.2.6) und mit (B.2.4), (B.2.3) und (B.2.5) die Darstellung

$$\begin{aligned}
\mathcal{B}_2 &\leq \frac{1}{p} \text{tr}\{\boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Sigma}_0\} \|\mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n\|_S E(\text{tr}\{\mathbf{X}\mathbf{1}\mathbf{1}^T \mathbf{H} \mathbf{1}\mathbf{1}^T \mathbf{X}^T\}), \\
&\leq \frac{1}{p} \text{tr}\{\boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Sigma}_0\} E(\|\mathbf{X}^T \mathbf{X}\|) \|\mathbf{1}\mathbf{1}^T\|^2 \|\mathbf{H}\|_S \|\mathbf{B}_n^T \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_n\|_S.
\end{aligned}$$

Mit (2.3.9), Annahme 2, (C.0.19), (B.5.2) und (C.0.5) erhalten wir dann die Rate

$$\mathcal{B}_2 = \frac{1}{p} O(p) O(nK) n^2 O(1) O(1) = O(n^3 K). \quad (C.0.31)$$

Für die Gesamtaberschätzung von (C.0.27) nutzen wir (C.0.30) und (C.0.31):

$$\frac{8}{(n-1)^2} \mathcal{B}_1 + \frac{8}{n^2(n-1)^2} \mathcal{B}_2 = \frac{8}{(n-1)^2} O(nK) + \frac{8}{n^2(n-1)^2} O(n^3 K) = O(n^{-1} K), \quad (C.0.32)$$

das Resultat (C.0.22) folgt, da beide Summanden die gleiche Konvergenzrate liefern. $\square$

Aus diesem Beweis können wir direkt den Ausdruck $\mathbf{B}_n \mathbf{X} \mathbf{E}^T$ aus der Darstellung (2.2.14) abschätzen.

Korollar C.0.7 *Unter den Annahmen 1 - 6 und 8 gelten die Aussagen*

$$E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|^2) = O(p^2 nK) \quad (C.0.33)$$

und

$$E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|_{\boldsymbol{\Sigma}_n}^2) = O(nK). \quad (C.0.34)$$

Beweis. Der erste Ausdruck (C.0.33) entspricht $\mathcal{A}_1$ aus dem Beweis von Lemma C.0.6. Nach (C.0.25) gilt

$$E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|^2) = O(p^2 nK).$$

Für den zweiten Ausdruck (C.0.34) betrachten wir erneut $\mathcal{B}_1$ aus (C.0.27) und setzen (2.2.9) ein:

$$E(\|\mathbf{B}_n \mathbf{X} \mathbf{X}^T \mathbf{C}_n^T\|_{\boldsymbol{\Sigma}_n}^2) = E(\|\mathbf{B}_n \mathbf{X} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{E}^T\|_{\boldsymbol{\Sigma}_n}^2) = E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|_{\boldsymbol{\Sigma}_n}^2).$$

Die Abschätzung (C.0.30) liefert daher schon die Aussage

$$E(\|\mathbf{B}_n \mathbf{X} \mathbf{E}^T\|_{\boldsymbol{\Sigma}_n}^2) = O(nK).$$

$\square$

Weiterhin können wir mit den Abschätzungen aus den bisher geführten Beweisen den Ausdruck $\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T$ abschätzen.

Korollar C.0.8 *Unter den Annahmen 1 - 6 und 8 gelten folgende Aussagen*

$$E(\|\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\|^2) = O(n^3p^2K) \quad (\text{C.0.35})$$

und

$$E(\|\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\|_{\Sigma_n}^2) = O(n^3K). \quad (\text{C.0.36})$$

Beweis. Wir schreiben den ersten Ausdruck mithilfe der iterierten Erwartung in der Darstellung

$$\begin{aligned} E(\|\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\|^2) &= E(\text{tr}\{\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\mathbf{E}\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T\}) \\ &= E(E(\text{tr}\{\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\mathbf{E}\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T|\mathbf{X}\})) = E(\text{tr}\{\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T E(\mathbf{E}^T\mathbf{E}|\mathbf{X})\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T\}). \end{aligned}$$

Nach (C.0.23) und mit (B.2.4) gilt weiter

$$\begin{aligned} E(\|\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\|^2) &= E(\text{tr}\{\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T \text{tr}\{\Sigma_0\}\mathbf{I}_n\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T\}) = \text{tr}\{\Sigma_0\}E(\text{tr}\{\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T\}) \\ &\leq \text{tr}\{\Sigma_0\}E(\|\mathbf{B}_n^T\mathbf{B}_n\| \|\mathbf{X}^T\mathbf{X}\| \|\mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T\|) \end{aligned}$$

und mithilfe der Abschätzungen (C.0.19), (C.0.24) und (C.0.3) folgt das Resultat

$$E(\|\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\|^2) = O(p)O(p)O(nK)n^2 = O(n^3p^2K).$$

Für den zweiten Ausdruck folgt mit der iterierten Erwartung und (C.0.28)

$$\begin{aligned} E(\|\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\|_{\Sigma_n}^2) &= \frac{1}{p}E(\|\Sigma_n^{-\frac{1}{2}}\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\Sigma_n^{-\frac{1}{2}}\|^2) \\ &= \frac{1}{p}E(\text{tr}\{\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\Sigma_n^{-1}\mathbf{E}\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T\Sigma_n^{-\frac{1}{2}}\Sigma_n^{-\frac{1}{2}}\mathbf{B}_n\}) \\ &= \frac{1}{p}E(E(\text{tr}\{\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\Sigma_n^{-1}\mathbf{E}\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T\Sigma_n^{-1}\mathbf{B}_n|\mathbf{X}\})) \\ &= \frac{1}{p}E(\text{tr}\{\mathbf{X}\mathbf{1}\mathbf{1}^T E(\mathbf{E}^T\Sigma_n^{-1}\mathbf{E}|\mathbf{X})\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T\Sigma_n^{-1}\mathbf{B}_n\}) \\ &= \frac{1}{p}E(\text{tr}\{\mathbf{X}\mathbf{1}\mathbf{1}^T \text{tr}\{\Sigma_n^{-1}\Sigma_0\}\mathbf{I}_n\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T\Sigma_n^{-1}\mathbf{B}_n\}) \\ &= \frac{1}{p}\text{tr}\{\Sigma_n^{-1}\Sigma_0\}E(\text{tr}\{\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T\mathbf{X}^T\mathbf{B}_n^T\Sigma_n^{-1}\mathbf{B}_n\}). \end{aligned}$$

Dann führt eine Anwendung von (B.2.6), (B.2.3) und (B.2.4) zu folgender Darstellung

$$E(\|\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\|_{\Sigma_n}^2) \leq \frac{1}{p}\text{tr}\{\Sigma_n^{-1}\Sigma_0\}E(\|\mathbf{X}^T\mathbf{X}\|)\|\mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T\|\|\mathbf{B}_n^T\Sigma_n^{-1}\mathbf{B}_n\|_S.$$

Mit (2.3.9) folgt

$$\text{tr}\{[\Sigma_0\Sigma_n^{-1}]^2\} = \text{tr}\{[O(1)\Sigma_n^{-1}]^2\} \leq O(1)\|\Sigma_n^{-1}\|^2 = O(p)$$

und mit den Abschätzungen (C.0.19) und (C.0.4) folgt das Resultat

$$E(\|\mathbf{B}_n\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{E}^T\|_{\Sigma_n}^2) = \frac{1}{p}O(p)O(nK)n^2O(1) = O(n^3K).$$

□

Lemma C.0.9 Für den Schätzfehler der Kovarianzmatrix der Fehler in der Darstellung des Faktormodells gelten unter den Annahmen 1 - 6 und 8 folgende Aussagen

$$E(\|\mathbf{F}_n\|^2) = O(n^{-1}p) + O(n^{-2}pK^2) \quad (\text{C.0.37})$$

und

$$E(\|\mathbf{F}_n\|_{\Sigma_n}^2) = O(n^{-1}) + O(n^{-2}K^2). \quad (\text{C.0.38})$$

Beweis. Wir beginnen mit der ersten Aussage (C.0.37). Durch Einsetzen der Definition (2.2.11) und Anwendung der Cauchy-Schwarz-Ungleichung erhalten wir

$$\begin{aligned} E(\|\mathbf{F}_n\|^2) &= E(\text{tr}\{\mathbf{F}_n^2\}) = E(\text{tr}\{[\mathbf{I}_p \circ \frac{1}{n}\mathbf{E}(\mathbf{I}_n - \mathbf{H})\mathbf{E}^T - \Sigma_0]^2\}) \\ &= E(\text{tr}\{[\frac{1}{n}\mathbf{I}_p \circ \mathbf{E}\mathbf{E}^T - \Sigma_0 - \frac{1}{n}\mathbf{I}_p \circ \mathbf{E}\mathbf{H}\mathbf{E}^T]^2\}) \\ &\leq 2E(\text{tr}\{[\frac{1}{n}\mathbf{I}_p \circ \mathbf{E}\mathbf{E}^T - \Sigma_0]^2 + [\frac{1}{n}\mathbf{I}_p \circ \mathbf{E}\mathbf{H}\mathbf{E}^T]^2\}) \\ &= 2E(\text{tr}\{[\frac{1}{n}\mathbf{I}_p \circ \mathbf{E}\mathbf{E}^T - \Sigma_0]^2\}) + \frac{2}{n^2}E(\text{tr}\{[\mathbf{I}_p \circ \mathbf{E}\mathbf{H}\mathbf{E}^T]^2\}) =: 2\mathcal{A}_1 + \frac{2}{n^2}\mathcal{A}_2. \end{aligned}$$

Die Rate für den ersten Ausdruck $\mathcal{A}_1$ folgt direkt aus der Annahme an die gleichmäßig beschränkten vierten Momente der Fehler durch folgende Umformung

$$\begin{aligned} \mathcal{A}_1 &= E(\text{tr}\{[\frac{1}{n}\mathbf{I}_p \circ \mathbf{E}\mathbf{E}^T - \Sigma_0]^2\}) = E(\text{tr}\{[\text{diag}(\frac{1}{n}\sum_{i=1}^n(\varepsilon_{i1}^2 - E(\varepsilon_{i1}^2|\mathbf{f})), \dots, \frac{1}{n}\sum_{i=1}^n(\varepsilon_{ip}^2 - E(\varepsilon_{ip}^2|\mathbf{f})))^2\}) \\ &= E(\text{tr}\{[\frac{1}{n}\sum_{i=1}^n \text{diag}(\varepsilon_{i1}^2 - E(\varepsilon_{i1}^2|\mathbf{f})), \dots, \varepsilon_{ip}^2 - E(\varepsilon_{ip}^2|\mathbf{f}))]^2\}) \\ &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n E(\text{tr}\{\text{diag}(\varepsilon_{i1}^2 - E(\varepsilon_{i1}^2|\mathbf{f})), \dots, \varepsilon_{ip}^2 - E(\varepsilon_{ip}^2|\mathbf{f}))\text{diag}(\varepsilon_{j1}^2 - E(\varepsilon_{j1}^2|\mathbf{f})), \dots, \varepsilon_{jp}^2 - E(\varepsilon_{jp}^2|\mathbf{f}))\}). \end{aligned}$$

Da die Fehler unkorreliert und insbesondere unabhängig identisch verteilt sind, fallen alle Summanden, bei denen $i \neq j$ ist weg und es folgt

$$\begin{aligned} \mathcal{A}_1 &= \frac{1}{n^2}\sum_{i=1}^n E(\text{tr}\{[\text{diag}(\varepsilon_{i1}^2 - E(\varepsilon_{i1}^2|\mathbf{f})), \dots, \varepsilon_{ip}^2 - E(\varepsilon_{ip}^2|\mathbf{f}))]^2\}) \\ &= \frac{1}{n}E(\text{tr}\{[\text{diag}(\varepsilon_{11}^2 - E(\varepsilon_{11}^2|\mathbf{f})), \dots, \varepsilon_{1p}^2 - E(\varepsilon_{1p}^2|\mathbf{f}))]^2\}) = \frac{1}{n}E(\text{tr}\{[\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2) - \Sigma_0]^2\}) \\ &= \frac{1}{n}E(\text{tr}\{[\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2)]^2 - \text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2)\Sigma_0 - \Sigma_0\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2) + \Sigma_0^2\}) \\ &= \frac{1}{n}E(\text{tr}\{[\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2)]^2\}) - \frac{1}{n}\text{tr}\{E(\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2))\Sigma_0 - \Sigma_0E(\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2)) + \Sigma_0^2\}. \end{aligned}$$

Durch Anwendung der iterierten Erwartung gilt

$$E(\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2)) = E(E(\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2)|\mathbf{f})) = E(\Sigma_0)$$

und daher folgt weiter mit Annahme 2

$$\mathcal{A}_1 = \frac{1}{n}\text{tr}\{E([\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2)]^2)\} - \text{tr}\{\Sigma_0^2\} \leq \frac{1}{n}\sum_{i=1}^p E(\varepsilon_{1i}^4) = O(n^{-1}p). \quad (\text{C.0.39})$$

Die Abschätzung von $\mathcal{A}_2$ erfolgt analog zum Beweis von Lemma C.0.4. Die Rate wird dabei durch die Eigenschaften der Spur der Hutmatrix nach (B.5.1) und die gleichmäßig beschränkten vierten Momente der Fehler beeinflusst. Durch Einsetzen von (C.0.8) erhalten wir durch eine Anwendung der iterierten Erwartung

$$\begin{aligned}\mathcal{A}_2 &= E(\text{tr}\{[\mathbf{I}_p \circ \mathbf{EHE}^T]^2\}) \\ &= E(\text{tr}\{\sum_{i=1}^n \sum_{j=1}^n H_{ij} \text{diag}(\varepsilon_{i1}\varepsilon_{j1}, \dots, \varepsilon_{ip}\varepsilon_{jp}) \sum_{k=1}^n \sum_{l=1}^n H_{kl} \text{diag}(\varepsilon_{k1}\varepsilon_{l1}, \dots, \varepsilon_{kp}\varepsilon_{lp})\}) \\ &= E(E(\text{tr}\{\sum_{i=1}^n \sum_{j=1}^n H_{ij} \text{diag}(\varepsilon_{i1}\varepsilon_{j1}, \dots, \varepsilon_{ip}\varepsilon_{jp}) \sum_{k=1}^n \sum_{l=1}^n H_{kl} \text{diag}(\varepsilon_{k1}\varepsilon_{l1}, \dots, \varepsilon_{kp}\varepsilon_{lp})\} | \mathbf{X})).\end{aligned}$$

Analog zu (C.0.9) folgt wegen $E(\varepsilon | \mathbf{f}) = 0$ weiter

$$\begin{aligned}\mathcal{A}_2 &= E(\text{tr}\{\sum_{i=1}^n H_{ii}^2 E([\text{diag}(\varepsilon_{i1}^2, \dots, \varepsilon_{ip}^2)]^2 | \mathbf{X})\}) \\ &\quad + E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ii} H_{jj} E(\text{diag}(\varepsilon_{i1}^2, \dots, \varepsilon_{ip}^2) \text{diag}(\varepsilon_{j1}^2, \dots, \varepsilon_{jp}^2) | \mathbf{X})\}) \\ &\quad + E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 E([\text{diag}(\varepsilon_{i1}\varepsilon_{j1}, \dots, \varepsilon_{ip}\varepsilon_{jp})]^2 | \mathbf{X})\}) \\ &\quad + E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij} H_{ji} E(\text{diag}(\varepsilon_{i1}\varepsilon_{j1}, \dots, \varepsilon_{ip}\varepsilon_{jp}) \text{diag}(\varepsilon_{j1}\varepsilon_{i1}, \dots, \varepsilon_{jp}\varepsilon_{ip}) | \mathbf{X})\}) \\ &=: \mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3 + \mathcal{B}_4.\end{aligned}$$

Für den ersten Ausdruck $\mathcal{B}_1$ gilt wegen der unabhängig identisch verteilten Fehlerterme und mit (C.0.10)

$$\begin{aligned}\mathcal{B}_1 &= E(\text{tr}\{\sum_{i=1}^n H_{ii}^2 E([\text{diag}(\varepsilon_{i1}^2, \dots, \varepsilon_{ip}^2)]^2 | \mathbf{X})\}) = E(\text{tr}\{\sum_{i=1}^n H_{ii}^2 E([\text{diag}(\varepsilon_{i1}^2, \dots, \varepsilon_{ip}^2)]^2 | \mathbf{X})\}) \\ &\leq K E(\text{tr}\{E([\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2)]^2 | \mathbf{X})\}) = K E(\text{tr}\{\text{diag}(\varepsilon_{11}^4, \dots, \varepsilon_{1p}^4)\}).\end{aligned}$$

Mit Annahme 2 folgt dann die Abschätzung

$$\mathcal{B}_1 \leq K \sum_{i=1}^p E(\varepsilon_i^4) = O(pK). \quad (\text{C.0.40})$$

Für den zweiten Ausdruck $\mathcal{B}_2$ folgt wegen der unabhängig identisch verteilten Fehlerterme und (C.0.12)

$$\mathcal{B}_2 = E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ii} H_{jj} E(\text{diag}(\varepsilon_{i1}^2, \dots, \varepsilon_{ip}^2) \text{diag}(\varepsilon_{j1}^2, \dots, \varepsilon_{jp}^2) | \mathbf{X})\})$$

$$\begin{aligned}
&= E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ii} H_{jj} E(\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2) | \mathbf{X}) E(\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2) | \mathbf{X})\}) \\
&\leq K^2 E(\text{tr}\{E(\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2) | \mathbf{X}) E(\text{diag}(\varepsilon_{11}^2, \dots, \varepsilon_{1p}^2) | \mathbf{X})\}) = K^2 \sum_{i=1}^p [E(\varepsilon_{1i}^2)]^2.
\end{aligned}$$

Nach Annahme 2 folgt dann

$$\mathcal{B}_2 = O(pK^2). \quad (\text{C.0.41})$$

Sei $\boldsymbol{\eta}$ eine unabhängige Ausprägung von $\boldsymbol{\varepsilon}$. Dann gilt für den dritten Ausdruck $\mathcal{B}_3$ wegen der unabhängig identisch verteilten Fehlerterme und mit (C.0.14)

$$\begin{aligned}
\mathcal{B}_3 &= E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 E([\text{diag}(\varepsilon_{i1}\varepsilon_{j1}, \dots, \varepsilon_{ip}\varepsilon_{jp})]^2 | \mathbf{X})\}) \\
&= E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij}^2 E([\text{diag}(\varepsilon_{11}\eta_{11}, \dots, \varepsilon_{1p}\eta_{1p})]^2 | \mathbf{X})\}) \\
&\leq KE(\text{tr}\{E([\text{diag}(\varepsilon_{11}\eta_{11}, \dots, \varepsilon_{1p}\eta_{1p})]^2 | \mathbf{X})\}) = K \sum_{i=1}^p E(\varepsilon_{1i}^2) E(\eta_{1i}^2).
\end{aligned}$$

Mit Annahme 2 folgt dann die Rate

$$\mathcal{B}_3 = O(pK). \quad (\text{C.0.42})$$

Für den letzten Ausdruck $\mathcal{B}_4$ gilt wegen der unabhängig identisch verteilten Fehlerterme und (C.0.16)

$$\begin{aligned}
\mathcal{B}_4 &= E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij} H_{ji} E(\text{diag}(\varepsilon_{i1}\varepsilon_{j1}, \dots, \varepsilon_{ip}\varepsilon_{jp}) \text{diag}(\varepsilon_{j1}\varepsilon_{i1}, \dots, \varepsilon_{jp}\varepsilon_{ip}) | \mathbf{X})\}) \\
&= E(\text{tr}\{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n H_{ij} H_{ji} E(\text{diag}(\varepsilon_{11}\eta_{11}, \dots, \varepsilon_{1p}\eta_{1p}) \text{diag}(\eta_{11}\varepsilon_{11}, \dots, \eta_{1p}\varepsilon_{1p}) | \mathbf{X})\}) \\
&\leq KE(\text{tr}\{E(\text{diag}(\varepsilon_{11}\eta_{11}, \dots, \varepsilon_{1p}\eta_{1p}) \text{diag}(\eta_{11}\varepsilon_{11}, \dots, \eta_{1p}\varepsilon_{1p}) | \mathbf{X})\}).
\end{aligned}$$

Wegen der Diagonalstruktur der Matrizen und mit Annahme 2 erhalten wir

$$\begin{aligned}
\mathcal{B}_4 &\leq KE(\text{tr}\{\text{diag}(\varepsilon_{11}, \dots, \varepsilon_{1p}) \text{diag}(\varepsilon_{11}, \dots, \varepsilon_{1p}) \text{diag}(\eta_{11}, \dots, \eta_{1p}) \text{diag}(\eta_{11}, \dots, \eta_{1p})\}) \\
&= K \text{tr}\{[\text{diag}(E(\varepsilon_{11}^2), \dots, E(\varepsilon_{1p}^2))]^2\} = K \sum_{i=1}^p [E(\varepsilon_{1i}^2)]^2 = O(pK). \quad (\text{C.0.43})
\end{aligned}$$

Für $\mathcal{A}_2$ gilt dann nach (C.0.40), (C.0.41), (C.0.42) und (C.0.43) die Dominanz des zweiten Summanden $\mathcal{B}_2$:

$$\mathcal{A}_2 = E(\text{tr}\{[\mathbf{I}_p \circ \mathbf{EHE}^T]^2\}) = O(pK) + O(pK^2) + O(pK) + O(pK) = O(pK^2) \quad (\text{C.0.44})$$

und insgesamt folgt mithilfe der Abschätzungen aus (C.0.39) und (C.0.44) für (C.0.37)

$$E(\|\mathbf{F}_n\|^2) \leq 2\mathcal{A}_1 + \frac{2}{n^2}\mathcal{A}_2 = O(n^{-1}p) + O(n^{-2}pK^2).$$

Für den zweiten Teil des Beweises nutzen wir (2.2.5):

$$\begin{aligned} E(\|\mathbf{F}_n\|_{\Sigma_n}^2) &\leq \frac{1}{p}\|\Sigma_n^{-1}\|_S^2 E(\|\mathbf{F}_n\|^2) \stackrel{(2.2.8)}{=} \frac{1}{p}O(1)(O(n^{-1}p) + O(n^{-2}pK^2)) \\ &= O(n^{-1}) + O(n^{-2}K^2). \end{aligned}$$

□

Lemma C.0.10 *Für die Stichprobenkovarianzmatrix der Fehlerterme gelten unter den Annahmen 1 - 6 und 8 folgende Aussagen*

$$E(\|\mathbf{G}_n\|^2) = O(n^{-1}p^2) \quad (\text{C.0.45})$$

und

$$E(\|\mathbf{G}_n\|_{\Sigma_n}^2) = O(n^{-1}p). \quad (\text{C.0.46})$$

Beweis. Die Fehler sind nach Annahme 8 unabhängig identisch verteilt. Außerdem besitzen die Komponenten von ϵ nach Annahme 2 gleichmäßig beschränkte vierte Momente. Daher folgt die erste Aussage mit Lemma A.1.1.

Für den Beweis des zweiten Teils (C.0.46) nutzen wir wieder (2.2.5) und (2.2.8):

$$E(\|\mathbf{G}_n\|_{\Sigma_n}^2) \leq \frac{1}{p}\|\Sigma_n^{-1}\|_S^2 E(\|\mathbf{G}_n\|^2) = \frac{1}{p}O(1)O(n^{-1}p^2) = O(n^{-1}p).$$

□

Lemma C.0.11 *Für den Kleinsten Quadrate Schätzer für die Matrix der Faktorladungen gilt unter den Annahmen 2, 6, 3 und 8:*

$$E(\|\hat{\mathbf{B}}_n - \mathbf{B}_n\|^2) = O_P(n^{-1}pK). \quad (\text{C.0.47})$$

Beweis. Für den Beweis betrachten wir den Schätzfehler in der Darstellung von (2.2.9) und nutzen die Ungleichung (B.2.5)

$$\begin{aligned} E(\|\hat{\mathbf{B}}_n - \mathbf{B}_n\|^2) &= E(\|\mathbf{E}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\|^2) = E(\text{tr}\{\mathbf{E}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{E}^T\}) \\ &= E(\text{tr}\{(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{E}^T\mathbf{E}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\}) = E(E(\text{tr}\{(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{E}^T\mathbf{E}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\}|\mathbf{X})) \\ &= E(\text{tr}\{(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}E(\mathbf{E}^T\mathbf{E}|\mathbf{X})\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\}) = \text{tr}\{\Sigma_0\}E(\text{tr}\{(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\}) \\ &= \text{tr}\{\Sigma_0\}E(\text{tr}\{(\mathbf{X}\mathbf{X}^T)^{-1}\}) \end{aligned}$$

Mit dem schwachen Gesetz der großen Zahlen gilt

$$\frac{1}{n}\mathbf{X}\mathbf{X}^T = \frac{1}{n}\sum_{i=1}^n \mathbf{f}_i\mathbf{f}_i^T \xrightarrow{P} E(\mathbf{f}\mathbf{f}^T)$$

und daher auch die Konvergenz

$$\lambda_K\left(\frac{1}{n}\mathbf{X}\mathbf{X}^T\right) \xrightarrow{P} \lambda_K(E(\mathbf{f}\mathbf{f}^T)).$$

Für die Kovarianzmatrix der Faktoren gilt

$$\text{cov}(\mathbf{f}) = E(\mathbf{f}\mathbf{f}^T) - E(\mathbf{f})E(\mathbf{f}^T) \leq E(\mathbf{f}\mathbf{f}^T)$$

und mit Annahme 3 folgt

$$c_1 < \lambda_K(\text{cov}(\mathbf{f})) \leq \lambda_K(E(\mathbf{f}\mathbf{f}^T)).$$

Daher erhalten wir

$$\|(\frac{1}{n}\mathbf{X}\mathbf{X}^T)^{-1}\|_S = \|n(\mathbf{X}\mathbf{X}^T)^{-1}\|_S = O(1).$$

Und nach (B.1.8) erhalten wir

$$E(\text{tr}\{(\mathbf{X}\mathbf{X}^T)^{-1}\}) = E\left(\sum_{i=1}^K \lambda_i((\mathbf{X}\mathbf{X}^T)^{-1})\right) \leq K \lambda_1((\mathbf{X}\mathbf{X}^T)^{-1}) = O(n^{-1}K).$$

Dann gilt mit (C.0.23)

$$E(\|\hat{\mathbf{B}}_n - \mathbf{B}_n\|^2) = O(n^{-1}pK).$$

□

Lemma C.0.12 *Unter den Annahmen 1 - 6 und 8 gelten für die Inverse der Kovarianzmatrix der Fehlerterme die folgenden Konvergenzraten*

$$\|\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}\| = O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(n^{-1}p^{\frac{1}{2}}K) \quad (\text{C.0.48})$$

und

$$\|\hat{\Sigma}_0^{-1}\| = O_P(p^{\frac{1}{2}}). \quad (\text{C.0.49})$$

Beweis. Für diesen Beweis nutzen wir die Notation $\hat{\Sigma}_0 = \text{diag}(\hat{\sigma}_1, \dots, \hat{\sigma}_p)$ und $\Sigma_0 = \text{diag}(\sigma_1, \dots, \sigma_p)$. Aus Lemma C.0.9 kennen wir bereits die Rate

$$\|\hat{\Sigma}_0 - \Sigma_0\| = O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(n^{-1}p^{\frac{1}{2}}K).$$

Wegen der Diagonalstruktur der Matrizen können wir schreiben

$$\begin{aligned} \|\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}\| &= \|\text{diag}\left(\frac{1}{\hat{\sigma}_1} - \frac{1}{\sigma_1}, \dots, \frac{1}{\hat{\sigma}_p} - \frac{1}{\sigma_p}\right)\| = \left[\sum_{i=1}^p \left(\frac{1}{\hat{\sigma}_i} - \frac{1}{\sigma_i}\right)^2\right]^{\frac{1}{2}} \\ &= \left[\sum_{i=1}^p \left(\frac{\sigma_i - \hat{\sigma}_i}{\hat{\sigma}_i \sigma_i}\right)^2\right]^{\frac{1}{2}} = \left[\sum_{i=1}^p (\sigma_i - \hat{\sigma}_i)^2 \frac{1}{\hat{\sigma}_i^2 \sigma_i^2}\right]^{\frac{1}{2}}. \end{aligned}$$

Nach Annahme 4 sind die σ_i von Null weg beschränkt und wegen der konsistenten Schätzung nach Lemma C.0.9 folgt auch hier die Rate

$$\|\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}\| = O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(n^{-1}p^{\frac{1}{2}}K).$$

Mit diesem Resultat können wir direkt die zweite Aussage des Lemmas beweisen. Dazu betrachten wir Annahme 4, nach der $\min_{1 \leq i \leq p} \sigma_i > c_1$ gilt. Dann folgt auch

$$\|\Sigma_0^{-1}\| = \left[\sum_{i=1}^p \frac{1}{\sigma_i^2} \right]^{\frac{1}{2}} \leq \left[p \max_{1 \leq i \leq p} \frac{1}{\sigma_i^2} \right]^{\frac{1}{2}} = O_P(p^{\frac{1}{2}}).$$

Damit können wir folgende Abschätzung vornehmen

$$\begin{aligned} \|\hat{\Sigma}_0^{-1}\| &= \|\hat{\Sigma}_0^{-1} - \Sigma_0^{-1} + \Sigma_0^{-1}\| \leq \|\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}\| + \|\Sigma_0^{-1}\| \\ &= O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(n^{-1}p^{\frac{1}{2}}K) + O_P(p^{\frac{1}{2}}). \end{aligned}$$

Unter der Annahme, dass die Anzahl der Faktoren kleiner ist als die Stichprobengröße folgt

$$\|\hat{\Sigma}_0^{-1}\| = O_P(p^{\frac{1}{2}}).$$

□

Lemma C.0.13 *Unter den Annahmen 1 - 6 und 8 gelten folgende Konvergenzraten*

$$\|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\mathbf{B}}_n[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1}\| = O_P(n^{-\frac{1}{2}}p) + O_P(n^{-1}pK)$$

und

$$\|\Sigma_0^{-1} \hat{\mathbf{B}}_n[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T (\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\| = O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(n^{-1}p^{\frac{1}{2}}K).$$

Beweis. Wir beginnen mit der ersten Aussage. Dazu teilen wir den Ausdruck mithilfe von (B.2.3) und (B.2.5) in drei Faktoren auf

$$\begin{aligned} &\|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\mathbf{B}}_n[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1}\| \\ &= \|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\Sigma}_0^{\frac{1}{2}} \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\Sigma}_0^{-\frac{1}{2}}\| \\ &\leq \|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\Sigma}_0^{\frac{1}{2}}\| \|\hat{\Sigma}_0^{-\frac{1}{2}}\| \|\hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}}\|_S \\ &=: \mathcal{A}_1 \mathcal{A}_2 \mathcal{A}_3. \end{aligned}$$

Wir beginnen mit der Betrachtung von $\mathcal{A}_2$. Nach (C.0.49) gilt

$$\|\hat{\Sigma}_0^{-1}\| = [\text{tr}\{\hat{\Sigma}_0^{-2}\}]^{\frac{1}{2}} = \left[\sum_{i=1}^p (\hat{\Sigma}_0^{-1})_{ii} (\hat{\Sigma}_0^{-1})_{ii} \right]^{\frac{1}{2}} = O_P(p^{\frac{1}{2}}),$$

woraus

$$\mathcal{A}_2 = \|\hat{\Sigma}_0^{-\frac{1}{2}}\| = [\text{tr}\{\hat{\Sigma}_0^{-1}\}]^{\frac{1}{2}} = \left[\sum_{i=1}^p (\hat{\Sigma}_0^{-1})_{ii} \right]^{\frac{1}{2}} = O_P(p^{\frac{1}{2}}) \quad (\text{C.0.50})$$

folgt. Beim ersten Faktor $\mathcal{A}_1$ gehen wir analog zum Beweis von Lemma C.0.12 vor. Nutzen wir die dort eingeführte Notation, so folgt

$$\mathcal{A}_1 = \|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\Sigma}_0^{\frac{1}{2}}\| = \left[\sum_{i=1}^p \left(\frac{1}{\hat{\sigma}_i} - \frac{1}{\sigma_i} \right)^2 \hat{\sigma}_i \right]^{\frac{1}{2}} = \left[\sum_{i=1}^p (\sigma_i - \hat{\sigma}_i)^2 \frac{\hat{\sigma}_i}{\hat{\sigma}_i^2 \sigma_i^2} \right]^{\frac{1}{2}} = \left[\sum_{i=1}^p (\sigma_i - \hat{\sigma}_i)^2 \frac{1}{\hat{\sigma}_i \sigma_i^2} \right]^{\frac{1}{2}}.$$

Da $\hat{\Sigma}_0$ nach Lemma C.0.9 ein konsistenter Schätzer ist und nach Annahme 4 die Einträge von Σ_0 von Null weg beschränkt sind, folgt mit dem analogen Argument zu Lemma C.0.12 und der Rate aus (C.0.37)

$$\mathcal{A}_1 = O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}}) + O_P(n^{-1} p^{\frac{1}{2}} K). \quad (\text{C.0.51})$$

Für die Abschätzung von $\mathcal{A}_3$ nutzen wir die Darstellung von $\hat{\Sigma}_n^{-1}$ nach (2.2.16)

$$\hat{\Sigma}_n^{-1} = \hat{\Sigma}_0^{-1} - \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1}.$$

Diesen Ausdruck schreiben wir um zu

$$\begin{aligned} \hat{\Sigma}_0^{-1} - \hat{\Sigma}_n^{-1} &= \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\Sigma}_0^{-\frac{1}{2}} \\ &\Rightarrow \hat{\Sigma}_0^{\frac{1}{2}} (\hat{\Sigma}_0^{-1} - \hat{\Sigma}_n^{-1}) \hat{\Sigma}_0^{\frac{1}{2}} = \hat{\Sigma}_0^{\frac{1}{2}} \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\Sigma}_0^{\frac{1}{2}}, \end{aligned}$$

wobei nun auf der rechten Seite der Ausdruck steht, den wir in der Frobeniusnorm abschätzen wollen:

$$\begin{aligned} \hat{\Sigma}_0^{\frac{1}{2}} (\hat{\Sigma}_0^{-1} - \hat{\Sigma}_n^{-1}) \hat{\Sigma}_0^{\frac{1}{2}} &= \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}} \\ \Rightarrow \hat{\Sigma}_0^{\frac{1}{2}} \hat{\Sigma}_0^{-1} \hat{\Sigma}_0^{\frac{1}{2}} - \hat{\Sigma}_0^{\frac{1}{2}} \hat{\Sigma}_n^{-1} \hat{\Sigma}_0^{\frac{1}{2}} &= \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}} \\ \Rightarrow \mathbf{I}_p &\geq \mathbf{I}_p - \hat{\Sigma}_0^{\frac{1}{2}} \hat{\Sigma}_n^{-1} \hat{\Sigma}_0^{\frac{1}{2}} = \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}}. \end{aligned}$$

Wegen (B.4.2) folgt daraus

$$1 = \|\mathbf{I}_p\|_S \geq \|\hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}}\|_S = \mathcal{A}_3. \quad (\text{C.0.52})$$

Aus den Abschätzungen (C.0.51), (C.0.50) und (C.0.52) folgt nun insgesamt:

$$\begin{aligned} \|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1}\| &\leq \mathcal{A}_1 \mathcal{A}_2 \mathcal{A}_3 \\ &= O_P(p^{\frac{1}{2}}) [O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}}) + O_P(n^{-1} p^{\frac{1}{2}} K)] 1 = O_P(n^{-\frac{1}{2}} p) + O_P(n^{-1} p K). \end{aligned}$$

Für den zweiten Teil des Beweises schreiben wir analog zum ersten Teil

$$\begin{aligned} &\|\Sigma_0^{-1} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T (\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\| \\ &= \|\Sigma_0^{-1} \hat{\Sigma}_0 \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T (\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\| \\ &\leq \|\Sigma_0^{-1}\| \|\hat{\Sigma}_0^{-\frac{1}{2}} \hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T (\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}) \hat{\Sigma}_0\| \end{aligned}$$

und mit (B.2.5) und der Submultiplikativität der Spektralnorm schätzen wir ab durch

$$\leq \|\Sigma_0^{-1}\|_S \|\hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}}\|_S \|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}) \hat{\Sigma}_0\|$$

$$=: \mathcal{B}_1 \|\hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}}\|_S \mathcal{B}_2.$$

Den zweiten Ausdruck haben wir bereits durch $\mathcal{A}_3 \leq 1$ abgeschätzt. Für $\mathcal{B}_1$ gilt nach (2.2.7)

$$\mathcal{B}_1 = \|\Sigma_0^{-1}\|_S = O(1). \quad (\text{C.0.53})$$

Für den letzten Ausdruck $\mathcal{B}_2$ nutzen wir wieder den Beweis von Lemma C.0.12 und die dort eingeführte Notation

$$\mathcal{B}_2 = \|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\hat{\Sigma}_0\| = \left[\sum_{i=1}^p (\sigma_i - \hat{\sigma}_i)^2 \frac{\hat{\sigma}_i^2}{\sigma_i^2 \hat{\sigma}_i^2} \right]^{\frac{1}{2}} = \left[\sum_{i=1}^p (\sigma_i - \hat{\sigma}_i)^2 \frac{1}{\sigma_i^2} \right]^{\frac{1}{2}}.$$

Da nach Annahme 4 die σ_i von Null weg beschränkt sind und nach Lemma C.0.9 $\hat{\Sigma}_0$ ein konsistenter Schätzer ist folgt

$$\mathcal{B}_2 = O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}}) + O_P(n^{-1} p^{\frac{1}{2}} K). \quad (\text{C.0.54})$$

Insgesamt folgt aus (C.0.52), (C.0.53) und (C.0.54)

$$\begin{aligned} & \|\Sigma_0^{-1} \hat{\mathbf{B}}_n [\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T (\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\| \\ & \leq \mathcal{B}_1 \|\hat{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{B}}_n [\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-\frac{1}{2}}\|_S \mathcal{B}_2 \\ & = O_P(1) [O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}}) + O_P(n^{-1} p^{\frac{1}{2}} K)] = O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}}) + O_P(n^{-1} p^{\frac{1}{2}} K). \end{aligned}$$

□

Lemma C.0.14 *Unter den Annahmen 1 - 6 und 8 gelten folgende Konvergenzraten:*

$$\|\Sigma_0^{-1} (\hat{\mathbf{B}}_n - \mathbf{B}_n) [\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \Sigma_0^{-1}\| = O_P(n^{-\frac{1}{2}} p K^{\frac{3}{2}})$$

und

$$\|\Sigma_0^{-1} \mathbf{B}_n [\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} (\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T) \Sigma_0^{-1}\| = O_P(n^{-\frac{1}{2}} p K^{\frac{3}{2}}).$$

Beweis. Für die erste Aussage betrachten wir folgende Zerlegung:

$$\begin{aligned} & \|\Sigma_0^{-1} (\hat{\mathbf{B}}_n - \mathbf{B}_n) [\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \Sigma_0^{-1}\| \\ & \stackrel{(\text{B.2.3})}{\leq} \|\Sigma_0^{-1} (\hat{\mathbf{B}}_n - \mathbf{B}_n)\| \|[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1}\| \|\hat{\mathbf{B}}_n^T \Sigma_0^{-1}\| =: \mathcal{A}_1 \mathcal{A}_2 \mathcal{A}_3. \end{aligned}$$

Für den ersten Ausdruck $\mathcal{A}_1$ gilt mit (B.2.5), Annahme 4 und Lemma C.0.11

$$\|\Sigma_0^{-1} (\hat{\mathbf{B}}_n - \mathbf{B}_n)\| \leq \|\Sigma_0^{-1}\|_S \|\hat{\mathbf{B}}_n - \mathbf{B}_n\| = O(1) O(n^{-\frac{1}{2}} p^{\frac{1}{2}} K^{\frac{1}{2}}) = O(n^{-\frac{1}{2}} p^{\frac{1}{2}} K^{\frac{1}{2}}). \quad (\text{C.0.55})$$

Wegen

$$\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n \geq \text{côv}(\mathbf{f})^{-1}$$

gilt auch

$$[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \leq \text{côv}(\mathbf{f})$$

und

$$\|[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1}\| \leq \|\text{côv}(\mathbf{f})\|.$$

Dann können wir $\mathcal{A}_2$ mithilfe der Stichprobenkovarianzmatrix der Faktoren abschätzen:

$$\begin{aligned} \|\text{côv}(\mathbf{f})\| &= \left\| \frac{1}{n-1} \mathbf{X} \mathbf{X}^T - \frac{1}{n(n-1)} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T \right\| \\ &= [\text{tr}\{\left[\frac{1}{n-1} \mathbf{X} \mathbf{X}^T - \frac{1}{n(n-1)} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T\right]^2\}]^{\frac{1}{2}}. \end{aligned}$$

Mit der Cauchy–Schwarz-Ungleichung folgt

$$\begin{aligned} \|\text{côv}(\mathbf{f})\| &\leq [2\text{tr}\{\frac{1}{(n-1)^2} [\mathbf{X} \mathbf{X}^T]^2 + \frac{1}{n^2(n-1)^2} [\mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T]^2\}]^{\frac{1}{2}} \\ &= [\frac{2}{(n-1)^2} \|\mathbf{X} \mathbf{X}^T\|^2 + \frac{2}{n^2(n-1)^2} \|\mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T\|^2]^{\frac{1}{2}} \end{aligned}$$

und mit (C.0.19) und (B.2.4) folgt die Abschätzung

$$\|\text{côv}(\mathbf{f})\| \leq [\frac{2}{(n-1)^2} \|\mathbf{X} \mathbf{X}^T\|^2 + \frac{2}{n^2(n-1)^2} \|\mathbf{X}^T \mathbf{X}\|^2 \|\mathbf{1} \mathbf{1}^T\|^2]^{\frac{1}{2}} = O_P(K). \quad (\text{C.0.56})$$

Dann folgt für $\mathcal{A}_2$:

$$\mathcal{A}_2 = \|[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1}\| \leq \|\text{côv}(\mathbf{f})\| = O_P(K). \quad (\text{C.0.57})$$

Für $\mathcal{A}_3$ gilt mit (B.2.5), (2.2.7), (2.2.3) und Lemma C.0.11

$$\begin{aligned} \mathcal{A}_3 &= \|\hat{\mathbf{B}}_n^T \Sigma_0^{-1}\| \leq \|\Sigma_0^{-1}\|_S \|\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T + \mathbf{B}_n^T\| \leq O(1)(\|\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T\| + \|\mathbf{B}_n^T\|) \\ &= O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}} K^{\frac{1}{2}}) + O_P(p^{\frac{1}{2}}). \end{aligned} \quad (\text{C.0.58})$$

Nehmen wir an, dass die Anzahl der Faktoren kleiner ist als die Stichprobengröße, so folgt

$$\mathcal{A}_3 = O_P(p^{\frac{1}{2}})$$

Daher erhalten wir insgesamt:

$$\begin{aligned} \|\Sigma_0^{-1}(\hat{\mathbf{B}}_n - \mathbf{B}_n)[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} \hat{\mathbf{B}}_n^T \Sigma_0^{-1}\| &\leq \mathcal{A}_1 \mathcal{A}_2 \mathcal{A}_3 \\ &= O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}} K^{\frac{1}{2}}) O_P(K) O(p^{\frac{1}{2}}) = O_P(n^{-\frac{1}{2}} p K^{\frac{3}{2}}). \end{aligned}$$

Für die zweite Aussage betrachten wir die Zerlegung

$$\begin{aligned} &\|\Sigma_0^{-1} \mathbf{B}_n [\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} (\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T) \Sigma_0^{-1}\| \\ &\stackrel{(\text{B.2.3})}{\leq} \|\Sigma_0^{-1} \mathbf{B}_n\| \|[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1}\| \|(\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T) \Sigma_0^{-1}\| \\ &=: \mathcal{B}_1 \|[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1}\| \mathcal{B}_2. \end{aligned}$$

Die Abschätzung des zweiten Terms kennen wir bereits von $\mathcal{A}_2$ in (C.0.57). Für den ersten Faktor $\mathcal{B}_1$ gilt wegen (B.2.5), Annahme 4 und (2.2.3)

$$\mathcal{B}_1 = \|\Sigma_0^{-1} \mathbf{B}_n\| \leq \|\Sigma_0^{-1}\|_S \|\mathbf{B}_n\| = O(1) O_P(p^{\frac{1}{2}}) = O_P(p^{\frac{1}{2}}). \quad (\text{C.0.59})$$

Für den dritten Faktor $\mathcal{B}_2$ gilt wegen $\|(\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T)\boldsymbol{\Sigma}_0^{-1}\| = \|\boldsymbol{\Sigma}_0^{-1}(\hat{\mathbf{B}}_n - \mathbf{B}_n)\|$ wie in der Abschätzung von $\mathcal{A}_1$ in (C.0.55)

$$\mathcal{B}_2 = \|(\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T)\boldsymbol{\Sigma}_0^{-1}\| = O(n^{-\frac{1}{2}}p^{\frac{1}{2}}K^{\frac{1}{2}}). \quad (\text{C.0.60})$$

Nach (C.0.59), (C.0.57) und (C.0.60) folgt dann die Behauptung:

$$\begin{aligned} & \|\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n]^{-1}(\hat{\mathbf{B}}_n^T - \mathbf{B}_n^T)\boldsymbol{\Sigma}_0^{-1}\| \leq \mathcal{B}_1 O_P(K)\mathcal{B}_2 \\ & = O_P(p^{\frac{1}{2}})O_P(K)O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}K^{\frac{1}{2}}) = O_P(n^{-\frac{1}{2}}pK^{\frac{3}{2}}). \end{aligned} \quad (\text{C.0.61})$$

□

Lemma C.0.15 *Unter den Annahmen 1 bis 8 gilt folgende Konvergenzrate*

$$\begin{aligned} & \|\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n([\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]^{-1})\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\| \\ & = O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}K^{\frac{1}{2}}). \end{aligned}$$

Beweis. Wir nehmen folgende Aufteilung vor

$$\begin{aligned} & \|\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n([\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]^{-1})\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\| \\ & = \|[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]^{-1}\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-2}\mathbf{B}_n\| \\ & \stackrel{(\text{B.2.3})}{\leq} \|[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]^{-1}\| \|\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-2}\mathbf{B}_n\|. \end{aligned} \quad (\text{C.0.62})$$

Für die Abschätzung des zweiten Ausdrucks betrachten wir

$$\|\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-2}\mathbf{B}_n\| = \|\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n\| \stackrel{(\text{B.2.3})}{\leq} \|\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\| \|\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n\| = \|\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\|^2$$

und aus der Abschätzung (C.0.59) folgt daher

$$\|\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-2}\mathbf{B}_n\| = O_P(p). \quad (\text{C.0.63})$$

Zur Abschätzung des ersten Ausdrucks verwenden wir Formel (B.3.4) mit $\mathbf{A}_1 := [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]$ und $\mathbf{A}_1 + \mathbf{A}_2 := [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n]$. Dann gilt $\mathbf{A}_2 = [\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n] - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]$ und es folgt

$$\begin{aligned} & \|[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]^{-1}\| \\ & \leq \frac{\|[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n] - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]\| \|[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]^{-1}\|_S^2}{1 - \|[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]^{-1}\|_S \|[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{B}}_n] - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]\|}. \end{aligned} \quad (\text{C.0.64})$$

Wir betrachten nun $\|[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n]^{-1}\|_S$ und nutzen dazu den minimalen Eigenwert von $\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n$. Da $\text{cov}(\mathbf{f})^{-1}$ und $\boldsymbol{\Sigma}_0^{-1}$ positiv semidefinit sind, gilt

$$\lambda_K(\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n) \geq \lambda_K(\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n).$$

Sei $\mathbf{v} \in \mathbb{R}^K$ der normierte Eigenvektor zum minimalen Eigenwert von $\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n$. Dann gilt

$$\mathbf{v}^T\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n\mathbf{v} = \mathbf{v}^T\mathbf{v} \cdot \lambda_K(\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n) = \lambda_K(\mathbf{B}_n^T\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_n).$$

Für die Betrachtung des minimalen Eigenwertes können wir also die Darstellung $\mathbf{v}^T \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n \mathbf{v}$ nutzen. Da Σ_0^{-1} symmetrisch ist, folgt mit (B.4.1)

$$\mathbf{v}^T \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n \mathbf{v} \geq \lambda_p(\Sigma_0^{-1}) \mathbf{v}^T \mathbf{B}_n^T \mathbf{B}_n \mathbf{v}$$

und mit dem gleichen Argument folgt

$$\mathbf{v}^T \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n \mathbf{v} \geq \lambda_p(\Sigma_0^{-1}) \lambda_K(\mathbf{B}_n^T \mathbf{B}_n) \mathbf{v}^T \mathbf{v} = \lambda_p(\Sigma_0^{-1}) \lambda_K(\mathbf{B}_n^T \mathbf{B}_n).$$

Mit den Annahmen 4 und 7 folgt für ein $C > 0$

$$\mathbf{v}^T \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n \mathbf{v} \geq p \cdot C.$$

Damit erhalten wir

$$\|[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n \Sigma_0^{-1} \mathbf{B}_n]^{-1}\|_S = O(p^{-1}). \quad (\text{C.0.65})$$

Die letzte Abschätzung, die für (C.0.64) fehlt, ist die von $\|\hat{\text{cov}}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n - \text{cov}(\mathbf{f})^{-1} - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\|$. Dazu nutzen wir die Dreiecksungleichung und schätzen $\|\hat{\text{cov}}(\mathbf{f})^{-1} - \text{cov}(\mathbf{f})^{-1}\|$ und $\|\hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\|$ jeweils einzeln ab.

Nach Annahme 3 ist die Kovarianzmatrix der Faktoren invertierbar mit $\|\text{cov}(\mathbf{f})^{-1}\|_S = O(1)$ und nach Lemma C.0.2 ist $\hat{\text{cov}}(\mathbf{f})$ ein konsistenter Schätzer. Daher erhalten wir die Konvergenzrate mit dem Ansatz (B.3.4) aus der Rate von Lemma C.0.2:

$$\|\hat{\text{cov}}(\mathbf{f})^{-1} - \text{cov}(\mathbf{f})^{-1}\| = \|(\text{cov}(\mathbf{f}) + \mathbf{D}_n)^{-1} - \text{cov}(\mathbf{f})^{-1}\| \leq \frac{\|\mathbf{D}_n\| \|\text{cov}(\mathbf{f})^{-1}\|_S^2}{1 - \|\text{cov}(\mathbf{f})^{-1}\|_S \|\mathbf{D}_n\|} = O_P(n^{-\frac{1}{2}} K). \quad (\text{C.0.66})$$

Weiter gilt mit der Darstellung (2.2.9), der Dreiecksungleichung und (B.2.4)

$$\begin{aligned} \|\hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\| &= \|(\mathbf{B}_n + \mathbf{C}_n) \hat{\Sigma}_0^{-1} (\mathbf{B}_n + \mathbf{C}_n) - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\| \\ &= \|\mathbf{B}_n (\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}) \mathbf{B}_n + \mathbf{B}_n \hat{\Sigma}_0^{-1} \mathbf{C}_n + \mathbf{C}_n \hat{\Sigma}_0^{-1} \mathbf{B}_n + \mathbf{C}_n \hat{\Sigma}_0^{-1} \mathbf{C}_n\| \\ &\leq \|\mathbf{B}_n (\hat{\Sigma}_0^{-1} - \Sigma_0^{-1}) \mathbf{B}_n\| + 2\|\mathbf{B}_n \hat{\Sigma}_0^{-1} \mathbf{C}_n\| + \|\mathbf{C}_n \hat{\Sigma}_0^{-1} \mathbf{C}_n\| \\ &\leq \|\mathbf{B}_n\| \|(\hat{\Sigma}_0^{-1} - \Sigma_0^{-1})\| \|\mathbf{B}_n\| + 2\|\mathbf{B}_n\| \|\hat{\Sigma}_0^{-1}\| \|\mathbf{C}_n\| + \|\mathbf{C}_n\| \|\hat{\Sigma}_0^{-1}\| \|\mathbf{C}_n\|. \end{aligned}$$

Nun setzen wir (2.2.3), (C.0.47) und das Resultat aus Lemma C.0.12 ein und erhalten die Rate

$$\begin{aligned} \|\hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\| &= (O_P(n^{-\frac{1}{2}} p^{\frac{3}{2}}) + O_P(n^{-1} p^{\frac{3}{2}} K)) + O_P(n^{-\frac{1}{2}} p^{\frac{3}{2}} K^{\frac{1}{2}}) + O_P(n^{-1} p^{\frac{3}{2}} K) \\ &= O_P(n^{-\frac{1}{2}} p^{\frac{3}{2}} K^{\frac{1}{2}}), \end{aligned} \quad (\text{C.0.67})$$

da die vorletzte Rate dominiert, falls die Stichprobengröße größer ist als die Anzahl der Faktoren. Aus (C.0.66) und (C.0.67) erhalten wir

$$\begin{aligned} \|\hat{\text{cov}}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n - \text{cov}(\mathbf{f})^{-1} - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\| \\ \leq \|\hat{\text{cov}}(\mathbf{f})^{-1} - \text{cov}(\mathbf{f})^{-1}\| + \|\hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\| = O_P(n^{-\frac{1}{2}} p^{\frac{3}{2}} K^{\frac{1}{2}}). \end{aligned} \quad (\text{C.0.68})$$

Der Nenner in (C.0.64) konvergiert dann für $n \rightarrow \infty$ gegen 1. Wir betrachten nun den Zähler und mit (C.0.65) und (C.0.67) gilt

$$\|[\hat{\text{cov}}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n] - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]\| \|[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1}\|_S^2$$

$$= O_P(n^{-\frac{1}{2}} p^{\frac{3}{2}} K^{\frac{1}{2}}) O_P(p^{-2}) = O_P(n^{-\frac{1}{2}} p^{-\frac{1}{2}} K^{\frac{1}{2}}).$$

Zusammen mit (C.0.63) folgt in (C.0.62)

$$\|\boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n([\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\boldsymbol{\Sigma}}_0^{-1} \hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1} \mathbf{B}_n]^{-1}) \mathbf{B}_n^T \boldsymbol{\Sigma}_0^{-1}\| = O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}} K^{\frac{1}{2}}).$$

□

D. Herleitung der Konvergenzraten für Abschnitt 2.4

Lemma D.0.16 *Unter den Annahmen 9, 10 und 11 gilt*

$$E(\|\mathbf{D}_n\|^2) = O\left(\frac{K^2 \log n}{n}\right).$$

Beweis. Um die Aussage zu beweisen zeigen wir, dass für ein $C > 0$ die Aussage $E(\|\mathbf{D}_n\|^2) \leq C \frac{K^2 \log n}{n}$ gilt. Dazu setzen wir $a_n := \frac{K^2 \log n}{n}$. Dann folgt

$$\frac{1}{C} E(\|\mathbf{D}_n\|^2) = a_n E\left(\frac{\|\mathbf{D}_n\|^2}{C \cdot a_n}\right),$$

dabei handelt es sich um eine positive Zufallsvariable, daher können wir schreiben

$$\frac{1}{C} E(\|\mathbf{D}_n\|^2) = a_n \int_0^\infty P\left(\frac{\|\mathbf{D}_n\|^2}{C \cdot a_n} \geq s\right) ds.$$

Wir zeigen nun, dass für eine Funktion f mit $f(s) \rightarrow 0$, für $s \rightarrow \infty$ gilt, dass

$$P\left(\frac{\|\mathbf{D}_n\|^2}{C \cdot a_n} \geq s\right) \leq f(s),$$

denn dann erhalten wir

$$\int_0^\infty P\left(\frac{\|\mathbf{D}_n\|^2}{C \cdot a_n} \geq s\right) ds \leq \int_0^\infty f(s) ds \leq C_1$$

für ein $C_1 > 0$ und insgesamt gilt

$$\frac{1}{C} E(\|\mathbf{D}_n\|^2) \leq a_n C_1.$$

Anstelle von $\mathbf{D}_n$ nutzen wir im weiteren Beweis die alternative Darstellung

$$\mathbf{D}_n^* = \frac{1}{n} \mathbf{X} \mathbf{X}^T - \frac{1}{n^2} \mathbf{X} \mathbf{1} \mathbf{1}^T \mathbf{X}^T - \text{cov}(\mathbf{f}).$$

Die Asymptotik, die wir für $\mathbf{D}_n^*$ herleiten, gilt dann auch für $\mathbf{D}_n$.
Zunächst schreiben wir

$$\|\mathbf{D}_n^*\|^2 = \sum_{i=1}^K \sum_{j=1}^K \left[\frac{1}{n} \sum_{t=1}^n f_{it} f_{jt} - E(f_{i1} f_{j1}) - \left(\frac{1}{n^2} \sum_{t=1}^n f_{it} \sum_{t=1}^n f_{jt} - E(f_{i1}) E(f_{j1}) \right) \right]^2$$

$$\begin{aligned}
 &\leq 2 \sum_{i=1}^K \sum_{j=1}^K \left[\frac{1}{n} \sum_{t=1}^n f_{it} f_{jt} - E(f_{i1} f_{j1}) \right]^2 + \left[\frac{1}{n^2} \sum_{t=1}^n f_{it} \sum_{t=1}^n f_{jt} - E(f_{i1}) E(f_{j1}) \right]^2 \quad (\text{D.0.1}) \\
 &\leq 2K^2 \max_{1 \leq i, j \leq K} \left[\frac{1}{n} \sum_{t=1}^n f_{it} f_{jt} - E(f_{i1} f_{j1}) \right]^2 + 2K^2 \max_{1 \leq i, j \leq K} \left[\frac{1}{n^2} \sum_{t=1}^n f_{it} \sum_{t=1}^n f_{jt} - E(f_{i1}) E(f_{j1}) \right]^2.
 \end{aligned}$$

Diese beiden Summanden betrachten wir getrennt voneinander. Der Beweis basiert auf Annahme 11, den exponentiell abfallenden Tails der Verteilung der Faktoren. Für $s > 0$ schreiben wir mit der Bonferroniungleichung

$$\begin{aligned}
 P\left(\max_{1 \leq i, j \leq K} \left| \sum_{t=1}^n f_{it} f_{jt} - E(f_{i1} f_{j1}) \right| > sn\right) &= P\left(\bigcup_{1 \leq i, j \leq K} \left\{ \left| \sum_{t=1}^n f_{it} f_{jt} - E(f_{i1} f_{j1}) \right| > sn \right\}\right) \\
 &\leq \sum_{i=1}^K \sum_{j=1}^K P\left(\left| \sum_{t=1}^n f_{it} f_{jt} - E(f_{i1} f_{j1}) \right| > sn\right) \leq K^2 \max_{1 \leq i, j \leq K} P\left(\left| \sum_{t=1}^n f_{it} f_{jt} - E(f_{i1} f_{j1}) \right| > sn\right). \quad (\text{D.0.2})
 \end{aligned}$$

Nach Annahme 11 existieren Konstanten $r_2, b_1 > 0$, sodass für alle $s > 0$ und $i \in \{1, \dots, K\}$

$$P(|f_{it}| > s) \leq \exp\left(-\left(\frac{s}{b_1}\right)^{r_2}\right) \leq \exp\left(1 - \left(\frac{s}{b_1}\right)^{r_2}\right)$$

gilt. Mit Lemma A.3.1 folgt dann, dass auch für das Produkt $f_{it} f_{jt}$, $i, j \in \{1, \dots, K\}$ Konstanten $r_3, b_2 > 0$ existieren, sodass für alle $s > 0$

$$P(|f_{it} f_{jt}| > s) \leq \exp\left(1 - \left(\frac{s}{b_2}\right)^{r_3}\right)$$

gilt. Wie im zugehörigen Beweis muss $0 < r_3 < \frac{r_2^2}{2r_2} = \frac{1}{2}r_2$ erfüllt sein. Wir wählen $r_3 = \frac{1}{3}r_2$ und setzen $r_4^{-1} := 3r_2^{-1} + r_1^{-1}$, wobei r_1 die Konstante aus Annahme 10 ist.

Mit Annahme 11 ist dann $r_4 < 1$ und wir können Satz A.3.2 anwenden und erhalten damit eine Bernsteinungleichung für die Summe unserer schwach abhängigen, zentrierten Zufallsvariablen: Es existieren Konstanten $C_1, \dots, C_5 > 0$, sodass für alle $s > 0$ gilt

$$\begin{aligned}
 \max_{1 \leq i, j \leq K} P\left(\left| \sum_{t=1}^n f_{it} f_{jt} - E(f_{i1} f_{j1}) \right| > sn\right) &\leq n \exp\left(-\frac{(ns)^{r_4}}{C_2}\right) + \exp\left(-\frac{(ns)^2}{C_2(1 + nC_3)}\right) \\
 &+ \exp\left(-\frac{(ns)^2}{C_4 n} \exp\left(\frac{(ns)^{r_4(1-r_4)}}{C_5 (\log(ns))^{r_4}}\right)\right) =: \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3. \quad (\text{D.0.3})
 \end{aligned}$$

Nun betrachten wir (D.0.2) und (D.0.3) und setzen $s := C\tau \sqrt{\frac{\log n}{n}}$.

$$\begin{aligned}
 K^2 \mathcal{A}_1 &= K^2 n \exp\left(-\frac{(n\tau C \sqrt{\frac{\log n}{n}})^{r_4}}{C_2}\right) \\
 &= K^2 n \exp\left(-\frac{C^{r_4}}{C_2} n^{\frac{r_4}{2}} \tau^{r_4} [\log n]^{\frac{r_4}{2}}\right) \\
 &= K^2 n \exp\left(-\frac{C^{r_4}}{C_2} n^{\frac{r_4}{2}} (\tau^{r_4} - 1) [\log n]^{\frac{r_4}{2}} - \frac{C^{r_4}}{C_2} n^{\frac{r_4}{2}} [\log n]^{\frac{r_4}{2}}\right) \\
 &= K^2 n \exp\left(-\frac{C^{r_4}}{C_2} n^{\frac{r_4}{2}} [\log n]^{\frac{r_4}{2}}\right) \exp\left(-\frac{C^{r_4}}{C_2} n^{\frac{r_4}{2}} (\tau^{r_4} - 1) [\log n]^{\frac{r_4}{2}}\right)
 \end{aligned}$$

$$= K^2 n^{1 - \frac{r_4}{2} \frac{C^{r_4}}{C_2} n^{\frac{r_4}{2}}} \exp\left(-\frac{C^{r_4}}{C_2} n^{\frac{r_4}{2}} (\tau^{r_4} - 1) [\log n]^{\frac{r_4}{2}}\right),$$

für große Werte von n und C folgt dann

$$K^2 \mathcal{A}_1 \leq C' f_1(\tau),$$

mit $f_1(\tau) \rightarrow 0$, $\tau \rightarrow \infty$ und $\tau > 1$.

Für den zweiten Ausdruck gilt

$$\begin{aligned} K^2 \mathcal{A}_2 &= K^2 \exp\left(-\frac{n^2 C^2 \tau^2 \log n}{n C_2 (1 + n C_3)}\right) = K^2 \exp\left(-\frac{C^2 \tau^2 \log n}{C_2 (\frac{1}{n} + C_3)}\right) \\ &= K^2 \exp\left(-\frac{C^2 \log n}{C_2 (\frac{1}{n} + C_3)}\right) \exp\left(-\frac{C^2 (\tau^2 - 1) \log n}{C_2 (\frac{1}{n} + C_3)}\right) \\ &= K^2 n^{-\frac{C^2}{C_2 (\frac{1}{n} + C_3)}} \exp\left(-\frac{C^2 (\tau^2 - 1) \log n}{C_2 (\frac{1}{n} + C_3)}\right), \end{aligned}$$

für große Werte von C folgt

$$K^2 \mathcal{A}_2 \leq C' f_2(\tau),$$

mit $f_2(\tau) \rightarrow 0$, $\tau \rightarrow \infty$ und $\tau > 1$.

Für den letzten Ausdruck schreiben wir

$$\begin{aligned} K^2 \mathcal{A}_3 &= K^2 \exp\left(-\frac{n^2 C^2 \tau^2 \log n}{n^2 C_4}\right) \exp\left(\frac{(nC\tau\sqrt{\log n})^{r_4(1-r_4)}}{n^{\frac{r_4(1-r_4)}{2}} C_5 [\log(nC\tau\sqrt{\frac{\log n}{n}})]^{r_4}}\right) \\ &= K^2 \exp\left(-\frac{C^2 \log n}{C_4}\right) \exp\left(\frac{(nC\tau\sqrt{\log n})^{r_4(1-r_4)}}{n^{\frac{r_4(1-r_4)}{2}} C_5 [\log(nC\tau\sqrt{\frac{\log n}{n}})]^{r_4}}\right) \\ &\quad \cdot \exp\left(-\frac{C^2 (\tau^2 - 1) \log n}{C_4}\right) \exp\left(\frac{(nC\tau\sqrt{\log n})^{r_4(1-r_4)}}{n^{\frac{r_4(1-r_4)}{2}} C_5 [\log(nC\tau\sqrt{\frac{\log n}{n}})]^{r_4}}\right) \end{aligned}$$

und für große Werte von C und n erhalten wir

$$K^2 \mathcal{A}_3 \leq C' f_3(\tau),$$

mit $f_3(\tau) \rightarrow 0$, $\tau \rightarrow \infty$ und $\tau > 1$. Nun betrachten wir den zweiten Summanden aus (D.0.1).

$$\begin{aligned} &\max_{1 \leq i, j \leq K} \left| \frac{1}{n^2} \sum_{t=1}^n f_{it} \sum_{t=1}^n f_{jt} - E(f_{i1}) E(f_{j1}) \right|^2 \\ &= \max_{1 \leq i, j \leq K} \left| \left(\frac{1}{n} \sum_{t=1}^n f_{it} - E(f_{i1}) \right) \frac{1}{n} \sum_{t=1}^n f_{jt} + E(f_{i1}) \frac{1}{n} \sum_{t=1}^n f_{jt} + E(f_{i1}) \left(\frac{1}{n} \sum_{t=1}^n f_{jt} - E(f_{j1}) \right) \right. \\ &\quad \left. - E(f_{i1}) \frac{1}{n} \sum_{t=1}^n f_{jt} \right|^2, \end{aligned}$$

wegen der Stationarität und Ergodizität der Faktoren gilt

$$= \max_{1 \leq i, j \leq K} \left| \left(\frac{1}{n} \sum_{t=1}^n f_{it} - E(f_{i1}) \right) \frac{1}{n} \sum_{t=1}^n f_{jt} + E(f_{i1}) \left(\frac{1}{n} \sum_{t=1}^n f_{jt} - E(f_{j1}) \right) \right|^2.$$

Nach Annahme 2 sind die Erwartungswerte der Faktoren gleichmäßig beschränkt und wegen der Ergodizität ist dann auch $\frac{1}{n} \sum_{t=1}^n f_{jt}$ beschränkt. Für $P(\max_{1 \leq i \leq K} |\frac{1}{n} \sum_{t=1}^n f_{it} - E(f_{i1})| > C\tau \sqrt{\frac{\log n}{n}})$ erhalten wir ein analoges Resultat zu den vorhergegangenen Rechnungen mithilfe von Satz A.3.2.

Insgesamt folgt dann mit (D.0.1)

$$P(\|\mathbf{D}_n\|^2 \geq C\tau a_n) \leq C'f(\tau)$$

für $f(\tau) \rightarrow 0, \tau \rightarrow \infty$. und damit auch die Behauptung. $\square$

Korollar D.0.17 *Unter den Annahmen 2 - 6 und 9 - 11 gelten folgende Aussagen*

$$E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|^2) = O(n^{-1}[\log n]p^2 K^2)$$

und

$$E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|_{\Sigma_n}^2) = O(n^{-1}[\log n]p^{-1}K^2).$$

Beweis. Wir können direkt den Beweis von Lemma C.0.3 anwenden und erhalten mit Lemma D.0.16 und (C.0.3)

$$E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|^2) \leq \|\mathbf{B}_n^T \mathbf{B}_n\|^2 E(\|\mathbf{D}_n\|^2) = O(p^2)O(n^{-1}[\log n]K^2) = O(n^{-1}[\log n]p^2 K^2).$$

Mit (C.0.4) und Lemma D.0.16 erhalten wir die zweite Aussage

$$E(\|\mathbf{B}_n \mathbf{D}_n \mathbf{B}_n^T\|_{\Sigma_n}^2) \leq \frac{1}{p} E(\|\mathbf{D}_n\|^2) \|\mathbf{B}_n^T \Sigma_n^{-1} \mathbf{B}_n\|_S^2 = \frac{1}{p} O(n^{-1}[\log n]K^2) O(1) = O(n^{-1}[\log n]p^{-1}K^2).$$

$\square$

Bemerkung D.0.18 In den Beweisen der Lemmata C.0.4 - C.0.10 nutzen wir das Argument der unabhängigen Faktoren nicht. Die Resultate lassen sich direkt auf den Fall mit stationären Faktoren übertragen.

Lemma D.0.19 *Für den Kleinste Quadrate Schätzer für die Matrix der Faktorladungen gilt unter den Annahmen 2 - 6 und 9:*

$$E(\|\hat{\mathbf{B}}_n - \mathbf{B}_n\|^2) = O(n^{-1}pK). \quad (\text{D.0.4})$$

Beweis. Wir betrachten die Konvergenz im quadratischen Mittel. Durch die Anwendung der iterierten Erwartung erhalten wir

$$\begin{aligned} E(\|\hat{\mathbf{B}}_n - \mathbf{B}_n\|^2) &= E(\|\mathbf{E}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\|^2) = E(\text{tr}\{\mathbf{E}\mathbf{X}(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{E}^T\}) \\ &= E(E(\text{tr}\{\mathbf{X}\mathbf{E}^T\mathbf{E}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{X}^T)^{-1}\}|\mathbf{X})) \\ &= E(\text{tr}\{\mathbf{X}E(\mathbf{E}^T\mathbf{E}|\mathbf{X})\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{X}^T)^{-1}\}|\mathbf{X}). \end{aligned}$$

Analog zu (C.0.23) folgt

$$E(\|\hat{\mathbf{B}}_n - \mathbf{B}_n\|^2) = \text{tr}\{\Sigma_0\}E(\text{tr}\{\mathbf{X}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{X}^T)^{-1}\}) = \text{tr}\{\Sigma_0\}E(\text{tr}\{(\mathbf{X}\mathbf{X}^T)^{-1}\}).$$

Die Faktoren sind nach Annahme 9 stationär und ergodisch, daher gilt

$$\frac{1}{n} \mathbf{X} \mathbf{X}^T = \frac{1}{n} \sum_{t=1}^n \mathbf{f}_t \mathbf{f}_t^T \xrightarrow{P} E(\mathbf{f} \mathbf{f}^T), \quad n \rightarrow \infty,$$

also insbesondere

$$\lambda_K\left(\frac{1}{n} \mathbf{X} \mathbf{X}^T\right) \xrightarrow{P} \lambda_K(E(\mathbf{f} \mathbf{f}^T)) \geq \lambda_K(\text{cov}(\mathbf{f})).$$

Wegen Annahme 3 folgt dann

$$\lambda_1(n(\mathbf{X} \mathbf{X}^T)^{-1}) = O(1),$$

und

$$\lambda_1((\mathbf{X} \mathbf{X}^T)^{-1}) = O(n^{-1}).$$

Insgesamt folgt dann mit (B.1.8)

$$E(\text{tr}\{(\mathbf{X} \mathbf{X}^T)^{-1}\}) = E\left(\sum_{i=1}^K \lambda_i((\mathbf{X} \mathbf{X}^T)^{-1})\right) \leq K \lambda_1((\mathbf{X} \mathbf{X}^T)^{-1}) = KO(n^{-1})$$

und mit (C.0.24)

$$E(\|\hat{\mathbf{B}}_n - \mathbf{B}_n\|^2) = O(n^{-1}pK).$$

□

Die Änderung der Konvergenzrate für die Stichprobenkovarianzmatrix der Faktoren wirkt sich auch auf die Rate ihrer Inversen aus. Daher muss der Beweis von Lemma C.0.15 angepasst werden.

Lemma D.0.20 *Unter den Annahmen 1 - 7 und 9 - 11 gilt folgende Konvergenzrate*

$$\begin{aligned} & \|\Sigma_0^{-1} \mathbf{B}_n([\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1}) \mathbf{B}_n^T \Sigma_0^{-1}\| \\ & = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} p^{-1} K) + O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}} K^{\frac{1}{2}}). \end{aligned}$$

Beweis. Wir nutzen den Beweis von Lemma C.0.15.

Mit (C.0.62) betrachten wir

$$\begin{aligned} & \|\Sigma_0^{-1} \mathbf{B}_n([\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1}) \mathbf{B}_n^T \Sigma_0^{-1}\| \\ & \leq \|[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1}\| \|\mathbf{B}_n^T \Sigma_0^{-2} \mathbf{B}_n\|. \end{aligned} \quad (\text{D.0.5})$$

Für den zweiten Faktor erhalten wir aus (C.0.63) das Resultat

$$\|\mathbf{B}_n^T \Sigma_0^{-2} \mathbf{B}_n\| = O_P(p). \quad (\text{D.0.6})$$

Für den ersten Faktor aus (D.0.5) können wir mit (C.0.64) folgenden Ausdruck behandeln

$$\|[\text{cov}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1}\| \quad (\text{D.0.7})$$

$$\leq \frac{\|[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n] - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]\| \|[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1}\|_S^2}{1 - \|[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1}\|_S \|[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n] - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]\|}.$$

Wir nutzen das Resultat aus (C.0.65)

$$\|[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n \Sigma_0^{-1} \mathbf{B}_n]^{-1}\|_S = O(p^{-1}). \quad (\text{D.0.8})$$

Für den verbleibenden Ausdruck $\|\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n - \text{cov}(\mathbf{f})^{-1} - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\|$ müssen wir die veränderte Rate der Stichprobenkovarianzmatrix der Faktoren beachten. Dafür schreiben wir mit dem Resultat aus Lemma D.0.16

$$\begin{aligned} \|\text{côv}(\mathbf{f})^{-1} - \text{cov}(\mathbf{f})^{-1}\| &= \|(\text{cov}(\mathbf{f}) + \mathbf{D}_n)^{-1} - \text{cov}(\mathbf{f})^{-1}\| \leq \frac{\|\mathbf{D}_n\| \|\text{cov}(\mathbf{f})^{-1}\|_S^2}{1 - \|\text{cov}(\mathbf{f})^{-1}\|_S \|\mathbf{D}_n\|} \\ &= O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} K). \end{aligned}$$

Wir erhalten dann mit (C.0.67)

$$\begin{aligned} &\|\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n - \text{cov}(\mathbf{f})^{-1} - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\| \\ &\leq \|\text{côv}(\mathbf{f})^{-1} - \text{cov}(\mathbf{f})^{-1}\| + \|\hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n - \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n\| = O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} K) + O_P(n^{-\frac{1}{2}} p^{\frac{3}{2}} K^{\frac{1}{2}}). \end{aligned}$$

Während der Nenner in (D.0.7) für $n \rightarrow \infty$ gegen 1 konvergiert, gilt mit (D.0.8) für den Zähler

$$\begin{aligned} &\|[\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n] - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]\| \|[\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1}\|_S^2 \\ &= (O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} K) + O_P(n^{-\frac{1}{2}} p^{\frac{3}{2}} K^{\frac{1}{2}})) O_P(p^{-2}) \\ &= O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} p^{-2} K) + O_P(n^{-\frac{1}{2}} p^{-\frac{1}{2}} K^{\frac{1}{2}}). \end{aligned}$$

Zusammen mit (D.0.6) erhalten wir das Resultat in (D.0.5)

$$\begin{aligned} &\|\Sigma_0^{-1} \mathbf{B}_n ([\text{côv}(\mathbf{f})^{-1} + \hat{\mathbf{B}}_n^T \hat{\Sigma}_0^{-1} \hat{\mathbf{B}}_n]^{-1} - [\text{cov}(\mathbf{f})^{-1} + \mathbf{B}_n^T \Sigma_0^{-1} \mathbf{B}_n]^{-1}) \mathbf{B}_n^T \Sigma_0^{-1}\| \\ &= O_P(n^{-\frac{1}{2}} [\log n]^{\frac{1}{2}} p^{-1} K) + O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}} K^{\frac{1}{2}}). \end{aligned}$$

□

Bemerkung D.0.21 Da wir in (D.0.4) die gleiche Rate für die Matrix der Faktorladungen erhalten wie in (C.0.47) und der Beweis von Lemma C.0.5 nicht von Annahme 8 abhängt, können wir auch für den Fall stationärer Faktoren die Raten der Lemmata C.0.12 - C.0.14 nutzen.

Literaturverzeichnis

- C. Alexander. *Market Models: A Guide to Financial Data Analysis*. John Wiley and Sons, Ltd, Chichester, 2001.
- H. Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter, Berlin, 2002. ISBN 3-11-017236-4.
- D. Bernstein. *Matrix Mathematics: Theory, Facts and Formulas*. Princeton University Press, 2009. ISBN 978-0-691-14039-1.
- P. Brockwell und R. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer, 2009. ISBN 9781441903198.
- E. Burmeister, R. Grinold, und R. N. Kahn. *Multiple-Factor Models for Portfolio Risk*, pages 59–80. Research Foundation of the Institute of Chartered Financial Analysts, 1994. doi: 10.2470/rf.v1994.n4.4445.3.
- T. Cai und W. Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011. doi: 10.1198/jasa.2011.tm10560. URL <http://www.tandfonline.com/doi/abs/10.1198/jasa.2011.tm10560>.
- T. Cai, W. Liu, und X. Luo. A constrained l1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011. doi: 10.1198/jasa.2011.tm10155. URL <http://amstat.tandfonline.com/doi/abs/10.1198/jasa.2011.tm10155>.
- J. Y. Campbell, A. W. Lo, und A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, New Jersey, 1997. ISBN 0-691-04301-9.
- G. Chamberlain und M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. Working Paper 996, National Bureau of Economic Research, 1984. URL <http://www.nber.org/papers/w0996>.
- E. F. Fama und K. R. French. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–65, June 1992. URL <http://ideas.repec.org/a/bla/jfinan/v47y1992i2p427-65.html>.
- E. F. Fama und K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, February 1993. URL <http://ideas.repec.org/a/eee/jfinec/v33y1993i1p3-56.html>.

- J. Fan, Y. Fan, und J. Lv. High Dimensional Covariance Matrix Estimation Using a Factor Model. *ArXiv Mathematics e-prints*, Jan. 2007.
- J. Fan, Y. Liao, und M. Mincheva. High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.*, 39:3320–3356, 2011. ISSN 0090-5364. doi: doi:10.1214/11-AOS944. URL <http://projecteuclid.org/euclid.aos/1330958681>.
- O. Forster. *Analysis 1 9.Auflage*. Vieweg + Teubner, Wiesbaden, 2008.
- I. Friend. *Risk and Return in Finance Vol.1*. Ballinger, Cambridge, MA, 1977. ISBN 0-88410-653-5.
- A. Goldberger. *A Course in Econometrics*. Harvard University Press, Cambridge, 1991. ISBN 0-674-17544-1.
- G. Golub und C. F. Van Loan. *Matrix Computations, Second edition*. The Johns Hopkins University Press, Baltimore and London, 1989.
- R. Horn und C. Johnson. *Matrix Analysis*. Cambridge University Press, 1999. ISBN 0-521-30587-X.
- W. James und C. Stein. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*, 1961.
- K. Jänich. *Lineare Algebra*. Springer, Berlin Heidelberg, 2008. ISBN 3-540-75502-0.
- T. Kollo und D. von Rosen. *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht, Berlin, Heidelberg, New York, 2005. ISBN 1-4020-3418-0.
- J.-P. Kreiß und G. Neuhaus. *Einführung in die Zeitreihenanalyse*. Springer Berlin Heidelberg, 2006.
- B. Lenze. *Basiswissen Analysis*. w3l GmbH Herdecke, 2006.
- H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1952.tb01525.x. URL <http://dx.doi.org/10.1111/j.1540-6261.1952.tb01525.x>.
- A. McNeil, R. Frey, und P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton Series in Finance, Princeton University Press, 2005.
- F. Merlevède, M. Peligrad, und E. Rio. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *ArXiv e-prints*, Feb. 2009.
- S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, December 1976. URL <http://ideas.repec.org/a/eee/jetheo/v13y1976i3p341-360.html>.
- S. A. Ross. Return, risk and arbitrage. Technical report, 1977.

- R. Tsay. *Analysis of Financial Time Series*. Wiley series in probability and statistics, 2010. ISBN 978-0-470-41435-4.
- A. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. ISBN 0-521-78450-6.
- E. Zivot und J. Wang. *Modeling Financial Time Series with S-Plus*. Springer New York, 2006.

Erklärung

Hiermit versichere ich, Anna Maria Leister, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Stellen, die im Wortlaut oder sinnesgemäß anderen Werken entnommen sind, habe ich unter Angabe der Quellen kenntlich gemacht. Diese Arbeit hat in dieser oder ähnlicher Form noch nicht im Rahmen einer anderen Prüfung vorgelegen.

Marburg, 17. Januar 2013