

Robuste ANOVA-Typ-Statistik in hochdimensionalen Mehrgruppendedesigns

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm Mathematical Sciences

der Georg-August University School of Science (GAUSS)

vorgelegt von

Benjamin Markus Becker

aus Stuttgart

Göttingen, 2020

Betreuungsausschuss

Prof. Dr. Edgar Brunner, Institut für Medizinische Statistik

Prof. Dr. Tatyana Krivobokova, Institut für Mathematische Stochastik

Mitglieder der Prüfungskommission

Prof. Dr. Edgar Brunner, Institut für Medizinische Statistik

Prof. Dr. Tatyana Krivobokova, Institut für Mathematische Stochastik

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Dorothea Bahns, Mathematisches Institut

Jun.-Prof. Dr. Madeleine Jotz Lean, Mathematisches Institut

Prof. Dr. Gerlind Plonka-Hoch, Institut für numerische und angewandte Mathematik

Prof. Dr. Ingo Frank Witt, Mathematisches Institut

Tag der mündlichen Prüfung:

03. September 2020

Danksagung

An erster Stelle gebührt Herrn Prof. Dr. Edgar Brunner der Dank, mich bei der Erstellung dieser Arbeit betreut zu haben mit vielen wertvollen Hinweisen, klärenden Fachgesprächen, Geduld und Vertrauen. Ich bin ihm auch sehr dankbar für sein „Nachwuchsendagement“, durch das ich überhaupt erst auf das faszinierende Gebiet der Biostatistik aufmerksam wurde. Später hat er mich wissenschaftlich vorangebracht durch die Ermutigung zu reger Konferenzteilnahme, durch die Schulung von Vortragsfähigkeiten und durch seinen Einsatz, daß mir materielle Arbeitsmöglichkeiten am Institut für Medizinische Statistik zum Zwecke der Forschung und der Übernahme von Lehr- und Beratungstätigkeiten bereitgestellt wurden.

Ich danke auch sehr Frau Prof. Dr. Tatyana Krivobokova für die Übernahme des Korreferats und manche inspirierende Hinweise und Fragen in den Betreuungsgesprächen.

Herrn Prof. Dr. Tim Friede danke ich ebenfalls für die wissenschaftliche Unterstützung als Institutsdirektor. Den übrigen ehemaligen und aktuellen Kollegen am IMS danke ich für die harmonische und anregende Arbeitsatmosphäre und die vielen kleinen Fachdiskussionen und Unterhaltungen zwischen Schreibtisch und Mensa.

Bei Frau Dr. Julia Hofhuis bedanke ich mich für die Bereitstellung des zweiten Beispieldatensatzes mit den Mäusen und die Erlaubnis, ihn nutzen zu dürfen.

Den Entwicklern von den diversen hilfreichen, freien Softwarepaketen, die mir ganz viel Arbeit erleichtert oder erst ermöglicht haben danke ich auch.

Und allen Dank, die mich ermutigt haben!

Inhaltsverzeichnis

1	Einleitung	1
1.1	Übersicht	1
1.2	Notationen	2
1.3	Modellnotation und Normalverteilung	3
1.4	Fragestellung und Hypothesen	4
1.5	Vorstellung zweier Anwendungsbeispiele	5
1.5.1	Gewichtsentwicklung von Ratten bei unterschiedlichen Medi- kationen	5
1.5.2	Laufleistung von Mäusen mit und ohne Gendefekt	6
1.6	Robustheit bezüglich der Normalverteilungsannahme	8
1.6.1	Vorbemerkungen	8
1.6.2	Rahmenmodell einer allgemeinen Regularitätsbedingung	10
1.6.3	Additiv-lineares Modell	10
1.6.4	Elliptisches Modell	12
1.7	Bekannte Statistiken als Ausgangspunkt	13
1.7.1	ANOVA und Geisser-Greenhouse-Statistik	13
1.7.2	ANOVA-Typ-Statistik bei Normalverteilung	15
1.7.3	Robustheit der ANOVA-Typ-Statistik im Einstichprobenfall	18
1.8	Andere Ansätze	20
1.8.1	g-Inverse der empirischen Kovarianzmatrix	20
1.8.2	Multivariater Ansatz	22
1.8.3	Asymptotische Ansätze	22
2	Robuste Mehrstichprobenstatistiken	25
2.1	Skizze des Vorgehens	25
2.2	Hochdimensionales Behrens-Fisher-Problem	26
2.3	Robuste Verteilungsapproximation des Quotienten	28
2.4	Kritische Werte	32
2.5	Simulationen	33
2.5.1	Simulationslayout	33
2.5.2	Simulation der Taylor-Approximation	38
2.5.3	Niveausimulation ohne Schätzer	39

3	Freiheitsgradschätzungen	43
3.1	Historischer Abriß bei hoher Dimensionalität	43
3.2	Dimensionsstabilität	45
3.3	Dimensionsstabile Schätzer aus U -Statistiken	47
3.4	Huynh-Feldt- und Plug-in-Schätzer	50
3.5	Simulationen mit den Schätzern	52
4	Korrektur für das Einsetzen eines Schätzers	56
4.1	Verallgemeinerung des Box'schen Epsilon	56
4.2	Korrekturformel für die kritischen Werte	57
4.3	Abschätzung für $\text{Var}_{\mathcal{N}}(B_n)$	59
4.4	Zweite Ableitung der Quantilfunktion nach dem ersten Freiheitsgrad .	61
4.5	Korrekturformel für die p -Werte und geeignete Ausgabe	64
5	Simulation mit Schätzern und Korrektur	66
5.1	Niveau	66
5.2	Power	68
6	Auswertung der Anwendungsbeispiele	71
6.1	Gewichtsentwicklung der Ratten	71
6.2	Laufleistung der Mäuse	72
7	Zusammenfassung, Diskussion und Ausblick	74
A	Beweise	77
A.1	Bekannte Sätze	77
A.2	Beweise bis zum Kapitel 2 Robuste Mehrstichprobenstatistiken	78
A.2.1	Beweise über strukturierte quadratische Formen	78
A.2.2	Beweise über Approximationen und Verteilungen	82
A.3	Beweise zum Kapitel 3 Freiheitsgradschätzungen	85
A.3.1	Sätze über beide Modelle zugleich	85
A.3.2	Sätze für das additiv-lineare Modell	87
A.3.3	Sätze für das elliptische Modell	90
A.3.4	Zusammensetzen der Dimensionsstabilitätsbeweise	93
A.3.5	Modifizierte Huynh-Feldt- und Plug-in-Schätzer	97

A.4	Beweise zum Kapitel 4 Korrektur	100
A.4.1	Exakte Berechnung von $\text{Var}_{\mathcal{N}}(B_n)$	100
A.4.2	Ableitungen der F_F -Quantilfunktion und Integrierbarkeit . . .	106
B	Simulierte Konstellationen	110
B.1	Konstellationen mit identischen Kovarianzmatrizen.	110
B.2	Konstellationen aus zwei Stichproben bei unterschiedlichen Kovari- anzmatrizen.	110
B.3	Konstellationen aus drei und vier Stichproben bei unterschiedlichen Kovarianzmatrizen.	111
C	Literatur	112

1 Einleitung

1.1 Übersicht

Bei vielen Experimenten in den Biowissenschaften werden an den Versuchseinheiten (Probanden, Versuchstiere, Ackerparzellen, Kulturen von Pflanzen, Bakterien oder Pilzen) mehrere gleichartige Messungen an unterschiedlichen Zeitpunkten oder zu unterschiedlichen Bedingungen erhoben, teils als primär interessierende Größe, teils zu Kontrollzwecken oder für sekundäre Fragestellungen. Zu solchen Versuchslayouts gehören *Split-Plot-Designs*, *Meßwiederholungen* (Repeated Measures) oder *longitudinale Daten*.

Hierbei ist es oft der Fall, daß die Zahl der abhängigen Versuchseinheiten nicht größer ist als die Zahl der untereinander stochastisch abhängigen Messungen. Dann spricht man von *hochdimensionalen Daten*. Hochdimensionale Daten erfordern spezielle Auswertungsverfahren, weil die Daten in einem echten Unterraum des kartesischen Produktes der Messungen streuen, Varianzen orthogonal zu diesem Unterraum also scheinbar verschwinden. Mathematisch zeigt sich das in singulären Kovarianzmatrizen, so daß viele etablierte Verfahren nicht mehr definiert sind.

Es gibt allerdings verwandte Ansätze, mit denen Hypothesen über hochdimensionale Daten getestet werden können, bei denen dieselbe Größe auf derselben Skala gemessen wurde. Die Tests verwenden die euklidische Norm der linear transformierten Mittelwertsvektoren, so daß nur die Summe der Varianzen benötigt wird. Technisch werden dazu die Tails von quadratischen Formen in den Beobachtungsvektoren durch Chiquadratverteilungen approximiert. Die Chiquadratverteilung wird durch die Modellannahme normalverteilter Daten gerechtfertigt. Diese Tests funktionieren in einem weiten Parameterbereich recht gut.

Sehr selten wurde in der Literatur behandelt, wie diese Verfahren sich verhalten, wenn in Wahrheit keine Normalverteilung vorliegt, was man in der Realität selten genau wissen kann. Statistiken, die sowohl in der idealen Modellannahme Optimalitätseigenschaften zu haben (z. B. beste Power bei Tests, minimaler mittlerer quadratischer Fehler bei Schätzern), und bei moderater Abweichung von den Annahmen noch akzeptabel sind, nennt man *robust*.

Etwas mehr Arbeiten gibt es zum *Behrens-Fisher-Problem* im Kontext hochdimensionaler Daten. Vom Behrens-Fisher-Problem spricht man, wenn zwischen unterschiedlich großen Stichproben unterschiedliche Kovarianzmatrizen vorliegen. In diesem Fall gibt es keine exakten Tests mit sinnvollen Eigenschaften. Statt dessen gibt es approximative Tests, die ebenfalls auf approximativen Verteilungen in den Tails beruhen.

Die vorliegende Arbeit befaßt sich mit dem Behrens-Fisher-Problem und zugleich mit Robustheit bei hochdimensionalen Daten. Insbesondere ist zu untersuchen, ob die Approximation für das Behrens-Fisher-Problem Robustheitseigenschaften fördert. Hierzu wird im nächsten Kapitel in Abschnitt 2.2 zunächst gefunden, welche

Form ein Test für das hochdimensionale Behrens-Fisher-Problem haben muß. In Abschnitt 2.3 wird dann eine approximative hochdimensionale Statistik in einem Modell hergeleitet, das das Normalverteilungsmodell verallgemeinert. Es stellt sich heraus, daß Störparameter aus dem verallgemeinerten Modell genau dann verschwinden, wenn der Test die für das Behrens-Fisher-Problem geeignete Form hat. Im Rest des Kapitels 2 wird gezeigt, daß mit der hergeleiteten robusten Statistik automatisch auch eine andere etablierte Statistik robust ist. Dies wird anschließend in Simulationen gezeigt, in denen allerdings Modellparameter dort eingesetzt werden, wo in der Realität Schätzer benötigt werden. Diese Parameter geben an, wie unterschiedlich die Komponenten der Beobachtungsvektoren streuen. Es sind daher sogenannte *Sphärizitätsparameter*.

Mit der robusten Schätzung dieser unbekanntem Sphärizitätsparameter bei hoher Dimensionalität befaßt sich Kapitel 3. Die betrachteten Nichtnormalverteilungen sind hier bestimmte Verallgemeinerungen der multivariaten Normalverteilung. Es wird mit zwei bekannten Schätzerpaaren gearbeitet. Das eine Paar von Schätzern hat die Form von U -Statistiken. Es ist bei Nichtnormalverteilung erwartungstreu und hat Variationskoeffizienten, die nicht von der Dimensionalität abhängen. Das andere Paar ist auf Normalverteilung ausgelegt und dort erwartungstreu. Mit den Schätzern aus U -Statistiken als existierender Majorante wird gezeigt, daß die Schätzer für Normalverteilung dort optimal hinsichtlich des mittleren quadratischen Fehlers sind. In Simulationen wird anschließend gezeigt, daß die Schätzer aus U -Statistiken bei Normalverteilung nur vernachlässigbar wenig schlechter sind, dafür aber bei Nichtnormalverteilung deutlich besser, also robust sind.

Das Kapitel 4 ist dem Phänomen gewidmet, daß die bekannten hochdimensionalen Tests –auch die robusten– mit erwartungstreuen Schätzern für die Sphärizitätsparameter in einem bestimmten Bereich leicht liberal sind. Dies liegt daran, daß die Schätzer zwar dimensionsstabil und konsistent sind, die Konvergenz aber von der liberalen Seite her geschieht, finit also die Schätzer zuviel Varianz haben. Da dies bei Nichtnormalverteilungen noch schlimmer werden kann, wird dieses Phänomen beschrieben und das Konzept einer Korrektur für das Einsetzen der robusten Schätzer ausgearbeitet.

Im Kapitel 5 wird mit Simulationen gezeigt, daß sich die robusten Schätzer bewähren und die Korrektur sowohl Niveauüberschreitungen dämpft, wie auch die Power nicht nachteilig beeinflusst. Danach werden in Kapitel 6 zwei Anwendungsbeispiele ausgewertet, die in diesem Kapitel zusammen mit der Notation, den interessierenden Hypothesen, den Modellen und einem Überblick der existierenden Literatur vorgestellt werden. Im Abschlußkapitel 7 wird das Erreichte zusammengefaßt und ein Ausblick auf mögliche Weiterentwicklungen gegeben.

1.2 Notationen

Matrizen werden mit fetten Großbuchstaben bezeichnet, Vektoren mit fetten Kleinbuchstaben und Skalare mit normalen Kleinbuchstaben. Skalare Zufallsvariablen

wiederum werden mit normalen Großbuchstaben bezeichnet, vektorwertige Zufallsvariablen mit fetten Großbuchstaben. Blockmatrizen werden folgendermaßen notiert:

$$[\mathbf{A}_{ii'}]_{\substack{i=1,\dots,a \\ i'=1,\dots,a'}} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{1a'} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & & \\ \vdots & & \ddots & \\ \mathbf{A}_{a1} & & & \mathbf{A}_{aa'} \end{bmatrix}$$

Der Operator für das *Kroneckerprodukt* wird bezeichnet durch \otimes und es besagt

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \\ \vdots & & \ddots \end{bmatrix}.$$

Der Operator für die *direkte Summe* ist \oplus :

$$\mathbf{A}_1 \oplus \mathbf{A}_2 = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix}$$

Besondere Matrizen sind die $k \times k$ -*Identitätsmatrix* \mathbf{I}_k mit dem Eintrag 1 auf der Hauptdiagonalen, sonst 0, die „Einsenmatrix“ $\mathbf{J}_k = [1]_{j,j'=1,\dots,k}$, eine $k \times k$ -Matrix nur aus 1. Eine $k_1 \times k_2$ -Matrix mit demselben Eintrag ist $\mathbf{J}_{k_1 \times k_2}$. Außerdem ist die sogenannte *zentrierende Matrix* $\mathbf{P}_k = \mathbf{I}_k - \frac{1}{k}\mathbf{J}_k$ wichtig.

Verschiedenheit von mehreren Indices wird statt durch $k \neq l, l \neq t, t \neq k$ in „polnischer Notation“ durch $\neq k, l, t$ geschrieben.

Das Zeichen $\dot{\sim}$ bedeutet, daß in den Tails approximativ die linke Seite der Verteilung auf der rechten Seite folgt. Mit dem Zeichen \doteq wird Gleichheit im Erwartungswert einer Taylorapproximation erster oder zweiter Ordnung bezeichnet.

1.3 Modellnotation und Normalverteilung

Allgemein gebe es a Stichproben mit n_i unabhängigen Versuchseinheiten in Stichprobe i . Zur Abkürzung sei $N = \sum_{i=1}^a n_i$ der Gesamtstichprobenumfang.

Das Verteilungsmodell für die *multivariat normalverteilte* Beobachtung k in Stichprobe i lautet damit

$$\mathbf{Y}_{ik} \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad (1)$$

wobei $\boldsymbol{\mu}_i \in \mathbb{R}^d$ der Erwartungswertvektor und $\boldsymbol{\Sigma}_i$ eine symmetrische positivsemidefinite Matrix, die Kovarianzmatrix, ist. $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{in_i}$ sind unabhängig identisch verteilte Vektoren von Beobachtungen.

Strenggenommen bedeutet multivariat normalverteilt, daß $\boldsymbol{\Sigma}_i$ positiv definit sein

muß. Dann ist die Dichte

$$\frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma}_i)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i (\mathbf{x} - \boldsymbol{\mu}_i)\right)$$

und jede Projektion der Zufallsvektoren auf einen Unterraum gehört derselben Verteilungsfamilie an. Dies ist eine weit verbreitete Definition der multivariaten Normalverteilung. Hier aber soll von multivariater Normalverteilung auch dann gesprochen werden, wenn eine Verteilung nur in einem Teilraum des \mathbb{R}^d multivariat normalverteilt ist.

Eine Voraussetzung für die weitere Arbeit ist, daß nicht alle $\boldsymbol{\Sigma}_i = \mathbf{0}$ sein dürfen, d. h. in mindestens einer Stichprobe muß wenigstens irgendeine Art von Varianz vorliegen. Diese Voraussetzung an die Kovarianzmatrizen ist sehr schwach, denn wenn sie nicht erfüllt ist, ist sowieso nichts zu testen.

Wenn einer Varianz- oder Erwartungswertformel ein normalverteiltes Modell zugrundeliegt, wird $\text{Var}_{\mathcal{N}}$ bzw. $E_{\mathcal{N}}$ geschrieben.

1.4 Fragestellung und Hypothesen

Die Faktoren, die über die $i = 1, \dots, a$ unabhängigen Stichproben variieren, werden *Wholeplotfaktoren* genannt. *Subplotfaktoren* sind Faktoren, die die d Messungen je Versuchseinheit strukturieren. Zu dem Modell verwandt sind longitudinale Daten und Repeated measures (Meßwiederholungen). Von Meßwiederholungen wird gesprochen, wenn $a = 1$ ist, also alle Beobachtungsvektoren identisch verteilt sind. Bei $i \geq 2$ wird das Versuchslayout ein Split-Plot-Layout genannt (Sattler und Pauly (2018)).

Mit der Schreibweise $\boldsymbol{\mu} := (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_a)'$ werden lineare Hypothesen der Form $\mathbf{H}\boldsymbol{\mu} = \mathbf{0}$ betrachtet. $\mathbf{H} := \mathbf{H}_a \otimes \mathbf{H}_d$ ist hierbei die Form der Hypothesenmatrix für die gekreuzte Struktur aus a Stufen(kombinationen) aller Wholeplotfaktoren und d Stufen(kombinationen) aller Subplotfaktoren. Die Matrix \mathbf{H}_a hat also a Spalten und \mathbf{H}_d hat d Spalten.

Lineare Hypothesen, deren Hypothesenmatrix nicht dergestalt als Kroneckerprodukt geschrieben werden kann, können mit den Methoden der vorliegenden Arbeit nicht getestet werden, denn solche Hypothesen würden nicht zur gekreuzten Struktur der Wholeplot- und Subplotfaktoren passen.

Stichprobe i nennt man eine *hochdimensionale Stichprobe*, wenn $n_i + 1 \leq d$ ist, wenn also ihre Beobachtungsvektoren einen echten linearen Unterraum des Raumes aufspannen, in dem die Beobachtungen liegen. Ein *hochdimensionaler Test* ist entsprechend ein Test zu einer Hypothese \mathbf{H} , wenn

$$1 + \sum_{i=1}^a n_i \leq \text{rg}(\mathbf{H})$$

ist, wenn also die Dimensionalität des Spaltenraumes von \mathbf{H} größer ist als die Dimensionalität des Raumes, der von den Beobachtungsvektoren aufgespannt wird. In dieser Arbeit sollten Tests ausgearbeitet werden, die hochdimensional sein können.

Es ist bekanntlich $\mathbf{T}_a = \mathbf{H}'_a (\mathbf{H}_a \mathbf{H}'_a)^{-1} \mathbf{H}_a$ symmetrisch und idempotent und eindeutig für alle \mathbf{H}_a , die Hypothesenmatrizen zu äquivalenten Hypothesen sind. Dasselbe gilt für $\mathbf{T}_d = \mathbf{H}'_d (\mathbf{H}_d \mathbf{H}'_d)^{-1} \mathbf{H}_d$, siehe Searle und Gruber (2017). $\mathbf{T} = \mathbf{T}_a \otimes \mathbf{T}_d$ ist dann automatisch auch idempotent und symmetrisch. Es werde $\mathbf{T}_a = [t_{ii'}]_{i,i'=1,\dots,a}$ für die Einzelelemente von \mathbf{T}_a geschrieben.

Typischerweise relevante Hypothesen sind in genau dieser Schreibweise in Happ, Harrar und Bathke (2016) aufgeführt:

1. „Kein Haupteffekt des Wholeplotfaktors (auf den Mittelwert über die Subplot-Faktorstufen)“ $H_0(A) : (\mathbf{P}_a \otimes \frac{1}{a} \mathbf{J}_d) \boldsymbol{\mu} = \mathbf{0}$ — Diese Hypothese ist nicht hochdimensional. Für sie wird daher nichts weiter ausgearbeitet.
2. „Kein Subploteffekt“ $H_0(B) : (\frac{1}{a} \mathbf{J}_a \otimes \mathbf{P}_d) \boldsymbol{\mu} = \mathbf{0}$
3. „Keine Wechselwirkung zwischen Wholeplot- und Subploteffekt“: $H_0(AB) : (\mathbf{P}_a \otimes \mathbf{P}_d) \boldsymbol{\mu} = \mathbf{0}$
4. „Kein Effekt der Wholeplotfaktoren auf die Subplots als Ganzes (Zeitverläufe)“: $H_0(A|B) : (\mathbf{P}_a \otimes \mathbf{I}_d) \boldsymbol{\mu} = \mathbf{0}$

Es sind noch mehr Hypothesen denkbar, wenn auf die Wholeplots nicht ein Faktor sondern mehrere wirken, ebenso können auch die Subplots mehrfaktoriell strukturiert sein. In diesen Fällen sind \mathbf{T}_a bzw. \mathbf{T}_d ihrerseits Kroneckerprodukte.

1.5 Vorstellung zweier Anwendungsbeispiele

1.5.1 Gewichtsentwicklung von Ratten bei unterschiedlichen Medikationen

In einem Experiment wurde die Gewichtsentwicklung von Wistarratten erhoben, denen von einem Wirkstoff eine unterschiedliche Dosis je Gramm ihres Körpergewichts täglich über einen Zeitraum von 22 Meßtagen verabreicht wurde. Die Versuchstiere waren hierbei in drei Gruppen zu jeweils derselben Dosis eingeteilt.

In diesem Layout ist die Dosisgruppe der Wholeplot-Faktor, während die Zeit einen Subplot-Faktor darstellt. Die zentrale Fragestellung ist, ob die Dosierung einen Unterschied auf die Gewichtsentwicklung hat. Dazu sind zwei verschiedene Hypothesen denkbar.

Wenn $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ und $\boldsymbol{\mu}_3$ die Vektoren der erwarteten Gewichtsverläufe in den drei Gruppen sind, kann zum einen auf deren Gleichheit getestet werden, also die Hypothese

$$H_0 : (\mathbf{P}_3 \otimes \mathbf{I}_{22}) \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_3 \end{pmatrix} = \mathbf{0}.$$

Diese Hypothese wird schon dann abgelehnt, wenn die Gruppen parallele Gewichtsverläufe haben, sich also nur in dem Baselinewerten unterscheiden. Will man eine Wechselwirkung zwischen Gewicht und Dosierung untersuchen, dann ist die Hypothese

$$H_0 : (\mathbf{P}_3 \otimes \mathbf{P}_{22}) \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_3 \end{pmatrix} = \mathbf{0}$$

geeigneter.

Ein geeigneter Test für diese Hypothesen muß mit den Eigenschaften verträglich sein, die die Verteilungen der Gewichtsmessungen haben. Gedanklich kann man die Eigenschaften zerlegen in die Abhängigkeitsstruktur zwischen den Meßwiederholungen und die Verteilung der einzelnen Messungen.

Zwischen den Meßwiederholungen an denselben Versuchseinheiten kann keine Unabhängigkeit angenommen werden. Unbekannte Abhängigkeitsparameter fallen also an. Weiterhin kann nicht ausgeschlossen werden, daß die Art der Abhängigkeiten sich in den Gruppen unterscheiden. Auch die Varianzen der einzelnen Gewichte müssen nicht identisch zwischen den Gruppen sein.

Bei normalverteilten Daten würde dieser Sachverhalt durch die Annahme unterschiedlicher Kovarianzen zwischen den Meßwiederholungen und wiederum unterschiedliche Kovarianzmatrizen zwischen den Gruppen abgebildet.

Bezüglich der Verteilung der einzelnen Messungen hat es eine gewisse Plausibilität, daß Gewichte in guter Näherung normalverteilt sind, weil argumentiert werden kann, daß das Gewicht zu einem festen Zeitpunkt die Summe aller vorhergehenden Gewichtszunahmen ist, welche ihrerseits Voraussetzungen für gewisse Formen von Grenzwertsätzen erfüllen können. Entsprechend wurde dieser Datensatz auch schon von Harden (2013) unter Annahme der Normalverteilung ausgewertet.

Allerdings müssen die Gewichtszunahmen ihrerseits nicht zwingend normalverteilt sein. Genau diese Zunahmen gehen aber bei der zweiten Hypothese in die Auswertung ein, so daß die Frage nach Verträglichkeit mit der Nichtnormalverteilungsannahme auch bei Gewichtsdaten von Belang ist.

1.5.2 Laufleistung von Mäusen mit und ohne Gendefekt

Dem nächsten Anwendungsbeispiel liegen die täglichen Laufleistungen von jungen Mäusen zweier Genotypen in einem Laufrad zugrunde (siehe Hofhuis (2013)). Die Kontrollgruppe waren gewöhnliche Labormäuse („Wildtyp“). Die anderen Mäuse waren Knock-out-Mäuse, bei denen ein Gen deaktiviert war, das für die Expression von Dysferlin wichtig ist, einem Muskelprotein. Fehler in Dysferlin-Genen spielen auch bei seltenen menschlichen Erberkrankungen eine Rolle, bei denen sich ab dem Jugendalter eine nicht-tödliche Muskeldystrophie entwickelt. Besonders an dieser speziellen Muskeldystrophie ist, daß sich in sehr jungen Jahren der Patienten zunächst eine geringere Ermüdungsneigung zeigt. Die Knock-out-Mäuse sollen als

Modellorganismen für diese Art von Erkrankungen dienen, und im Versuch soll untersucht werden, ob sich das Krankheitsbild menschlicher Patienten bei den Mäusen phänotypisch widerspiegelt.

Aus Kapazitätsgründen konnte nicht allen Mäusen für die gesamte Versuchsdauer von 93 Wochen ein eigener Käfig mit Laufrad und Umdrehungszähler vorgehalten werden. Deswegen wurden die Mäuse in größeren Zeitabständen ab Geburt für jeweils drei Wochen in einen Käfig gesetzt, in denen ihre Lauflistung gemessen werden konnte. Die Meßwiederholungen können deswegen durch zwei feste und miteinander gekreuzte Faktoren modelliert werden: Ein Faktor ist der Tag nach dem Einsetzen in den Laufradkäfig, der andere Faktor ist der „Langzeitfaktor“, in welcher Lebenswoche die Maus das Laufrad bekam.

Bei dem Versuch traten fehlende Werte auf, deren Ursachen allesamt geklärt werden konnten und als vollständig zufällig (MCAR) vom Versuchsziel betrachtet werden konnten, also stochastisch unabhängig vom Mausphänotyp sind; meist handelte es sich um Ausfälle der Meßapparatur. An 9 KO-Mäusen und 10 Mäusen vom Wildtyp liegen jedoch vollständige Beobachtungen über jeweils 18 Tage vor, die im Alter von 4, 12 und 40 Wochen erhoben wurden. Die Rohdaten der 19 Mäuse sind in Abbildung 1 abgebildet, indem die Lauflistung jeder Maus gegen den Versuchstag (näherungsweise) ab Geburt der Maus aufgetragen ist. Die längeren Erhebungspausen nach den Enden der Dreiwochenabschnitte sind mit gestrichelten Vertikalen markiert.

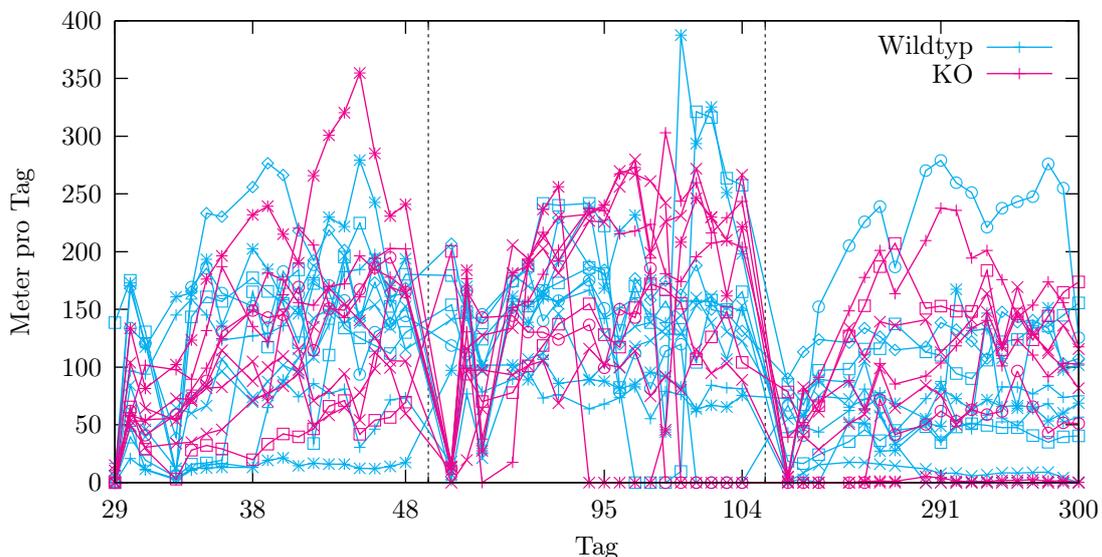


Abbildung 1: Tägliche Lauflistung der Mäuse

Die generelle Fragestellung ist, ob der Genotyp einen Effekt auf die Änderung der Lauflistung hat. Dies kann man am besten testen, indem man alle 54 Meßzeitpunkte einfaktoruell gegliedert betrachtet.

Wenn man mit $d_1 = 3$ die drei langen Meßzeiträume und mit $d_2 = 18$ die Tagesmessungen innerhalb der drei Zeiträume bezeichnet, dann kann man auch eine Wechselwirkung zwischen den Meßzeiträumen und den täglichen Änderungen betrachten

durch eine Hypothesenmatrix $\mathbf{T}_3 \otimes \mathbf{T}_{18}$, also in zweifaktorieller Gliederung der Messungen. Dies wäre eine sekundäre Fragestellung nach einem Gewöhnungseffekt, d. h. ob die tägliche Änderung der Laufleistung sich zwischen den Meßzeiträumen unterscheidet.

Tägliche Laufleistungen können nicht als Bestandsgrößen wie z. B. das Körpergewicht aufgefaßt werden, für die sich eine plausible Begründung der Normalverteilungsannahme finden ließe. Gegen Normalverteilung spricht außerdem eine Dominanz der Null, denn vor allem ab dem Tag 94 bewegten einige Tiere ihr Laufrad mindestens einen ganzen Tag lang nicht. Dies ist ein deutliches Zeichen, daß keine normalverteilten Daten vorliegen, also unbedingt mit robusten Verfahren getestet werden sollte.

1.6 Robustheit bezüglich der Normalverteilungsannahme

1.6.1 Vorbemerkungen

Die Intention des Robustheitsbegriffs ist es, Aussagen zu treffen, wie empfindlich Tests (oder Schätzer) auf Abweichungen von ihren ursprünglichen Herleitungsannahmen reagieren. Diese Frage ist für die Praxis sehr wichtig, weil in der Realität nur sehr selten argumentiert werden kann, ob und warum bestimmte Verteilungsannahmen bei einem Datensatz zwingend zutreffen.

Robustheit soll anschaulich besagen, daß „ein Verfahren im ursprünglichen Modell genau / effizient ist und kleine Abweichungen von den Modellannahmen nur zu kleinem Verlust an Genauigkeit / Effizienz führen“ (Huber (1981), S. 5). Arbeiten über die Robustheit statistischer Verfahren konstruieren gewöhnlich keine völlig neuen Methoden, sondern beweisen eher die Robustheitseigenschaft eines bekannten Verfahrens, von dem schon im seinem ursprünglichen Modell eine gewisse Optimalität bekannt ist. Für die mathematische Behandlung ist aber eine genauere Definition der Robustheitseigenschaft nötig. Davon gibt es in der Literatur mehrere.

Eine Möglichkeit, Robustheit zu behaupten liegt darin, Modellerweiterungen zu finden, in denen die Statistik ihre Eigenschaften behält, die sie im ursprünglichen, „normalen“ Modell hat. Beispiele sind Kariya und Sinha (1989) und insbesondere Kariya (1981). In letzterer Arbeit wird die multivariate Normalverteilung durch ein sogenanntes elliptisches Modell –siehe unten im Abschnitt 1.6.4– verallgemeinert und gezeigt, daß Hotellings T^2 -Statistik ihre Eigenschaften im allgemeineren Modell behält.

Ein anderer Robustheitsbegriff ist „qualitative robustness“ (Huber (1981)). Sie liegt vor, wenn die Verteilung einer Statistik stetig in der Verteilung der Daten ist, d. h. leicht abweichende Verteilung im Sinne der schwachen Topologie verändert die Verteilung der Statistik nur leicht. Praktisch können Aussagen dieser Art über Konvergenzsätze gewonnen werden.

„Quantitative robustness“ (ebenda) soll im Gegensatz dazu Fehlerschranken für die Zielgröße (z. B. Typ-I-Fehler oder Power) in Abhängigkeit von einem Maß angeben, wie „nichtnormal“ die Daten sind. Dies ist allgemein sehr schwierig zu zeigen. Es setzt z. B. voraus, daß für die Stetigkeit bei der qualitativen Robustheit metrisierbare Topologien existieren und analytisch zugänglich sind.

All diese Kriterien haben nur Sinn, wenn das „normale“ Modell exakte Statistiken ermöglicht, also etwa Tests, die den Typ-I-Fehler exakt einhalten. Die Robustheitsaussage besteht dann darin, daß die Abweichung von dieser exakten Situation gering ist.

Im vorliegenden Fall mit ungleichen Kovarianzmatrizen und unterschiedlichen Stichprobenumfängen, dem sogenannten Behrens-Fisher-Problem, gibt es allerdings schon bei normalverteilten Daten keine exakte Prozedur, deren Ablehnbereich keine Unstetigkeitsstellen hat (siehe Linnik (1968), S. 145). Solche Unstetigkeitsstellen sind aus Anwendersicht problematisch, weil aus dem Testergebnis auf eine allgemeine qualitative Aussage zur Fragestellung geschlossen werden soll. Dafür müssen die konkreten Daten als eine Art Repräsentation zumindest einer ihrer offenen Umgebungen verstanden werden können, über die ähnliche Aussagen getroffen werden können. Wo der Rand des Ablehnbereichs unstetig ist, gibt es solche offenen Umgebungen nicht.

Statt exakter Tests gibt es aber Absätze für sehr brauchbare approximative Teststatistiken für das Behrens-Fisher-Problem. In Abschnitt 1.7 werden davon die Ansätze referiert, die für die vorliegende Arbeit am wichtigsten sind.

Zu untersuchen ist, ob solche Verfahren als robust gelten können. Da aber bei diesen nichtexakten Verfahren die Wahrscheinlichkeiten für den α -Fehler schon bei Normalverteilung nicht exakt zugänglich sind, kann die Robustheit auch nicht auf der Grundlage beurteilt werden, wie sehr bei Nichtnormalverteilung diese Fehlerwahrscheinlichkeiten abweichen.

Aus diesem Grund soll hier eine *Robustheit der Herleitung* betrachtet werden, d. h. eine Approximationsprozedur, die bei Normalverteilung funktioniert, soll auch bei erweiterten Modellen, die gewissen, realistischen Bedingungen genügen, angewendet werden können.

Robustheit der Herleitung ist insofern ähnlich dem Robustheitsbeweis durch Modelerweiterung, nur mit dem Unterschied, daß von der robust hergeleiteten Statistik keine Exaktheit erwartet werden muß, die schon bei Normalverteilung nicht erreichbar ist.

Vor allem sollen nicht die charakteristischen Eigenschaften der Normalverteilung benutzt werden. Diese Eigenschaften sind zum einen die Besonderheit, daß aus Unkorreliertheit linearer Funktionale von multivariaten Normalverteilungen stets ihre stochastische Unabhängigkeit folgt, und zum anderen die Momente von quadratischen Formen.

Unvermeidliche Eigenschaften, die die Verteilungen zu erfüllen haben, sind im Rest dieses Abschnittes dargestellt. Für die Statistikapproximationen im nächsten Kapitel

genügt ein gröberes Rahmenmodell. Für Aussagen über die zugehörigen Schätzer werden Verfeinerungen benötigt: Das additiv-lineare und ein elliptisches Modell in den Abschnitten 1.6.3 und 1.6.4.

1.6.2 Rahmenmodell einer allgemeinen Regularitätsbedingung

Für die Herleitung der Statistiken bis zum nächsten Kapitel wird nur eine sehr schwache Regularitätsbedingung benötigt, die das normalverteilte Modell verallgemeinert. Die Simulationen und die Herleitung passender Parameterschätzungen ab Kapitel 3 werden mit zwei Sonderfällen dieses Modells durchgeführt, die zusätzliche Eigenschaften haben. Deswegen soll das Modell, das sowohl diese beide Sonderfälle wie auch das normalverteilte Modell umfaßt, *Rahmenmodell* genannt werden. Die Parameter des Erwartungswertes und der Kovarianzmatrix, wie sie schon im normalverteilten Modell in Abschnitt 1.3 vorgekommen sind, behalten ihre Bedeutung im Rahmenmodell. Um Trivialfälle zu vermeiden, muß in mindestens einer Stichprobe tatsächlich Varianz vorliegen.

Definition 1. Im *Rahmenmodell* folgt der k -te Beobachtungsvektor \mathbf{Y}_{ik} aus Stichprobe i der Verteilung einer Zufallsvariablen, deren Erwartungswert $E(\mathbf{Y}_{ik}) = \boldsymbol{\mu}_i$ ist und deren Kovarianzmatrix $\text{Cov}(\mathbf{Y}_{ik}) = \boldsymbol{\Sigma}_i$ existiert und positiv semidefinit ist, wobei für mindestens eine Stichprobe i' die Kovarianzmatrix $\boldsymbol{\Sigma}_{i'}$ positiv definit ist. Alle $N = \sum_{i=1}^a n_i$ Beobachtungsvektoren sind voneinander unabhängig und die Vektoren mit übereinstimmendem Gruppenindex i sind identisch verteilt. Es gilt dabei die Regularitätsbedingung

$$\text{Var}(\mathbf{Y}'_{ik} \mathbf{Y}_{ik}) < \infty. \quad (2)$$

Anschaulich ist die Bedingung (2) erfüllt, wenn die euklidische Norm der Beobachtungsvektoren endliche Varianz hat. Da in der Realität Meßbereiche praktisch beschränkt sind, ist auch diese Varianz gewöhnlich endlich.

Das Rahmenmodell ist allerdings zu vage, um es zu simulieren. Auch ist eine so extreme Art von Nichtnormalverteilung für die Untersuchung der Robustheit überflüssig: Eine Teststatistik gilt dann als robust, wenn sie „gut“ (präzise im Typ-I-Fehler und zulässig oder zumindest unverzerrt in der Powerfunktion) im ursprünglichen Modell ist –hier der Normalverteilung–, diese Eigenschaften bei einem moderat erweiterten Modell behält und bei einem stärker erweiterten Modell keine „Katastrophe“ auftritt (Huber (1981)). Die Bedingung (2) wird beispielsweise auch von dichotomen Beobachtungen erfüllt, die reellwertig codiert sind. Diese Art von Verteilung wird nicht als nahe an der Normalverteilung verstanden.

1.6.3 Additiv-lineares Modell

Das *additiv-lineare Modell* ist eine Verallgemeinerung des Normalverteilungsmodells, bei dem die Eigenschaft beibehalten wird, daß die Randverteilungen der Beobachtungsvektoren in jeder Stichprobe als Linearkombinationen unabhängiger, unbeobachtbarer Zufallsvariablen aufgefaßt werden können.

Die unbeobachtbaren Zufallsvariablen können unterschiedliche Verteilungen haben. Die Vektoren der unbeobachtbaren Zufallsvariablen, die zu den Beobachtungsvektoren gehören, müssen aber innerhalb jeder Stichprobe identische gemeinsame Verteilungen haben, damit in jeder Stichprobe die Beobachtungsvektoren identische Verteilungen haben.

Die Koeffizienten der Linearkombinationen in Stichprobe i werden durch die Matrix $\mathbf{\Gamma}_i$ in der folgenden Definition dargestellt. Dann ist $\mathbf{\Gamma}_i \mathbf{\Gamma}'_i = \mathbf{\Sigma}_i$ die Kovarianzmatrix. Damit d -dimensionale Beobachtungsvektoren entstehen, muß $\mathbf{\Gamma}_i$ stets d Zeilen haben. Die Zahl der Spalten werde r_i genannt. Die Zahl r_i entspricht zugleich der Zahl der unbeobachtbaren Zufallsvariablen, die zu einem Beobachtungsvektor der Stichprobe i kombiniert werden. An r_i wird nichts weiter vorausgesetzt, als daß r_i endlich und größer als 0 ist. Insgesamt ergibt sich folgende Definition:

Definition 2. Im additiv-linearen Modell folgt der k -te Beobachtungsvektor \mathbf{Y}_{ik} aus Stichprobe i einer Darstellung

$$\mathbf{Y}_{ik} = \mathbf{\Gamma}_i \mathbf{Z}_{ik} + \boldsymbol{\mu}_i, \quad (3)$$

wobei

1. die Matrix $\mathbf{\Gamma}_i \in \mathbb{R}^{d \times r_i}$ mit $r_i \in \mathbb{N}$ so gestaltet ist, daß $\mathbf{\Gamma}_i \mathbf{\Gamma}'_i = \mathbf{\Sigma}_i$ positiv semidefinit ist und für mindestens eine Stichprobe i' im Datensatz muß $\mathbf{\Sigma}_{i'}$ positiv definit sein,
2. $\mathbf{Z}_{ik} = (Z_{ik1}, \dots, Z_{ikr_i}) \in \mathbb{R}^{r_i}$ Zufallsvektoren der Länge r_i mit unabhängigen, zentrierten und standardisierten Komponenten sind, also $E Z_{ikj} = 0$ und $\text{Var}(Z_{ikj}) = 1$, und
3. die Verteilungen der Vektoren $\mathbf{Z}_{ik}, \dots, \mathbf{Z}_{in_i}$ jeweils identisch sind für übereinstimmendes $i = 1, \dots, a$.
4. Die Ferner gilt die zusätzliche Regularitätsbedingung

$$E Z_{ikj}^4 = \delta_{ij} < \infty. \quad (4)$$

Von der Unabhängigkeit wird nur

$$E(Z_{ijj_1} Z_{ijj_2} Z_{ijj_3} Z_{ijj_4}) = E Z_{ijj_1} E Z_{ijj_2} E Z_{ijj_3} E Z_{ijj_4}, \text{ für } j_1 \neq j_2, \dots, j_4$$

in den Beweisen der vorliegenden Arbeit gebraucht.

Mit diesen Voraussetzungen ist zwangsläufig $\text{Cov}(\mathbf{Y}_{ik}) = \mathbf{\Sigma}_i$. Die Bedingung (4) an die vierten Momente sichert, daß die Varianzen quadratischer Formen in den Zufallsvariablen \mathbf{Y}_{ik} existieren, also Bedingung (2) aus dem Rahmenmodell erfüllt ist.

Satz 3. Das additiv-lineare Modell gehorcht mit der Bedingung (4) auch der Bedingung (2).

Beweis. Siehe den Beweis zu Lemma 34 im Anhang auf Seite 87, wenn man dort \mathbf{I}_d statt \mathbf{T}_d liest. \square

Das additiv-lineare Modell hat mit der Normalverteilung gemein, daß die Abhängigkeiten zwischen den Komponenten genau durch die Korrelation/Kovarianz beschrieben wird. Die Randverteilungen können hingegen unterschiedlich nichtnormalverteilt sein und die Dichtekonturen können nicht-elliptisch sein.

1.6.4 Elliptisches Modell

Um elliptische Modelle zu erklären, wird in der Literatur üblicherweise von einer Definition über die charakteristische Funktion ausgegangen (siehe z. B. Kariya und Sinha (1989) und Mathai, Provost und Hayakawa (1995)). Dies wird hier abgekürzt, indem eine handlichere und anschaulichere Definition verwendet wird, die nach Cambanis, Huang und Simons (1981), Theorem 1, äquivalent zur Erklärung über die charakteristische Funktion ist.

Definition 4. \mathbf{Y}_{ik} folgt einem *elliptischen Modell*, falls es eine Darstellung der Form

$$\mathbf{Y}_{ik} = W_{ik}\mathbf{\Gamma}_i\mathbf{U}_{ik} + \boldsymbol{\mu}_i, \quad (5)$$

wobei $\mathbf{\Gamma}_i \in \mathbb{R}^{d \times r_i}$ eine Matrix mit $\text{rg}(\mathbf{\Gamma}_i) = r_i \leq d$ ist, $\mathbf{\Gamma}_i$ also vollen Spaltenrang hat. Für mindestens eine Stichprobe i' im Datensatz muß gelten $r_{i'} > 0$. W_{ik} ist eine nicht-negative reellwertige Zufallsvariable, die unabhängig von der Zufallsvariable \mathbf{U}_{ik} ist, welche einer Gleichverteilung auf der Einheitssphäre S^{r_i-1} folgt. $\mathbf{U}_{i1}, \dots, \mathbf{U}_{in_i}$ und W_{i1}, \dots, W_{in_i} sind jeweils identisch verteilt und alle untereinander unabhängig. Als *Regularitätsbedingung* wird

$$E(W_{ik}^4) = \delta_i < \infty \quad (6)$$

gefordert und zusätzlich

$$E(W_{ik}^2) = r_i. \quad (7)$$

Die Bedingung an den Spaltenrang macht die Bedingung (7) eindeutig, welche wiederum garantiert, daß auch im elliptischen Modell $\mathbf{\Gamma}_i\mathbf{\Gamma}_i' = \boldsymbol{\Sigma}_i = \text{Cov}(\mathbf{Y}_{ik})$ ist. Das vereinheitlicht die Notation zwischen elliptischen und additiv-linearen Modell und vereinfacht die Beweise, schränkt aber nicht ein, welche Verteilungen die linke Seite von (5) haben kann.

Durch die Regularitätsbedingung (6) fügt sich das elliptische Modell in das Rahmenmodell ein.

Satz 5. *Das elliptische Modell genügt mit der Regularitätsbedingung (6) auch der Regularitätsbedingung (2) aus dem Rahmenmodell.*

Beweis. Der Beweis ist identisch zu dem Beweis von Lemma 36.1 im Anhang auf Seite 90, wenn man dort $\mathbf{T}_d = \mathbf{I}_d$ setzt. \square

Das elliptische Modell hat mit der Normalverteilung gemein, daß die Dichtekonturen elliptisch sind, d. h. alle Projektionen der Zufallsvektoren auf eine Dimension haben bis auf lineare Transformationen dieselben Verteilungen. Die Abhängigkeit der Komponenten läßt sich im Unterschied zur Normalverteilung und zum additiv-linearen Modell nicht mehr allein durch Kovarianzen / Korrelationen beschreiben.

Aus dem Modell (5) ergibt sich unmittelbar, wie man elliptische Verteilungen simulieren kann. Simulationsalgorithmen für spezielle elliptische Verteilungen, mit denen man zugleich die Kovarianzmatrix einfach unter Kontrolle halten kann, sogenannte Pearson-Typ-II und -Typ-VII-Verteilungen, finden sich bei Johnson (1987), ab S. 110.

1.7 Bekannte Statistiken als Ausgangspunkt

1.7.1 ANOVA und Geisser-Greenhouse-Statistik

Die Geisser-Greenhouse-Statistik (Geisser und Greenhouse 1958; Greenhouse und Geisser 1959) wendet eine Approximationsprozedur auf die quadratischen Formen an, die auch in der zweifaktoriellen ANOVA vorkommen. Die Geisser-Greenhouse-Statistik kennt hierbei nur zwei hochdimensionale Hypothesen: „Kein Zeiteffekt“ $H_0(B)$ und „keine Interaktionseffekt zwischen einem Zeit- und einem Gruppenfaktor“ $H_0(AB)$. Mehr als ein Zeit- oder Gruppenfaktor ist in den Originalarbeiten nicht vorgesehen. Auch in kommerzieller Statistiksoftware ist nur dieses zweifaktorielle Modell implementiert (siehe z. B. *IBM SPSS Statistics for Windows, Version 25.0* (2017), *SAS/STAT (R) 9.3 User's Guide* (2011) und *Statistica 13.3.1* (2018)).

Mit

$$\mathbf{Y} = (\mathbf{Y}'_{11}, \dots, \mathbf{Y}'_{1n_1}, \dots, \mathbf{Y}'_{an_a})'$$

lauten die quadratischen Formen für diese Tests $Q_1 := \mathbf{Y}'\mathbf{M}_1\mathbf{Y}$, $Q_4 := \mathbf{Y}'\mathbf{M}_4\mathbf{Y}$ und $Q_5 := \mathbf{Y}'\mathbf{M}_5\mathbf{Y}$, wobei die Matrizen lauten

$$\mathbf{M}_1 = \frac{1}{\sum_{i=1}^a n_i^{-1}} \left[\frac{1}{n_i n_{i'}} \mathbf{J}_{n_i \times n_{i'}} \otimes \mathbf{P}_d \right]_{i, i'=1, \dots, a} \quad (8)$$

$$\mathbf{M}_4 = \left(\bigoplus_{i=1}^a n_i^{-1} \mathbf{J}_{n_i} - N^{-1} \mathbf{J}_N \right) \otimes \mathbf{P}_d \quad (9)$$

$$\mathbf{M}_5 = \bigoplus_{i=1}^a \mathbf{P}_{n_i} \otimes \mathbf{P}_d. \quad (10)$$

Sie sind idempotent. Zu beachten ist, daß wie in der ANOVA üblich in der Matrix \mathbf{M}_4 durch $(\bigoplus_{i=1}^a n_i^{-1} \mathbf{J}_{n_i} - N^{-1} \mathbf{J}_N)$ mit gewichteten Mittelwerten gerechnet wird.

Die Teststatistik zur Hypothese $H_0(B)$ —kein Zeiteffekt— ist

$$\frac{\text{rg}(\mathbf{M}_5) \mathbf{Y}' \mathbf{M}_1 \mathbf{Y}}{\text{rg}(\mathbf{M}_1) \mathbf{Y}' \mathbf{M}_5 \mathbf{Y}}$$

und die zur Hypothese $H_0(AB)$ —keine Wechselwirkung— ist

$$\frac{\text{rg}(\mathbf{M}_5) \mathbf{Y}' \mathbf{M}_4 \mathbf{Y}}{\text{rg}(\mathbf{M}_4) \mathbf{Y}' \mathbf{M}_5 \mathbf{Y}}.$$

Bei der ANOVA wird die exakte Verteilung dieser Statistiken ausgehend von der Annahme gefunden, daß die Kovarianzen alle identisch seien. Dieselbe Konsequenz hat auch das allgemeinere Kriterium der sogenannten *Sphärizität*, wenn $\mathbf{T}_d \boldsymbol{\Sigma}_1 \mathbf{T}_d = \dots = \mathbf{T}_d \boldsymbol{\Sigma}_a \mathbf{T}_d$ nur die Eigenwerte 0 und 1 haben, siehe Huynh und Feldt (1970), daher auch Huynh-Feldt-Bedingung genannt. Die Verteilung ist dann eine F -Verteilung, deren Freiheitsgrade an den Rängen abgelesen werden können, die in der Statistik stehen.

Um die Beschränkung der Sphärizität zu überwinden, wurden schon früh sogenannte Varianzkomponentenverfahren eingeführt (Searle, Casella und McCulloch (1992)). Sie verwenden eine als bekannt vorausgesetzte Struktur der Kovarianzen, für die dann nur wenige Parameter geschätzt werden müssen. Mit dem Resultat können verallgemeinerte Kleinste-Quadrate-Schätzer (Generalized least squares) ausgerechnet werden, aus denen dann die Teststatistik besteht.

Dies funktioniert aber nur bei niedrigdimensionalen Situationen zuverlässig, weil bei vielen Subplotfaktorstufen stets auch viele Parameter falsch geschätzt werden können und die Schätzungen bei demgegenüber geringem Stichprobenumfang zu schlecht sein können.

Dagegen gibt es fast ebenso alte Ansätze, die von der Verteilung quadratischer Formen des gewöhnlichen Kleinste-Quadrate-Schätzers bei beliebigen Strukturen der Kovarianzmatrix ausgehen. Anders als bei der Sphärizität der ANOVA haben $\mathbf{Y}' \mathbf{M}_1 \mathbf{Y}$, $\mathbf{Y}' \mathbf{M}_4 \mathbf{Y}$ und $\mathbf{Y}' \mathbf{M}_5 \mathbf{Y}$ unter der Annahme einer gemeinsamen aber beliebigen Kovarianzmatrix, also $\boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_a$, Verteilungen, die Summen aus gewichteten und unabhängigen Chiquadratverteilungen sind. Die Gewichte sind hierbei die Eigenwerte der typischerweise unbekannt gemeinsamen Kovarianzmatrix.

Die Box-Patnaik-Approximation (Box (1954a,b) und Patnaik (1949)) bietet für diesen Fall eine Möglichkeit, die Tails der Verteilung einer quadratischen Form in zentriert multivariat normalverteilten Zufallsvariablen anzunähern, ohne die Kovarianzmatrix im Einzelnen kennen zu müssen, sondern nur zwei Funktionale von ihr.

Hierzu wird die quadratische Form durch eine gestreckte Chiquadratverteilung approximiert, deren Freiheitsgrad und Streckungsparameter so gewählt sind, daß in Erwartungswert und Varianz die quadratische Form mit der gestreckten Chiquadratverteilung übereinstimmt, d. h. es ist mit einer symmetrischen und idempotenten Matrix \mathbf{M} und $\boldsymbol{\Sigma}_Y := \text{Cov}(\mathbf{Y})$

$$\mathbf{Y}' \mathbf{M} \mathbf{Y} \dot{\sim} g \chi_f^2,$$

wenn g und f die Gleichungen

$$\begin{aligned} E(\mathbf{Y}'\mathbf{M}\mathbf{Y}) &= \text{Sp}(\mathbf{M}'\boldsymbol{\Sigma}_Y\mathbf{M}) = gf \\ \text{Var}_{\mathcal{N}}(\mathbf{Y}'\mathbf{M}\mathbf{Y}) &= 2\text{Sp}(\mathbf{M}'\boldsymbol{\Sigma}_Y\mathbf{M})^2 = 2g^2f \end{aligned}$$

erfüllen. Man erhält als Approximation also

$$\frac{\mathbf{Y}'\mathbf{M}\mathbf{Y}}{\text{Sp}(\mathbf{M}'\boldsymbol{\Sigma}_Y\mathbf{M})} \overset{\cdot}{\sim} f^{-1}\chi_f^2.$$

Da bei normalverteilten Daten die quadratischen Formen Q_1 und Q_4 stochastisch unabhängig zu Q_5 sind und unter der Nullhypothese denselben Erwartungswert $\text{Sp}(\mathbf{M}'\boldsymbol{\Sigma}_Y\mathbf{M})$ haben, kann daraus eine näherungsweise F -verteilte Statistik mit den Freiheitsgraden $\text{erg}\mathbf{T}$ und $\epsilon(N - a)$ gewonnen werden, wobei

$$\epsilon := \frac{\text{Sp}^2(\boldsymbol{\Sigma}_1)}{(d - 1)\text{Sp}(\boldsymbol{\Sigma}_1^2)} \quad (11)$$

das sogenannte *Box'sche Epsilon*, als unbekannter Parameter verbleibt.

Die Geisser-Greenhouse-Statistik ist eine der ältesten Anwendungen der Box-Patnaik-Approximation Box (1954a,b) und Patnaik (1949). Noch heute ist sie populär in unterschiedlicher Statistik-Software (z.B. *SAS/STAT* [®] 9.3 *User's Guide* (2011), *IBM SPSS Statistics for Windows, Version 25.0* (2017)), obwohl bekannt ist, daß sie nicht für unterschiedliche Kovarianzmatrizen bei gleichzeitig unterschiedlichen Stichprobenumfängen ausgelegt ist (siehe Brunner, Bathke und Placzek 2012). Beides zusammen reicht bekanntermaßen aus (z.B. Oberfeld und Franke (2013)), daß die Statistik den Typ-I-Fehler erheblich verfehlt, auch schon bei Normalverteilung.

Weil der Hauptunterschied der Geisser-Greenhouse-Statistik zur gewöhnlichen zweifaktoriellen ANOVA lediglich modifizierte Freiheitsgrade sind, die der beliebigen Kovarianzstruktur von $\boldsymbol{\Sigma}_1$ Rechnung tragen, sind auch keine besonderen Vorkehrungen zu erkennen, die die Robustheitseigenschaften bei Nichtnormalverteilung verbessern könnten.

Es bleibt daher auch für die Geisser-Greenhouse-Statistik ein Resultat von Atiqullah (1962) gültig, wonach quadratische unbalancierte Designs nicht robust sind. Quadratische Balanciertheit würde vorliegen, wenn auf der Hauptdiagonalen von \mathbf{M}_5 identische Einträge stünden. Quadratische Balanciertheit ist in einem Layout mit nur einem Wholeplotfaktor gleichbedeutend mit identischen Stichprobenumfängen.

1.7.2 ANOVA-Typ-Statistik bei Normalverteilung

Die ANOVA-Typ-Statistik (ATS) wurde als Verallgemeinerung der ANOVA für unbalancierte und heteroskedastische normalverteilte Daten entwickelt, insbesondere für den Zweck, eine Rangtransformation der Daten aufzunehmen, die bei Unbalanciertheit immer Heteroskedastizität erzeugt, siehe Brunner, Dette und Munk (1997). Der

ursprüngliche Gedanke der ATS geht von einer Wald-Typ-Statistik der Art

$$\bar{\mathbf{Y}}' \mathbf{H}' \left(\mathbf{H} \hat{\Sigma} \mathbf{H}' \right)^+ \mathbf{H} \bar{\mathbf{Y}} \rightarrow \chi_{\text{rg}(\mathbf{H})}^2 \text{ für } n_i \rightarrow \infty, \frac{n_i}{N} \rightarrow \eta_i \in (0, 1)$$

für eine Hypothesenmatrix \mathbf{H} aus, wobei $(\cdot)^+$ die Moore-Penrose-Inverse ist. Die Schätzung $\hat{\Sigma}$ führt speziell bei hochdimensionalen Daten aber zu einer sehr langsamen Konvergenz aus der liberalen Richtung. Als Abhilfe wird $\hat{\Sigma}$ durch die Identitätsmatrix \mathbf{I}_{Nd} ersetzt und anschließend eine geeignete Näherung für die Verteilung von

$$\bar{\mathbf{Y}}' \mathbf{H}' (\mathbf{H} \mathbf{H}')^+ \mathbf{H} \bar{\mathbf{Y}}$$

gesucht.

Sie wird gefunden in der Box-Patnaik-Approximation. Diese Approximationsprozedur soll als nächstes für das unbalancierte Mehrstichprobenlayout wie es bei Becker (2010), Brunner, Bathke und Placzek (2012), Happ, Harrar und Bathke (2016) und Harden (2013) für Split-Plot- bzw. Meßwiederholungen vorgestellt werden.

Es sei daran erinnert, daß $\mathbf{H}' (\mathbf{H} \mathbf{H}')^{-1} \mathbf{H} = \mathbf{T} = \mathbf{T}_a \otimes \mathbf{T}_d$ ist und $\mathbf{T}_a = [t_{ii}]_{i,i'=1,\dots,a}$ notiert wird. Ausgegangen wird von der quadratischen Form

$$\bar{\mathbf{Y}}' \mathbf{H}' (\mathbf{H} \mathbf{H}')^{-1} \mathbf{H} \bar{\mathbf{Y}} = \bar{\mathbf{Y}}' \mathbf{T} \bar{\mathbf{Y}} =: Q_n. \quad (12)$$

Die Zufallsvariable Q_n ist damit eine quadratische Form, die mit der Matrix

$$\mathbf{M}_n := \left(\bigoplus_{i=1}^a n_i^{-1} \mathbf{1}_{n_i} \otimes \mathbf{I}_d \right) \mathbf{T} \left(\bigoplus_{i=1}^a n_i^{-1} \mathbf{1}'_{n_i} \otimes \mathbf{I}_d \right) \quad (13)$$

auch geschrieben werden kann als $Q_n = \mathbf{Y}' \mathbf{M}_n \mathbf{Y}$. Bei Normalverteilung und unter der Nullhypothese folgt sie einer Linearkombination von unabhängigen zentrierten χ^2 -Verteilungen mit einem Freiheitsgrad, also

$$Q_n \sim \sum \lambda_j \chi_1^2,$$

wobei λ_j die Eigenwerte von $\left(\bigoplus_{i=1}^a \left(\mathbf{I}_{n_i} \otimes \Sigma_i^{\frac{1}{2}} \right) \right) \mathbf{M}_n \left(\bigoplus_{i=1}^a \left(\mathbf{I}_{n_i} \otimes \Sigma_i^{\frac{1}{2}} \right) \right)$ sind. Diese Linearkombination kann nach dem Vorschlag von Box und Patnaik Box 1954a,b; Patnaik 1949 durch eine gestreckte χ^2 -Verteilung approximiert werden, die in Erwartungswert und Varianz mit der Linearkombination übereinstimmt, also

$$Q_n \dot{\sim} g \chi_{f_n}^2,$$

Wobei „ $\dot{\sim}$ “ für „ist approximativ verteilt wie“ zu lesen ist.

Der Streckungsparameter g hat die Lösung $g = E_{H_0}(Q_n) = \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\mathbf{T}_d \Sigma_i)$, so daß sich anbietet, analog zur ANOVA oder zur Geisser-Greenhouse-Statistik einen Quotienten aus Q_n und einer anderen, unabhängigen quadratischen Formen zu bilden, die denselben Erwartungswert hat. Eine solche quadratische Form entsteht sehr

einfach, indem man den Plug-in-Schätzer für $E_{H_0}(Q_n)$ durch Einsetzen der empirischen Kovarianzmatrizen wählt, also

$$Q_e := \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\mathbf{T}_d \hat{\Sigma}_i). \quad (14)$$

Mit

$$\mathbf{M}_e := \bigoplus_{i=1}^a \frac{t_{ii}}{n_i(n_i-1)} \mathbf{P}_{n_i} \otimes \mathbf{T}_d \quad (15)$$

kann $Q_e = \mathbf{Y}' \mathbf{M}_e \mathbf{Y}$ geschrieben werden. Damit sieht man, daß \mathbf{M}_n und \mathbf{M}_e orthogonal zueinander sind im Sinne von $\mathbf{M}_n (\bigoplus \mathbf{I}_{n_i} \otimes \Sigma_i) \mathbf{M}_e = \mathbf{0}$ (Satz von Craig und Sakamoto). Daraus folgt schon die Unabhängigkeit von Q_n und Q_e unter Normalverteilung.

Für Q_e kann nach demselben Schema wie für Q_n eine approximative Chiquadratverteilung gefunden werden. Da Q_n und Q_e unabhängig sind, ergibt sich als Statistik

$$F := \frac{Q_n}{Q_e} \dot{\sim} F(f_n, f_e). \quad (16)$$

Die Lösungen für die Freiheitsgrade sind

$$f_n = \frac{2E^2(Q_n)}{\text{Var}_{\mathcal{N}}(Q_n)} \quad \text{und} \quad f_e = \frac{2E^2(Q_e)}{\text{Var}_{\mathcal{N}}(Q_e)}$$

und die Streckungsparameter sind $\text{Var}_{\mathcal{N}}(Q_n)/E_{H_0}(Q_n)$ bzw. $\text{Var}_{\mathcal{N}}(Q_e)/E_{H_0}(Q_e)$. Erweitert man mit $E_{H_0}Q_n = EQ_e$, kürzen sich im Quotienten F die Streckungsparameter weg.

Die Erwartungswerte und Varianzen für die Freiheitsgrade kann man direkt ausrechnen. Für die Erwartungswerte $E^2(Q_n)$ und $E^2(Q_e)$ sei auf Satz 28 auf Seite 78 verwiesen. Die Varianz $\text{Var}_{\mathcal{N}}(Q_n)$ wird in Harden 2013 mit einem Satz über die Varianz quadratischer Formen in Normalverteilungen berechnet, ergibt sich aber auch als Sonderfall aus Satz 9 auf Seite 30. Für die Werte dieser Parameter erhält man mit den Abkürzungen

$$b_1 := E^2(Q_n) = \sum_{i=1}^a \sum_{i'=1}^a \frac{t_{ii'}^2}{n_i n_{i'}} \text{Sp}(\mathbf{T}_d \Sigma_i) \text{Sp}(\mathbf{T}_d \Sigma_{i'}) \quad (17)$$

$$b_n := \frac{1}{2} \text{Var}_{\mathcal{N}}(Q_n) = \sum_{i=1}^a \sum_{i'=1}^a \frac{t_{ii'} t_{i'i}}{n_i n_{i'}} \text{Sp}(\mathbf{T}_d \Sigma_i \mathbf{T}_d \Sigma_{i'}) \quad (18)$$

$$b_e := \frac{1}{2} \text{Var}_{\mathcal{N}}(Q_e) = \sum_{i=1}^a \frac{t_{ii}^2}{n_i^2 (n_i - 1)} \text{Sp}(\mathbf{T}_d \Sigma_i)^2 \quad (19)$$

die endgültige Form der Lösungen

$$f_n = \frac{b_1}{b_n} \quad \text{und} \quad f_e = \frac{b_1}{b_e}. \quad (20)$$

Es ist immer

$$f_n \in [1, \text{rg}(\mathbf{T}_a) \text{rg}(\mathbf{T}_d)], \quad (21)$$

insbes. ist f_n in der Praxis endlich. $f_n = 1$ kann nur vorkommen, wenn in genau einer Stichprobe positive Varianz in lediglich einer Messung vorliegt. Ist diese Stichprobe diejenige mit dem kleinsten Stichprobenumfang, dann nimmt auch f_e sein Minimum an:

$$f_e \in \left[\min_{i \in \{1, \dots, a\}} n_i - 1, \text{rg}(\mathbf{T}_d) \cdot (n - a) \right]. \quad (22)$$

Die Obergrenzen werden bei $\mathbf{T}_d \boldsymbol{\Sigma}_1 \mathbf{T}_d = \dots = \mathbf{T}_d \boldsymbol{\Sigma}_a \mathbf{T}_d = \mathbf{T}_d$ verwirklicht, also falls perfekte Sphärizität, d. h. die Modellvoraussetzungen der klassischen ANOVA gilt.

Die linke Seite der Geisser-Greenhouse-Statistik unterscheidet sich bei den beiden Hypothesen $H_0(AB)$ und $H_0(B)$ von der ATS, wenn die Stichprobenumfänge nicht identisch sind, indem die gepoolte Kovarianzmatrix für Q_e verwendet wird ((Bathke, Schabenberger u. a. 2009)) und indem für den Test auf die Wechselwirkungshypothese $H_0(AB)$ mit den gewichteten Mittelwerten gerechnet wird. Bei identischen Stichprobenumfängen fallen diese Unterschiede zusammen.

Dadurch „erbt“ die ATS von der Geisser-Greenhouse-Statistik, genau dort mit der klassischen ANOVA übereinzustimmen, wo die ANOVA exakt ist, d. h. wo die sogenannte Huynh-Feldt-Bedingung (Huynh und Feldt (1970)) erfüllt ist. Man muß also den Gewinn, eine gute Näherungsstatistik im Behrens-Fisher-Problem oder bei Nichterfüllung der Huynh-Feldt-Bedingung nicht damit bezahlen, dort ungenau zu testen, wo die ANOVA exakt wäre. Ob die Huynh-Feldt-Bedingung erfüllt ist, ist in der praktischen Anwendung gewöhnlich unbekannt.

1.7.3 Robustheit der ANOVA-Typ-Statistik im Einstichprobenfall

Bei Atiqullah (1962) wird Robustheit der ANOVA gefolgert, wenn gezeigt werden kann, daß die Exzess-Kurtosis der Beobachtungen keine Rolle spielt. Von Helms (2010) wurde die ATS für den Fall einer Stichprobe mit unbekannter Kovarianzmatrix so konstruiert, daß sie nach diesem Vorbild robust hergeleitet werden kann. Verwendet wird vereinfacht gesagt ein additiv-lineares Modell wie in Abschnitt 1.6.3, nur für eine einzige Stichprobe.

Für die Herleitung werden die quadratischen Formen

$$\begin{aligned} Q_N^1 &:= \bar{\mathbf{Y}}' \mathbf{T}_d \bar{\mathbf{Y}} \\ Q_e^1 &:= \mathbf{Y}' (\mathbf{P}_N \otimes \mathbf{T}_d) \mathbf{Y} = \text{Sp}(\mathbf{T}_d \hat{\boldsymbol{\Sigma}}) \end{aligned}$$

definiert und als Teststatistik wird

$$F^1 := \frac{Q_N^1}{Q_e^1}$$

gewählt.

In den Tails wird die Approximation

$$F^1 \underset{H_0}{\simeq} g_1 \chi_{f_1}^2$$

vorgenommen, wenn Freiheitsgrad f_1 und Streckungsparameter g_1 so gewählt sind, daß Varianz und Erwartungswert von F^1 mit denen der $g_1^{-1} \chi_{f_1}^2$ -Verteilung übereinstimmen. Hier wird also der gesamte Quotient der quadratischen Formen approximiert. Hintergrund ist, daß Q_N^1 und Q_e^1 zwar orthogonale quadratische Formen sind, allerdings wegen des über die Normalverteilung hinausgehenden Modells nicht mehr als stochastisch unabhängig gelten können. Die Abhängigkeit kann das Verhalten von F^1 so verändern, daß man es für eine vernünftige Statistikapproximation nicht mehr ignorieren kann.

Die Abhängigkeit wird durch approximative Formeln für Varianz und Erwartungswert eines Quotienten von korrelierten Zufallsvariablen einbezogen. Diese Formeln beruhen auf den Taylorapproximationen aus Korollar 31 auf Seite 84, nämlich

$$\begin{aligned} E \left(\frac{Q_N^1}{Q_e^1} \right) &\doteq \frac{EQ_N^1}{EQ_e^1} \left(1 + \frac{\text{Var}(Q_e^1)}{E^2Q_e^1} - \frac{\text{Cov}(Q_N^1, Q_e^1)}{EQ_N^1 EQ_e^1} \right) \\ &\stackrel{H_0}{=} 1 + \frac{\text{Var}(Q_e^1) - \text{Cov}(Q_N^1, Q_e^1)}{E^2Q_e^1} \end{aligned} \quad (23)$$

und

$$\begin{aligned} \text{Var} \left(\frac{Q_N^1}{Q_e^1} \right) &\doteq \frac{E^2Q_N^1}{E^2Q_e^1} \left(\frac{\text{Var}(Q_N^1)}{E^2Q_N^1} + \frac{\text{Var}(Q_e^1)}{E^2Q_e^1} - \frac{2\text{Cov}(Q_N^1, Q_e^1)}{EQ_N^1 EQ_e^1} \right) \\ &\stackrel{H_0}{=} \frac{\text{Var}(Q_N^1) + \text{Var}(Q_e^1) - 2\text{Cov}(Q_N^1, Q_e^1)}{E^2Q_e^1}. \end{aligned} \quad (24)$$

In einem Modell wie in Modellgleichung (3), nur für eine Stichprobe, wird ausgerechnet

$$\begin{aligned} \text{Var}(Q_N^1) &= 2\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma})^2 + \kappa \\ \text{Var}(Q_e^1) &= \frac{2}{N-1} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma})^2 + \kappa \\ \text{Cov}(Q_N^1, Q_e^1) &= \kappa \end{aligned}$$

wobei κ von der Kovarianzmatrix abhängt und im additiv-linearen Modell (3) zusätzlich von den vierten Momenten der Randverteilungen in den \mathbf{Z} -Vektoren. Die

κ -Terme fallen auf diese Weise in den Formeln (23) und (24) weg, wenn diese vierten Momente endlich sind. Der Streckungsparameter g und der Freiheitsgrad f hängen damit nur noch von Termen ab, die aus $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma})^2$, $\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma})$ und dem Stichprobenumfang bestehen, nicht von Termen, die wie κ von der Verteilungsfamilie abhängen.

Die Herleitung der Statistik ist also robust, insofern die vierten Momente in den Rändern der \mathbf{Z} -Vektoren aus Modellgleichung (3) anders sein können als im normalverteilten Modell, wenn nur die Regularitätsbedingung (4) eingehalten wird. Bei Helms (2010) wurde mit Simulationen bestätigt, daß die Statistik für praktische Anwendungen robust genug sein kann.

Ungeklärt durch die Arbeit von Helms (2010) bleibt, ob ein analoges Vorgehen wirklich zur Robustheit bei unbalancierten Stichproben führen kann, weil das Einstichprobenlayout bei vollständigen Beobachtungen ohnehin quadratisch balanciert ist. Daher kann es nach Atiqullah (1962) ohnehin als robust gelten.

1.8 Andere Ansätze

1.8.1 g-Inverse der empirischen Kovarianzmatrix

Statistiken auf der Grundlage von Wald-Typ-Statistiken wie

$$\bar{\mathbf{Y}}' \mathbf{H}' \left(\mathbf{H} \hat{\boldsymbol{\Sigma}} \mathbf{H}' \right)^+ \mathbf{H} \bar{\mathbf{Y}}$$

wurden von Secchi, Stamm und Vantini (2013) sowie Srivastava und Fujikoshi (2006) vorgeschlagen. Im Niedrigdimensionalen können daraus Wald-Typ-Statistiken gewonnen werden, indem bei großen Stichprobenumfängen dieser Ausdruck einer Chi-Quadratverteilung folgt, deren Freiheitsgrade dem Rang von \mathbf{H} entsprechen.

Im Hochdimensionalen hat diese Statistik allerdings den Nachteil, daß für Alternativen, die im Kern von $\mathbf{H} \hat{\boldsymbol{\Sigma}} \mathbf{H}'$ liegen, nur eine sehr schlechte Power erreichbar ist, wie auch Secchi, Stamm und Vantini (2013) einräumen. Die Dimensionalität des Kerns wiederum wächst mit der Differenz zwischen der Zahl der Meßwiederholungen und dem Gesamtstichprobenumfang.

Man kann sich das Komplement des Ablehnbereich eines solchen Tests hinsichtlich $\mathbf{H} \bar{\mathbf{Y}}$ bei fest gedachtem $\hat{\boldsymbol{\Sigma}}$ geometrisch wie eine Säule vorstellen, die orthogonal zum Kern von $\mathbf{H} \hat{\boldsymbol{\Sigma}} \mathbf{H}'$ ellipsoidal ist und entlang des Kerns unendlich lang ist.

Die empirische Kovarianzmatrix $\hat{\boldsymbol{\Sigma}}$ ist aber selbst eine Zufallsvariable, deren Eigenräume von den Vektoren $\mathbf{H}_d (\mathbf{Y}_{ik} - \bar{\mathbf{Y}}_i)$, $i = 1, \dots, n_i$, $k = 1, \dots, n_i$, aufgespannt werden. Es ist daher plausibel, wenn auch noch nicht bewiesen, daß die Richtungen, in denen die Verteilung am wenigsten streut –also die Richtungen der Eigenvektoren von $\mathbf{H} \boldsymbol{\Sigma} \mathbf{H}'$ mit den kleinsten oder gar verschwindenden Eigenwerten– am ehesten den Kern von $\mathbf{H} \hat{\boldsymbol{\Sigma}} \mathbf{H}'$ aufspannen. Das Komplement des Ablehnbereichs hat also gerade da den kleinsten Durchmesser, wo die Daten am meisten streuen.

Dieses Verhalten ist unerwünscht. So könnte ein solcher Test eine bessere Power bekommen, indem man irrelevante Daten mit einer hohen Varianz hinzufügt. Solche Probleme mit der Powerfunktion sollen durch Simulationen belegt, deren Ergebnisse in Abbildung 2 gezeigt werden.

Simuliert wurde mit festem Stichprobenumfang $n = 20$ und wachsender Dimensionalität die Statistik aus Vermutung 6 von (Secchi, Stamm und Vantini 2013), weil diese Statistik im Niedrigdimensionalen identisch zur T^2 -Statistik ist. Betrachtet wurden zwei Alternativen mit demselben Mahalanobis-Abstand zur Null, so daß die Power beim T^2 -Test dieselbe sein muß. Allerdings liegt der Erwartungswertvektor der ersten Alternative auf dem Eigenvektor zum größten Eigenwert der Kovarianzmatrix von Σ_d . Bei der zweiten Alternative liegt er auf dem Eigenvektor zum kleinsten Eigenwert, also in der Richtung der geringsten Streuung der simulierten multivariaten Normalverteilung. Zum Vergleich wurde auch die Nullhypothese mitsimuliert.

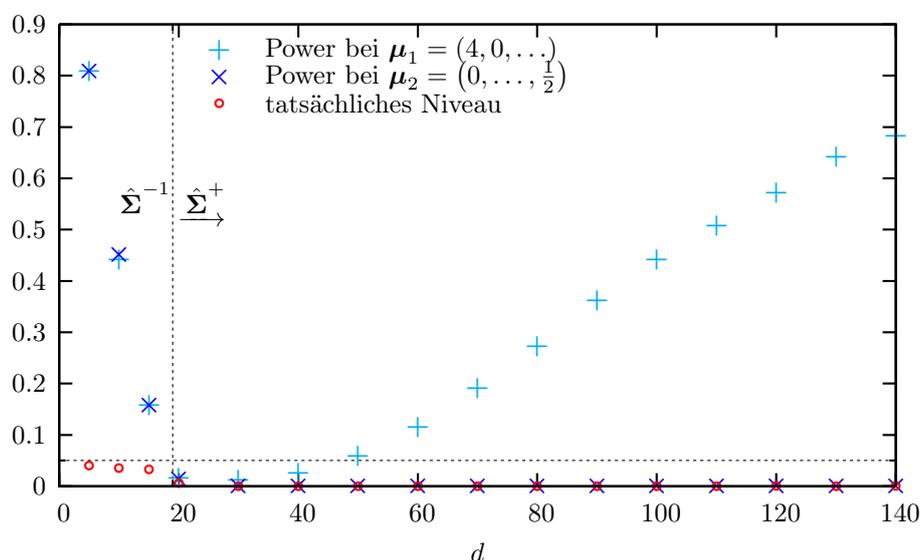


Abbildung 2: Power und Niveau des vermuteten Tests aus (Secchi, Stamm und Vantini 2013) bei $\Sigma_d = \text{diag}(16, 8, 4, 2, \mathbf{1}'_{d-5}, \frac{1}{4})$, $n = 20$ und $H_0 : \boldsymbol{\mu} = \mathbf{0}$ simuliert mit 10000 Wiederholungen. Von links bis an die gepunktete Linie bei $d = 19$ ist dieser Test identisch zu Hotellings T^2 -Test.

Man sieht, daß der T^{2+} -Test bei den gewählten Parametern im Hochdimensionalen noch sehr konservativ ist. Offensichtlich ist die Verteilung von T^{2+} -Statistiken insbesondere im Finiten noch längst nicht gut verstanden. Gravierend ist, daß wie vorausgesagt im Hochdimensionalen der T^{2+} -Test die Alternative entlang der größten Streuung besser aufdeckt, während bei der Alternative entlang der geringsten Streuung die Wahrscheinlichkeit für den Typ-I-Fehler von der Power kaum überschritten wird. Man berührt also die Grenze zur Verzerrtheit eines statistischen Tests. Der Powerunterschied zwischen beiden Alternativen vergrößert sich sogar mit wachsendem d , indem die Power bei der ersten Alternative beständig zunimmt. Das

kontrastiert zu der ebenfalls erwarteten identischen Power bei beiden Alternativen im Falle der –niedrigdimensionalen– T^2 -Statistik links von $d = 20$.

1.8.2 Multivariater Ansatz

Bei Srivastava, Katayama und Kano (2013a,b) und Feng u. a. (2015) wird ein hochdimensionaler Test für die Hypothese $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ ausgehend von

$$(\bar{\mathbf{Y}}_1. - \bar{\mathbf{Y}}_2.)' \mathbf{D}^{-1} (\bar{\mathbf{Y}}_1. - \bar{\mathbf{Y}}_2.) \text{ mit } \mathbf{D} = \mathbf{I}_d \# \left(n_1^{-1} \hat{\boldsymbol{\Sigma}}_1 + n_2^{-1} \hat{\boldsymbol{\Sigma}}_2 \right)$$

vorgeschlagen. Diese Metrik wird auch Karl-Pearson-Abstand genannt (siehe Chiang (2005)). Eine niedrigdimensionale Version wird „modified ATS“ genannt, siehe Friedrich und Pauly (2018). Mit Statistiken dieser Art kann nur ein Test für die Hypothese „kein Effekt der beiden Gruppen auf die Zeitverläufe“ durchgeführt werden. Dabei hängt die Powerfunktion von den komponentenweise standardisierten Differenzen der Beobachtungsvektoren ab. Solche Tests sind also nur geeignet, wenn es sachlogisch Sinn hat, die Alternative dergestalt als Vektor normierter Verschiebungen anzugeben.

1.8.3 Asymptotische Ansätze

Es wurde in Abschnitt 1.6 angedeutet, daß man auch durch Konvergenzsätze qualitative Robustheit belegen kann. Oftmals gilt dieselbe Verteilungsaussage über eine Statistik unter der Annahme einer bestimmten Verteilung der Daten wie unter den Voraussetzungen eines Konvergenzsatzes. Ein wohlbekanntes Beispiel ist der Gaußtest, dessen Teststatistik bei normalverteilten Daten standardnormal ist. Dieselbe Verteilung hat diese Teststatistik bei hinreichend großem Stichprobenumfang, wenn die Daten nur den Voraussetzungen des zentralen Grenzwertsatzes gehorchen, ohne normalverteilt zu sein.

Dieses Konzept der Asymptotik über den Stichprobenumfang wurde häufig angewendet, ist aber nicht mit dem Paradigma hoher Dimensionalität vereinbar, weil eine Asymptotik über den Stichprobenumfang bei fixiertem d auf niedrigdimensionale Szenarien hinausläuft. Man kann d auch nicht beliebig wachsen lassen, weil $d = o(n^2)$ sei muß, damit der Zentrale Grenzwertsatz greifen kann (Portnoy (1986)).

In der Literatur wird daher oft eine Asymptotik mit festen Stichprobenumfängen und $d \rightarrow \infty$ oder eine sogenannte *Verhältniskonsistenz* gefordert, d. h.

$$\frac{d}{N} \xrightarrow{N, d \rightarrow \infty} c > 0$$

und eine Bedingung an die asymptotischen Kovarianzmatrizen (Bai und Saranadasa (1996), S. X. Chen und Qin (2010), Feng u. a. (2015), Happ, Harrar und Bathke (2016), Katayama, Kano und Srivastava (2013), Park und Ayyala (2013), Pauly, Ellenberger und Brunner (2015), Srivastava und Du (2008), Srivastava und Fujikoshi

(2006) und Srivastava, Katayama und Kano (2013a) oder an die asymptotische Abhängigkeit zwischen den Messungen (Wang und Akritas (2010)). Meistens besagen diese Bedingungen, daß die asymptotischen Eigenwerte der Kovarianzmatrizen nicht zu unterschiedlich oder die Abhängigkeit nicht zu groß sein soll.

Eine derartige Asymptotik steht allerdings im Widerspruch zur Bedeutung von festen Faktoren (insbes. Subplotfaktoren) und zufälligen Faktoren. Zufällige Faktorstufen können per Definition vermindert oder vermehrt werden, ohne die Bedeutung zu berühren, die die Versuchsergebnisse haben können (Verallgemeinerungsregel). Feste Faktoren hängen hingegen direkt mit der Aussage zusammen, die ein Versuch haben soll.

Nichtsdestotrotz ist dieser Ansatz populär. Dabei ist die Asymptotik über die asymptotische Gestalt der Kovarianzmatrizen schlecht überprüfbar, wenn schon die finite Kovarianzstruktur unbekannt ist. Auch sachlogisch ist keine Eindeutigkeit gegeben: Wenn Meßzeitpunkte enger gewählt werden, nimmt die Korrelation zwischen ihnen gewöhnlich zu. Wenn der Meßzeitraum verlängert wird, nimmt die Korrelation zwischen der ersten und der letzten Messung gewöhnlich ab. Beides führt zu grundverschiedenen asymptotischen Kovarianzmatrizen.

Wie entscheidend deren unterschiedliches asymptotisches Verhalten ist, wurde von Pauly, Ellenberger und Brunner (2015) gezeigt. Bei ungünstiger asymptotischer Kovarianzmatrix und normalverteilten Daten folgen die Statistiken, die ansonsten als asymptotisch standardnormalverteilt gelten, einer zentrierten und standardisierten Chiquadratverteilung mit Freiheitsgrad 1.

Würde man vorsichtshalber daher nur noch die kritischen Werte aus dieser Chiquadratverteilung benutzen für diese Art von Statistiken, dann erreicht man im ungünstigsten Fall nur einen Typ-I-Fehler von 0,0223 beim Signifikanzniveau $\alpha = 0,05$, also weniger als die Hälfte¹. Aus diesem Grunde sind Tests mit einer von einer Asymptotik abhängigen Verteilung für hochdimensionale Daten unbrauchbar. Eine asymptotische Pivotgröße gibt es bei beliebigen Abhängigkeitsstrukturen nicht.

In Pauly, Ellenberger und Brunner (2015) wird eine „Kompromißlösung“ vorgeschlagen, mit einem anderen Parameter als dem Box'schen Epsilon die „richtige“ Grenzverteilung auszuwählen, siehe auch Sattler und Pauly (2018). Dieser Parameter involviert die Schiefe der quadratischen Form in der Teststatistik. Diese ist zwangsläufig abhängig von der Normalverteilungsannahme und ist bei mehreren Stichproben überdies nur mit sehr viel Rechenaufwand zu schätzen.

Im weitesten Sinne ebenfalls asymptotisch ist der parametrische Bootstrap, siehe z. B. die Anwendung bei Krishnamoorthy, Lu und Mathew (2007). Der parame-

¹Ausgabe vom R-Befehl

```
1 - pnorm((qchisq(0.95,1)-1)/sqrt(2))
```

oder in SAS

```
1 - probnorm(QUANTILE('CHISQUARE', 0.95, 1));
```

trische Bootstrap findet die kritischen Werte zu einer Teststatistik, indem die Statistik mehrfach mit (pseudo-)Zufallszahlen ausgerechnet wird, die –bedingt auf die Nullhypothese– der empirischen Verteilungsfunktion folgen. Letztere konvergiert gewöhnlich gegen die wahre Verteilungsfunktion, so daß die so simulierte Verteilung gegen die wahre Verteilung der Teststatistik unter der Nullhypothese konvergieren muß. Im Hochdimensionalen ist diese Konvergenz der empirischen Verteilungsfunktion aber wertlos, da insbesondere die Streuung nur in einem echten Unterraum des d -dimensionalen Beobachtungsraumes beobachtet werden kann.

2 Robuste Mehrstichprobenstatistiken

2.1 Skizze des Vorgehens

Ziel dieses Kapitels ist es, eine *robuste Statistik* für unbalancierte Daten bei Kovarianzheterogenität zu entwickeln. Wie in Abschnitt 1.6.1 angesprochen, soll in dieser Arbeit ein Test als robust gelten, wenn er zunächst bei Normalverteilung den Typ-I-Fehler gut einhält, das heißt, daß seine Abweichungen vom vorgegebenen Signifikanzniveau nicht größer als bei der ANOVA-Typ-Statistik sind. Weitere Voraussetzungen sind, daß die approximative Verteilung der Teststatistik in den Modellen aus den Abschnitten 1.6.2 bis 1.6.4 hergeleitet werden kann, daß durch den robusten Test die Power sich nicht gegenüber dem bekannten Test verschlechtert und daß bei moderater Nichtnormalverteilung das vorgegebene Fehlerniveau nicht zu stark verfehlt wird. Bei Bradley 1978 wird ein tatsächlicher Typ-I-Fehler im Bereich jenseits von $[\alpha/2, 3\alpha/2]$ als zu stark verfehlt betrachtet.

Vor allem aber soll der robuste Test für den Anwendungszweck der ANOVA-Typ-Statistik geeignet sein. Daraus ergeben sich folgende Eigenschaften:

1. Es sollen lineare Hypothesen der Form aus Abschnitt 1.4 getestet werden können.
2. Der Nichtablehnbereich des Tests soll kompakt sein im Unterschied zu den Methoden mit einer verallgemeinerten Inversen aus Abschnitt 1.8.1.
3. Der Test soll invariant unter Skalentransformationen sein, d. h. wenn alle die Messungen mit einer Transformation $x \mapsto ax, a \in \mathbb{R}^*$ transformiert werden, soll sich das Ergebnis nicht ändern.
4. Der Test soll auf Meßwiederholungen bzw. Split-Plot-Designs ausgelegt sein, also auf der euklidischen Norm von $T\boldsymbol{\mu}$ basieren und nicht auf der Pearson-Distanz aus Abschnitt 1.8.2.
5. Auf den Gebrauch von Asymptotiken soll für die Herleitung wegen der damit einhergehenden Probleme bei hoher Dimensionalität, die in Abschnitt 1.8.3 berichtet wurden, möglichst verzichtet werden.

Die ersten vier Anforderungen können durch eine Form der ATS erfüllt werden, wenn Q_n aus Definition (12) in Abschnitt 1.7.2 als Grundlage genommen wird. Die entscheidende Frage bleibt dann nur noch, welche quadratische Form im Nenner des Quotienten stehen soll. Sie ist notwendig, um die Skaleninvarianz, die dritte geforderte Eigenschaft, sicherzustellen. Sie soll dafür unabhängig von der tatsächlich vorliegenden Alternative denselben Erwartungswert haben, den Q_n unter der Nullhypothese hätte.

Um dem Behrens-Fisher-Problem gerecht zu werden, soll diese Erwartungsgleichheit für alle Kombinationen aus Stichprobenumfängen und Kovarianzmatrizen gelten,

damit die Kovarianzmatrixparameter sich weggürzen wie in Abschnitt 1.7.2. Die unter diesen Anforderungen eindeutige zulässige Gestalt der quadratischen Form des Nenners wird im nächsten Abschnitt herausgearbeitet.

Im übernächsten Abschnitt wird untersucht, wann Statistiken, die Quotienten aus Q_n und anderen quadratischen Formen sind, approximative Verteilungen haben, die unabhängig von speziellen Eigenschaften des Normalverteilungsmodells sind. Diese sind zum einen die Folge der Momente, insbesondere der Momente von quadratischen Formen in normalverteilten Zufallsvektoren. Zum Anderen würde das normalverteilte Modell die Eigenschaft garantieren, daß Orthogonalität von quadratischen Formen in denselben Zufallsvektoren auch deren Unabhängigkeit bedeutet. In diesem Abschnitt wird darauf eingegangen, wie sich Robustheit von der Normalverteilung zum Behrens-Fisher-Problem verhält.

Um die Herleitung einer Statistik von diesen Besonderheiten des normalverteilten Modells zu befreien, soll eine Approximation des gesamten Quotienten wie in Abschnitt 1.7.3 vorgenommen werden mit entsprechenden Approximationen für Erwartungswert und Varianz des Quotienten. Dadurch wird die Unabhängigkeit nicht mehr benötigt. Wenn die so entstandene Verteilungsapproximation auch noch frei von Parametern ist, die von den höheren Momenten abhängen, dann kann die Robustheit folgert werden.

Es stellt sich heraus, daß genau der Test hergeleitet werden kann, ohne Eigenschaften der Normalverteilung zu benutzen, der auch auf das Behrens-Fisher-Problem ausgelegt ist.

Zur besseren Vergleichbarkeit mit den bekannten Statistiken aus Abschnitt 1.7, die vor allem F -verteilt sind, sollen anschließend F -Verteilungen gefunden werden, die ähnliche oder zumindest nicht liberalere kritische Werte liefern wie die Verteilungsapproximation.

2.2 Hochdimensionales Behrens-Fisher-Problem

Ausgangspunkt soll wieder die quadratische Form

$$Q_n = \bar{\mathbf{Y}}' \mathbf{T} \bar{\mathbf{Y}} = \bar{\mathbf{Y}}' (\mathbf{T}_a \otimes \mathbf{T}_d) \bar{\mathbf{Y}}$$

sein. Unverändert zum normalverteilten Modell bleiben die Erwartungswerte von Q_n :

Satz 6. *Es ist*

$$E(Q_n) = \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\boldsymbol{\Sigma}_i \mathbf{T}_d) + \boldsymbol{\mu}' \mathbf{T} \boldsymbol{\mu} \quad (25)$$

$$\stackrel{H_0}{=} \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\boldsymbol{\Sigma}_i \mathbf{T}_d)$$

Beweis. Dies ist auszurechnen mit dem Satz von Lancaster, der nicht die Normalverteilung voraussetzt, siehe im Anhang auf Seite 78. \square

Q_n soll durch einen erwartungstreuen Schätzer Q_s für $E_{H_0}Q_n$ dividiert werden, damit sich dieser unbekannte Erwartungswert herauskürzt. Auf diese Weise wird ein Schritt der ATS-Herleitung aus Abschnitt 1.7.2 imitiert, da $E_{H_0}Q_n = EQ_s$ sich in der Statistik (16) herauskürzte.

Q_s soll möglichst effizient sein, d. h. $\text{Var}(Q_s)$ soll minimal sein. Außerdem soll Q_s orthogonal zu Q_n sein, d. h. wenn $Q_n = \mathbf{Y}'\mathbf{M}_n\mathbf{Y}$ und $Q_s = \mathbf{Y}'\mathbf{M}_s\mathbf{Y}$ geschrieben werden kann, soll $\mathbf{M}_n \left(\bigoplus_{i=1}^a \mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}_i \right) \mathbf{M}_s = \mathbf{0}$ sein. Diese beiden Bedingungen erfüllen die ATS und die ANOVA, wo sie exakt ist. In diesem Falle sind diese Bedingungen also ein notwendiges Kriterium dafür, daß die zu findende Statistik mit der ANOVA übereinstimmen kann.

Der nächste Satz erweitert Theorem 1 aus Atiqullah (1962). Dort wurde die Effizienz des Residuenschätzers aus der ANOVA für quadratische Balanciertheit oder Normalverteilung (genauer: Kurtosis wie bei Normalverteilung) gezeigt. Weil hier eine Aussage jenseits quadratischer Balanciertheit und Normalverteilung angestrebt wird, muß auch der Satz neu formuliert werden, indem er auf $E_{H_0}Q_n$ ausgelegt wird anstatt auf die quadratische Form aus der ANOVA. $E_{H_0}Q_n$ unterscheidet sich durch die ungewichteten Mittelwerte von den quadratischen Formen aus der ANOVA bzw. der Arbeit von Atiqullah (1962).

Satz 7. *Unter allen positiv semidefiniten quadratischen Formen ist*

$$\mathbf{Y}'_i \left(\frac{t_{ii}}{n_i} \mathbf{P}_{n_i} \otimes \mathbf{T}_d \right) \mathbf{Y}_i$$

unverzerrter Schätzer für $\frac{t_{ii}}{n_i} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)$ mit minimaler Varianz im Rahmenmodell mit $E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})^2 \leq E(\mathbf{Y}'_{ik} \mathbf{Y}_{ik})^2 < \infty$.

Beweis. Im Anhang A wird der Beweis auf Seite 79 geführt. Er beruht darauf, daß der optimale Schätzer invariant unter Permutation der Beobachtungsvektoren sein muß. \square

Damit ist zwangsläufig Q_n durch

$$Q_s = \mathbf{Y}' \left(\bigoplus_{i=1}^a s_i \mathbf{P}_{n_i} \otimes \mathbf{T}_d \right) \mathbf{Y}$$

zu dividieren. $E_{H_0}Q_n = EQ_s$ wird dann durch Koeffizientenvergleich der a unabhängigen Stichproben erreicht, wenn $s_i = \frac{t_{ii}}{n_i}$ ist. Es ist also $Q_s = Q_e$ mit

$$Q_e = \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\hat{\boldsymbol{\Sigma}}_i \mathbf{T}_d)$$

wie in Abschnitt 1.7.2. Die Teststatistik, die auf das Behrens-Fisher-Problem ausgelegt ist, ist daher wie in Abschnitt 1.7.2

$$F = \frac{Q_n}{Q_e}.$$

Diese Statistik hat zu gegebenen Q_n das optimale $Q_s = Q_e$. Das eben bewiesene Resultat kann in folgendem Satz festgehalten werden.

Satz 8. *Unter allen Teststatistiken für die Hypothese $\mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, die*

1. *auf der euklidischen Norm von $\mathbf{T}\bar{\mathbf{Y}}$ basieren,*
2. *aus einem Quotienten quadratischer Formen bestehen,*
3. *bei unterschiedlichen Kovarianzmatrizen und Unbalanciertheit keine Kenntnis der $\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_a$ voraussetzen,*

hat

$$F = \frac{Q_n}{Q_e}$$

die geringste Varianz.

Zu diesem F ist im verallgemeinerten Modell mit Regularitätsbedingung (2) eine geeignete Verteilungsapproximation zu finden. Wichtig ist, daß man $\text{Cov}(Q_n, Q_e) = 0$ nicht mehr verwenden darf, wie für den Sonderfall $a = 1$ in Abschnitt 1.7.3 gesehen.

2.3 Robuste Verteilungsapproximation des Quotienten

Es sei wieder Q_s eine quadratische Form mit

$$E_{H_0}(Q_n) = E(Q_s)$$

und

$$Q_s = \mathbf{Y}' \left(\bigoplus_{i=1}^a s_i \mathbf{P}_{n_i} \otimes \mathbf{T}_d \right) \mathbf{Y}. \quad (26)$$

Die Approximation für die ATS kann ohne Normalverteilungsannahme hergeleitet werden, indem ein f gefunden wird, so daß in

$$\frac{Q_n}{Q_s} \dot{\sim} g\chi_f^2 \quad (27)$$

beide Seiten in Varianz und Erwartungswert übereinstimmen. Auf diese Weise wird nicht mehr benötigt, daß bei Normalverteilung Zähler und Nenner stochastisch un-

abhängig sind. Man erhält das Gleichungssystem

$$\begin{aligned} E\left(\frac{Q_n}{Q_s}\right) &= gf \\ \text{Var}\left(\frac{Q_n}{Q_s}\right) &= 2g^2 f. \end{aligned}$$

Dessen Lösung ist

$$g = \frac{\text{Var}\left(\frac{Q_n}{Q_s}\right)}{2E\left(\frac{Q_n}{Q_s}\right)} \quad f = \frac{2E^2\left(\frac{Q_n}{Q_s}\right)}{\text{Var}\left(\frac{Q_n}{Q_s}\right)}. \quad (28)$$

Da exakte Formeln für diese Varianz und diesen Erwartungswert von den unbekannt-ten Verteilungen der Daten abhängen, haben sie für die weitere Herleitung keinen Sinn.

Stattdessen wird eine Taylorapproximation für Erwartungswert und Varianz eines Quotienten von Zufallsvariablen verwendet. Sie beruht auf dem Taylorpolynom erster Ordnung für die Funktion $(x_1, x_2) \mapsto \frac{x_1}{x_2}$, entwickelt um $(E_{H_0}Q_n, E_{H_0}Q_s)$. Für Details der Herleitung siehe im Anhang Korollar 31 auf Seite 84. Den Erwartungswert erhält man näherungsweise durch

$$E\left(\frac{Q_n}{Q_s}\right) \doteq \frac{E(Q_n)}{E(Q_s)} \stackrel{H_0}{=} 1, \quad (29)$$

weil $E(Q_n) \stackrel{H_0}{=} E(Q_s)$ gefordert wurde. Setzt man diese Approximation für den Erwartungswert ein, ist automatisch der Streckungsparameter

$$g = f^{-1} = \frac{1}{2} \text{Var}\left(\frac{Q_n}{Q_s}\right).$$

Außerdem vereinfacht sich die Varianz in der Taylorapproximation zu

$$\begin{aligned} \text{Var}\left(\frac{Q_n}{Q_s}\right) &\doteq \left(\frac{EQ_n}{EQ_s}\right)^2 \left(\frac{\text{Var}(Q_n)}{E^2Q_n} + \frac{\text{Var}(Q_s)}{E^2Q_s} - \frac{\text{Cov}(Q_n, Q_s)}{EQ_nEQ_s}\right) \\ &\stackrel{H_0}{=} \frac{\text{Var}(Q_n) + \text{Var}(Q_s) - 2\text{Cov}(Q_n, Q_s)}{E_{H_0}^2Q_n} \end{aligned} \quad (30)$$

$$= \frac{\text{Var}(Q_n - Q_s)}{E^2Q_s} \quad (31)$$

Die Lösung für den Streckungsparameter und den Freiheitsgrad in der Box-Approximation des gesamten Quotienten läuft also auf die Statistik

$$F \sim f^{-1}\chi_f^2 \quad \text{mit } f = \frac{2E^2Q_s}{\text{Var}(Q_n - Q_s)}$$

hinaus.

Anzumerken ist, daß man hier wie in 1.7.3 für $E(Q_n/Q_s)$ eine Taylorapproximation zweiter Ordnung benutzen könnte. Die zusätzliche Genauigkeit durch die höhere Ordnung geht später aber durch schlechter schätzbare Parameter wieder verloren.

In der Varianz $\text{Var}(Q_n - Q_s)$ kommt ein neuer Störparameter hinzu, der sich durch die Abhängigkeit von Q_n und Q_s bei Nichtnormalverteilung ergibt:

Satz 9. *Mit der Notation $\kappa_i(\mathbf{Y}_{i1}) := E(\mathbf{Y}'_{i1} \mathbf{T}_d \mathbf{Y}_{i1})^2 - 2\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 - \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)$ gilt unter H_0 :*

1.

$$\text{Var}_{H_0}(Q_n) = 2 \sum_i^a \sum_{i'}^a t_{ii'}^2 n_i^{-1} n_{i'}^{-1} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'}) + \sum_{i=1}^a \frac{t_{ii}^2}{n_i^3} \kappa_i(\mathbf{Y}_{i1}) \quad (32)$$

2.

$$\text{Var}(Q_s) = 2 \sum_{i=1}^a (n_i - 1) s_i^2 \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 + \sum_{i=1}^a \frac{s_i^2 (n_i - 1)^2}{n_i} \kappa_i(\mathbf{Y}_{i1})$$

3.

$$\text{Cov}_{H_0}(Q_n, Q_s) = \sum_{i=1}^a \frac{t_{ii} s_i (n_i - 1)}{n_i^2} \kappa_i(\mathbf{Y}_{i1})$$

Beweis. Der Beweis beruht auf Korollar 2 in Bathke und Harrar (2008), siehe im Anhang auf Seite 80. \square

Da bei Normalverteilung $E(\mathbf{Y}'_{i1} \mathbf{T}_d \mathbf{Y}_{i1})^2 = 2\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 + \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)$ ist (siehe Mathai und Provost 1992, S. 53) verschwinden in diesem Fall die κ_i -Terme. Dies ist im Einklang mit der Eigenschaft der Normalverteilung, daß orthogonale quadratische Formen stets unabhängig zueinander sind, denn dadurch verschwindet auch $\text{Cov}_{H_0}(Q_n, Q_s)$. Bei Nichtnormalverteilung gilt dies nur in Ausnahmefällen. Daher müßte man zusätzlich zu den Termen der Art $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})$, die man auch bei Normalverteilung schätzt –siehe Becker (2010), Brunner, Bathke und Placzek (2012) und Harden (2013)– auch die κ_i schätzen. Damit würde man aber Genauigkeit im normalverteilten Fall verlieren.

Die einzige Lösung dieser Situation ist, Q_s so festzulegen, daß die κ_i -Summen in der Summe $\text{Var}_{H_0}(Q_n) + \text{Var}(Q_s) - 2\text{Cov}_{H_0}(Q_n, Q_s)$ sich aufheben. Setzt man die Resultate aus Satz 9 in die Varianzapproximation des Quotienten (30) ein, dann verschwinden die κ_i genau dann, wenn

$$s_i = \frac{t_{ii}}{n_i (n_i - 1)}$$

ist, also genau im Fall der auf das Behrens-Fisher-Problem ausgelegten Statistik. Dieses Ergebnis kann man als folgenden Satz formulieren.

Satz 10. Eine approximative Statistik aus einem Quotienten aus Q_n und einer quadratischen Form Q_s mit $Q_s = \bigoplus_{i=1}^a s_i \mathbf{P}_{n_i} \otimes \mathbf{T}_d$ hängt genau dann nicht von $E(\mathbf{Y}'_{i1} \mathbf{T}_d \mathbf{Y}_{i1})^2$ mit $i = 1, \dots, a$ ab, wenn

$$s_i = \frac{t_{ii}}{n_i(n_i - 1)}$$

ist, also wenn die linke Seite der Statistik mit der ANOVA-Typ-Statistik (16) übereinstimmt.

Beweis. Die detaillierte Rechnung findet sich auf S. 82. □

Anschaulich entscheidet sich die Robustheit an den Koeffizienten vor Termen der Art $\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik}$ in Q_n und Q_e . Treten diese Terme in beiden quadratischen Formen in übereinstimmenden Linearkombinationen auf, also weil die Matrizen \mathbf{M}_n und \mathbf{M}_e identische Hauptdiagonalen haben, dann erscheinen die Varianzen dieser Terme nicht mehr in der Näherungsformel 31 für die Varianz des Quotienten, genauer gesagt in $\text{Var}_{H_0}(Q_n - Q_s) = \text{Var}_{H_0}(Q_n) + \text{Var}(Q_s) - 2\text{Cov}_{H_0}(Q_n, Q_s)$ laut dem Resultat von Satz 9. Aber nur die Varianz von Termen der Art $\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik}$ hängt von den unbekanntenen Termen der Art $E(\mathbf{Y}'_{i1} \mathbf{T}_d \mathbf{Y}_{i1})^2$ ab, so daß dieser Störparameter wegfällt.

In der Geisser-Greenhouse-Statistik hingegen stimmen die Hauptdiagonalen bei unbalancierten Layouts nicht überein. Daher ist die Geisser-Greenhouse-Statistik im Allgemeinen nicht robust.

Abschließend kann noch eine Bemerkung zu asymptotischen Ansätzen gemacht werden. Für $n_1, \dots, n_a \rightarrow \infty$ verschwinden die störenden κ_i -Terme gegenüber den Varianzen der quadratischen Formen. Allerdings ist eine Asymptotik über die Stichprobenumfänge nicht mit hochdimensionalen Daten vereinbar.

Arbeiten, die Asymptotiken nur über die Dimensionalität benutzen, können an der Stelle eine Voraussetzung an die asymptotische Abhängigkeitsstruktur zwischen den Subplotstufen fordern, die in letzter Konsequenz auf $\kappa_i \xrightarrow{d \rightarrow \infty} 0$ hinausläuft. Kann die Verhältniskonsistenz gezeigt werden, die in Abschnitt 22 erwähnt wurde, würde in letzter Konsequenz auch $\kappa_i n_i^{-3} \rightarrow 0$ gelten. Auf beide Weisen wäre eine asymptotische Herleitung einer Statistik aus Q_n und Q_e ebenfalls robust.

Die Achillesferse asymptotischer Ansätze ist aber, was bei Nichtzutreffen der Annahmen geschieht und was dies für die Anwendung bedeutet. In der vorliegenden Arbeit soll eine Statistik im Vorhinein so konstruiert werden, daß die Störparameter sich finit aufheben, wenn noch keine Asymptotik greift.

2.4 Kritische Werte

Als Resultat des vorherigen Abschnittes kann im nichtnormalverteilten Rahmenmodell mit der Eigenschaft (2) die Statistik

$$F \dot{\sim} f^{-1}\chi_f^2 \quad (33)$$

approximiert werden mit

$$f = \frac{b_1}{b_n + b_e},$$

wobei b_1 , b_n und b_e wie in (17), (18) und (19) in Abschnitt 1.7.2 auf Seite 17 definiert sind. Im fertigen Test würden also die kritischen Werte einer $f^{-1}\chi_f^2$ -Verteilung verwendet werden.

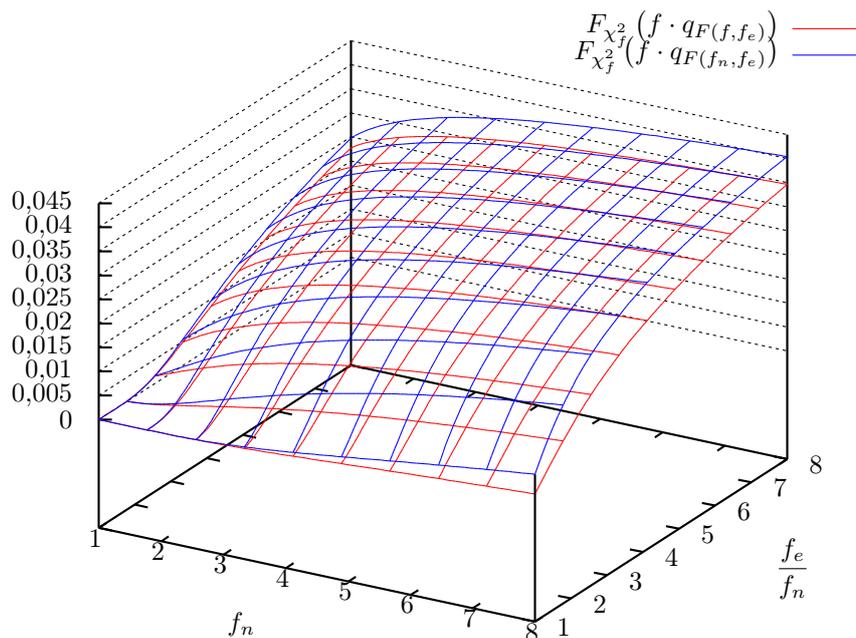
Dies wurde in dem Rahmenmodell hergeleitet, das das multivariat normalverteilte Modell verallgemeinert. Gleichzeitig ist die linke Seite von Statistik (33) dieselbe wie von Statistik (16). Man möchte daher erwarten, daß sich die kritischen Werte nicht verändern, ob nun die Daten normalverteilt sind oder nicht. Tatsächlich liegt genau darin der praktische Wert der Robustheitsaussage. Deswegen soll als nächstes argumentiert werden, daß der Unterschied zwischen den oberen Quantilen einer $f^{-1}\chi_f^2$ -Verteilung und einer $F(f_n, f_e)$ -Verteilung praktisch nicht ins Gewicht fallen, die $F(f_n, f_e)$ -Verteilung aber zu bevorzugen ist, weil sie konservativer ist.

Satz 11 (Approximation durch F -Verteilung). *Die Quantile einer $f^{-1}\chi_f^2$ -Verteilung konvergieren an die Quantile einer $F(f_n, f_e)$ -Verteilung, wenn $n_i \rightarrow \infty$ ist.*

Beweis. Siehe im Anhang auf Seite 84. □

Numerisch berechnete Unterschiede zwischen den Quantilen bzw. den Verteilungsfunktionen sollen zur Illustration für kleine Werte der Freiheitsgrade dargestellt werden. In Abbildung 3 auf der nächsten Seite ist dafür ein geeigneter Plot zu finden. Abgebildet ist mit dem blauen Flächengitter der von 1 abgezogene Wert der Verteilungsfunktion einer $f^{-1}\chi_f^2$ -Verteilung auf der z-Achse, wenn das 0,95-Quantils einer $F(f_n, f_e)$ -Verteilung eingesetzt wird. Mit dem roten Flächengitter ist derselbe Wert abgebildet, wenn dasselbe Quantil einer $F(f, f_e)$ -Verteilung eingesetzt wird. Auf der x- und der y-Achse sind f_n und f_e/f_n jeweils zwischen 1 und 8 aufgetragen, da nur $f_n \leq f_e$ von Interesse ist. Der Parameter f_e ist bei realistisch großen Datensätzen sogar um ein Vielfaches größer als f_n .

Auf diese Weise kann man also erkennen, welche Typ-I-Fehler zustandekommen, wenn für einen Test der kritische Wert aus einer F -Verteilung mit den jeweiligen Freiheitsgraden genommen wird, obwohl die Verteilung der Teststatistik einer $f^{-1}\chi_f^2$ -Verteilung folgt. Wo das Flächengitter unterhalb von 0,05 auf der z-Achse, liegt ist der Test konservativ.

Abbildung 3: Wert der F_F -Verteilungsfunktion

Man erkennt auf Abbildung 3, daß ein Test mit den kritischen Werten sowohl der einer $F(f_n, f_e)$ -Verteilung wie auch der einer $F(f, f_e)$ -Verteilung konservativ ist, wenn die wahre Verteilung der Teststatistik einer $f^{-1}\chi_f^2$ -Verteilung folgt: Beide Flächengitter bleiben im Plot unterhalb von 0,05. Sie nähern sich aber mit wachsendem f_n und f_e von unten an 0,05 an. In Bezug auf f_e wird dies durch Satz 11 angekündigt. Für sehr kleines f_e , d. h. $f_e = f_n$, ist die Approximation mit den beiden F -Verteilungen sehr konservativ, weil man dann noch sehr weit von der Asymptotik aus Satz 11 entfernt ist.

Die am wenigsten konservativen Testentscheidungen kommen mit den kritischen Werten einer $F(f_n, f_e)$ -Verteilung zustande, da das blaue Flächengitter oberhalb des roten Flächengitters liegt. Daher sind diese kritischen Werte zu bevorzugen. Am Ende erhält man auf diese Weise exakt denselben Test wie in Abschnitt 1.7.2, allerdings ohne Rückgriff auf die Normalverteilungsannahme.

2.5 Simulationen

2.5.1 Simulationslayout

Nur mit Monte-Carlo-Simulationen kann gezeigt werden, inwieweit die Taylor-Approximationen geeignet sind, wie sich die robuste Statistik insbesondere im Vergleich zum etablierten Verfahren, der Geisser-Greenhouse-Statistik, verhält, und wie robust die Prozedur wirklich ist.

Die Parameterkonstellationen, die zu simulieren sind, sollen einerseits umfassend genug sein, damit keine Szenarien übersehen werden, in denen die Statistiken bzw. die Approximationen sich sehr anders verhalten als sonst. Andererseits soll auf Szenarien verzichtet werden, die etwa durch die Invarianzeigenschaften der Tests redundant sind. Außerdem sollen auch starke Abweichungen von dem Normalverteilungsmodell einbezogen werden, solange sie noch den Modellvoraussetzungen aus den Abschnitt 1.6 genügen.

Die Hypothesen, zu denen die Tests und die quadratischen Formen simuliert werden, sind die drei hochdimensionalen Hypothesen in der Aufzählung auf Seite 5. Noch mehr andere Hypothesen, etwa für mehrere Wholeplot- und Subplotfaktoren müssen nicht simuliert werden, weil es einerlei ist, ob die Unterschiedlichkeit der denkbaren Verteilungen von $\mathbf{T}\bar{\mathbf{Y}}$ durch Variationen in den Hypothesenmatrizen oder in den Kovarianzmatrizen zustandekommt.

Die Zahl a der Wholeplotfaktorstufen variiert zwischen zwei und vier. Die Zahl d der Subplotfaktorstufen wird mit

$$d \in \{4, 8, 16, 32, 64, 128, 256, 512\}$$

festgelegt.

Die Stichprobenumfänge liegen zwischen 10 und 30. Kleinere Stichprobenumfänge werden von der Anwenderseite bei Versuchsplanungen selten als ausreichend erachtet. Größere Stichprobenumfänge sind für diese Arbeit mit ihrem Fokus auf hoher Dimensionalität weniger interessant.

Zunächst ist das „Standardmodell“ der Normalverteilung zu simulieren, um den Effekt von Unbalanciertheit und Heteroskedastizität vom Effekt durch die Nichtnormalverteilung unterscheiden zu können. Bei Normalverteilung und allgemein im elliptischen Modell reduziert sich zugleich der Simulationsaufwand, weil Invarianz unter linearen Transformationen ausgenutzt werden können, siehe Satz 32 auf Seite 85 im Anhang. Dieser Satz besagt anschaulich, daß sich bei elliptischen Verteilungen durch lineare Transformationen nur die Parameter Kovarianzmatrix und Erwartungswert verändern, nicht die „Form“ der Verteilung.

Speziell im Fall von identischen Kovarianzmatrizen ist der Vektor der Eigenwerte der Kovarianzmatrix der einzig wichtige Parameter neben den Stichprobenumfängen, weil die Statistiken invariant unter orthogonalen Transformationen der d -dimensionalen Beobachtungsvektoren sind. Die genaue Struktur der Kovarianzmatrix spielt in diesem Fall keine Rolle. So ist es keine Einschränkung, wenn nur Toeplitz-Matrizen simuliert werden, also Matrizen, bei denen innerhalb der Diagonalen jeweils identische Werte stehen. Die Toeplitz-Matrizen haben dadurch den anschaulichen Vorteil, daß die Abhängigkeit aller Messungen untereinander komplett durch die erste Zeile der Matrix verstanden werden kann. Ferner sind die Kovarianzstrukturen, die in Varianzkomponentenverfahren geschätzt werden, meistens Toeplitz-Matrizen (siehe Searle, Casella und McCulloch (1992)).

Es werden vier verschiedene Abhängigkeitsverläufe zwischen den Subplot-Faktoren benutzt: Konstante Korrelation, stark und weniger stark geometrisch abnehmende Korrelationen und linear abfallende Korrelationen. Damit sind auch die Kovarianzmatrizen unterschiedlich sphärisch, d. h. die nach Größe sortierten Eigenwerte fallen unterschiedlich stark ab.

Für das additiv-lineare Modell sollen zusätzlich Kovarianzstrukturen verglichen werden können, die unter einer Nullhypothese zu unterschiedlichen additiv-linearen Verteilungen bei denselben Kovarianzparametern führen. Der Grund dafür ist, daß im additiv-linearen Modell die Invarianz im Allgemeinen nicht mehr gilt, siehe wiederum Satz 32. Zu diesem Zweck werden die Identitätsmatrix und die sogenannte Compound-symmetry-Struktur verwendet.

Die einbezogenen Strukturen sind im Detail somit folgendermaßen definiert:

1. Die *Identitätsmatrix* \mathbf{I}_d , d. h. die unveränderte Übernahme der \mathbf{Z}_{ik} -Vektoren
b. Hier ist der größte Einfluß der Nichtnormalverteilung in additiv-linearen Modell zu erwarten. Andererseits liegt hier immer Sphärizität vor, die ANOVA wäre also bei Normalverteilung exakt und
2. Eine *Compound-symmetry-Struktur* $CS := \frac{1}{2}\mathbf{I}_d + \frac{1}{2}\mathbf{J}_d$. Zusammen mit $\mathbf{T}_d = \mathbf{P}_d$ ergibt sich dieselbe Kovarianzmatrix wie bei der Identität, allerdings sind die Komponenten der \mathbf{Y}_{ik} nunmehr Faltungen. Dies macht für das additiv-lineare Modell einen Unterschied. Die Korrelationen zwischen zwei verschiedenen Zeitpunkten sind stets dieselben. Hier liegt mit den Hypothesen $H_0(AB)$ und $H_0(B)$ Sphärizität vor.
3. Im Zeitverlauf geometrisch abnehmende Korrelationen treten in einer *autoregressiven Struktur mit geringen Korrelationen*

$$AR(0, 6) := \left[0, 6^{|j-j'|} \right]_{j,j'=1,\dots,d}$$

auf. Sie führt zu einem etwas kleineren f_n als die vorherigen beiden Strukturen. Außerdem führt diese Struktur und die folgenden bei den interessierenden Hypothesenmatrizen zu Nichtsphärizität.

4. Bei einer *autoregressiven Struktur mit hohen Korrelationen*

$$AR(0, 9) := \left[0, 9^{|j-j'|} \right]_{j,j'=1,\dots,d}$$

ist f_n noch kleiner.

5. $linToe := \left[1 - \frac{|j-j'|}{d} \right]_{j,j'=1,\dots,d}$ ist eine *Toeplitzstruktur mit linear abfallenden Korrelationen*. Dadurch hat man bei hoher Dimensionalität größere Abhängigkeiten als bei den autoregressiven Strukturen. Gleichzeitig führt es im Gegensatz zur *CS*-Struktur unter den genannten Hypothesen zu Nichtsphärizität.

6. $harToe := \left[\min \left\{ 1, \frac{1}{2^{|j-j'|}} \right\} \right]_{j,j'=1,\dots,d}$ ist eine *Toeplitzstruktur mit harmonisch abfallenden Korrelationen*. Sie steht zwischen der autoregressiven Struktur und der Struktur mit linear abfallenden Korrelationen.

Neben der Normalverteilung werden im additiv-linearen Modell folgende Verteilungen der Z_{ikj} simuliert:

1. Die *Rademacher-Verteilung*, also die Verteilung mit $P(X = -1) = P(X = 1) = \frac{1}{2}$ als extrem diskretes Beispiel.
2. Eine zentrierte Exponentialverteilung mit Parameter 1 als Beispiel für eine extrem schiefe Verteilung.
3. Die Gleichverteilung auf $\left[-\frac{\sqrt{12}}{2}, \frac{\sqrt{12}}{2} \right]$ als Beispiel für eine stetige Verteilung mit kompaktem Träger.
4. Eine zentrierte Lognormalverteilung mit den Parametern $\mu = 0$ und $\sigma^2 = \log\left(\frac{1+\sqrt{5}}{2}\right)$. Sie kann erzeugt werden aus $U \sim \mathcal{N}(0, \sigma^2)$ durch $\exp(U) - \exp\left(\frac{\sigma^2}{2}\right)$ und ist ein anderes Beispiel für eine schiefe Verteilung.
5. Eine $\sqrt{3/5} \cdot t_5$ -Verteilung, also eine standardisierte t_5 -Verteilung als Beispiel, bei dem das fünfte Momente nicht mehr existiert.

Für einen Zufallsvektor \mathbf{Y}_{ik} wird erst ein d -dimensionaler Vektor von Pseudozufallszahlen unabhängiger Gleichverteilungen auf dem Einheitsintervall erzeugt. Der Algorithmus für diese Pseudozufallszahlen ist der Mersenne-Twister, der in R (R Core Team (2018)) voreingestellt ist.

Anschließend wird \mathbf{Z}_{ik} erzeugt, indem die Pseudozufallszahlen aus dem Intervall $(0, 1)$ komponentenweise auf die entsprechende Verteilung mit der Quantilmethode monoton transformiert werden. Die Pseudozufallszahlen werden dabei für die unterschiedlichen Verteilungen wiederverwertet. Das spart Rechenzeit und garantiert, daß unterschiedliche Ergebnisse zwischen den Verteilungen nur auf diese zurückgeführt werden können und nicht zufällig durch die Pseudozufallszahlen entstehen können.

Zuletzt wird der \mathbf{Z}_{ik} -Vektor mit dem $\mathbf{\Gamma}_i$ aus Modellgleichung (3) multipliziert, das durch eine Cholesky-Zerlegung der vorgegebenen Kovarianzmatrix $\mathbf{\Sigma}_i$ erreicht wird.

Im elliptischen Modell genügt es, die Verteilung der skalaren Zufallsvariablen W_{ik} aus der Modellgleichung (5) zu beschreiben. Aus der Modellgleichung (5) ergibt sich damit unmittelbar die Erzeugungsmethode der Zufallsvektoren. Folgende Verteilungen werden einbezogen:

1. $|W_{ik}|$ ist exponentialverteilt, d. h. es resultiert eine Verteilung, deren Dichte ein singuläres Maximum beim Erwartungswert hat („spitzgipflig“).

2. \mathbf{Y}_{ik} ist Pearson-II-verteilt mit Parameter $m = \frac{1}{4}$, siehe Johnson (1987), S. 111. Diese Verteilung ist nicht unimodal. Die Wahrscheinlichkeitsdichte ist an den Rändern eines Ellipsoids um den Erwartungswert größer als um den Erwartungswert selber. Der Zweck dieser Verteilung ist, nicht nur unimodale oder Gleichverteilungen zu simulieren.
3. \mathbf{Y}_{ik} ist gleichverteilt auf der Menge $\left\{ \mathbf{x} \mid \|\mathbf{x}\|_2 \leq \frac{\sqrt{12}}{2} \right\}$. Dies ist ebenfalls eine Pearson-II-Verteilung.
4. \mathbf{Y}_{ik} ist Pearson-II-verteilt mit Parameter $m = \frac{1}{2}$. Die Dichte entspricht einer „Halbkugel“, siehe Johnson (1987), S. 114.
5. \mathbf{Y}_{ik} ist Pearson-II-verteilt mit Parameter $m = 1$, so daß die Dichte einer Glocke mit kompaktem Träger gleicht, siehe Johnson (1987), S. 115.
6. $12d/5 \cdot W_{ik}^2$ folgt einer durch normierten t -Verteilung mit fünf Freiheitsgraden. Dies ist ein Beispiel für eine Verteilung, bei der das fünfte Moment nicht mehr existiert.
7. $W_{ik}^2 + 1$ folgt einer Pareto-Verteilung mit der Dichtefunktion

$$x \mapsto \frac{4,08}{x^{5,08}}.$$

Dadurch hat W_{ik} mit 444,889 eine größere Kurtosis als im Falle einer t_5 -Verteilung, bei der es nur 9 sind.

Die elliptisch verteilten Zufallsvektoren werden im Detail erzeugt, indem eine Pseudozufallszahl aus dem $(0, 1)$ -Intervall mit der Quantilmethode auf die gewünschte Verteilung von W_{ik} bzw. $|W_{ik}|$ transformiert wird. Bei den Pearson-II-Verteilungen werden dafür Beta-Verteilungen verwendet, siehe Johnson (1987) S. 117, dort auch die Beschreibung der parameterabhängigen Standardisierung, damit die Kovarianzmatrizen dieselben Vorfaktoren haben.

Die Gleichverteilung auf der Einheitssphäre, also die Vektoren \mathbf{U}_{ik} aus der Modellgleichung (5) wird mit dem Algorithmus von Tashiro (1977) am schnellsten erzeugt. Die $\mathbf{\Gamma}_i$ für die beabsichtigte Kovarianzstruktur stammen wieder aus der Cholesky-Zerlegung von $\mathbf{\Sigma}_i$.

Die Simulationen bestehen aus jeweils 20.000 Wiederholungen.

Die genauen Kombinationen aus n_i und $\mathbf{\Sigma}_i$ werden im Anhang B auf Seite 110 tabelliert, getrennt nach Konstellationen mit identischen und mit unterschiedlichen Kovarianzmatrizen. In der Darstellung der Simulationsergebnisse wird auf eine Zuordnung zwischen individuellem Simulationsergebnis und zugehörigen Kovarianzmatrizen verzichtet, um übersichtlicher zu bleiben und weil die Matrizen sowieso als unbekannt vorausgesetzt werden. Ferner werden keine balancierten Szenarien simuliert, weil dann ATS und Geisser-Greenhouse-Statistik ohnehin übereinstimmen (Bathke, Schabenberger u. a. (2009)).

2.5.2 Simulation der Taylor-Approximation

Zur Erörterung der Taylor-Approximation sollen die relativen Fehler der Erwartungswert- und Varianzapproximation des Quotienten Q_n/Q_e unter der Nullhypothese für die verschiedenen Szenarien gegeneinander aufgeplottet werden. Dies gibt sowohl einen Eindruck von deren Größenordnung wie auch vom Verhältnis zueinander bei verschiedenen Szenarien. Da in der Lösung (28) nur Quotienten aus Erwartungswert und Varianz vorkommen, können sich Approximationsfehler, die für beide Parameter etwa gleich groß sind, gegenseitig aufheben.

Zunächst werden in Abbildung 4 die relativen Fehler für die heteroskedastischen und unbalancierten Szenarien aus dem additiv-linearen Modell getrennt nach Verteilungsfamilie aufgeplottet. Der Übersichtlichkeit halber sind je drei Familien auf eine Hälfte der Graphik aufgeteilt. Zur Orientierung ist die Größe der Signaturen entsprechend den Gesamtstichprobenumfängen gewählt. Außerdem ist die Gerade $y = x$ durch die Plots gelegt.

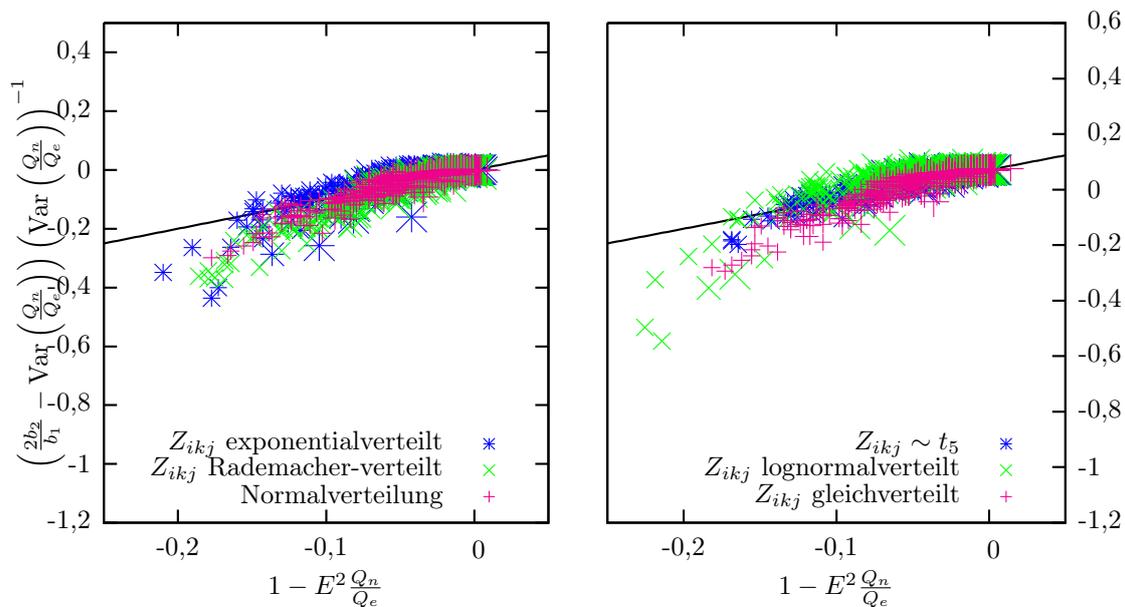


Abbildung 4: Relativer Fehler der Approximation von Erwartungswert und Varianz, add.-lin. Modell

Man erkennt in Abbildung 4, daß die relativen Fehler der Varianzapproximationen im Bereich von 0 bis $-0,4$ liegen, die tatsächlichen Varianzen sind also leicht größer als die Approximation. Ähnliches gilt für die Erwartungswertapproximation, die ebenfalls kaum überschätzt wird. Dort ist der Approximationsfehler im extremsten Fall sogar nur halb so groß. Die relativen Fehler sind grob proportional zueinander. Man kann also erwarten, daß sie sich zum Teil ausgleichen werden. Nur in wenigen Fällen ist der relative Fehler der Varianzapproximation doppelt so groß wie der der Erwartungswertapproximation.

In Abbildung 5 wird dasselbe für die elliptischen Verteilungen gezeigt.

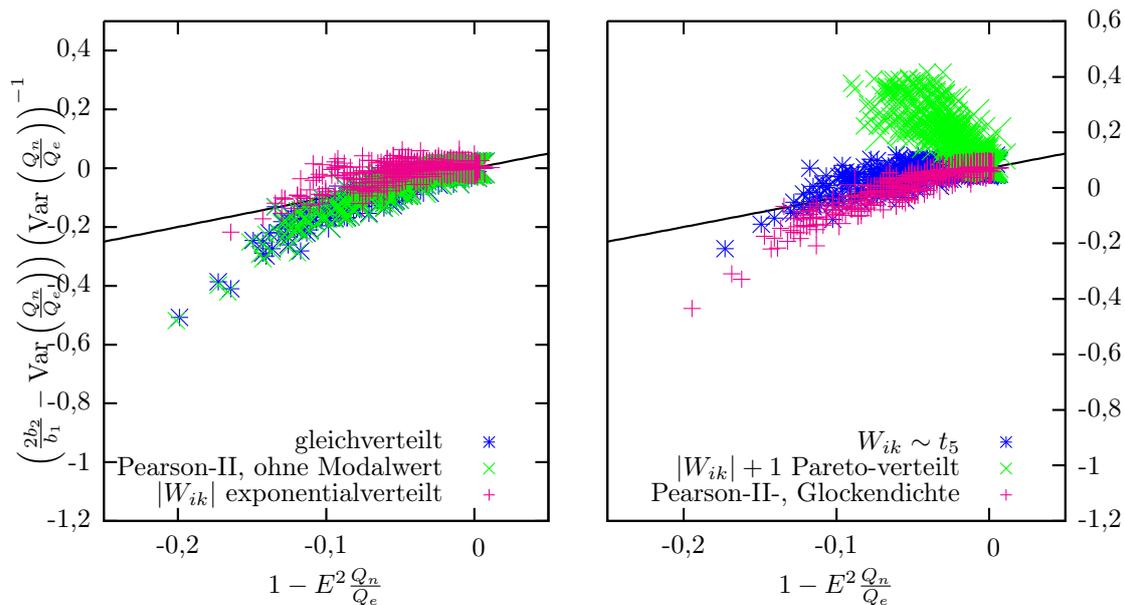


Abbildung 5: Relative Fehler der Approximation von Erwartungswert und Varianz, ellipt. Modell

Abbildung 5 zeigt für die elliptischen Verteilungen etwa dasselbe Bild wie für die additiv-linearen Verteilungen. Einzige Ausnahme ist die Pareto-Verteilung von $|W_{ik}| + 1$ mit der besonders großen Kurtosis. Bei ihr ist die tatsächliche Varianz des Quotienten kleiner als die Approximation voraussagt. In den übrigen Simulationsergebnissen werden die Folgen davon genauer betrachtet.

Insgesamt zeigen sich keine Anhaltspunkte, daß die Taylorapproximationen zu grob wären oder zu stark von den Verteilungen abhängen, um sie zur Grundlage einer robusten Approximation zu machen.

2.5.3 Niveausimulation ohne Schätzer

In den folgenden Abbildungen soll gezeigt werden, wie sich die Geisser-Greenhouse-Statistik im Vergleich zur ANOVA-Typ-Statistik verhält, wenn nichtnormalverteilte Daten vorliegen und, separat davon, wenn die Kovarianzmatrizen normalverteilter Stichproben nicht übereinstimmen. Die in der Realität unbekanntes Freiheitsgradparameter werden dazu aus den simulierten Modellverteilungen berechnet, um die Approximationsansätze unabhängig von den Schätzungen der Parameter zu vergleichen.

Dazu werden zunächst in Abbildung 6 beide Statistiken im Normalverteilungsfall und exemplarisch im Fall lognormalverteilter Z_{ikj} des additiv-linearen Modelles verglichen.

Der Plot ist so aufgebaut, daß auf der Ordinate das simulierte Niveau aufgetragen wird. Die gestrichelten Markierungen geben dabei ein 99%-Zufallsintervall für das

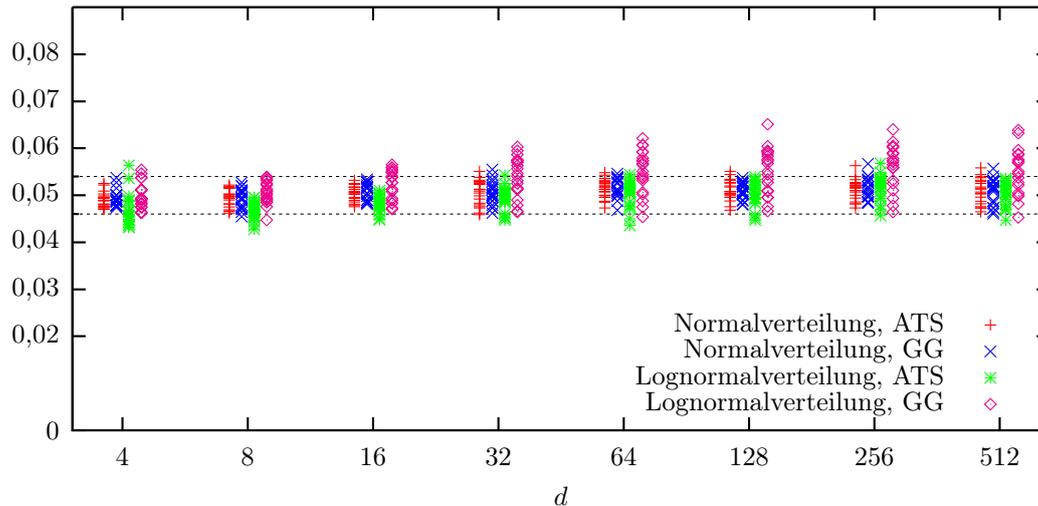


Abbildung 6: Vergleich von ANOVA-Typ-Statistik (ATS) und Geisser-Greenhouse-Statistik (GG) bei übereinstimmenden Kovarianzmatrizen, Normal- und lognormalverteilte Z_{ikj} im additiv-linearen Modell.

nominelle Niveau α mit der Formel

$$\alpha \pm u_{0,995} \sqrt{\frac{\alpha(1-\alpha)}{20000}} \quad (34)$$

an. In der Darstellung der Ergebnisse wird stets als nominelles Niveau $\alpha = 0,05$ verwendet, weil dies in der Anwendung am meisten benutzt wird. Es wurden auch die Niveaus $\alpha = 0,01$ und $\alpha = 0,1$ betrachtet, aber bei ihnen können dieselben Schlußfolgerungen wie bei $\alpha = 0,05$ gezogen werden, bis auf das Detail, daß bekanntermaßen die Box-Approximation und alle Statistiken, die daraus abgeleitet werden, vor allem bei $\alpha > 0,05$ zunehmend schlechter werden.

Auf der Abszisse sind diskret die Dimensionalitäten von 4 bis 512 aufgetragen. Leicht versetzt entlang dieser Achse sind die zugehörigen simulierten Niveaus der Simulationsszenarien als Punkte aufgeplottet.

Man erkennt in Abbildung 6, daß bei Normalverteilung beide Statistiken sehr genau das Niveau einhalten: Bei der gewählten Anzahl von 20000 Wiederholungen kann ein Genauigkeitsunterschied zu einer exakten Statistik nicht aufgedeckt werden. Bei Lognormalverteilung der Z_{ikj} aus dem additiv-linearen Modell sieht man hingegen mit der Geisser-Greenhouse-Statistik eine liberale Tendenz, wo die ANOVA-Typ-Statistik etwa so genau wie bei Normalverteilung bleibt. Bei den übrigen Verteilungen im additiv-linearen Modell ergeben sich Resultate, die zwischen diesen beiden liegen und deswegen nicht gezeigt werden.

Für den Fall zweier exemplarisch ausgewählter elliptischer Verteilungen in Abbildung 7, exponentialverteiltes $|W_{ik}|$ und Pareto-verteiltes $|W_{ik}| + 1$ ist das Ergebnis extremer.

Die ANOVA-Typ-Statistik hält hier ihr Niveau bei der Exponentialverteilung noch

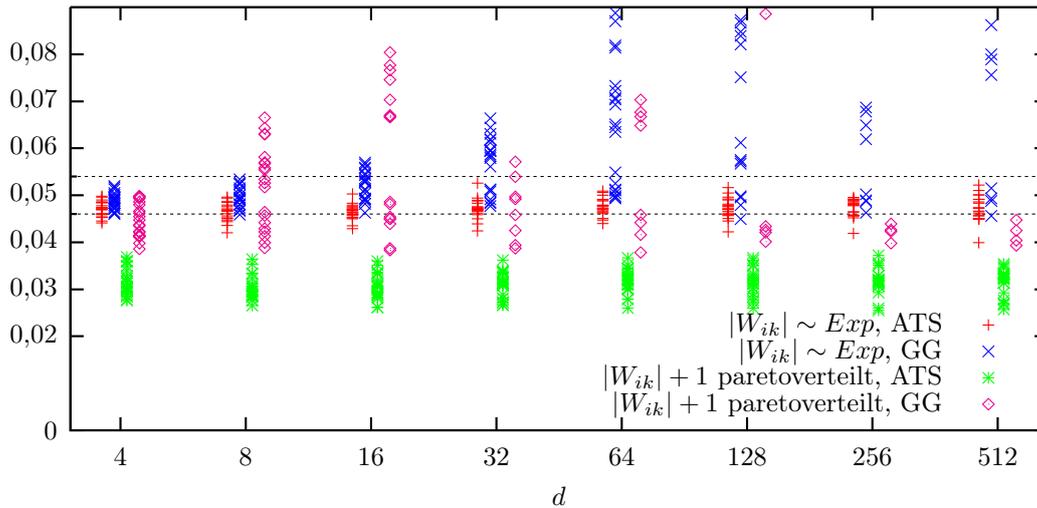


Abbildung 7: Vergleich von ANOVA-Typ-Statistik (ATS) und Geisser-Greenhouse-Statistik (GG) bei übereinstimmenden Kovarianzmatrizen mit zwei elliptischen Verteilungen.

sehr gut ein. Dieses Bild wiederholt sich auch bei den anderen elliptischen Verteilungen, die nicht gezeigt sind. Bei der Pareto-Verteilung wird die ANOVA-Typ-Statistik konservativ. Offensichtlich ist das die Folge davon, daß bei dieser Verteilung $\text{Var}(Q_n/Q_e)$ deutlich kleiner ist als durch die Taylorapproximation erwartet.

Das Niveau der Geisser-Greenhouse-Statistik hängt im Kontrast dazu sehr viel extremer von der Verteilung ab. Ab $d = 16$ liegt das tatsächliche Niveau bei beiden Verteilungen teils deutlich über 7,5%.

In Abbildung 8 wird das Ergebnis für die Szenarien unterschiedlicher Kovarianzmatrizen bei Normalverteilung gezeigt.

Der Effekt unterschiedlicher Kovarianzmatrizen beeinflusst das Niveau der Geisser-Greenhouse-Statistik mehr als der Effekt von Nichtnormalverteilung. Dieses Ergebnis ist bekannt (Oberfeld und Franke 2013). Die Genauigkeit der ANOVA-Typ-Statistik hingegen wird davon nicht beeinträchtigt.

Damit kann die Schlußfolgerung gezogen werden, daß die Geisser-Greenhouse-Statistik weder robust noch geeignet für unterschiedliche Kovarianzmatrizen ist, und offensichtlich hängt beides miteinander zusammen. Die Gefahr, in der Anwendung keine normalverteilten Daten zu haben, ist damit zugleich ein Argument gegen die Praxis, von einem Test auf Gleichheit der (Ko-)Varianzen abhängig zu machen, ob ein Verfahren verwendet werden kann, das auf identische Kovarianzmatrizen ausgelegt ist, denn ein solches Verfahren kann schlechtere Robustheitseigenschaften haben, selbst wenn die Varianzannahmen stimmen.

Die Auslegung der ANOVA-Typ-Statistik auf unterschiedliche Kovarianzmatrizen wirkt sich hingegen auch für die Robustheit positiv aus. Diese Tatsache ist eine Voraussetzung dafür, daß es sich lohnt im nächsten Abschnitt Schätzer für die unbekannteren Freiheitsgrade zu finden.

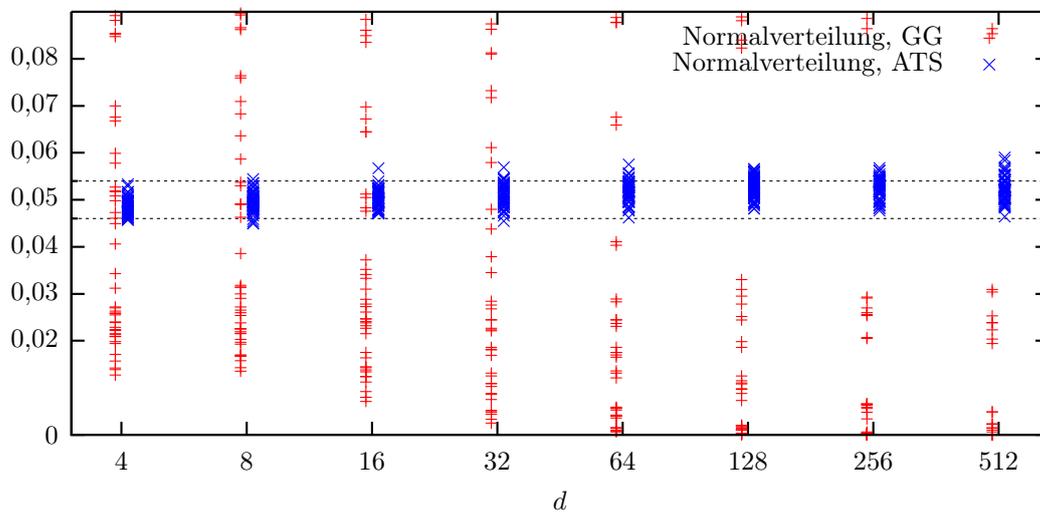


Abbildung 8: Vergleich von ANOVA-Typ-Statistik (ATS) und Geisser-Greenhouse-Statistik (GG) bei Normalverteilung und Szenarien mit unterschiedlichen Kovarianzmatrizen

3 Freiheitsgradschätzungen

3.1 Historischer Abriss bei hoher Dimensionalität

Weil die neuen Statistiken im vorherigen Kapitel genau wie die bekannten Statistiken aus Abschnitt 1.7 auf der Box-Approximation fußen, liegen schon Ansätze vor, Parameter ähnlich zu f_e , f_n zu schätzen. Am umfangreichsten und ältesten ist hierzu die Literatur zur Schätzung des Box'schen ϵ für die Geisser-Greenhouse-Statistik aus Abschnitt 1.7.1.

Das Box'sche ϵ spiegelt wieder, wie sehr sich die Freiheitsgrade im Vergleich zur (exakten) Statistik bei Unabhängigkeit der Meßwiederholungen vermindern. Es ist daher ein Maß für die Verschiedenheit der Eigenwerte von $\mathbf{T}_d \boldsymbol{\Sigma}_1 \mathbf{T}_d$ (siehe (Hoekstra und Chenier 1994)) mit dem Maximum 1 bei identischen Eigenwerten und dem Minimum $1/\text{rg}(\mathbf{T}_d)$ bei einem einzigen positiven Eigenwert.

Die Untergrenze von ϵ bietet sich als sehr einfache, aber konservative Schätzung für ϵ an, die schon von Geisser und Greenhouse (1958; 1959) vorgeschlagen wurde. Die Autoren empfahlen sie, weil für sie noch nicht absehbar war, wie sich der Plug-in-Schätzer durch Einsetzen der empirischen Kovarianzmatrix verhält. Man solle lieber, so der Vorschlag aus (Greenhouse und Geisser 1959), nacheinander

1. die Hypothese nicht ablehnen, falls man mit der Obergrenze $\epsilon = 1$ nicht ablehnen würde, andernfalls
2. die Hypothese ablehnen, falls mit der konservativen Untergrenze $\epsilon = (\text{rg}(\mathbf{T}_d))^{-1}$ abgelehnt würde, andernfalls
3. den Test durchführen mit dem Plug-in-Schätzer

$$\hat{\epsilon} := \frac{\text{Sp}^2 \mathbf{T}_d \hat{\boldsymbol{\Sigma}}}{\text{rg}(\mathbf{T}_d) \text{Sp}(\mathbf{T}_d \hat{\boldsymbol{\Sigma}})^2},$$

also unter Verwendung der (gepoolten) empirischen Kovarianzmatrix.

Weil $\hat{\epsilon}$ denselben Wertebereich wie ϵ hat, kann $\hat{\epsilon}$ nie einen Wert annehmen, durch den eine Entscheidung getroffen würde, die den Ergebnissen der Stufen 1 oder 2 widersprechen würde. Insofern könnte sofort mit dem Plug-in-Schätzer $\hat{\epsilon}$ getestet werden. Die Statistik mit dem Einsetzen von $\hat{\epsilon}$ wird in der Literatur als Geisser-Greenhouse-Test bezeichnet, obgleich genau davor die beiden Autoren warnten.

Die Schätzung durch $\hat{\epsilon}$ hat in Simulationen von Collier u. a. (1967) und Stoloff (1970) bei einer kleinen Anzahl von Meßwiederholungen überwiegend gute Resultate für den Typ-I-Fehler gebracht. Sofern der doppelte Stichprobenumfang allerdings kleiner als das die Zahl der Meßwiederholungen ist –also lange vor hoher Dimensionalität– und ϵ eher bei 1 liegt, zeigt sich eine konservative Tendenz (Huynh und Feldt 1976). Bei

hoher Dimensionalität wird dies noch schlimmer mit entsprechendem Powerverlust. Offensichtlich liegt die Ursache darin, daß $\hat{\epsilon}$ tendenziell zu klein ist: Die kritischen Werte sind zu groß, weil sie aus einer F -Verteilung mit zu kleinen Freiheitsgraden berechnet werden.

Man kann dieses Phänomen im Falle identischer positiver Eigenwerte von $\mathbf{T}_d \boldsymbol{\Sigma}_1 \mathbf{T}_d$ anschaulich erklären, also bei $\epsilon = 1$. Da die positiven Eigenwerte der empirischen Version $\mathbf{T}_d \hat{\boldsymbol{\Sigma}}_1 \mathbf{T}_d$ fast sicher nicht alle identisch sind, muß das Box'sche Epsilon als Maß für die Verschiedenheit der Eigenwerte bei der empirischen Kovarianzmatrix strikt kleiner sein als bei der tatsächlichen. Daraus folgt, daß fast sicher mit größeren kritischen Werten getestet wird, als eigentlich von der Box-Approximation vorgesehen wäre.

Die erste Arbeit, die sich damit befaßte, $\hat{\epsilon}$ für hochdimensionale normalverteilte Daten zu korrigieren, um einen weniger konservativen Schätzer zu erhalten, ist von Huynh und Feldt (1976). Sie berechnen für Nenner und Zähler jeweils unverzerrte Schätzer, indem die Verzerrung im Normalverteilungsmodell ausgerechnet und abgezogen wird. Der Huynh-Feldt-Schätzer² $\tilde{\epsilon}$ wird also als Quotient von unverzerrten Schätzern gewonnen.

Dieser Test hat allerdings zwei Nachteile: Zum einen ist aus Simulationen bekannt, daß er liberal wird, falls $\epsilon \leq 0,75$ ist, siehe Lane (2016), Maxwell und Arvey (1982) und Quintana und Maxwell (1994) und die Originalarbeit Huynh und Feldt (1976). In der Literatur ist daher die Meinung weitgehend einhellig, daß der Huynh-Feldt-Schätzer nur für große $\epsilon > 0,75$ geeignet ist (Huynh (1978), Kirk (1995), Lane (2016), Maxwell und Arvey (1982), Quintana und Maxwell (1994) und Rogan, Kesselman und Mendoza (1979)). Teils gibt man sich damit zufrieden, weil Fälle von kleinem Box'schen ϵ selten seien, teils empfiehlt man die Prozedur von Quintana und Maxwell (1994), in der vorgeschlagen wird, den Huynh-Feldt-Schätzer einzusetzen, falls dieser kleiner als 0,75 ist, und andernfalls den Plug-in-Schätzer zu benutzen. Diese Prozedur wurde erst von Lane (2016) als Computerprogramm implementiert. Es gibt auch die Empfehlung, nur den Plug-in-Schätzer anzuwenden Kirk (1995).

Der andere Nachteil ist ein Erbe der Geisser-Greenhouse-Approximation, nämlich die Beschränkung auf identische Kovarianzmatrizen oder identische Stichprobenumfänge. Dies kann behoben werden, indem die ATS verwendet wird und auf ihre Freiheitsgrade (16) der Ansatz der Huynh-Feldt-Schätzer übertragen wird. Dies geschieht in den Arbeiten von Becker (2010), Harden (2013) sowie Brunner, Bathke und Placzek (2012) und Happ, Harrar und Bathke (2016).

Becker (2010) –für den Zweistichprobenfall– und Harden (2013) –für den Mehrstichprobenfall– verwenden dafür neue erwartungstreue Schätzer auf der Basis von quadratischen und bilinearen Formen für Parameter der Art $\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)$ und $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2$ sowie für die gemischten Terme $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i) \cdot \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_{i'})$ und $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})$ mit

²In der Originalpublikation von Huynh und Feldt (1976) ist bei der Formel für den Schätzer ein Fehler enthalten, der von Lecoutre (1991) korrigiert wurde. Seither wird mit Huynh-Feldt-Schätzer immer die korrigierte Version gemeint.

$i \neq i'$. Aus diesen Teilschätzern setzt man Schätzer für f_n und f_e so zusammen, wie es analog aus den Gleichungen (17) bis (20) hervorgeht.

Näher an der Idee der Huynh-Feldt-Schätzer, die Verzerrung abzuziehen, sind die Schätzer bei Brunner, Bathke und Placzek (2012), zunächst nur für den Zweistichprobenfall. Bei ihnen wird die Verzerrung der Plug-in-Schätzer unter Normalverteilung berechnet, geschätzt und abgezogen. Die Schätzungen für die Parameter f_n und f_e werden anschließend wie zuvor erreicht. In Happ, Harrar und Bathke (2016) wird dieses Konzept für den Mehrstichprobenfall verallgemeinert und gezeigt, daß diese Schätzer verhältniskonsistent sind. Diese Schätzer werden im Folgenden *modifizierte Huynh-Feldt-Schätzer* genannt.

In diesen Arbeiten wurde stets die Normalverteilungsannahme verwendet. Bei Bono u. a. (2016) wurde aber mit Simulationen beobachtet, daß die Verzerrung des Plug-in-Schätzers für das Box'sche Epsilon nicht nur mit der Zahl der Meßwiederholungen wächst, sondern auch von der Verteilung der Zufallsvariablen abhängt. Es ist deswegen naheliegend, daß davon auch die originalen und modifizierten Huynh-Feldt-Schätzer betroffen sind.

Die Nichtrobustheit der originalen Huynh-Feldt-Statistik wurde in Simulationen schon von Oberfeld und Franke (2013) gezeigt. Da zu diesem Ergebnis aber die gezeigte Nichtrobustheit der Geisser-Greenhouse-Statistik beiträgt, ist für die ATS mit den modifizierten Huynh-Feldt-Schätzern damit noch nichts ausgesagt. Auch die Schätzer aus quadratischen und bilinearen Formen sind auf ihr Verhalten bei Nichtnormalverteilung anhand geeigneter Kriterien zu untersuchen und

3.2 Dimensionsstabilität

Statt der Konsistenz aus der klassischen niedrigdimensionalen Statistik und der Verhältniskonsistenz aus der Mehrheit der hochdimensionalen Literatur wurde in der Arbeit von Werner (2002) das Kriterium der *Dimensionsstabilität* eingeführt, um hochdimensional geeignete Schätzer zu charakterisieren. Wenn ein Funktional eines mehrdimensionalen Verteilungsparameters zu schätzen ist –etwa vom Verteilungsparameter Σ_i die Funktionale $\text{Sp}(\mathbf{T}_d \Sigma_i)^2$, $\text{Sp}^2(\mathbf{T}_d \Sigma_i)$ oder das Box'sche Epsilon–, dann soll Dimensionsstabilität aussagen, daß Verzerrung und Varianz des Schätzers nur vom Wert des Funktionals und dem Stichprobenumfang abhängen, nicht von der Dimensionalität des Verteilungsparameters.

Definition 12. Ein Feld von Schätzern $\hat{\theta}_{n,d}$ für einen Parameter θ_d heißt *dimensionsstabil*, wenn eine Nullfolge z_n und eine Folge v_n existiert, so dass für alle d gilt

1.
$$\left| E \left(\frac{\hat{\theta}_{n,d}}{\theta_d} - 1 \right) \right| \leq z_n$$

2.

$$\theta_d^{-2} \text{Var} \left(\hat{\theta}_{n,d} \right) \leq v_n.$$

Bei dimensionsstabilen Schätzern gibt es gleichmäßig für alle d eine obere Schranke für den mittleren quadratischen Fehler. Hat man erwartungstreue Schätzer, dann ist dieser Fehler gleichmäßig durch v_n beschränkt. Ist v_n eine Nullfolge, dann konvergiert der Schätzer $\hat{\theta}_{n,d}$ gleichmäßig gegen θ_d im Sinne der L_2 -Norm und ist insbesondere damit konsistent.

Dies ist deutlich stärker als die Verhältniskonsistenz, die in den asymptotischen Statistiken aus Abschnitt 1.8.3 vorkommt. Erstens ist L_2 -Konvergenz stärker als Konvergenz in Wahrscheinlichkeit. Zweitens würde in der Asymptotik $N, d \rightarrow \infty$ mit $\frac{d}{N} \rightarrow c > 0$ bei dimensionsstabiler Konsistenz sogar $c = 0$ zulässig sein, und drittens enthält dieses Kriterium keine Annahme über die asymptotische Gestalt der Kovarianzmatrix, weder asymptotisch noch finit.

Bei Werner (2002) wurden für den Einstichprobenfall der ATS, der in Abschnitt 1.7.3 beschrieben ist, unter Normalverteilung dimensionsstabile, konsistente und unverzerrte Schätzer für $\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_1)$ und $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_1)^2$ gefunden. Sie setzen voraus, daß

$$E(\mathbf{T}_d \mathbf{Y}_{1k}) = \mathbf{0} \text{ für } k = 1, \dots, n \quad (35)$$

gilt, was unter der Nullhypothese im Einstichprobenfall zutrifft. Es sind im Einzelnen

$$\frac{1}{n(n-1)} \sum_{k \neq l} \mathbf{Y}'_k \mathbf{T}_d \mathbf{Y}_k \mathbf{Y}'_l \mathbf{T}_d \mathbf{Y}_l$$

der Schätzer für $\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_1)$ und

$$\frac{1}{n(n-1)} \sum_{k \neq l} (\mathbf{Y}'_k \mathbf{T}_d \mathbf{Y}_l)^2$$

der Schätzer für $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_1)^2$.

Bei Werner (2002) wird die Erwartungstreue mit

$$E(\mathbf{Y}'_k \mathbf{T}_d \mathbf{Y}_k \mathbf{Y}'_l \mathbf{T}_d \mathbf{Y}_l) = \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_1)$$

und

$$E\left((\mathbf{Y}'_k \mathbf{T}_d \mathbf{Y}_l)^2\right) = \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_1)^2$$

gezeigt. Diese Schätzer sind damit sogenannte *U-Statistiken*, siehe Serfling (1980), S. 172. *U-Statistiken* beruhen darauf, daß eine Funktion einer Anzahl von unterschiedlichen Beobachtungen –der *Kern der U-Statistik*, hier $\mathbf{Y}'_k \mathbf{T}_d \mathbf{Y}_k \mathbf{Y}'_l \mathbf{T}_d \mathbf{Y}_l$ und $(\mathbf{Y}'_k \mathbf{T}_d \mathbf{Y}_l)^2$ – erwartungstreu für den unbekannt Parameter ist. Die *U-Statistik* mittelt über all die Funktionswerte, die man erhält, wenn man die unabhängig identisch verteilten Beobachtungen aus derselben Stichprobe durchtauscht. Dadurch sind *U-Statistiken* stets konsistent.

Von Helms (2010) wird gezeigt, daß diese Schätzer auch in einem additiv-linearen Modell ähnlich zu (3) dimensionsstabil, konsistent und erwartungstreu sind. Im Mehrstichprobenfall gilt eine Bedingung wie (35) nicht mehr, weil auch unter der Nullhypothese $E(\mathbf{T}_d \mathbf{Y}_{ik}) = \mathbf{0}$ für $i = 1, \dots, a$ und $k = 1, \dots, n_i$ nicht mehr vorausgesetzt werden kann. Daher müssen modifizierte Schätzer verwendet werden.

Eine solche Modifikation ist es bei Becker (2010) und Harden (2013), die Zentrierung durch eine Differenz aus zwei Beobachtungsvektoren derselben Stichprobe zu erreichen, und daraus analoge U -Statistiken zu konstruieren.

Offen ist bei diesen Schätzern, ob Dimensionsstabilität, Konsistenz und Erwartungstreue auch bei Nichtnormalverteilung gelten. Für das additiv-lineare Modell ist dies wegen der Ähnlichkeit zu den Einstichprobenschätzern aus Helms (2010) plausibel, aber nicht bewiesen. Für das elliptische Modell ist noch nichts dergleichen ausgearbeitet.

Dasselbe ist auch für die modifizierten Huynh-Feldt-Schätzer zu zeigen. Ebenso ist zu untersuchen, welche Schätzer bei Normalverteilung minimale Varianz haben, da es zum Sinn robuster Verfahren gehört, zunächst im „normalen“ Modell optimal oder so gut wie optimal zu sein (Huber (1981), S. 5).

3.3 Dimensionsstabile Schätzer aus U -Statistiken

Schätzer nach Becker (2010) und Harden (2013) beruhen auf den quadratischen Formen

$$A_{i(k,l)} := (\mathbf{Y}_{ik} - \mathbf{Y}_{il})' \mathbf{T}_d (\mathbf{Y}_{ik} - \mathbf{Y}_{il}) \quad (36)$$

und den Bilinearformen

$$A_{i(k,l)l'(s,t)} := (\mathbf{Y}_{ik} - \mathbf{Y}_{il})' \mathbf{T}_d (\mathbf{Y}_{i's} - \mathbf{Y}_{i't}), \quad (37)$$

indem

$$\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i) \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_{i'}) \text{ durch } B_1^{(i,i')} := (4n_i(n_i - 1)n_{i'}(n_{i'} - 1))^{-1} \sum_{\neq k,l}^{n_i} \sum_{\neq s,t}^{n_{i'}} A_{i(k,l)} A_{i'(s,t)} \quad (38)$$

$$\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'}) \text{ durch } B_2^{(i,i')} := (4n_i(n_i - 1)n_{i'}(n_{i'} - 1))^{-1} \sum_{\neq k,l}^{n_i} \sum_{\neq s,t}^{n_{i'}} A_{i(k,l)l'(s,t)}^2, \quad (39)$$

$$\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i) \text{ durch } B_1^{(i)} := (4)^{-1} \sum_{\neq k,l,s,t}^{n_i} A_{i(k,l)} A_{i(s,t)} \text{ und} \quad (40)$$

$$\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 \text{ durch } B_2^{(i)} := \left(4 \frac{n_i!}{(n_i - 4)!}\right)^{-1} \sum_{\neq k,l,s,t}^{n_i} A_{i(k,l)l'(s,t)}^2 \quad (41)$$

geschätzt wird. Für $i \neq i'$ sind dies genau die Plug-in-Schätzer, siehe Becker (2010), S. 65 und S. 67.

Satz 13. *Sind k, l, s und t untereinander verschieden und $k' \neq l'$, dann gilt:*

$$E(A_{i(kl)}A_{i(st)}) = 4 \operatorname{Sp}^2 \mathbf{T}_d \boldsymbol{\Sigma}_i$$

$$E(A_{i(kl)}A_{i'(k'l')}) = 4 \operatorname{Sp} \mathbf{T}_d \boldsymbol{\Sigma}_1 \cdot \operatorname{Sp} \mathbf{T}_d \boldsymbol{\Sigma}_2$$

$$E(A_{i(kl)i(st)})^2 = 4 \operatorname{Sp} \mathbf{T}_d \boldsymbol{\Sigma}_i^2$$

$$E(A_{i(kl)i'(k'l')})^2 = 4 \operatorname{Sp} \mathbf{T}_d \boldsymbol{\Sigma}_1 \mathbf{T}_d \boldsymbol{\Sigma}_2$$

Beweis. Es wird nur benötigt, daß die Kovarianzmatrix $\boldsymbol{\Sigma}_i$ überhaupt existiert. Insofern ist der Beweis nicht neu gegenüber den Arbeiten Becker (2010) und Harden (2013), wo zwar die Normalverteilung vorausgesetzt, aber nicht gebraucht wird. Der Vollständigkeit halber ist der Beweis trotzdem im Anhang zu finden, und zwar ab Seite 86 im 5. und 6. Teil von Lemma 33. \square

Wegen der Unverzerrtheit sind auch diese Schätzer U -Statistiken. Damit steht fest, daß $B_1^{(i)}$ minimale Varianz hat unter den Schätzern, die nur aus Termen der Art $A_{i(k,l)i(s,t)}$ bestehen, siehe Lehmann (2004). Entsprechendes gilt für die anderen drei Schätzer.

Insgesamt ergeben sich die Schätzer

$$B_1 = \sum_{i=1}^a \frac{t_{ii}^2}{n_i^2} B_1^{(i)} + \sum_{i \neq i'}^a \frac{t_{ii'}}{n_i n_{i'}} B_1^{(i,i')} \quad \text{für } b_1 \quad (42)$$

$$B_n = \sum_{i=1}^a \frac{t_{ii}^2}{n_i^2} B_2^{(i)} + \sum_{i \neq i'}^a \frac{t_{ii} t_{i'i'}}{n_i n_{i'}} B_2^{(i,i')} \quad \text{für } b_n \quad (43)$$

$$B_e = \sum_{i=1}^a \frac{t_{ii}^2}{n_i^2 (n_i - 1)} B_2^{(i)} \quad \text{für } b_e.$$

Diese Schätzer sind bei stetigen Verteilungen fast sicher positiv, weil fast sicher nicht alle Vektoren der Form $\mathbf{T}_d(\mathbf{Y}_{ik} - \mathbf{Y}_{il})$ orthogonal zueinander sind oder verschwinden. Lediglich bei diskreten Verteilungen kann eine geringe Wahrscheinlichkeit vorhanden sein, daß B_n und B_e verschwinden und damit die Freiheitsgradschätzer nicht existieren. Dies geschieht nur dann, wenn alle Differenzvektoren der Art $\mathbf{T}_d(\mathbf{Y}_{ik} - \mathbf{Y}_{il})$ aus allen Stichproben orthogonal zueinander sind.

Zu zeigen ist die Dimensionsstabilität dieser Schätzer im Sinne von Definition 12 im additiv-linearen Modell von Definition 2 und im elliptischen Modell von Definition 4. Weil $B_1^{(i)}$ und $B_2^{(2)}$ als U -Statistiken betrachtet werden können, genügt es hierbei, zu zeigen, daß für die Summanden in der Darstellung (40) bei $\neq k, l, s, t$ gilt

$$\operatorname{Var}(A_{i(k,l)}A_{i(s,t)}) \leq c \operatorname{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i),$$

mit einer Konstanten c , die zwar vom Modell abhängt, nicht aber von d . Für die übrigen Summanden in den Gleichungen (38) bis (41) ist entsprechendes zu zeigen.

Bei Becker (2010) und Harden (2013) wurde dies im normalverteilten Modell durch Sätze über die Momente quadratischer bzw. bilinearer Formen in normalverteilten Zufallsvariablen gezeigt. Diese Sätze stehen in verallgemeinerten Modellen nicht mehr zur Verfügung. Stattdessen ist folgender Satz neu:

Satz 14. *Die Schätzer $B_1^{(i)}$, $B_2^{(i)}$, $B_1^{(i,i')}$ und $B_2^{(i,i')}$ aus den Gleichungen (40) bis (38) sind im additiv-linearen Modell und im elliptischen Modell dimensionsstabil und konsistent.*

Beweis. Der Beweis ist im Anhang ab Seite 95 zu finden und benötigt einige Lemmata, die davor im Anhang ab 85 stehen. \square

Ausgehend von diesem Satz können geeignete Schätzer für f_n und f_e gefunden werden durch

$$\hat{f}_n^* := \frac{B_1}{B_n}$$

bzw.

$$\hat{f}_e^* := \frac{B_1}{B_e}.$$

Da Quotientenbildung keine lineare Abbildung ist, können diese Schätzer im Allgemeinen nicht exakt erwartungstreu sein, sondern nur approximativ in der Taylorapproximation erster Ordnung für den Erwartungswert

$$E\left(\hat{f}_n^*\right) \doteq \frac{E(B_1)}{E(B_n)} \quad \text{und} \quad E\left(\hat{f}_e^*\right) \doteq \frac{E(B_1)}{E(B_e)},$$

die analog schon in Gleichung (29) verwendet wurde. Erwartungstreue der Quotienten ist allerdings erstens nicht so wichtig, da die Schätzer später in die Quantilfunktion der F -Verteilung eingesetzt werden, die wiederum nicht linear ist.

Entscheidender ist folgende Aussage:

Satz 15. *\hat{f}_n^* und \hat{f}_e^* sind konsistent und näherungsweise dimensionsstabil in dem Sinne, daß die Varianz einer Taylorapproximation ersten Grades der Quotienten*

$$\hat{f}_n^* = \frac{B_1}{B_n}$$

bzw.

$$\hat{f}_e^* = \frac{B_1}{B_e}$$

nicht von d abhängt.

Beweis. Siehe Anhang S. 97 \square

Damit steht fest, daß die Freiheitsgrade auch bei hochdimensionalen Daten, die nicht multivariat normalverteilt sind, genauso sinnvoll geschätzt werden können wie bei Normalverteilung.

Diese Freiheitsgradschätzer sind allerdings nicht bereichserhaltend. Sie können ihren Parameterbereich, der in (21) und (22) angegeben ist, überschreiten. Bei Überschreitung sollen sie einfach abgeschnitten werden. Es entstehen die endgültigen Schätzer

$$\hat{f}_n := \max \left\{ 1, \min \left\{ \hat{f}_n^*, \text{rg} \mathbf{T} \right\} \right\} \quad (44)$$

und

$$\hat{f}_e := \max \left\{ \min_{i \in \{1, \dots, a\}} n_i - 1, \min \left\{ \hat{f}_n^*, \text{rg}(\mathbf{T}_d) \cdot (N - a) \right\} \right\}. \quad (45)$$

3.4 Huynh-Feldt- und Plug-in-Schätzer

Für die modifizierten Huynh-Feldt-Schätzer nach Brunner, Bathke und Placzek (2012) und Happ, Harrar und Bathke (2016), die auf unbalancierte Layouts bei nicht notwendigerweise identischen Kovarianzmatrizen ausgelegt sind, wird von den Plug-in-Schätzern ihre Verzerrung unter Normalverteilung abgezogen. Statt $B_1^{(i)}$ und $B_2^{(i)}$ ergibt sich

$$\begin{aligned} H_1^{(i)} &= \frac{n_i(n_i - 1)}{(n_i - 2)(n_i + 1)} \left(\text{Sp}^2(\mathbf{T}_d \hat{\Sigma}_i) - \frac{2}{n_i} \text{Sp}(\mathbf{T}_d \hat{\Sigma}_i)^2 \right) \\ H_2^{(i)} &= \frac{(n_i - 1)^2}{(n_i - 2)(n_i + 1)} \left(\text{Sp}(\mathbf{T}_d \hat{\Sigma}_i)^2 - \frac{1}{n_i - 1} \text{Sp}^2(\mathbf{T}_d \hat{\Sigma}_i) \right), \end{aligned}$$

Die Schätzer für die gemischten Terme $\text{Sp}(\mathbf{T}_d \Sigma_i) \text{Sp}(\mathbf{T}_d \Sigma_{i'})$ und $\text{Sp}(\mathbf{T}_d \Sigma_i \mathbf{T}_d \Sigma_{i'})$ sind dieselben wie zuvor. Schätzer für Zähler und Nenner der Freiheitsgrade sind also

$$\begin{aligned} H_1 &= \sum_{i=1}^a \frac{t_{ii}^2}{n_i^2} H_1^{(i)} + \sum_{i \neq i'}^a \frac{t_{ii'}^2}{n_i n_{i'}} B_1^{(i,i')} && \text{für } b_1, \\ H_n &= \sum_{i=1}^a \frac{t_{ii}^2}{n_i^2} H_2^{(i)} + \sum_{i \neq i'}^a \frac{t_{ii'} t_{i'i}}{n_i n_{i'}} B_2^{(i,i')} && \text{für } b_n, \\ \text{und} \quad H_e &= \sum_{i=1}^a \frac{t_{ii}^2}{n_i^2 (n_i - 1)} H_2^{(i)} && \text{für } b_e. \end{aligned}$$

Die Freiheitsgradschätzer werden daraus analog durch

$$\tilde{f}_n^* = \frac{H_1}{H_n}$$

bzw.

$$\tilde{f}_e^* = \frac{H_1}{H_e}$$

gewonnen.

$H_1^{(i)}$ und $H_2^{(i)}$ sind bei Normalverteilung per Konstruktion zwar unverzerrt, aber trotzdem keine U -Statistiken. Es kommen Terme der Art $(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})^2$ vor, wenn man $H_1^{(i)}$ und $H_2^{(i)}$ in Summen über vier Indices ausmultipliziert. Dies verkompliziert, die Varianz dieser Schätzer auszurechnen.

Neu kann dank der Schätzer $B_1^{(i)}$ und $B_2^{(i)}$ aus (40) und (41) auf Seite 47 mit dem Satz von Lehmann-Scheffé gezeigt werden, daß $H_1^{(i)}$ und $H_2^{(i)}$ bei Normalverteilung ebenfalls dimensionsstabil sind und dann sogar besser sind als $B_1^{(i)}$ und $B_2^{(i)}$.

Satz 16. *Im normalverteilten Modell haben die modifizierten Huynh-Feldt-Spurschätzer kleinste Varianz unter allen erwartungstreuen Schätzern. Insbesondere sind die modifizierten Huynh-Feldt-Spurschätzer bei Normalverteilung ebenfalls dimensionsstabil und die Freiheitsgradschätzer näherungsweise dimensionsstabil in dem Sinne, daß eine Taylorapproximation ersten Grades ihrer Varianz nicht von d abhängt.*

Beweis. Siehe im Anhang auf Seite 97. □

Diese Aussage ist angewiesen auf die Normalverteilungsannahme. Gilt hingegen im additiv-linearen Modell (3)-(4) für ein i, j die Bedingung $\delta_{ij} \neq 3$, dann sind die Schätzer nicht mal mehr erwartungstreu.

Satz 17. *(Verzerrung der Huynh-Feldt-Schätzer bei Nichtnormalverteilung) Die Erwartungswerte der Plug-in-Spurschätzer von $\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)$ sowie $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2$ sind im additiv-linearen Modell (3) aus Definition 2 und im elliptischen Modell (5) aus Definition 4*

$$E\left(\text{Sp}^2\left(\mathbf{T}_d \hat{\boldsymbol{\Sigma}}_i\right)\right) = \text{Sp}^2\left(\mathbf{T}_d \boldsymbol{\Sigma}_i\right) + \frac{2}{n_i - 1} \text{Sp}\left(\mathbf{T}_d \boldsymbol{\Sigma}_i\right)^2 + \frac{\zeta_i}{n_i - 1} \quad (46)$$

$$E\left(\text{Sp}\left(\mathbf{T}_d \hat{\boldsymbol{\Sigma}}_i\right)^2\right) = \frac{n_i}{n_i - 1} \text{Sp}\left(\mathbf{T}_d \boldsymbol{\Sigma}_i\right)^2 + \frac{1}{n_i - 1} \text{Sp}^2\left(\mathbf{T}_d \boldsymbol{\Sigma}_i\right) + \frac{\zeta_i}{n_i}, \quad (47)$$

wobei ζ_i je nach Modell bedeutet:

1. Im additiv-linearen Modell (3) bedeutet ζ_i mit der Abkürzung $\mathbf{A} = [a_{jj'}]_{j, j'=1, \dots, r_i} = \boldsymbol{\Gamma}'_i \mathbf{T}_d \boldsymbol{\Gamma}_i$

$$\zeta_i := \sum_{j=1}^{r_i} (\delta_{ij} - 3) a_{jj}^2. \quad (48)$$

2. Im elliptischen Modell (5) und mit der Abkürzung $\text{rg}(\boldsymbol{\Gamma}'_i \mathbf{T}_d \boldsymbol{\Gamma}_i) = \tilde{r}_i$ bedeutet ζ_i

$$\zeta_i := \frac{\delta_i - 3}{\tilde{r}_i (\tilde{r}_i + 2)} (2\text{Sp}\left(\mathbf{T}_d \boldsymbol{\Sigma}_i\right)^2 + \text{Sp}^2\left(\mathbf{T}_d \boldsymbol{\Sigma}_i\right)) \quad (49)$$

Beweis. Siehe im Anhang A.3.5 auf Seite 97. □

Daraus ergibt sich, daß die Schätzer $H_1^{(i)}$ und $H_2^{(i)}$ die Verzerrungen

$$EH_1^{(i)} - \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i) = \frac{(n_i^2 - 2n_i + 2)}{n_i(n_i + 1)(n_i - 2)} \zeta_i \quad (50)$$

$$EH_2^{(i)} - \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 = \frac{(n_i^2 - n_i + 1)}{n_i(n_i + 1)(n_i - 2)} \zeta_i \quad (51)$$

haben. Die Verzerrungen kommen daher, daß in $E(H_1^{(i)})$ und $E(H_2^{(i)})$ Terme der Art $E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})^2$ vorkommen, siehe den Beweis zu Satz 17. Die vierten Momente kommen als Störparameter nur wegen dieser Terme in der Verzerrung vor. Dadurch werden die Spuren bei Verteilungen mit negativer Exzess-Kurtosis unterschätzt ($\zeta_i < 0$), bei positiver Exzess-Kurtosis überschätzt. Da diese Terme in $B_1^{(i)}$ und $B_2^{(i)}$ nicht auftreten, sind die U -Statistiken robuste Schätzer im Gegensatz zu den modifizierten Huynh-Feldt-Schätzern.

Immerhin verschwinden für $n_i \rightarrow \infty$ die relativen Verzerrungen gleichmäßig über d .

Satz 18. *Im additiv-linearen und im elliptischen Modell gibt es eine Nullfolge c_{n_i} , die nicht von d abhängt, so daß gilt*

$$\frac{EH_1^{(i)} - \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)}{\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)} \leq c_n.$$

Analoges gilt für $H_2^{(i)}$.

Beweis. Siehe im Anhang auf Seite 99. □

Bleibt noch die Varianz der modifizierten Huynh-Feldt-Schätzer bei Nichtnormalverteilung zu diskutieren. Hier kann aus dem Vorhergehenden geschlossen werden, daß deren Varianz von den achten Momenten abhängen müssen, weil die Varianz von Termen der Art $(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})^2$ unweigerlich in der Zerlegung von $\text{Var}(H_1^{(i)})$ und $\text{Var}(H_2^{(i)})$ als Summe vorkommen muß. Exakte Varianzberechnungen der Schätzer sind ohne zusätzliche Annahmen über höhere Momente nicht möglich. Deswegen beschränkt sich der Vergleich der Varianzen auf die simulierten Verteilungen im Zusammenhang mit dem Monte-Carlo-Simulationen zum Niveau der ATS mit den eingesetzten Schätzern. Hierbei werden analog zu (44) und (45) die Schätzer \tilde{f}_n^* und \tilde{f}_e^* abgeschnitten an ihren Bereichsgrenzen.

3.5 Simulationen mit den Schätzern

Die ATS mit den eingesetzten hochdimensionalen Schätzern und deren Erwartungswert und Varianz wurden in denselben Layouts und mit denselben Verteilungen durchgeführt wie in Abschnitt 2.5 auf Seite 33 beschrieben. Zunächst wird $\text{Var}(B_e)$

mit $\text{Var}(H_e)$ verglichen, indem in Abbildung 9 mittels eines Scatterplots $\text{Var}(H_e) b_e^{-2}$ gegen $\text{Var}(B_e) b_e^{-2}$ aus den simulierten Szenarien getrennt nach Verteilung aufgeplottet werden. Dafür sind vier Verteilungsmodelle exemplarisch ausgewählt: Die multivariate Normalverteilung, das additiv-lineare Modell mit exponentialverteilten Z_{ikj} , die elliptische Gleichverteilung und die elliptische Verteilung mit $\sqrt{5/12d}W_{ik}^2 \sim t_5$.

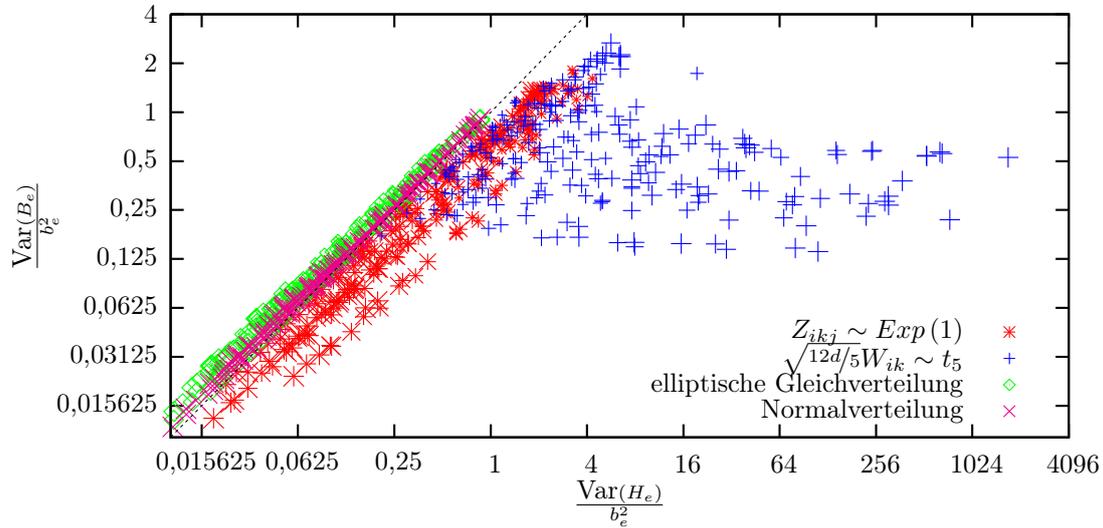


Abbildung 9: Quadrierte Variationskoeffizienten von B_e und H_e gegeneinander aufgeplottet bei logarithmischer Skalierung.

Man erkennt auf Abbildung 9, daß wie bewiesen $\text{Var}_{\mathcal{N}}(B_e) \geq \text{Var}_{\mathcal{N}}(H_e)$ ist, der Unterschied aber sehr gering ist. Nur bei den abgebildeten Ergebnissen unter elliptischer Gleichverteilung sind die Unterschiede stärker. Andersherum ist in den Simulationen $\text{Var}(B_e) \leq \text{Var}(H_e)$ beim additiv-linearen Modell mit unabhängig exponentialverteilten Z_{ikj} . Wesentlich größer ist der Unterschied bei der elliptischen Verteilung mit $\sqrt{12d/5}W_{ik} \sim t_5$.

Mit einer Ausnahme liegen die Ergebnisse mit den anderen simulierten Verteilungen im Rahmen dessen, was in Abbildung 9 gezeigt wird. Die Ausnahme ist die elliptische Verteilung mit Pareto-verteiltem $|W_{ik}| + 1$, die gewählt wurde, um eine elliptische Verteilung mit besonders großen vierten Momenten zu haben. Bei ihr ist $\text{Var}(H_e) - \text{Var}(B_e)$ noch größer.

Ähnliche Ergebnisse ergeben sich für den Vergleich von $\text{Var}(B_n)$ mit $\text{Var}(H_n)$. Sie werden deswegen nicht abgebildet. Ebenfalls nicht abgebildet wird der Vergleich von $\text{Var}(B_1)$ mit $\text{Var}(H_1)$, weil dort für alle Verteilungen nur geringe Unterschiede zwischen beiden Varianzen zu erkennen sind.

Man kann an dieser Stelle schließen, daß die dimensionsstabilen Schätzer aus U -Statistiken bei Normalverteilung zwar nicht exakt optimale Schätzer sind, den Optimalen aber sehr nahe kommen. Umgekehrt sind die modifizierte Huynh-Feldt-Schätzer zwar optimal unter Normalverteilung, bei Verteilungen mit schwereren Tails aber wesentlich schlechter als die U -Statistiken. Letztere können daher als robuste Schätzer bezeichnet werden.

Daß letztere daher den besten Kompromiß bieten, wird in Abbildung 11 ersichtlich. Dort ist das simulierte Niveau der ATS bei normalverteilten Daten aufgeplottet, wobei die modifizierten Huynh-Feldt-Schätzer, die robusten Schätzer und zum Vergleich die wahren Parameter eingesetzt wurden.

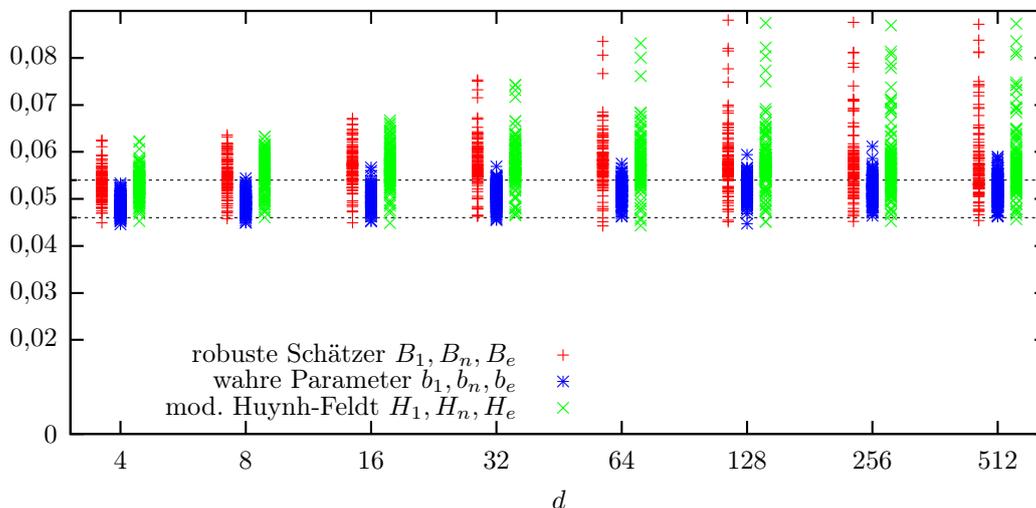


Abbildung 10: Niveau der ATS unter Normalverteilung mit den robusten Schätzern, den modifizierten Huynh-Feldt-Schätzern und zum Vergleich mit den wahren Parametern.

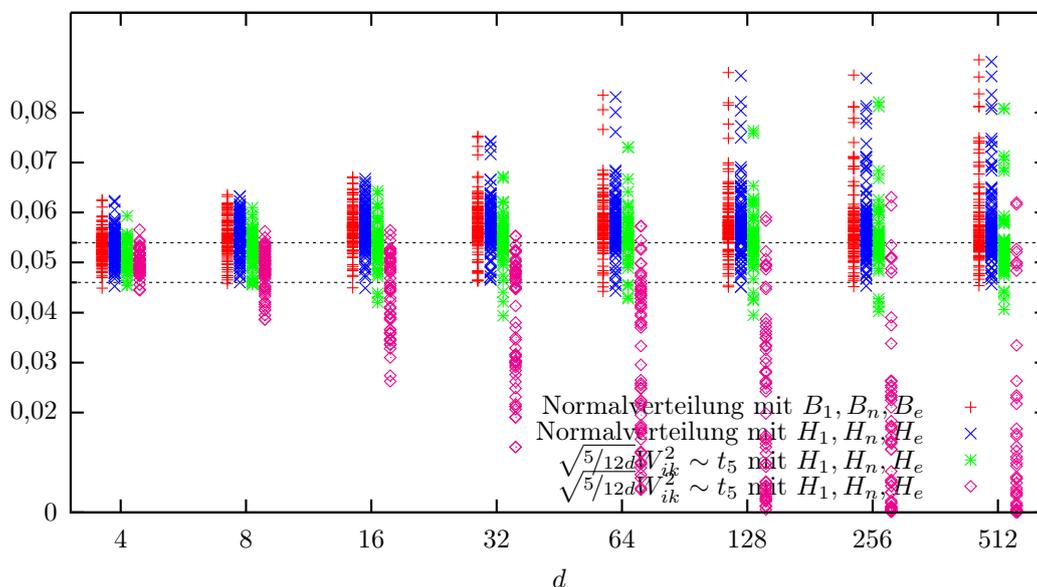


Abbildung 11: ANOVA-Typ-Statistik mit jeweils den robusten Schätzern B_1, B_n, B_e und den modifizierten Huynh-Feldt-Schätzern H_1, H_n, H_e unter Normalverteilung und elliptischer Verteilung mit $5/12d^{-1}W_{ik}^2 \sim t_5$.

In Abbildung 10 ist kaum ein Unterschied auf das Niveau feststellbar, der durch die unterschiedlichen Schätzer kommt. Offensichtlich hat man tatsächlich keinen Genauigkeitsverlust durch die robusten Schätzer gegenüber den bei Normalverteilung

optimalen modifizierten Huynh-Feldt-Schätzern. Das Niveau wird mit eingesetzten Schätzern häufiger überschritten als wenn die wahren Parameter eingesetzt werden. Dies ist nicht verwunderlich und in einem gewissen Maße erträglich, da es sich um approximative Verfahren handelt. Allerdings gibt es in einigen Szenarien ab $d = 32$ gravierende Niveauüberschreitungen.

Dies erinnert an die bekannte Eigenschaft der Huynh-Feldt-Tests, bei sehr kleinem Box'schen Epsilon liberal zu werden. Offensichtlich tritt ein ähnlicher Effekt auch bei der ATS durch das Einsetzen unverzerrter Spurschätzer auf, wurde aber bislang in seinem Ausmaß übersehen. Dies liegt auch daran, daß die Niveauüberschreitung bei festem d nicht in jeder Konstellation aus Kovarianzmatrizen und Stichprobenumfängen gleich ist. Im nächsten Kapitel soll deshalb die Ursache für die Niveauüberschreitungen beschrieben werden und daraus eine Korrektur ausgearbeitet werden. Diese ist auf die robusten Schätzer ausgelegt, weil deren Erwartungstreue benötigt wird. Durch die Verzerrung der modifizierte Huynh-Feldt-Schätzer in nichtnormalverteilten Modellen ist es kaum möglich eine Korrektur für diese Schätzer zu finden.

4 Korrektur für das Einsetzen eines Schätzers

4.1 Verallgemeinerung des Box'schen Epsilons

Da beim Huynh-Feldt-Test Niveauüberschreitungen mit kleinem Box'schen Epsilon $\epsilon < 0,75$ in Zusammenhang gebracht wurden, soll an dieser Stelle die Definition eines Parameters motiviert werden, der als eine Übertragung des Box'schen Epsilons auf das Behrens-Fisher-Problem betrachtet werden kann.

Das Box'sche Epsilon ist in Gleichung (11) als Maßzahl für die Nichtsphärizität, d. h. die Unterschiedlichkeit der d Eigenwerte der gemeinsamen Kovarianzmatrix beschrieben worden, siehe Hoekstra und Chenier (1994). Läßt man unterschiedliche Kovarianzmatrizen zu, dann können ad Eigenwerte unterschiedlich sein.

Dadurch entsteht eine zusätzliche Art von Nichtsphärizität allein wegen Unterschiede zwischen den a Gruppen. Selbst wenn die verschiedenen Kovarianzmatrizen jeweils perfekt sphärisch sind, so daß

$$\mathbf{T}'_d \boldsymbol{\Sigma}_i \mathbf{T}_d = s_i (\mathbf{I}_{\text{rg}(\mathbf{T}_d)} \oplus \mathbf{0}_{d-\text{rg}(\mathbf{T}_d)})$$

ist, können die Faktoren s_1, \dots, s_a unterschiedlich sein. Bei der Wechselwirkungshypothese $H_0(AB)$ und bei $H_0(A|B)$ kann dies zur Nichtsphärizität von $\text{Cov}(\mathbf{T}\bar{\mathbf{Y}})$ führen, bei der Hypothese $H_0(B)$ hingegen nicht.

Eine geeignete Verallgemeinerung des Box'schen Epsilon für den ersten Freiheitsgrad der ATS muß deswegen bei $H_0(B)$ denselben Parameterwert bringen wie das gewöhnliche Box'sche Epsilon, bei den anderen beiden Hypothesen aber einen anderen Wert. Mit f_n , dem ersten Freiheitsgrad der ATS wie in (20) ist eine solche Verallgemeinerung

$$\epsilon = \frac{f_n}{\text{rg}(\mathbf{T})}. \quad (52)$$

Sie stimmt mit dem Box'schen Epsilon aus (11) überein, wenn der Sonderfall einer Stichprobe betrachtet wird.

Für den zweiten Freiheitsgrad der ATS kann ein ähnlicher Parameter gefunden werden, allerdings ist er im Allgemeinen nicht eine lineare Funktion von ϵ . Es genügt in dieser Arbeit, sich nur mit dem ersten Freiheitsgrad zu befassen.

Daß auch bei der ATS eine Niveauinflation bei kleinem Box'schen Epsilon beobachtbar ist, wenn erwartungstreue Spurschätzer eingesetzt werden, kann man in Abbildung (12) erkennen. Dort ist der simulierte Typ-I-Fehler der ATS mit den robusten Schätzern gegen das verallgemeinerte Box'sche Epsilon aus (52) aufgeplottet für die simulierten Szenarien bei Normalverteilung. Die Punktgröße ist monoton in $b_n^{-2} \text{Var}(B_n)$ gewählt. Es sind außerdem mit gestrichelten Linien die Zufallsintervalle aus Gleichung (34) eingezeichnet.

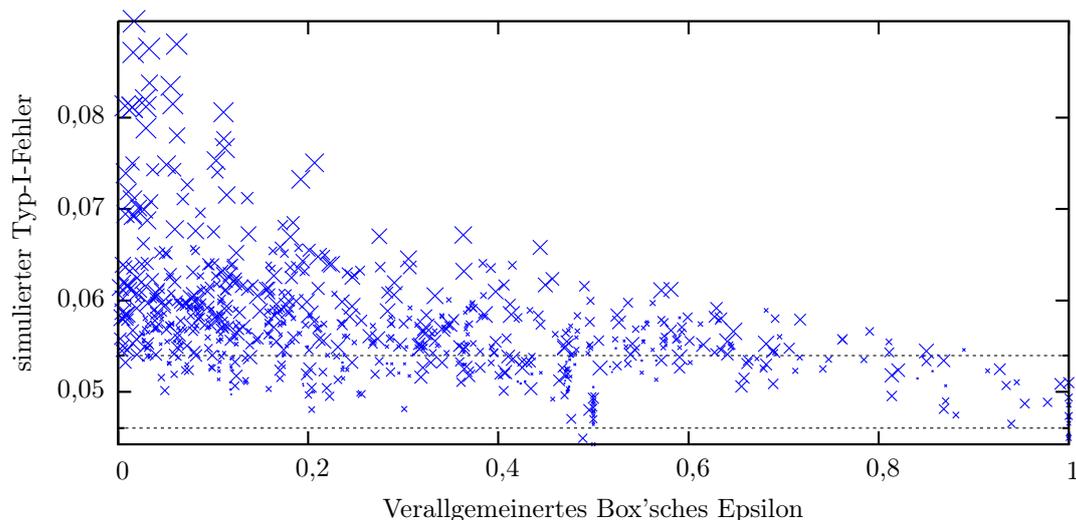


Abbildung 12: Simulierter Typ-I-Fehlers der ANOVA-Typ-Statistik mit den robusten Schätzern gegen

In Abbildung (12) erkennt man, daß bei $\epsilon \leq 0,75$ die erreichten Typ-I-Fehler mehrheitlich oberhalb des Zufallsintervalls liegen. Es ist also auch bei der ATS mit unverzerrten Spurschätzern derselbe Effekt wie bei der Geisser-Greenhouse-Statistik mit den unverzerrten Huynh-Feldt-Spurschätzern erkennbar.

Für den Huynh-Feldt-Test wurde von Quintana und Maxwell (1994) vorgeschlagen, bei kleinem geschätztem Box'schen Epsilon $\tilde{\epsilon}$ den Plug-in-Schätzer zu nehmen. Von Lane (2016) wurde dies aufgegriffen. Analog könnte man bei $\hat{f}_n/\text{rg}(\mathcal{T}) \leq 0,75$ die Plug-in-Schätzer für die ATS nehmen.

Dieser Ansatz führt aber offensichtlich zu Tests, deren Ablehnbereich einen unstetigen Rand hat. Was das für die Anwendung bedeutet, ist in den genannten Arbeiten nicht diskutiert worden. Auch ist denkbar, daß man damit am Ende wieder konservativ wird, wenn auch nicht so stark wie mit den Plug-in-Schätzern. Es ist deswegen besser, die Ursache für die Niveauüberschreitungen zu beschreiben und daraus einen Korrekturansatz zu folgern.

In Abbildung (12) kann man zu diesem Zweck zusätzlich sehen, daß besonders große Niveauüberschreitungen mit besonders großem $b_n^{-2}\text{Var}(B_n)$ einhergehen. Dies gibt einen Hinweis darauf, daß das Einsetzen von B_n eine Hauptursache für die Niveauüberschreitungen ist, zum einen, weil mit endlichen Stichprobenumfängen $\text{Var}(B_n)$ „zu groß“ ist, zum anderen wegen der Abhängigkeit von den Kovarianzmatrizen. Es wird ausreichen, im Folgenden eine Korrektur nur für das Einsetzen dieses einen Schätzers auszuarbeiten.

4.2 Korrekturformel für die kritischen Werte

Bei der ATS werden die kritischen Werte aus den Quantilen der F -Verteilung gewonnen. Sei also $q_{1-\alpha}(\nu_1, \nu_2)$ die Quantilfunktion zum $1 - \alpha$ -Quantil der F -Verteilung

mit den Freiheitsgraden ν_1 und ν_2 als Argumente. Die kritischen Werte zum Niveau α sollen mit $q_{1-\alpha}(B_1/B_n, B_1/B_e)$ bzw. $q_{1-\alpha}(b_1/b_n, b_1/b_e)$ notiert werden, wenn sie Schätzer B_1, B_n, B_e aus Abschnitt 3.3 berechnet werden bzw. auf Grundlage der der wahren Parameter b_1, b_n, b_e , wie sie in Abschnitt 1.7.2 in den Gleichungen (17)-(19) definiert sind.

Die Spürschätzer wurden auf Unverzerrtheit hin konstruiert, um die Verzerrung durch die Plug-in-Schätzer, durch die die ATS konservativ wird, zu beheben. Unverzerrtheit ist bei B_n aber strenggenommen nicht das benötigte Kriterium, da mit Unverzerrtem B_n die streng konvexe Funktion b_1/B_n nach der Jensen'schen Ungleichung wieder verzerrt ist. Es ist aber auch keine ausreichende Lösung, die Verzerrung des Quotienten zu beheben, etwa durch eine approximative Korrektur –nach Bickel und Lehmann (1969) kann es keine exakt erwartungstreuen Schätzer für derartige Funktionen geben–, weil der Quotient seinerseits wieder in die Quantilfunktion eingesetzt wird, die wiederum nichtlinear im ersten Freiheitsgrad ist.

Es soll daher die Verzerrung von $q_{1-\alpha}(b_1/B_n, b_1/b_e)$ als Schätzer für den „wahren kritischen Wert“ $q_{1-\alpha}(b_1/b_n, b_1/b_e)$ gefunden werden. Weil B_n erwartungstreu ist, kann hierzu eine Taylorapproximation zweiter Ordnung für den Erwartungswert verwendet werden (siehe im Anhang 30 auf Seite 82):

$$\begin{aligned} E\left(q_{1-\alpha}\left(\frac{b_1}{B_n}\right)\right) &\doteq q_{1-\alpha}\left(\frac{b_1}{b_n}\right) + \frac{1}{2}\text{Var}_{\mathcal{N}}(B_n) \cdot \frac{\partial^2\left(q_{1-\alpha}\left(\frac{b_1}{b_n}\right)\right)}{\partial b_n^2} \\ &= q_{1-\alpha}\left(\frac{b_1}{b_n}\right) + \frac{1}{2}\text{Var}_{\mathcal{N}}(B_n) \left(\frac{\partial^2 q_{1-\alpha}}{\partial \nu_1^2} \cdot \frac{b_1^2}{b_n^4} + 2\frac{\partial q_{1-\alpha}}{\partial \nu_1} \cdot \frac{b_1}{b_n^3}\right) \quad (53) \end{aligned}$$

Für Testentscheidungen oder globale Konfidenzintervalle kann man daraus eine Korrektur des kritischen Wertes machen, die die Verzerrung durch das Einsetzen von B_n behebt, indem man von $q_{1-\alpha}(B_1/B_n, B_1/B_e)$ den zweiten Summand in (53) abzieht. Hierfür müssen die neuen unbekannt Parameter $\text{Var}(B_n)$, die beiden Ableitungen $\partial^2 q_{1-\alpha}/\partial \nu_1^2$ und $\partial q_{1-\alpha}/\partial \nu_1$ sowie b_1^2/b_n^4 und b_1/b_n^3 durch ausreichend genaue Schätzer ersetzt werden.

„Ausreichend genau“ ist hierbei ein äußerst großer Bereich. Am wichtigsten ist, daß keine positive Korrektur vom unkorrigierten kritischen Wert abgezogen wird, um den Test nicht noch liberaler zu machen. Dem kann sehr einfach vorgebeugt werden, indem der Korrekturterm stets bei 0 abgeschnitten wird. Hierdurch entstehen auch keine Unstetigkeiten wie bei Lane (2016).

Der Absolutbetrag der gesamten Korrektur verträgt einen recht großen relativen Fehler, wobei eine Überkompensation, die zu einem konservativen Test führen würde weniger problematisch wäre. Deswegen sollen in Gleichung (53) die Parameter b_1 und b_n durch ihre erwartungstreuen, dimensionsstabilen und konsistenten Schätzer B_1 und B_n ersetzt werden, um konservative Schätzungen der Quotienten b_1^2/b_n^4 und b_1/b_n^3 zu erhalten. Dabei sind etwaige Bereichsüberschreitungen von B_1/B_n und B_1/B_e

aber zu berücksichtigen. Es soll deswegen so geklammert werden, daß die Freiheitsgradschätzer \hat{f}_n und \hat{f}_e , die auf ihren Wertebereich abgeschnitten sind, einbezogen werden. In die Ableitungen der Quantilfunktion nach dem ersten Freiheitsgrad werden ebenfalls die Freiheitsgradschätzer \hat{f}_n und \hat{f}_e eingesetzt.

Die grobe Berechnung eines Schätzer für $\text{Var}_{\mathcal{N}}(B_n)$, der mit V_n abgekürzt werde, wird im folgenden Abschnitt erklärt. Die praktische Berechnung der Ableitungen wird in Abschnitt 4.4 auf Seite 61 erklärt. Hierbei ist zu beachten, daß die Ableitungen nach dem ersten Freiheitsgrad nur in einem offenen Bereich berechnet werden kann, numerisch muß dieser von den Grenzen weg beschränkt sein.

Da im Bereich großer Box'sche Epsilon nahe bei 1 kein Anlaß für eine Korrektur besteht, soll die hier vorgeschlagene Korrektur einen Faktor bekommen, der bei einem geschätzten Box'schen Epsilon von $\hat{f}/\text{rg}(\mathbf{T}) = 1$ verschwindet. Damit dadurch im Bereich $\epsilon \leq 0,75$ die Korrektur nicht zu schwach ist, wird als Ausgleich nochmal der Faktor 4 eingefügt. Es ergibt sich als korrigierter kritischer Wert

$$q_{1-\alpha}^* = q_{1-\alpha}(\hat{f}_n, \hat{f}_e) + \underbrace{4 \left(1 - \frac{\hat{f}_n}{\text{rg}(\mathbf{T})}\right) \frac{V_n}{B_n^2} \min \left\{ 0, \left(\hat{f}_n^2 \frac{\partial^2 q_{1-\alpha}(\hat{f}_n, \hat{f}_e)}{\partial \nu_1^2} + \hat{f}_n \frac{\partial q_{1-\alpha}(\hat{f}_n, \hat{f}_e)}{\partial \nu_1} \right) \right\}}_{=: k_{1-\alpha}^*}, \quad (54)$$

der überall dort zu berechnen ist, wo $\hat{f}_n > 1 + \epsilon_{\text{Maschine}}$ und $q_{1-\alpha}(\hat{f}_n, \hat{f}_e) > \epsilon_{\text{Maschine}}$ ist, wobei $\epsilon_{\text{Maschine}}$ von der Maschinengenauigkeit nach unten begrenzt wird. Für praktische Anwendungen ist $\epsilon_{\text{Maschine}} = 0,001$ völlig ausreichend. Die Korrektur $k_{1-\alpha}^*$ ist nie positiv.

4.3 Abschätzung für $\text{Var}_{\mathcal{N}}(B_n)$

Man kann mit etwas Geduld $\text{Var}_{\mathcal{N}}(B_n)$ exakt berechnen.

Satz 19. *Es ist*

$$\begin{aligned} \text{Var}_{\mathcal{N}}(B_n) = & \sum_{i=1}^a t_{ii}^4 \left(\frac{4(2n_i^3 - 9n_i^2 + 9n_i + 4)}{n_i^5 (n_i - 1)(n_i - 2)(n_i - 3)} \text{Sp}(\boldsymbol{\Sigma}_i)^4 \right. \\ & \left. + \frac{4}{n_i^5 (n_i - 3)} \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 \right) \\ & + 4 \sum_{i \neq i'} t_{ii'}^4 \left(\frac{(n_i + n_{i'} - 1)}{n_i^2 (n_i - 1) n_{i'}^2 (n_{i'} - 2)} \text{Sp}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i'})^2 \right. \\ & \left. + \frac{4}{n_i^2 (n_i - 1) n_{i'}^2 (n_{i'} - 1)} \text{Sp}^2(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i'}) \right) \end{aligned}$$

$$\begin{aligned}
& + 4 \sum_{\substack{i,i',i'' \\ \neq i,i',i''}} t_{ii'}^2 t_{ii''}^2 \frac{2}{n_i^2 (n_i - 1) n_{i'} n_{i''}} \text{Sp} (\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i'} \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i''}) \\
& + 4 \sum_{i \neq i'} t_{ii}^2 t_{ii'}^2 \frac{4}{n_i^3 (n_i - 1) n_{i'}} \text{Sp} (\boldsymbol{\Sigma}_i^3 \boldsymbol{\Sigma}_{i'}).
\end{aligned}$$

Beweis. Siehe den Beweis auf Seite 102 in Abschnitt A.4.1. \square

Die Terme $\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2$ und $\text{Sp}^2(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i'})$ können durch $(B_2^{(i)})^2$ bzw. $(B_2^{(i,i')})^2$ konsistent geschätzt werden.

Es soll vermieden werden, für Terme der Art $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^4$, $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})^2$, und $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^3 \mathbf{T}_d \boldsymbol{\Sigma}_{i'}$ sowie $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'} \mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i''})$ vollkommen neue Schätzer zu entwickeln. Da keine große Genauigkeit nötig ist, genügen Behelfsparameter, die in der Hauptsache abhängig von den $\boldsymbol{\Sigma}_i$ dieselben Extremstellen haben sollen wie die interessierenden Terme und für die aus den bereits entwickelten Spurschätzern neue Schätzer kombiniert werden können. Durch diesen Umweg werden die interessierenden Terme zwar sehr ungenau, dafür aber sehr recheneffizient geschätzt.

Wie bei der Einführung des Box'schen Epsilon schon in Abschnitt 1.7.1 gesehen, kann $\text{Sp}(\mathbf{S}^2)/\text{Sp}^2(\mathbf{S})$ als Sphärizitätsparameter für eine symmetrisch positiv semidefinite $d \times d$ -Matrix \mathbf{S} interpretiert werden. Ein verwandter Sphärizitätsparameter ist $\text{Sp}(\mathbf{S}^4)/\text{Sp}^2(\mathbf{S}^2)$. Es ist einfach zu zeigen, daß er dieselben Extrema an derselben Stelle hat wie $\text{Sp}(\mathbf{S}^2)/\text{Sp}^2(\mathbf{S})$, siehe Lemma 20. Deswegen wird statt für

$$\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^4 = \frac{\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^4}{\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2} \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2$$

ein Schätzer für $\frac{\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2}{\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)} \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2$ benutzt, der mit den vorhandenen Schätzern durch $(B_2^{(i)})^3 (B_1^{(i)})^{-1}$ konsistent und in dem Sinne dimensionsstabil geschätzt werden kann, daß die Taylorapproximation der Varianz dieses Bruches nicht vom zu schätzenden Parameter abhängt, analog zu Satz 15, siehe Korollar 31 auf Seite 84 im Anhang.

Ähnlich kann mit den übrigen Spuren verfahren werden dank folgendem Lemma:

Lemma 20. *Seien $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3$ positiv semidefinite Matrizen. Dann haben*

1.

$$\frac{\text{Sp}(\boldsymbol{\Sigma}_1)^4}{\text{Sp}^2(\boldsymbol{\Sigma}_1)^2} \text{ und } \frac{\text{Sp}(\boldsymbol{\Sigma}_1)^2}{\text{Sp}^2(\boldsymbol{\Sigma}_1)}$$

2.

$$\frac{\text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)^2}{\text{Sp}^2(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)} \text{ und } \frac{\text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)}{\text{Sp}(\boldsymbol{\Sigma}_1) \text{Sp}(\boldsymbol{\Sigma}_2)}$$

3.

$$\frac{\text{Sp}(\Sigma_1 \Sigma_2 \Sigma_1 \Sigma_3)}{\text{Sp}(\Sigma_1 \Sigma_2) \text{Sp}(\Sigma_1 \Sigma_3)} \text{ und } \frac{\text{Sp}(\Sigma_1 \Sigma_2) \text{Sp}(\Sigma_1 \Sigma_3)}{\text{Sp}^2(\Sigma_1) \text{Sp}(\Sigma_2) \text{Sp}(\Sigma_3)}$$

4.

$$\frac{\text{Sp}(\Sigma_1^2 \Sigma_3)}{\text{Sp}(\Sigma_1^2) \text{Sp}(\Sigma_1 \Sigma_3)} \text{ und } \frac{\text{Sp}(\Sigma_1^2) \text{Sp}(\Sigma_1 \Sigma_3)}{\text{Sp}^3(\Sigma_1) \text{Sp}(\Sigma_3)}$$

jeweils dieselben Extrema für dieselben $\Sigma_1, \Sigma_2, \Sigma_3$.

Beweis. Siehe im Anhang auf Seite 105. □

Die so gewonnenen ungenauen, recheneffizienten Schätzer sind insgesamt also

$$\begin{aligned} & \left(B_2^{(i)}\right)^3 \left(B_1^{(i)}\right)^{-1} \text{ für } \text{Sp}(\mathbf{T}_d \Sigma_i)^4 \\ & \left(B_2^{(i,i')}\right)^3 \left(B_1^{(i,i')}\right)^{-1} \text{ für } \text{Sp}(\mathbf{T}_d \Sigma_i \mathbf{T}_d \Sigma_{i'})^2 \\ & \left(B_2^{(i,i')} B_2^{(i,i'')} B_2^{(i',i'')}\right) \cdot \left(B_1^{(i)} B_1^{(i',i'')}\right)^{-1} \text{ für } \text{Sp}(\mathbf{T}_d \Sigma_i \mathbf{T}_d \Sigma_{i'} \mathbf{T}_d \Sigma_i \mathbf{T}_d \Sigma_{i''}) \\ & \left(B_2^{(1)} B_2^{(1,3)}\right)^{\frac{3}{2}} \left(B_1^{(1)} B_1^{(1,3)}\right)^{-1} \text{ für } \text{Sp}((\mathbf{T}_d \Sigma_1)^3 \mathbf{T}_d \Sigma_2) \end{aligned}$$

Diese Schätzer ersetzen die entsprechenden unbekanntenen Terme in $\text{Var}_{\mathcal{N}}(B_n)$ aus Satz 19, um V_n zu erhalten. Die Korrekturformel (54) wird durch derartig grobe Schätzungen nicht zu schlecht. Erstens kürzen sich Schätzfehler durch den Faktor $V_n B_n^{-2}$ grob weg, zweitens reicht die Korrektur in den meisten Fällen schon bei einem recht großen relativen Fehler von 50% –also wenn nur halb korrigiert wird–, sehr gut in den nach Bradleys Kriterium Bradley (1978) definierten tolerierbaren Bereich von unter $\frac{3}{2}\alpha$, da der extremste simulierte unkorrigierte Typ-I-Fehler bei 2α lag.

Man kann Konsistenz und Dimensionsstabilität der bekannten Schätzer heranziehen, um zu rechtfertigen, daß man sie einsetzen kann, um $\text{Var}_{\mathcal{N}}(B_n)$ zu schätzen.

4.4 Zweite Ableitung der Quantilfunktion nach dem ersten Freiheitsgrad

Die Dichtefunktion der F -Verteilung mit ν_1 und ν_2 Freiheitsgraden sei notiert als

$$F'_F(x, \nu_1, \nu_2) = \frac{\nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} x^{\frac{\nu_1-2}{2}} (\nu_2 + \nu_1 x)^{-\frac{\nu_1+\nu_2}{2}}, \quad x \geq 0$$

wobei B (ohne Indices) die Beta-Funktion ist. Die Verteilungsfunktion ist somit

$$F_F(x, \nu_1, \nu_2) = \int_0^x F'_F(x, \nu_1, \nu_2) dx.$$

Numerisch wird die Quantilfunktion $q(p, \nu_1, \nu_2)$ zu einem Wahrscheinlichkeitslevel p und konstanten Freiheitsgraden ν_1 und ν_2 gewöhnlich durch ein Newton-Verfahren berechnet, das die implizite Funktionsdefinition von $q_F(p, \nu_1, \nu_2)$ als Nullstelle in x der Funktion

$$g(x, \nu_1, \nu_2) := \int_0^x F'_F(t, \nu_1, \nu_2) dt - p = 0 \quad (55)$$

nutzt. So wird etwa für die R-Funktion q_F ein Newton-Verfahren verwendet, siehe R Core Team (2018). Hierbei wird das Quantil bis auf eine voreingestellte Fehlertoleranz berechnet.

Leider liegt darin auch der Grund, warum derartig implementierte Quantilfunktionen unbrauchbar sind, um sie numerisch zu differenzieren. Numerisches Differenzieren beruht stets auf der Annäherung durch finite Differenzenquotienten der Art

$$\frac{q_F(p+h, \nu_1, \nu_2) - q_F(p-h, \nu_1, \nu_2)}{2h}$$

mit kleinem h . Damit überhaupt das Vorzeichen der numerischen Ableitung richtig ist, wenn q_F mit einem Newton-Verfahren mit absoluten Fehler Δq_F berechnet wird, muß $\Delta q_F < (q_F(p+h, \nu_1, \nu_2) - q_F(p-h, \nu_1, \nu_2))/2$ sein. Um also die Fehlertoleranz für das Newton-Verfahren zur Berechnung von $q_F(p+h, \nu_1, \nu_2)$ und $q_F(p-h, \nu_1, \nu_2)$ im ersten Schritt einzustellen, wird das Ergebnis aus der Differenz $q_F(p+h, \nu_1, \nu_2) - q_F(p-h, \nu_1, \nu_2)$ benötigt, die erst im zweiten Schritt bekannt ist. Solche Probleme sind zu vermeiden.

Vom Rechenaufwand würde dies darüber hinaus mehrere Anwendungen des Newton-Verfahrens erfordern, welche wiederum mehrmals das Integral $\int_0^x F'_F(t, \nu_1, \nu_2) dt$ auswerten. Insgesamt stehen Genauigkeit und Rechenaufwand in keinem erträglichen Verhältnis. Für die eigentlich interessierende zweite Ableitung der Quantilfunktion gilt dies erst recht.

Stattdessen kann mit der Darstellung in Gleichung (55) der Satz über die Ableitung implizit definierter Funktionen angewendet werden, siehe z. B. Heuser (2000), Satz 170. Heraus kommt eine Integraldarstellung der ersten Ableitung, wie der folgende Satz besagt. Auf diese Weise wird die Berechnung einer Ableitung in die Berechnung eines Integrals umgewandelt, was numerisch wesentlich günstiger ist.

Satz 21. *Die erste Ableitung der in Gleichung (55) implizit definierten Quantilfunktion nach dem ersten Freiheitsgrad $\nu_1 \geq 1$ lautet*

$$\frac{\partial q_F(p, \nu_1, \nu_2)}{\partial \nu_1} = \omega(p, \nu_1, \nu_2) \cdot \int_0^{q_F(p, \nu_1, \nu_2)} \tau_1(t, \nu_1, \nu_2) dt, \quad (56)$$

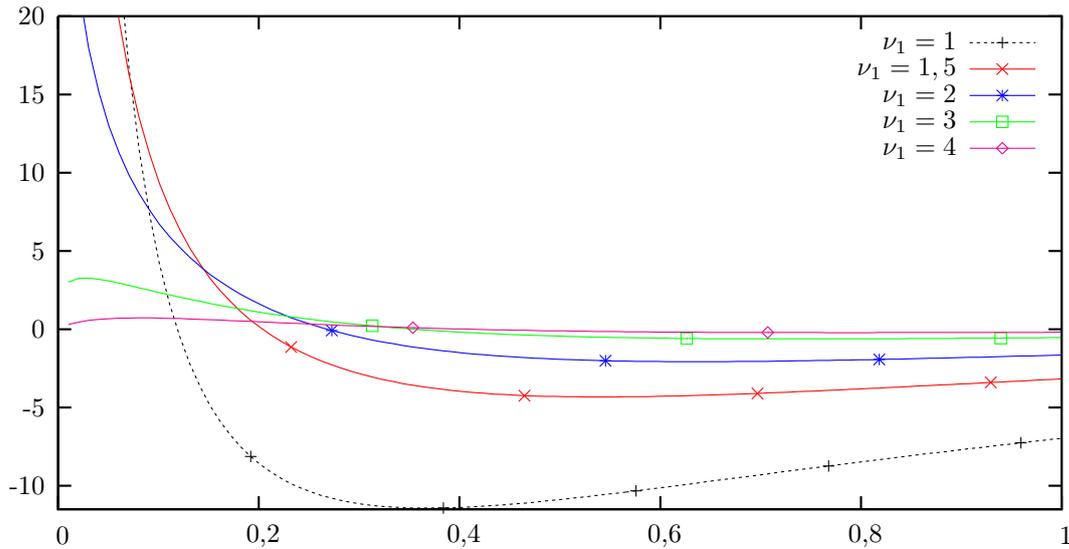


Abbildung 13: Integrand für die Integraldarstellung von $\frac{\partial q_F}{\partial \nu_1}$ bei $\nu_2 = 10$ und $p = 0,95$.

wobei die Abkürzungen

$$\tau_1(t, \nu_1, \nu_2) := t^{\frac{\nu_1}{2}-1} \left(\frac{\nu_1 t + \nu_2}{\nu_2} \right)^{-\frac{\nu_1 + \nu_2 + 2}{2}} \left(\nu_2(t-1) + (\nu_1 t + \nu_2) \cdot \left[\log \left(\frac{\nu_1 t + \nu_2}{\nu_1 t} \right) - \psi \left(\frac{\nu_1 + \nu_2}{2} \right) + \psi \left(\frac{\nu_1}{2} \right) \right] \right)$$

und

$$\omega(p, \nu_1, \nu_2) := \frac{1}{2} (q_F(p, \nu_1, \nu_2))^{1-\frac{\nu_1}{2}} (\nu_1 q_F(p, \nu_1, \nu_2) + \nu_2)^{\frac{\nu_1 + \nu_2}{2}}$$

gelten und ψ für die Digammafunktion $\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ steht (siehe Abramowitz und Stegun (1964), Nr. 6.3). Insbesondere existiert das Integral (56).

Beweis. Siehe Seite 106. □

Das Integral in Gleichung (56) ist numerisch auszuwerten. Zu beachten ist hierbei die Polstelle $t = 0$ bei kleinem ersten Freiheitsgraden $1 \leq \nu_1 \leq 2$, deren Integrierbarkeit im Beweis gezeigt wurde. Zur Illustration dieser Polstelle ist der Integrand für verschiedene Werte des ersten Freiheitsgrades ν_1 in Abbildung 13 aufgeplottet. Es wird $p = 0,95$ und $\nu_2 = 10$ vorgegeben.

Die numerische Integration an integrierbaren Polstellen ist nur mit speziellen Quadraturverfahren möglich. Geeignet für den gegebenen Fall ist die Routine `qaws` aus der GNU Scientific Library (Galassi u. a. (2011)). Sie verwendet Tschebyscheff-Polynome zu Gewichtsfunktionen der Art $t^\nu \log(t)$ mit $\nu > -1$ für die Polstelle $t = 0$.

Für die zweite Ableitung des Quantils nach dem ersten Freiheitsgrad kann numerisch differenziert werden (z. B. mit dem Algorithmus `gsl_deriv_central` aus der GNU Scientific Library). Numerisch effizienter und voraussichtlich besser konditioniert wäre es, die Gleichung (56) nochmals symbolisch zu differenzieren. Dann sind zur Berechnung von $\frac{\partial^2 q_F}{\partial \nu_1^2}$ nur zwei Integrale auszuwerten, nämlich das Integral aus Gleichung (56) und das Integral aus dessen Ableitung nach ν_1 . Daß letzteres Integral existiert, zeigt Lemma 41 auf Seite 106.

Leider treten in diesem Integranden aber Terme der Art $h(t)t^{\frac{\nu_1}{2}}\log^2(t)$ auf, wobei $h(t)$ eine beschränkte Funktion ist. Dadurch liegt bei kleinem ν_1 an $t = 0$ eine integrierbare Polstelle vor, in deren Umgebung mit Integrationsroutinen aus der GNU Scientific Library kein Integral ausgerechnet werden kann. Geeignete Verfahren müßten erst verfügbar gemacht werden.

4.5 Korrekturformel für die p -Werte und geeignete Ausgabe

Konfidenzintervalle, die dual zum Test sind, berichtet man vor allem dann, wenn auf der Skala der Messungen der Nichtablehnbereich abgebildet werden soll. Für Globaltests ist das nicht mehr möglich, da der Nichtablehnbereich typischerweise mehrdimensional ist. Deswegen werden hier eher p -Werte berichtet. Statt korrigierter kritischer Werte zu einem festen α müssen also p -Werte korrigiert werden.

Man kann p -Werte auch als das α betrachten, bei dem der kritische Wert gerade mit der Teststatistik übereinstimmt. Eine Möglichkeit wäre daher, die nichtlineare Gleichung

$$p = 1 - F_F \left(F - k_p^*, \hat{f}_n \right)$$

durch ein Newton-Verfahren zu lösen. Weil k_p^* konzeptbedingt recht grob berechnet wird und für ein Newton-Verfahren mehrere Auswertungen nötig sind, soll darauf verzichtet werden und ein anderer Ansatz bevorzugt werden, der im Endeffekt auf einer linearen Approximation beruht, also gewissermaßen einem einzigen Newton-Schritt.

Wenn ein Test exakt ist, wie in diesem Kapitel für die ATS mit bekannten Freiheitsgraden angenommen, dann gilt nach Definition (siehe Lehmann und Romano (2005))

$$\begin{aligned} \alpha &= E \left(\mathbb{1}_{F > q_{1-\alpha}(b_1 b_n^{-1})} \right) \\ &= 1 - F_F \left(q_{1-\alpha}(b_1 b_n^{-1}) \right). \end{aligned} \tag{57}$$

Außerdem ist der zugehörige p -Wert $1 - F_F(F, b_1 b_n^{-1})$ durch dieselbe Annahme gleichverteilt auf $[0, 1]$, siehe Lehmann und Romano (2005), Lemma 3.3.1. Wenn man daher unkorrigiert

$$p := 1 - F_F \left(F, \hat{f}_n \right)$$

schreibt, kann man zu einem korrigierten p -Wert p^* gelangen durch eine lineare

Approximation (54)

$$p^* := p + F'_F \left(F, \hat{f}_n \right) \cdot k_p^*.$$

Auch hier wird auf die Berechnung von k_p^* verzichtet, wenn $p \leq 0,001$ ist. Mit dieser Korrekturformel für den p -Wert wurden auch die Monte-Carlo-Simulationen im folgenden Abschnitt durchgeführt.

Wenn Testergebnisse der ANOVA oder der ATS berichtet werden, werden konventionell auch der Wert der Teststatistik, die Freiheitsgrade und die p -Werte angegeben. Gibt man die korrigierten p -Werte aus, dann ließen sie sich nicht mehr mit den angegebenen Freiheitsgradschätzern \hat{f}_n und \hat{f}_e bei der beobachteten Teststatistik F reproduzieren. Es erspart in der Anwendung daher Irritationen, wenn \hat{f}_n so modifiziert ausgegeben wird, daß man damit und mit der Teststatistik F und \hat{f}_e den korrigierten p -Wert erhält. Dies ist mit einem einfachen Newton-Verfahren möglich.

5 Simulation mit Schätzern und Korrektur

5.1 Niveau

Als erstes ist die Qualität der Korrektur bei Normalverteilung hinsichtlich des Typ-I-Fehlers zu untersuchen. Dies geschieht mit den bekannten Simulationsszenarien in Abbildung 14. Zum Vergleich werden die Ergebnisse der unkorrigierten ATS mit den robusten Schätzern und der ATS mit den eingesetzten Parametern der Simulationen gezeigt. Es ist nicht zu erwarten, daß die Korrektur zu genaueren Ergebnissen führt als das Einsetzen der wahren Parameter. Weiterhin sind Simulationsergebnisse dargestellt, die sich ergeben, wenn bei $\hat{f}_n \leq 0,75$ der Plug-in-Schätzer verwendet wird, in Analogie zur Arbeit von Lane (2016) und der Idee von Quintana und Maxwell (1994).

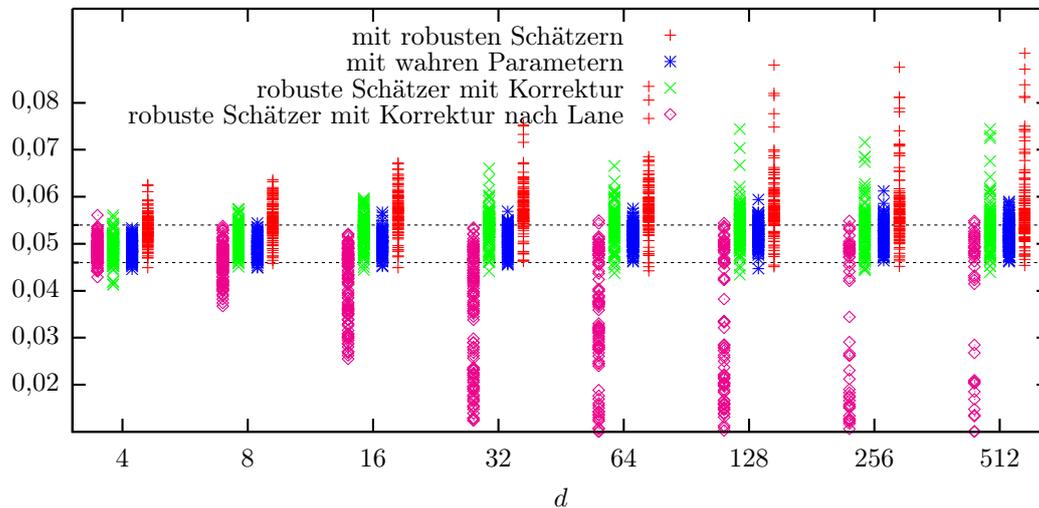


Abbildung 14: ATS mit und ohne Korrektur und wahren Parametern bei Normalverteilung

Man sieht in Abbildung 14, daß die größten Niveauüberschreitungen mit der Korrektur halb so groß sind wie mit des unkorrigierten Tests. Die kleinsten beobachteten Typ-I-Fehler der Korrektur sind nur sehr geringfügig unter denen des unkorrigierten Tests, ebenfalls sehr geringfügig und obendrein selten ist die Korrektur konservativer als mit den wahren Parametern. Diese Unterschiede sind aber geringer als die „natürliche“ Fehlertoleranz, die durch die begrenzte Zahl von 20000 Simulationswiederholungen berücksichtigt werden muß.

Durch das „Umschalten“ auf den Plug-in-Schätzer bei kleinem \hat{f}_n entsprechend Lane (2016) und Quintana und Maxwell (1994) wird der Test zwar nie mehr liberal, allerdings wird er bei hoher Dimensionalität zunehmend konservativ. Durch diesen Ansatz ist also bei wirklich hochdimensionalen Daten nichts gewonnen.

Exemplarisch für die anderen Nichtnormalverteilungen sind in Abbildung 15 und 16

die Ergebnisse für das additiv-lineare Modell mit Rademacher-verteilten Z_{ikj} und lognormalverteilten Z_{ikj} gezeigt.

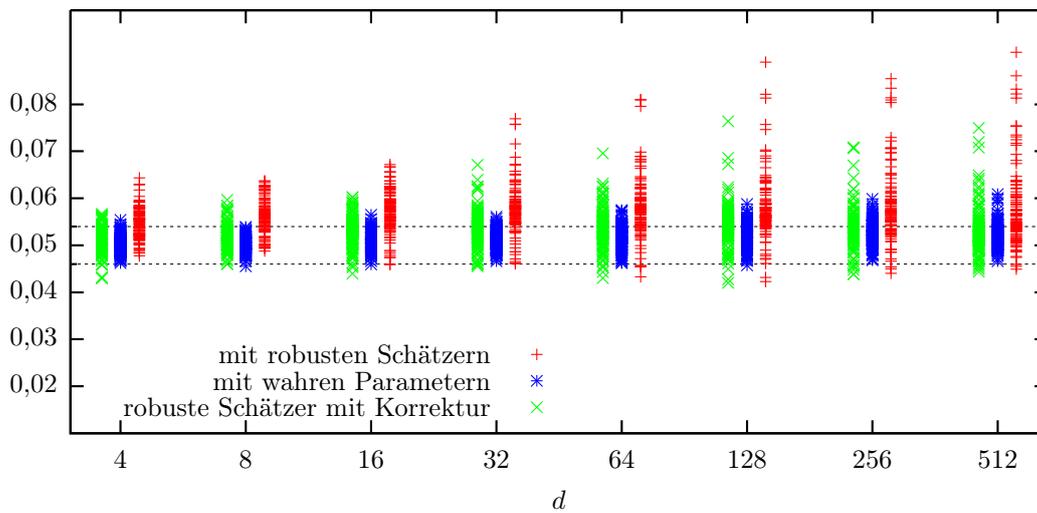


Abbildung 15: ATS mit und ohne Korrektur und wahren Parametern bei Rademacher-Verteilung im additiv-linearen Modell

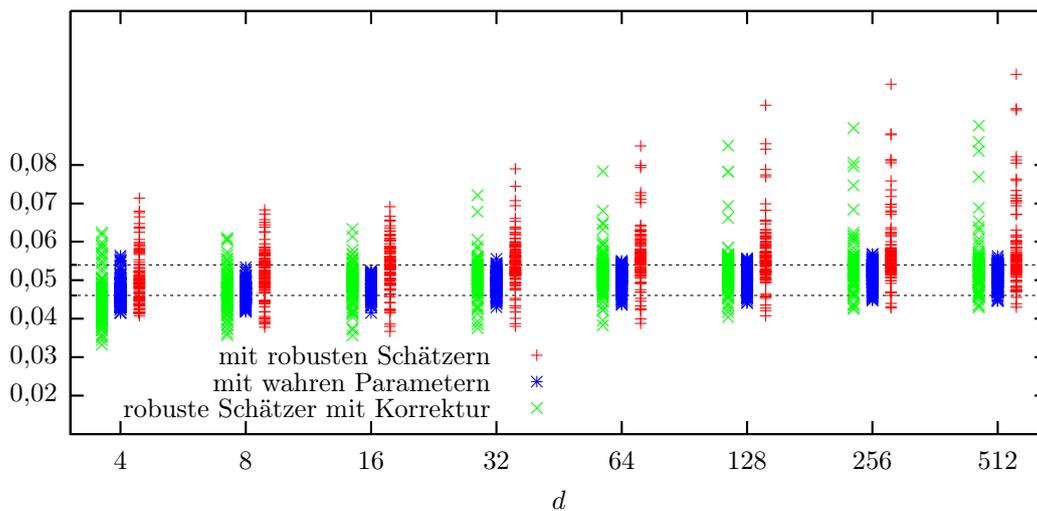


Abbildung 16: ATS mit und ohne Korrektur und wahren Parametern bei Lognormalverteilung im additiv-linearen Modell

Bei Rademacher-Verteilung in Abbildung 15 ist ein ähnliches Ergebnis zu beobachten wie bei Normalverteilung. Bei Lognormalverteilung in Abbildung 16 zeigt sich wieder die größere Ungenauigkeit, die man schon bei den Simulationen mit den wahren Parametern gesehen hat. Die Korrektur ist hier öfter und stärker konservativ. Auf der anderen Seite ist die Korrektur durchaus noch liberal, wenn auch nicht so extrem wie der unkorrigierte Test.

Mit dem wahren Parametern haben sich bei der elliptischen Verteilung mit Pareto-verteiltem $|W_{ik}| + 1$ sehr konservative Ergebnisse gezeigt. In Abbildung (17) wird

dasselbe mit korrigierter und unkorrigierter Schätzung betrachtet.

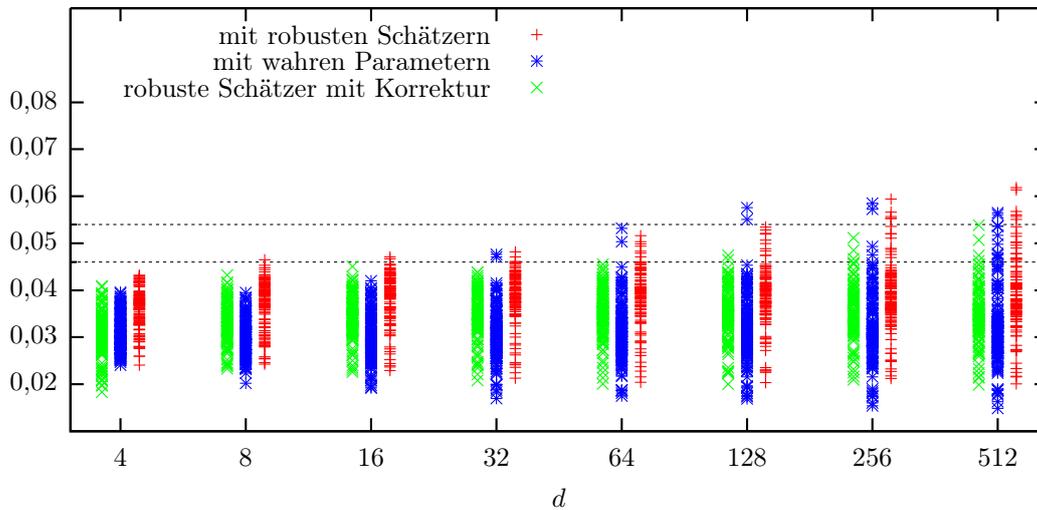


Abbildung 17: ATS mit und ohne Korrektur und wahren Parametern bei elliptischer Paretoverteilung.

In Abbildung 17 sind zwar auch die ATSen mit den Schätzern konservativ, man sieht aber nicht, daß die Korrektur dies verschlimmert.

Mit den anderen Verteilungen, insbesondere den übrigen elliptischen Verteilungen ändert sich das Bild nicht grundlegend. Man kann also schlußfolgern, daß die robusten Schätzer tatsächlich zu robusten Tests führen. Es zeigt sich außerdem, daß eine Korrektur für das Einsetzen von erwartungstreuen Schätzern bei kleinem Box'schen Epsilon durchaus Verbesserungen bringt.

5.2 Power

Weil die behandelten Spurschätzer alle translationsinvariant sind, ist nicht zu erwarten, daß sich bei der Power etwas grundlegend anderes zeigt als von der ATS bislang bekannt. Insbesondere die Poweruntersuchungen in Becker (2010), Brunner, Bathke und Placzek (2012), Happ, Harrar und Bathke (2016) und Harden (2013) würden sich hier reproduzieren. Es ist aber nicht mehr selbstverständlich, daß die Korrektur translationsinvariant ist, wenn die p -Werte wie in Abschnitt 4.5 beschrieben korrigiert werden. Aus diesem Grunde sind für Normalverteilung in Abbildung 18 Powerkurven für $d \in \{20, 50, 100, 200\}$ aufgeplottet.

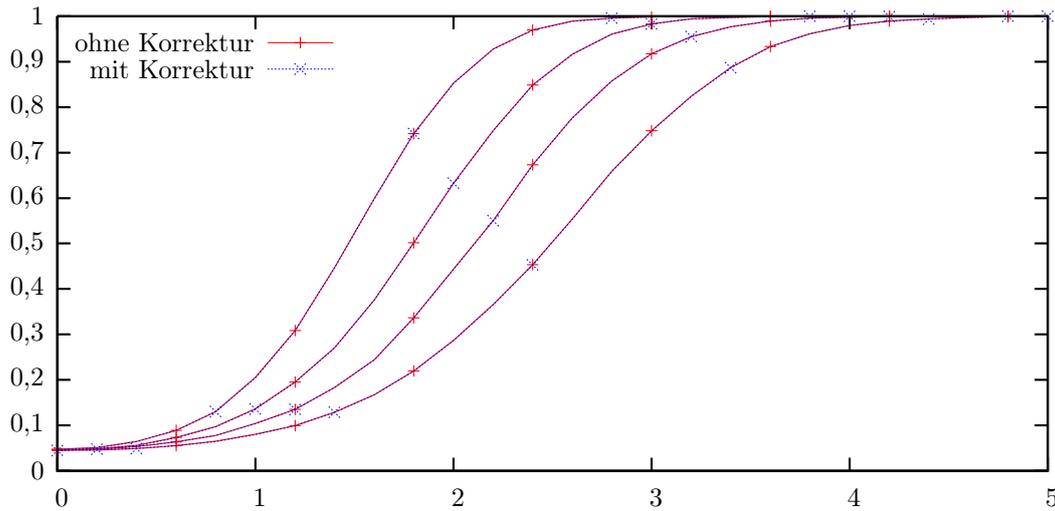


Abbildung 18: Die Powerkurven mit und ohne Korrektur sind fast identisch bei Normalverteilung, $n_1 = 10$, $n_2 = 15$, $n_3 = 25$, $\Sigma_1 = \text{linToe}$, $\Sigma_2 = 1,25\Sigma_1$, $\Sigma_3 = 1,5\Sigma_1$ und von unten nach oben mit $d \in \{20, 50, 100, 200\}$.

Als Kovarianzmatrixstruktur wurde die Toeplitzstruktur mit linear abfallenden Kovarianzen gewählt (*linToe*), weil bei ihr das Box'sche Epsilon besonders klein werden kann, so daß die Korrektur eher greift. Es wurden unterschiedlich große Stichprobenumfänge gewählt ($n_1 = 10$, $n_2 = 15$, $n_3 = 25$) und die Kovarianzmatrizen in der zweiten bzw. dritten Stichprobe sind das 1,25-fache bzw. das 1,5-fache der Matrix aus der ersten Stichprobe. Die kleinste Stichprobe hat also auch die kleinste Varianz. Getestet wurde die Wechselwirkungshypothese und die Alternative ist wie bei Happ, Harrar und Bathke (2016) gewählt: $\mu_1 = \mu_2 = \mathbf{0}$ und $\mu_3 = (0, \dots, 0, \mu)$ mit $\mu \in [0, 5]$.

Man erkennt in Abbildung 18 keinen Powerunterschied, der sich nicht auch schon im Typ-I-Fehler gezeigt hätte. Die Korrektur beeinträchtigt die Power also nicht.

Die Nichtnormalverteilungsmodelle unterscheiden sich nicht im Lokationsparameter μ , sondern nur in den Streuungen. Man kann daher erwarten, daß die Powerkurven sich nur in dem Maße unterscheiden, wie es sich von den Typ-I-Fehlern her fortsetzt, die Powerkurven verschiedener Verteilungen also nicht unterschiedlich flach ausfallen. Exemplarisch ist die Power bei drei elliptischen Verteilungen in Abbildung 19 zu sehen, insbesondere der Pareto-Verteilung, die in den Abschnitten 2.5 und 3.5 zu konservativen Tests führte. Powerfunktionen bei Verteilungen im additiv-linearen Modell sind nicht abgebildet, weil sie fast deckungsgleich sind. Zwischen korrigierter und unkorrigierter ATS wird in dem Plot auch kein Unterschied gemacht, weil er so gering ist, daß er im Plot optisch nicht auflösbar wäre.

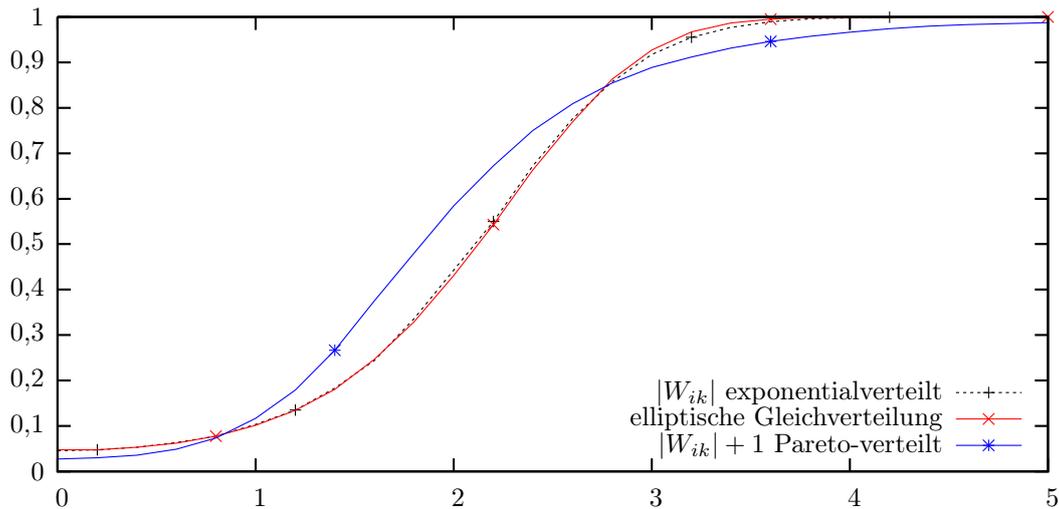


Abbildung 19: Power der ATS bei ausgewählten elliptischen Verteilungen und $d = 100$.

Die elliptische Gleichverteilung und die Exponentialverteilung haben sehr ähnliche Powerkurven. Einzig bei der Pareto-Verteilung, bei der die ATS bekanntermaßen konservativ ist, erkennt man eine stark abweichende Powerkurve. Interessanterweise ist im mittleren Bereich bis zu einer Power von 0,8 die Power der Pareto-Verteilung sogar höher. Das relativiert erheblich, daß in den Niveausimulationen bei dieser Verteilung die ATS konservativ war wegen der „heavy tails“ dieser Verteilung.

Insgesamt kann man schließen, daß die Power der ATS weder bemerkenswert nachteilig von der Korrektur noch von Nichtnormalverteilung beeinträchtigt wird.

6 Auswertung der Anwendungsbeispiele

6.1 Gewichtsentwicklung der Ratten

Die Testergebnisse für den Gruppeneffekt, den Zeiteffekt, die Wechselwirkung und den Zeitverlaufseffekt sind für die ATS mit den robusten, unkorrigierten Schätzern in Tabelle 1 zu sehen. In den Spalten sind der geschätzte erste Freiheitsgrad \hat{f}_n bezeichnet mit ndf, der zweite Freiheitsgrad \hat{f}_e (ddf), die Teststatistik F , der p -Wert und –nachrichtlich– das verallgemeinerte Box’sche Epsilon auf Grundlage von \hat{f}_n eingetragen.

Hypothese	ndf	ddf	Teststatistik	p-Wert	Box’sches ϵ
Gruppe	1,872	30,447	10,589	0,0004	0,9361
Zeit	4,002	107,151	274,691	0,0000	0,1906
Gruppe * Zeit	6,827	107,151	8,694	0,0000	0,1625
Gruppe Zeit	1,961	31,915	10,540	0,0003	0,0446

Tabelle 1: Testergebnis zur Gewichtsentwicklung mit robusten Schätzern

Qualitativ gleichen die Ergebnisse in Tabelle 1 denen in Harden (2013): Zum 5%-Niveau signifikant sind alle getesteten Effekte, insbesondere unterscheiden sich die Verlaufskurven zwischen den drei Gruppen (vierte Zeile). Am wichtigsten ist die signifikante Wechselwirkung zwischen Dosisstufe und Zeit (dritte Zeile), denn ihrerwegen können die beiden Haupteffekte (Zeit und Gruppenzugehörigkeit) nicht isoliert voneinander interpretiert werden.

Bemerkenswert ist, daß die Box’schen Epsilon außer beim Gruppeneffekt –der ohnehin nicht hochdimensional ist– deutlich unter 0,75 liegen. Deswegen ist die Benutzung der Korrektur angebracht. Die Ergebnisse davon stehen in Tabelle 2. Dort ist statt dem Box’schen Epsilon in einer Spalte vermerkt, ob die Korrektur wirklich angewendet wurde. In den Zeilen „Ja“ wurde korrigiert. In den Zeilen „Rand“ wurde keine Korrektur vorgenommen, weil der Rand des Definitionsbereichs der involvierten Funktionen zu nah war.

Hypothese	ndf	ddf	Teststatistik	p-Wert	Korrektur
Gruppe	1,872	30,447	10,589	0,0004	Rand
Zeit	4,002	107,151	274,691	0,0000	Rand
Gruppe * Zeit	6,827	107,151	8,694	0,0000	Rand
Gruppe Zeit	1,373	31,915	10,540	0,0011	Ja

Tabelle 2: Testergebnis zur Gewichtsentwicklung Schätzern mit robusten Schätzern und korrigiertem Test.

In fast allen Fällen waren die unkorrigierten p -Werte zu klein. Eine Berechnung hätte am qualitativen Ergebnis allerdings nichts geändert. Dies ist auch der Fall bei dem Test auf einen Gruppeneffekt auf die Zeitverlaufskurven. Auch der korrigierte

p -Wert liegt noch unterhalb des Signifikanzniveaus von 5%. Der ersten Freiheitsgrad beträgt nach der Korrektur aber nur noch 1,373 und nicht mehr 1,961. Die Interpretation, daß es eine Wechselwirkung zwischen Dosis und Gewichtszunahme gibt, bleibt bestehen.

Der Datensatz über die Gewichtsentwicklung von Wistar-Ratten ist also ein Beleg dafür, daß Datensätze mit kleinem Box'schen Epsilon durchaus vorkommen können.

6.2 Lauffleistung der Mäuse

Die Testergebnisse der ANOVA-Typ-Statistik mit den robusten Schätzern sind in Tabelle 3 zu sehen.

Hypothese	ndf	ddf	Teststatistik	p-Wert	Box'sches ϵ
Genotyp	1,000	17,000	0,352	0,5608	1,0000
Woche	1,961	28,877	5,818	0,0078	0,9806
Tag	5,781	124,093	15,465	0,0000	0,3401
Genotyp * Woche	1,961	28,877	0,019	0,9802	0,9806
Genotyp * Tag	5,781	124,093	1,607	0,1533	0,3401
Woche * Tag	15,525	252,036	2,818	0,0004	0,4566
Genotyp * Woche * Tag	15,525	252,036	0,652	0,8343	0,4566
Genotyp Woche	3,000	51,000	0,164	0,9199	1,0000
Genotyp Tag	2,115	37,686	0,832	0,4488	0,1175
Genotyp Woche , Tag	7,215	121,832	0,519	0,8236	0,1336

Tabelle 3: Unkorrigiertes Testergebnis mit robusten Schätzern

Bei zwei niedrigdimensionalen Tests ist es nötig gewesen, die Freiheitsgradschätzungen auf ihren Bereich zu beschneiden, nämlich beim Test auf den Haupteffekt des Genotyps (Zeile „Genotyp“) und den multivariaten Effekt des Genotyps auf den Verlauf der Gesamtleistungen in den drei großen Zeitabschnitten (Zeilen „Genotyp | Woche“).

Wirklich hochdimensionale Tests sind nur in den Zeilen mit „Tag“ zu finden. In all diesen Fällen werden allerdings verallgemeinerte Box'sche Epsilone mit den unverzerrten Spurschätzern geschätzt, die kleiner sind als 0,75. Die Korrektur würde also einen Effekt haben können.

In Tabelle 4 sind die Testergebnisse tabelliert, wenn für das Einsetzen der erwartungstreuen Schätzer korrigiert wird. In den Zeilen mit „kons.“ wurde der p -Wert konstant gelassen, weil die Korrektur ihn verringert hätte.

Hypothese	ndf	ddf	Teststatistik	p-Wert	Korrektur
Genotyp	1,000	17,000	0,352	0,5608	Rand
Woche	1,943	28,877	5,818	0,0080	Ja
Tag	5,781	124,093	15,465	0,0000	Rand
Genotyp * Woche	1,961	28,877	0,019	0,9824	Ja
Genotyp * Tag	5,460	124,093	1,607	0,1572	Ja
Woche * Tag	14,684	252,036	2,818	0,0005	Ja
Genotyp * Woche * Tag	15,525	252,036	0,652	0,8343	kons.
Genotyp Woche	3,000	51,000	0,164	0,9199	Ja
Genotyp Tag	2,115	37,686	0,832	0,4488	kons.
Genotyp Woche , Tag	7,215	121,832	0,519	0,8236	kons.

Tabelle 4: Testergebnis mit robusten Schätzern und korrigiertem Test.

In drei Fällen hat die Korrektur den unkorrigierten p -Wert konstant gelassen. In fünf Fällen hat die Korrektur gewirkt, wovon bei einem, „Genotyp | Woche“, die korrigierte Differenz unter der Rundungsgenauigkeit lag. Die ersten Freiheitsgrade haben sich aber durch die Korrektur nur wenig geändert.

Hypothese	ndf	ddf	Teststatistik	p-Wert	Korrektur
Genotyp	1,000	17,000	0,352	0,5608	Rand
Zeitpunkt	8,428	129,175	6,791	0,0000	Rand
Genotyp * Zeitpunkt	8,428	129,175	0,576	0,8045	kons.
Genotyp Zeitpunkt	7,215	121,832	0,519	0,8236	kons.

Tabelle 5: Hauptfragestellung mit Korrektur

Betreffs der Hauptfragestellung nach dem Effekt des Genotyps ist in den Daten der bis zu 43 Wochen alten Mäusen (noch) nichts aufzufinden. Beispielsweise ist der Genotypeneffekt weder signifikant hinsichtlich des Zeitverlaufs aus den 54 Meßtagen, (Zeile Genotyp | Woche, Tag), noch hinsichtlich einer Wechselwirkung mit den Zeitfaktoren. Falls die KO-Mäuse anfänglich schneller sind und dann binnen 40 Wochen von den Wildtyp-Mäusen überholt würden, wäre eine solche Wechselwirkung sichtbar gewesen. Der signifikante Haupt-Zeiteffekt ist erwartet worden.

Stattdessen zeigt sich ein sehr deutlicher Zeiteffekt nach dem Einsetzen in den Lauf radkäfing in Wechselwirkung mit dem jeweiligen Alter in Wochen in der sekundären Fragestellung.

Zum Schluß kann festgehalten werden, daß auch dieser Datensatz ein Beispiel für kleine Box'sche Epsilon ist, bei denen die Korrektur angebracht ist.

7 Zusammenfassung, Diskussion und Ausblick

In der vorliegenden Arbeit wurde eine Statistikapproximation auf Basis der ANOVA-Typ-Statistik hergeleitet, deren Robustheit nach dem Vorbild von Atiqullah (Atiqullah 1962) gilt, bei dem also die vierten Momente der Ausgangsdaten sich nicht auf die ersten beiden Momente der Prüfgrößen durchschlagen. Hierbei hat sich gezeigt, daß die Robustheit nicht wie bei der ANOVA von der quadratischen Balanciertheit abhängt, sondern von der Auslegung der quadratischen Formen auf das Behrens-Fisher-Problem.

Die vorliegende Arbeit hat ferner gezeigt, daß keine anderen etablierten Prüfgrößen als die Statistik aus den quadratischen Formen Q_n und Q_e für die parametrische Analyse von hochdimensionalen Meßwiederholungen geeignet sind.

Begleitend zu den Statistiken wurden die beiden bekannten Schätzervarianten untersucht, die für die Freiheitsgradschätzung bei hochdimensionalen Daten im Behrens-Fisher-Problem geeignet sind. So haben sich in der vorliegenden Arbeit die Schätzer nach Brunner, Bathke und Placzek (2012) und Happ, Harrar und Bathke (2016) bei multivariater Normalverteilung als optimal und insbesondere als dimensionsstabil herausgestellt. Als robuste Schätzer können trotzdem die Schätzer auf Basis von U -Statistiken aus Becker (2010) und Harden (2013) betrachtet werden. Sie haben nur geringfügig schlechtere Varianz bei Normalverteilung, sind aber im elliptischen und additiv-linearen Verteilungsmodell dimensionsstabil, wenn dort nur die vierten Momente existieren.

Mit der Korrektur für das Einsetzen von B_n konnte die Behebbarkeit eines alten Problems des Huynh-Feldt-Tests gezeigt werden. Auf diese Weise kann möglicherweise auch in anderen Anwendungsfällen überwunden werden, stets auf Konsistenz und den Satz von Slutsky rekurren zu müssen, um Schätzer für unbekannte Störparameter einsetzen zu dürfen, sofern nur die ungefähre Varianz und Erwartungstreue eines Schätzers bekannt ist.

Die Korrektur selber kann durchaus noch verbessert werden. Beispielsweise kann etwas rigoroser ausgehend von

$$\begin{aligned}\alpha &= E\left(\mathbb{1}_{F > q_{1-\alpha}(b_1 b_n^{-1})}\right) \\ &= 1 - F\left(q_{1-\alpha}(b_1 b_n^{-1})\right)\end{aligned}$$

der vergrößerte Typ-I-Fehler durch Einsetzen des Schätzers B_n mit

$$\begin{aligned}
1 - E(F_F \circ q_{1-\alpha}(b_1 B_n^{-1})) &\approx 1 - F_F \circ q_{1-\alpha}(b_1 b_n^{-1}) - \\
&\quad \frac{1}{2} \text{Var}(B_n) \cdot \frac{\partial^2 (F_F \circ q_{1-\alpha}(b_1 b_n^{-1}))}{\partial b_n^2} \\
&= 1 - \alpha - \frac{1}{2} \text{Var}(B_n) \cdot \left(F_F''(q_{1-\alpha}(b_1 b_n^{-1})) \left(\frac{\partial q_{1-\alpha}}{\partial \nu_1} \frac{b_1}{b_n^2} \right)^2 \right. \\
&\quad \left. + F_F' \left(\frac{\partial^2 q_{1-\alpha}}{\partial \nu_1^2} \frac{b_1^2}{b_n^4} - 2 \frac{\partial q_{1-\alpha}}{\partial \nu_1} \frac{b_1}{b_n^3} \right) \right)
\end{aligned}$$

beschrieben werden. Diese läuft auf eine etwas andere Korrektur hinaus, die nicht untersucht wurde.

Weiterhin ist denkbar, das additiv-lineare und das elliptische Modell weiter zu fassen oder multivariate Verteilungen allgemein zu charakterisieren, in denen die robusten Spurschätzer dimensionsstabil bleiben. Eine gründlichere Möglichkeit könnte auch sein, die Abhängigkeiten durch Copulae zu modellieren, siehe zu dem Begriff (Nelsen (1999)). Die Frage wäre dann, durch welche Bedingungen an Copula und Randverteilungen der Verlust bei den Schätzern begrenzt wird, wenn er an Abweichungen vom vorgegebenen Signifikanzniveau gemessen wird und bei den Randverteilungen die vierten Momente beschränkt sind.

Diese Fragestellung könnte überleiten zu einer Anwendung der ANOVA-Typ-Statistik für hochdimensionale rangtransformierte Daten in der Nichtparametrik. Für die nichtparametrische Auswertung von Meßwiederholungen (siehe Brunner, Dette und Munk (1997)) und für die nichtparametrische Auswertung von verschiedenen Größen wie in Brunner, Munzel und Puri (2002) werden rangtransformierte Daten in die ANOVA-Typ-Statistik eingesetzt, deren unbekannte Parameter mit den im Hochdimensionalen konservativen Plug-in-Schätzern geschätzt werden. Dimensionsstabiles Schätzen auf der Grundlage rangtransformierter Daten hat bislang einer Lösung widerstanden, weil durch die Rangtransformation die Unabhängigkeit der Beobachtungsvektoren verlorengeht, aus denen die robusten Schätzer konstruiert wurden.

Für einen rigorosen Beweis der Robustheit wäre es wünschenswert gewesen, wenn eine Abschätzung der Art

$$|H(x) - G(x)| \leq c(x)$$

angewendet werden könnte, bei der H und G die Verteilungsfunktionen derselben Statistik sind, die mal mit, mal ohne normalverteilte Daten berechnet wird. Wenn es gelingt, eine Schranke für $c(x)$ zu finden, die von der Art der Nichtnormalverteilung abhängen, hätte man ein rigoroses Robustheitskriterium in dem Sinne, daß man abhängig von diesen Funktionalen garantieren kann, sich nicht zu weit vom vorgegebenen Typ-I-Fehler zu entfernen.

In der Literatur werden diese Abschätzungen manchmal „nichtgleichmäßiger Berry-Esséen-Abstand“ genannt (siehe z. B. Neammanee (2005), aber auch Akhiezer (1965),

L. H. Chen und Shao (2001) und Lindsay und Basak (2000)). Die Herausforderung ist, für die rechte Seite solcher Abschätzungen möglichst scharfe Grenzen zu finden, die von zugänglichen Funktionalen der multivariaten Nichtnormalverteilungen abhängen und nicht nur für die Verteilungen der Mittelwerte gelten sondern auch für Statistiken wie der ATS nutzbar sind.

Als weitere Fortentwicklung wäre eine Methode für (zumindest) vollständig zufällig fehlende Werte wünschenswert. Da das Konzept, die Statistiken mittels Übereinstimmung der ersten beiden Momente herzuleiten, für unbalancierte Stichproben gedacht ist, ist es logisch kein weiter Schritt, auch „Unbalanciertheit der Meßwiederholungen“ zuzulassen, wenn also bei manchen Versuchseinheiten einzelne Messungen fehlen. Q_n kann ohne weiteres bei fehlenden Werten ausgerechnet werden, nur mit etwas unhandlicherer Notation. Die Herleitung einer Entsprechung für Q_e mit möglichst geringer Varianz und die Konstruktion von passenden Spurschätzern wäre anschließend zu bewältigen. Unproblematisch wären die Spurschätzer, wenn je Zeitpunktpaar in jeder Stichprobe mindestens vier Versuchseinheiten vorliegen, doch dies wäre kaum zu erwarten.

Ein weiteres wichtiges Thema sind Globaltests, wenn nichtkommensurable Daten vorkommen. Haben alle Messungen an einer Versuchseinheit unterschiedliche physikalische Größen, dann sind Verteilungen für Prüfgrößen wie in Abschnitt 1.8.2 zu finden. In der Praxis kommen jedoch auch *Mischungen aus Meßwiederholungen und nichtkommensurablen Daten* vor. In der letzten Arbeit zum Thema nichtkommensurable Daten, Friedrich und Pauly (2018), wurde eine Mischung mit Meßwiederholungen noch ausgeschlossen, obwohl Mischungen sogar besonders leicht zu hoher Dimensionalität führen.

A Beweise

A.1 Bekannte Sätze

Satz 22. *Seien $\mathbf{A}, \mathbf{B}, \mathbf{C}$ konforme Matrizen und $x \in \mathbb{R}$. Dann gelten folgende Eigenschaften der Spur:*

1. *Linearität:* $\text{Sp}(x\mathbf{A}) = x\text{Sp}(\mathbf{A})$
2. *Invarianz unter zyklischen Vertauschungen:* $\text{Sp}(\mathbf{ABC}) = \text{Sp}(\mathbf{BCA})$
3. *Die Spur einer symmetrischen Matrix ist die Summe ihrer Eigenwerte.*
4. $\text{Sp}(\mathbf{AB}) = \mathbf{1}'_d (\mathbf{A} \# \mathbf{B}') \mathbf{1}_d$

Beweis. Die Aussagen 1, 2 und 4 finden sich in Brunner und Munzel 2013 in Satz B.14 auf Seite 265, die Aussage 3 in Satz B.34 auf Seite 270 derselben Publikation. \square

Satz 23. *[Satz von Lancaster] Sei \mathbf{Y} ein Zufallsvektor, dessen zweites Moment existiert, so daß $\text{Cov}(\mathbf{Y}) = \mathbf{\Sigma}$ und $E(\mathbf{Y}) = \boldsymbol{\mu}$ ist. Sei ferner $\mathbf{A} = \mathbf{A}'$. Dann ist*

$$E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = \text{Sp}(\mathbf{A}\mathbf{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

Beweis. Siehe (Mathai und Provost 1992), S. 56. \square

Satz 24 (Satz von Craig und Sakamoto). *Sei $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ und seien \mathbf{A} und \mathbf{B} positiv semidefinit und \mathbf{b} ein konstanter Vektor.*

1. *Dann ist $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ von $\mathbf{Y}'\mathbf{B}\mathbf{Y}$ stochastisch unabhängig, falls $\mathbf{B}\mathbf{\Sigma}\mathbf{A} = \mathbf{0}$ ist.*
2. *$\mathbf{A}\mathbf{X}$ und $\mathbf{b}'\mathbf{X}$ sind stochastisch unabhängig, falls $\mathbf{b}'\mathbf{\Sigma}\mathbf{A} = \mathbf{0}$ ist. Ohne Normalverteilung gilt hier zumindest Unkorreliertheit.*

Lemma 25. *(Bathke-Harrar-Lemma) Seien \mathbf{M}_A und \mathbf{M}_B symmetrische $n \times n$ Matrizen und sei $\mathcal{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ eine Zufallsmatrix mit unabhängig verteilten und zentrierten Zufallsvektoren \mathbf{Y}_k , deren Kovarianzmatrix jeweils $\mathbf{\Sigma}_k$ sei. Dann ist*

$$\text{Cov}(\text{Sp}(\mathcal{Y}\mathbf{M}_A\mathcal{Y}'), \text{Sp}(\mathcal{Y}\mathbf{M}_B\mathcal{Y}')) = \sum_{k=1}^n \sum_{k'=1}^n 2a_{kk'}b_{kk'}\text{Sp}(\mathbf{\Sigma}_k\mathbf{\Sigma}_{k'}) + \sum_{k=1}^n a_{kk}b_{kk}\kappa_4(\mathbf{Y}_k)$$

mit $\kappa_4(\mathbf{Y}_k) = E(\mathbf{Y}'_k\mathbf{Y}_k)^2 - 2\text{Sp}\mathbf{\Sigma}_k - \text{Sp}^2\mathbf{\Sigma}_k$.

Beweis. Siehe Bathke und Harrar (2008) \square

Satz 26 (Satz von Slutsky). *Es sei $(\mathbf{Y}_k)_{k=1, \dots, \infty}$ eine Folge von Zufallsvektoren mit $\mathbf{Y}_k \in \mathbb{R}^d$ und $\mathbf{Y}_k \rightarrow \mathbf{Y}_0 \in \mathbb{R}^d$ in Wahrscheinlichkeit und es sei f sei eine in \mathbf{Y}_0 stetige Abbildung. Dann konvergiert auch $f(\mathbf{Y}_k) \rightarrow f(\mathbf{Y}_0)$ in Wahrscheinlichkeit.*

Beweis. Siehe z. B. Lehmann (2004). \square

A.2 Beweise bis zum Kapitel 2 Robuste Mehrstichprobenstatistiken

A.2.1 Beweise über strukturierte quadratische Formen

Lemma 27. Sei $\mathbf{T} \in \mathbb{R}^{d \times d}$ symmetrisch und idempotent und $\mathbf{Y} \in \mathbb{R}^d$ eine Zufallsvariable, für die $\text{Var}(\mathbf{Y})$ existiert. Dann ist $E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})^2 \leq E(\mathbf{Y}'_{ik} \mathbf{Y}_{ik})^2$.

Beweis. Sei $\mathbf{Q} \in \mathbb{R}^{d \times d}$ eine orthogonale Matrix, so daß \mathbf{T} diagonalisiert wird zu

$$\mathbf{Q}' \mathbf{T} \mathbf{Q} = \mathbf{I}_{\text{rg}(\mathbf{T})} \oplus \mathbf{0}_{d-\text{rg}(\mathbf{T})}.$$

Wegen der Invarianz des Skalarprodukts unter orthogonalen Transformationen ist dann

$$\begin{aligned} E(\mathbf{Y}' \mathbf{T} \mathbf{Y})^2 &= E(\mathbf{Y}' \mathbf{Q}' \mathbf{T} \mathbf{Q} \mathbf{Y})^2 \\ &\leq E(\mathbf{Y}' (\mathbf{I}_{\text{rg}(\mathbf{T})} \oplus \mathbf{0}_{d-\text{rg}(\mathbf{T})}) \mathbf{Y} + \mathbf{Y}' (\mathbf{0}_{\text{rg}(\mathbf{T})} \oplus \mathbf{I}_{d-\text{rg}(\mathbf{T})}) \mathbf{Y}) \\ &= E(\mathbf{Y}' \mathbf{Y})^2. \end{aligned}$$

□

Satz 28. Die Erwartungswerte von Q_n und Q_e lauten

$$\begin{aligned} EQ_n &= \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i) + \boldsymbol{\mu}' \mathbf{T} \boldsymbol{\mu} \\ EQ_s &= \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i) \\ EQ_e &= \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i) \end{aligned}$$

Beweis. Als Abkürzung sei

$$\underline{\boldsymbol{\mu}} := (\mathbf{1}'_{n_1} \otimes \boldsymbol{\mu}'_1, \dots, \mathbf{1}'_{n_a} \otimes \boldsymbol{\mu}'_a)' = E(\mathbf{Y}) \quad (58)$$

der Erwartungswert des Vektors aller Beobachtungen hintereinander. Zur Anwendung kommt der Satz von Lancaster, der nichts anderes voraussetzt, als die Existenz von Erwartungswertvektor und Kovarianzmatrix, siehe z. B. Searle und Gruber (2017), S. 77.

$$\begin{aligned} EQ_n &= \text{Sp}(\mathbf{M}_n \boldsymbol{\Sigma}) + \underline{\boldsymbol{\mu}}' \mathbf{M}_n \underline{\boldsymbol{\mu}} = \sum_{i=1}^a n_i^{-2} t_{ii} \text{Sp}(\mathbf{J}_{n_i \times n_i} \otimes \boldsymbol{\Gamma}'_i \mathbf{T}_d \boldsymbol{\Gamma}_i) + \boldsymbol{\mu}' \mathbf{T} \boldsymbol{\mu} \\ &= \sum_{i=1}^a n_i^{-2} t_{ii} \text{Sp}(\mathbf{J}_{n_i \times n_i}) \text{Sp}(\boldsymbol{\Gamma}'_i \mathbf{T}_d \boldsymbol{\Gamma}_i) + \boldsymbol{\mu}' \mathbf{T} \boldsymbol{\mu} \end{aligned}$$

$$= \sum_{i=1}^a n_i^{-1} t_{ii} \text{Sp}(\Gamma_i' \mathbf{T}_d \Gamma_i) = \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp} \Sigma_i \mathbf{T}_d + \underline{\boldsymbol{\mu}}' \mathbf{T} \underline{\boldsymbol{\mu}}$$

$$\begin{aligned} EQ_e &= \text{Sp}(\mathbf{M}_e \boldsymbol{\Sigma}) + \underline{\boldsymbol{\mu}}' \mathbf{M}_e \underline{\boldsymbol{\mu}} = \sum_{i=1}^a \frac{t_{ii}}{n_i (n_i - 1)} \text{Sp}(\mathbf{P}_{n_i} \otimes \Gamma_i' \mathbf{T}_d \Gamma_i) + 0 \\ &= \sum_{i=1}^a \frac{t_{ii}}{n_i (n_i - 1)} \text{Sp} \mathbf{P}_{n_i} \cdot \text{Sp}(\Gamma_i' \mathbf{T}_d \Gamma_i) \\ &= \sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\Sigma_i \mathbf{T}_d) \end{aligned}$$

□

Zum Beweis von Satz 7 wird ein Lemma benötigt:

Lemma 29. *Seien $\mathbf{Y}_1, \dots, \mathbf{Y}_n \stackrel{iid}{\sim} F_\Theta$ Zufallsvariablen im \mathbb{R}^d , deren Verteilung über $\theta \in F \subset \mathbb{R}$ parametrisiert ist. Die geringste Varianz von allen unverzerrten Schätzern für θ wird von den Schätzern angenommen, die invariant unter Permutationen der n Zufallsvariablen sind.*

Beweis. Dies ist eine Verkürzung des Beweises von Theorem 5 aus (Halmos 1946) um die Teile, die mit der Eindeutigkeit des permutationsinvarianten Schätzers zu tun haben. Dieser Verzicht ermöglicht, auch Zufallsvariablen zu betrachten, die nicht reellwertig und nicht auf einem „finit abgeschlossen“ Maßraum definiert sind.

Sei $\hat{\theta}$ ein beliebiger, möglicherweise nicht permutationsinvariant erwartungstreuer Schätzer für θ . Bezeichne S_n die Menge aller Permutationen über n Elemente; es ist also $|S_n| = n!$. Dann ist

$$\hat{\theta}_* := \frac{1}{n!} \sum_{\tau \in S_n} \hat{\theta}(\tau(\mathbf{y}_1, \dots, \mathbf{y}_n))$$

ein permutationsinvarianter und ebenfalls unverzerrter Schätzer. Mit der Jensen'schen Ungleichung kann man sein zweites Moment abschätzen durch

$$E(\hat{\theta}_*^2) = E\left(\frac{1}{n!} \sum_{\tau \in S_n} \hat{\theta}(\tau(\mathbf{y}_1, \dots, \mathbf{y}_n))\right)^2 \leq \frac{1}{n!} \sum_{\tau \in S_n} E\left(\hat{\theta}(\tau(\mathbf{y}_1, \dots, \mathbf{y}_n))\right)^2 = E(\hat{\theta}^2).$$

Die letzte Gleichheit gilt, da wegen der unabhängig identischen Verteilung die Varianz von $\hat{\theta}$ unabhängig von der Permutation der Beobachtungen dieselbe ist. Wegen der Unverzerrtheit folgt daraus sofort die Behauptung über die Varianz. □

Beweis von Satz 7. Es genügt, nur für den Teil $\text{Sp} \mathbf{T}_d \hat{\boldsymbol{\Sigma}}_i$ nachzuweisen, daß es die beste unverzerrt schätzende positiv-semidefinite quadratische Form ist. Diese Eigenschaft vererbt sich dann auf $\sum_{i=1}^a \frac{t_{ii}}{n_i} \text{Sp}(\mathbf{T}_d \hat{\boldsymbol{\Sigma}}_i)$, weil die Stichproben voneinan-

der stochastisch unabhängig sind und deswegen die Varianz eines erwartungstreuen Schätzers sich wie der Erwartungswert in eine Summe zerlegen läßt.

Da wegen der Regularitätsbedingung im Rahmenmodell und Lemma 27 die Ungleichungen $E(\mathbf{Y}'_{ik}\mathbf{T}_d\mathbf{Y}_{ik})^2 \leq E(\mathbf{Y}'_{ik}\mathbf{Y}_{ik})^2 < \infty$ gelten, wird die zu untersuchende quadratische Form endliche Varianz haben.

Sei $Q_W := \mathbf{Y}'\mathbf{W}\mathbf{Y}$ eine unverzerrte quadratische Form mit symmetrisch positiv semidefiniter Matrix \mathbf{W} . Diese Matrix kann in $d \times d$ -Blöcke partitioniert werden:

$$Q_W = \sum_{k,l=1}^{n_i} \mathbf{Y}'_k \mathbf{W}_{kl} \mathbf{Y}_l$$

Damit Q_W die geringstmögliche Varianz aller unverzerrter Schätzer hat, muß Q_W laut vorhergehendem Lemma 29 invariant unter Permutationen der $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ sein. Dies ist erfüllt, falls $\mathbf{W}_{kk} = \mathbf{W}_{11}$ für alle $k = 1, \dots, n$ und $\mathbf{W}_{kl} = \mathbf{W}_{12}$ für alle $k \neq l$ ist.

Aus der Erwartungstreue für beliebige Erwartungswertvektoren $\boldsymbol{\mu}_i$ kann für \mathbf{W}_{11} und \mathbf{W}_{12} gefolgert werden, daß sie in

$$EQ_W = n_i (\text{Sp}(\mathbf{W}_{11}\boldsymbol{\Sigma}_i) + \boldsymbol{\mu}'_i \mathbf{W}_{11} \boldsymbol{\mu}_i) + n_i (n_i - 1) \boldsymbol{\mu}'_i \mathbf{W}_{12} \boldsymbol{\mu}_i$$

die Terme mit $\boldsymbol{\mu}_i$ verschwinden lassen und tatsächlich $n_i \text{Sp}(\mathbf{W}_{11}\boldsymbol{\Sigma}_i) = \text{Sp}(\mathbf{T}_d\boldsymbol{\Sigma}_i)$ für alle $\boldsymbol{\Sigma}_i$ ist. Letzteres wird erreicht durch $\mathbf{W}_{11} = n_i^{-2}\mathbf{T}_d$, ersteres durch $\mathbf{W}_{12} = n_i^{-2}(n_i - 1)^{-1}\mathbf{T}_d$. Insgesamt ist damit durch die Matrix $\mathbf{W} = \mathbf{P}_{n_i} \otimes \mathbf{T}_d$ eine unverzerrte positiv semidefinite quadratische Form als Schätzer von $n_i^{-1}\text{Sp}\mathbf{T}_d\boldsymbol{\Sigma}_i$ definiert, die die geringstmögliche Varianz unter allen Schätzern dieser Klasse annimmt. Mit diesem \mathbf{W} ist der Schätzer aber gerade $n_i^{-1}\text{Sp}\mathbf{T}_d\hat{\boldsymbol{\Sigma}}_i$. \square

Beweis von Satz 9. Mit \mathbf{M}_n wie in (13) und $\mathbf{M}_s = (\bigoplus_{i=1}^a s_i \mathbf{P}_{n_i} \otimes \mathbf{T}_d)$ als Abkürzung für die Matrix in (26) ist unter der Nullhypothese $\mathbf{M}_n \underline{\boldsymbol{\mu}} = \mathbf{0}$ und immer $\mathbf{M}_s \underline{\boldsymbol{\mu}} = \mathbf{0}$, wobei $\underline{\boldsymbol{\mu}} = E(\mathbf{Y})$ wie in (58) definiert sei. Daher hängen auch alle Momente der quadratischen Formen nicht mehr vom Erwartungswertvektor ab:

$$\begin{aligned} E(\mathbf{Y}'(\mathbf{M}_n \otimes \mathbf{T}_d)\mathbf{Y})^h &= E((\mathbf{Y}_0 + \boldsymbol{\mu})'(\mathbf{M}_n)(\mathbf{Y}_0 + \boldsymbol{\mu}))^h \\ &= E(\mathbf{Y}'_0 \mathbf{M}_n \mathbf{Y}_0 + 2\mathbf{Y}'_0 \mathbf{M}_n \boldsymbol{\mu} + \boldsymbol{\mu}' \mathbf{M}_n \boldsymbol{\mu})^h \\ &= E(\mathbf{Y}'_0 \mathbf{M}_n \mathbf{Y}_0)^h. \end{aligned}$$

Aus diesem Grund kann ohne Einschränkung der Allgemeinheit unterstellt werden, daß die Erwartungswertvektoren in jeder Stichprobe verschwinden, also $E(\mathbf{Y}) = \mathbf{0}$ ist.

Es soll das Lemma 25 auf Seite 77 angewendet werden. Dort sind die Daten als Matrix geschrieben, während im vorliegenden Satz die Daten im Vektor \mathbf{Y} angeordnet sind. Beides kann ineinander überführt werden. Im konkreten Fall wird benötigt, daß allgemein mit einem nd -Vektor $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_n)'$, einer $d \times n$ -Matrix $\mathcal{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)$, einer $n \times n$ -Matrix \mathbf{M} und einer idempotenten $d \times d$ -Matrix \mathbf{T} stets

$$\text{Sp}(\mathbf{T}\mathcal{Y}\mathbf{M}\mathcal{Y}'\mathbf{T}') = \mathbf{Y}'(\mathbf{M} \otimes \mathbf{T})\mathbf{Y}$$

gilt.

1. $\text{Var}Q_n$: Wenn die Beobachtungen k und k' der Stichprobe i und i' angehören, dann ist $m_{kk'} = t_{ii'}n_i^{-1}n_{i'}^{-1}$. Dieses kommt jeweils $n_i n_{i'}$ mal in der Doppelsumme vor. Es ergibt sich:

$$\begin{aligned} \text{Var}(Q_n) &= 2 \sum_{k=1}^n \sum_{k'=1}^n m_{kk'}^2 \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_k \mathbf{T}_d \boldsymbol{\Sigma}_{k'}) + \sum_{k=1}^n m_{kk}^2 \kappa_i(\mathbf{Y}_{i1}) \\ &= 2 \sum_{i=1}^a \sum_{i'=1}^a t_{ii'}^2 n_i^{-1} n_{i'}^{-1} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'}) + \sum_{i=1}^a t_{ii}^2 n_i^{-3} \kappa_i(\mathbf{Y}_{i1}) \\ &= 2 \sum_{i \neq i'}^a t_{ii'}^2 n_i^{-1} n_{i'}^{-1} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'}) + 2 \sum_{i=1}^a t_{ii}^2 \frac{n_i - 1}{n_i^3} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 \\ &\quad + \sum_i^a t_{ii} n_i^{-3} \kappa_i(\mathbf{Y}_{i1}) \end{aligned}$$

2. $\text{Var}(Q_s)$: Die Einträge der Matrix $\mathbf{M}_s = \bigoplus s_i \mathbf{P}_{n_i} \otimes \mathbf{T}_d$ haben die Form

$$m_{kk'} = \begin{cases} 0 & k, k' \text{ aus unterschiedlichen Stichproben} \\ -\frac{s_i}{n_i} & k, k' \text{ aus Stichprobe } i, \text{ aber } k \neq k' \\ \frac{s_i(n_i-1)}{n_i} & k = k' \end{cases}$$

Da die letzten beiden Fälle jeweils $n_i(n_i - 1)$ mal bzw. n_i mal auftreten, ergibt sich

$$\begin{aligned} \text{Var}(Q_s) &= 2 \sum_{k=1}^n \sum_{k'=1}^n m_{kk'}^2 \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_k \mathbf{T}_d \boldsymbol{\Sigma}_{k'}) \\ &\quad + \sum_{k=1}^n m_{kk}^2 \left(E(\mathbf{Y}'_k \mathbf{T}_d \mathbf{Y}_k)^2 - 2\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_k)^2 - \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_k) \right) \\ &= 2 \sum_{i=1}^a \left(n_i(n_i - 1) \frac{s_i^2}{n_i^2} + n_i \frac{s_i^2(n_i - 1)^2}{n_i^2} \right) \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 + \\ &\quad \sum_{i=1}^a n_i \frac{s_i^2(n_i - 1)^2}{n_i^2} \left(E(\mathbf{Y}'_{i1} \mathbf{T}_d \mathbf{Y}_{i1})^2 - 2\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_k)^2 - \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_k) \right) \\ &= 2 \sum_{i=1}^a (n_i - 1) s_i^2 \text{Sp} \boldsymbol{\Sigma}_i^2 \end{aligned}$$

$$+ \sum_{i=1}^a \frac{s_i^2 (n_i - 1)^2}{n_i} \left(E(\mathbf{Y}'_{i1} \mathbf{T}_d \mathbf{Y}_{i1})^2 - 2\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_k)^2 - \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_k) \right)$$

3. Für die Kovarianz der quadratischen Formen ergibt sich mit den Bemerkungen zu den Einträgen der Matrizen \mathbf{M}_s und \mathbf{M}_n :

$$\begin{aligned} \text{Cov}(Q_n, Q_s) &= 2 \sum_{i=1}^a \left(n_i (n_i - 1) \left(\frac{t_{ii} s_i}{n_i^2 n_i} \right) + n_i \left(\frac{t_i s_i (n_i - 1)}{n_i^2 n_i} \right) \right) \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 + \\ &\quad \sum_{i=1}^a n_i \frac{t_{ii} s_i (n_i - 1)}{n_i^2 n_i} \left(E(\mathbf{Y}'_{i1} \mathbf{Y}_{i1})^2 - 2\text{Sp}(\boldsymbol{\Sigma}_i^2) - \text{Sp}^2 \boldsymbol{\Sigma}_i \right) \\ &= \sum_{i=1}^a \frac{(n_i - 1) s_i t_{ii}}{n_i^2} \left(E(\mathbf{Y}'_{i1} \mathbf{Y}_{i1})^2 - 2\text{Sp}(\boldsymbol{\Sigma}_i^2) - \text{Sp}^2 \boldsymbol{\Sigma}_i \right) \end{aligned}$$

□

Beweis von Satz 10. Mit den Resultaten aus Satz 9 ist

$$\begin{aligned} \text{Var}_{H_0}(Q_n - Q_s) &= \text{Var}_{H_0}(Q_n) + \text{Var}(Q_s) - 2\text{Cov}_{H_0}(Q_n, Q_s) \\ &= 2 \left(\sum_i^a \sum_{i'}^a t_{ii'}^2 n_i^{-1} n_{i'}^{-1} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'}) + \sum_{i=1}^a \frac{t_{ii}^2}{n_i^3} \kappa_i(\mathbf{Y}_{i1}) \right. \\ &\quad \left. + \sum_{i=1}^a (n_i - 1) s_i^2 \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 + \sum_{i=1}^a \frac{s_i^2 (n_i - 1)^2}{n_i} \kappa_i(\mathbf{Y}_{i1}) \right. \\ &\quad \left. - \sum_{i=1}^a \sum_{i=1}^a \frac{t_{ii} s_i (n_i - 1)}{n_i^2} \kappa_i(\mathbf{Y}_{i1}) \right) \\ &= 2 \left(\sum_i^a \sum_{i'}^a t_{ii'}^2 n_i^{-1} n_{i'}^{-1} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'}) \right. \\ &\quad \left. + \sum_i^a (n_i - 1) s_i^2 \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 \right. \\ &\quad \left. + \sum_{i=1}^a \left(\frac{s_i^2 n_i^2 (n_i - 1)^2 + t_{ii}^2 - 2t_{ii} n_i s_i (n_i - 1)}{n_i^3} \right) \kappa_i(\mathbf{Y}_{i1}) \right). \end{aligned}$$

Es hängt $\text{Var}_{H_0}(Q_n - Q_s)$ genau dann nicht von Ausdrücken der Art $E(\mathbf{Y}'_{i1} \mathbf{T}_d \mathbf{Y}_{i1})^2$ mit $i = 1, \dots, a$ ab, falls $s_i^2 n_i^2 (n_i - 1)^2 + t_{ii}^2 - 2t_{ii} n_i s_i (n_i - 1) = 0$ ist für alle $i = 1, \dots, a$. Die Lösung dieser Gleichungen ist jeweils $s_i = \frac{t_{ii}}{n_i(n_i - 1)}$. □

A.2.2 Beweise über Approximationen und Verteilungen

Satz 30. Sei $h : \mathbb{R}^k \rightarrow \mathbb{R}$, $k = 1, 2$, zweimal stetig differenzierbar mit ersten beiden Ableitungen Dh und D^2h . Sei ferner \mathbf{X} eine Zufallsvariable im \mathbb{R}^k , deren erste vier

(Produkt-)Momente existieren, insbesondere sei $E(\mathbf{X}) = \boldsymbol{\xi}$. Falls dann bei $k = 1$ die Wahrscheinlichkeit $P(\{|X_1 - \xi_1| \geq \epsilon\})$ bzw. bei $k = 2$

$$P(\{|X_1 - \xi_1| \geq \epsilon_1\} \cup \{|X_2 - \xi_2| \geq \epsilon_2\})$$

klein ist, gelten die Taylor-Approximationen erster Ordnung

$$E(h(\mathbf{X})) \doteq h(\boldsymbol{\xi})$$

und

$$\text{Var}(h(\mathbf{X})) \doteq \sum_{i,j=1}^k \frac{\partial h(\boldsymbol{\xi})}{\partial \xi_i} \cdot \frac{\partial h(\boldsymbol{\xi})}{\partial \xi_j} \cdot \text{Cov}(X_i, X_j)$$

sowie die Taylorapproximation zweiter Ordnung

$$E(h(\mathbf{X})) \doteq h(\boldsymbol{\xi}) + \frac{1}{2} \sum_{i,j=1}^k \frac{\partial^2 h(\boldsymbol{\xi})}{\partial \xi_i \partial \xi_j} \cdot \text{Cov}(X_i, X_j).$$

Beweis. Falls $P(\{|X_1| \geq \epsilon\})$ bzw. $P(\{|X_1| \geq \epsilon_1\} \cup \{|X_2| \geq \epsilon_2\})$ klein sind, ist nur der Wert von h in diesen ϵ -Umgebungen um $\boldsymbol{\xi}$ relevant. Die Taylorentwicklung der Funktion h bis zu Termen zweiter Ordnung an der Stelle \mathbf{x} innerhalb einer solchen ϵ -Umgebung um $\boldsymbol{\xi}$ ist

$$h(\mathbf{x}) \doteq h(\boldsymbol{\xi}) + Dh(\boldsymbol{\xi})(\boldsymbol{\xi} - \mathbf{x}) + \frac{1}{2}(\boldsymbol{\xi} - \mathbf{x})' D^2 h(\boldsymbol{\xi})(\boldsymbol{\xi} - \mathbf{x}). \quad (59)$$

Man erhält nun die Näherung für $E(h(\mathbf{X}))$, indem man \mathbf{X} für \mathbf{x} einsetzt und auf beiden Seiten zum Erwartungswert aufintegriert; also erst

$$h(\mathbf{X}) \approx h(\boldsymbol{\xi}) + Dh(\boldsymbol{\xi})(\boldsymbol{\xi} - \mathbf{X}) + \frac{1}{2}(\boldsymbol{\xi} - \mathbf{X})' D^2 h(\boldsymbol{\xi})(\boldsymbol{\xi} - \mathbf{X}) \quad (60)$$

und dann

$$\begin{aligned} E(h(\mathbf{X})) &\approx h(\boldsymbol{\xi}) + Dh(\boldsymbol{\xi}) \underbrace{E(\boldsymbol{\xi} - \mathbf{X})}_{=0} + \frac{1}{2} E[(\boldsymbol{\xi} - \mathbf{X})' D^2 h(\boldsymbol{\xi})(\boldsymbol{\xi} - \mathbf{X})] \\ &= h(\boldsymbol{\xi}) + \frac{1}{2} \sum_{i,j=1}^k \frac{\partial^2 h(\boldsymbol{\xi})}{\partial \xi_i \partial \xi_j} \cdot E[(\xi_i - X_i)(\xi_j - X_j)] \\ &= h(\boldsymbol{\xi}) + \frac{1}{2} \sum_{i,j=1}^k \frac{\partial^2 h(\boldsymbol{\xi})}{\partial \xi_i \partial \xi_j} \cdot \text{Cov}(X_i, X_j). \end{aligned} \quad (61)$$

Die Taylorapproximation erster Ordnung des Erwartungswertes ergibt sich, indem der Summand mit den zweiten Ableitungen von Anfang an weggelassen wird und somit auch die Summe mit den Kovarianzen nicht mehr in Erscheinung tritt.

Die Varianzapproximation erster Ordnung erhält man, indem von der Taylorformel

erster Ordnung

$$h(\mathbf{X}) \approx h(\boldsymbol{\xi}) + Dh(\boldsymbol{\xi}) \cdot (\boldsymbol{\xi} - \mathbf{X})$$

die Erwartungswertapproximation erster Ordnung abgezogen wird, diese Differenz –es ist $Dh(\boldsymbol{\xi}) \cdot (\boldsymbol{\xi} - \mathbf{X})$ – quadriert wird und anschließend der Erwartungswert genommen wird:

$$\begin{aligned} \text{Var}(h(\mathbf{X})) &\approx E(Dh(\boldsymbol{\xi}) \cdot (\boldsymbol{\xi} - \mathbf{X}))^2 \\ &= E\left(\sum_{i=1}^k \frac{\partial h(\boldsymbol{\xi})}{\partial \xi_i} \cdot (\xi_i - X_i)\right)^2 \\ &= \sum_{i=1}^k \frac{\partial h(\boldsymbol{\xi})}{\partial \xi_i} \frac{\partial h(\boldsymbol{\xi})}{\partial \xi_{i'}} \cdot \text{Cov}(X_i, X_{i'}) \end{aligned}$$

□

Korollar 31. Für $Q_s > 0$ fast sicher gelten die Erwartungswertapproximation erster Ordnung

$$E\left(\frac{Q_n}{Q_s}\right) \doteq \frac{E(Q_n)}{E(Q_s)}$$

und die Varianzapproximationen erster Ordnung

$$\text{Var}\left(\frac{Q_n}{Q_s}\right) \doteq \left(\frac{EQ_n}{EQ_s}\right)^2 \left(\frac{\text{Var}(Q_n)}{E^2Q_n} + \frac{\text{Var}(Q_s)}{E^2Q_s} - \frac{2\text{Cov}(Q_n, Q_s)}{EQ_nEQ_s}\right)$$

sowie mit $B_1^{(i)} > 0$ fast sicher $(B_2^{(i)})^3 (B_1^{(i)})^{-1}$

$$\begin{aligned} \text{Var}\left(\left(B_2^{(i)}\right)^3 \left(B_1^{(i)}\right)^{-1}\right) &\doteq \frac{E^6(B_2^{(i)})}{E^2(B_1^{(i)})} \cdot \left(9 \frac{\text{Var}(B_2^{(i)})}{E^2B_2^{(i)}} + \frac{\text{Var}(B_1^{(i)})}{E^2B_1^{(i)}}\right. \\ &\quad \left.- 3 \frac{\text{Cov}(B_1^{(i)}, B_2^{(i)})}{EB_1^{(i)}EB_2^{(i)}}\right) \end{aligned}$$

Beweis von Satz 11 (Approximation durch F -Verteilung). Sei $q_F(1 - \alpha | \nu_1, \nu_2)$, die Quantilfunktion der F -Verteilung am $1 - \alpha$ -Quantil mit den Freiheitsgraden ν_1 und ν_2 und $q_{\chi^2}(1 - \alpha | \nu_1)$ die Quantilfunktion der $\chi^2_{\nu_2}$ -Verteilung an demselben Quantil. Nach Abramowitz und Stegun (1964) Nr. 26.6.11, ist

$$q_F(1 - \alpha | \nu_1, \nu_2) \xrightarrow{\nu_2 \rightarrow \infty} \nu_1^{-1} q_{\chi^2}(1 - \alpha | \nu_1).$$

Nach der Definition von f_e ist außerdem $f_e \rightarrow \infty$, falls $n_i \rightarrow \infty$ für mindestens ein

$i = 1, \dots, a$. und in dem Falle auch

$$f = \frac{b_1}{b_n + b_e} = \frac{1}{f_n^{-1} + f_e^{-1}} \xrightarrow{n_i \rightarrow \infty} f_n.$$

Daraus folgt $q_F(1 - \alpha | f, f_e) - q_F(1 - \alpha | f_n, f_e) \xrightarrow{n_i \rightarrow \infty} 0$.

Für Kombinationen aus kleinen f_n und f_e muß auf numerische Berechnungsergebnisse verwiesen werden, wie sie in Abbildung 3 auf Seite 33 graphisch dargestellt sind. Offensichtlich sind für finite Freiheitsgrade die $F(f_n, f_e)$ -Quantile größer als die $f^{-1}\chi_f^2$ -Quantile. \square

Satz 32. *Das elliptische Modell (5) ist invariant unter regulären linearen Transformationen $\mathbf{A} \in GL_d$ in dem Sinne, daß wenn \mathbf{Y}_{ik} eine elliptische Verteilung mit Erwartungswert $\boldsymbol{\mu}_i$ und Kovarianzmatrix $\boldsymbol{\Sigma}_i$ hat, $\mathbf{A}\mathbf{Y}_{ik}$ dieselbe Verteilung mit Erwartungswert $\mathbf{A}\boldsymbol{\mu}_i$ und Kovarianzmatrix $\mathbf{A}\boldsymbol{\Sigma}_i\mathbf{A}'$ hat.*

Das additiv-lineare Modell (3) hat im Allgemeinen nicht diese Invarianzeigenschaft.

Beweis. Für das elliptische Modell folgt die Behauptung sofort aus der Definition 4.

Für die Aussage zum additiv-linearen Modell genügt ein Gegenbeispiel. Dafür folge Z_{ikj} jeweils einer Rademacherverteilung, also einer Verteilung mit $P(Z_{ijk} = 1) = P(Z_{ijk} = -1) = \frac{1}{2}$ und die Z_{ik1}, \dots, Z_{ikd} seien unabhängig. Dann ist $E\mathbf{Z}_{ik} = \mathbf{0}$ und $\text{Cov}(\mathbf{Z}_{ik}) = \mathbf{I}_d$ und die Verteilung von $\mathbf{Y}_{ik} = \mathbf{Z}_{ik}$ hat eine positive Wahrscheinlichkeitsmasse auf der Menge $\{-1, 1\}^d$. Sei nun

$$\mathbf{U} = \begin{bmatrix} \cos \frac{\pi}{6} & -\sin \frac{\pi}{6} \\ \sin \frac{\pi}{6} & \cos \frac{\pi}{6} \end{bmatrix} \oplus \mathbf{I}_{d-2}.$$

Dann haben \mathbf{Y}_{ik} und $\mathbf{U}\mathbf{Y}_{ik}$ jeweils denselben Erwartungswert und dieselbe Kovarianzmatrix, weil \mathbf{U} orthogonal ist. $\mathbf{U}\mathbf{Y}_{ik}$ hat hingegen nur auf der Menge

$$\left\{ \frac{1 + \sqrt{3}}{2}, \frac{\sqrt{3} - 1}{2}, \frac{1 - \sqrt{3}}{2}, -\frac{1 + \sqrt{3}}{2} \right\}^2 \otimes \{-1, 1\}^d$$

positive Wahrscheinlichkeitsmasse. Wären alle additiv-linearen Modelle invariant unter regulären linearen Transformationen, dann hätte mit der Übereinstimmung in Erwartungswert und Kovarianzmatrix auch Gleichheit der Verteilungen folgen müssen. \square

A.3 Beweise zum Kapitel 3 Freiheitsgradschätzungen

A.3.1 Sätze über beide Modelle zugleich

Lemma 33. *Seien \mathbf{Y}_{ik} unabhängige, in Stichprobe i identisch verteilte Zufallsvektoren, deren Kovarianzmatrix existiert. Dann gelten mit $\text{Cov}(\mathbf{Y}_{ik}) = \boldsymbol{\Sigma}_i$, $i = 1, \dots, n_i$,*

$k = 1, \dots, n_i$ folgende Gleichungen, falls die Vektoren zentriert sind, d. h. $E(\mathbf{Y}_{ik}) = \mathbf{0}$:

1. $E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})(\mathbf{Y}'_{i's} \mathbf{T}_d \mathbf{Y}_{i's}) = \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i) \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_{i'})$ für $k \neq s$ oder $i \neq i'$ mit $s = 1, \dots, n_{i'}$.
2. $E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's})^2 = \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})^2$ für $k \neq s$ oder $i \neq i'$.
3. $E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{il} \mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{il} \mathbf{T}_d \mathbf{Y}_{i't} \mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i't}) = \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})^2$ für $k \neq l$ und $t \neq s$ und $i \neq i'$ oder $\neq k, s, t, l$ mit $t = 1, \dots, n_{i'}$, $l = 1, \dots, n_i$.
4. $E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's})^2 (\mathbf{Y}'_{il} \mathbf{T}_d \mathbf{Y}_{i't})^2 = \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})$ für $k \neq l$ und $t \neq s$ und $i \neq i'$ oder $\neq k, s, t, l$.

Auch, wenn die Vektoren nicht zentriert sind, gilt:

5. Für $k \neq l$ und $t \neq s$ und $i \neq i'$ oder $\neq k, s, t, l$

$$\begin{aligned} E(\mathbf{Y}_{ik} - \mathbf{Y}_{il})' \mathbf{T}_d (\mathbf{Y}_{ik} - \mathbf{Y}_{il}) (\mathbf{Y}_{i's} - \mathbf{Y}_{i't})' \mathbf{T}_d (\mathbf{Y}_{i's} - \mathbf{Y}_{i't}) \\ = 4 \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i) \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_{i'}) \end{aligned}$$

6. $E((\mathbf{Y}_{ik} - \mathbf{Y}_{il})' \mathbf{T}_d (\mathbf{Y}_{i's} - \mathbf{Y}_{i't}))^2 = 4 \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})$ für $k \neq l$ und $t \neq s$ und $i \neq i'$ oder $\neq k, s, t, l$.

Beweis. Die Gleichungen können allesamt mit dem Satz von Lancaster bewiesen werden.

- 1.

$$\begin{aligned} E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})(\mathbf{Y}'_{i's} \mathbf{T}_d \mathbf{Y}_{i's}) &= E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik}) E(\mathbf{Y}'_{i's} \mathbf{T}_d \mathbf{Y}_{i's}) \\ &= \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i) \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_{i'}) \end{aligned}$$

- 2.

$$\begin{aligned} E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's})^2 &= \text{Sp} E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's})^2 \\ &= E \text{Sp}(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's}) \\ &= E \text{Sp}(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{i's} \mathbf{T}_d \mathbf{Y}_{ik}) \\ &= E \text{Sp}(\mathbf{T}_d \mathbf{Y}_{ik} \mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{i's}) \\ &= \text{Sp} E(\mathbf{T}_d \mathbf{Y}_{ik} \mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{i's}) \\ &= \text{Sp} E(\mathbf{T}_d \mathbf{Y}_{ik} \mathbf{Y}'_{ik}) E(\mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{i's}) \\ &= \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'}) \end{aligned}$$

3. Es werden dieselben Rechenschritte wie zuvor gemacht, nur bei einem längern Term:

$$\begin{aligned}
E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{il} \mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{il} \mathbf{T}_d \mathbf{Y}_{i't} \mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i't}) &= \text{Sp} E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{il} \mathbf{T}_d \mathbf{Y}_{i's} \\
&\quad \mathbf{Y}'_{il} \mathbf{T}_d \mathbf{Y}_{i't} \mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i't}) \\
&= \text{Sp} (E(\mathbf{T}_d \mathbf{Y}_{ik} \mathbf{Y}'_{ik}) E(\mathbf{T}_d \mathbf{Y}_{i's} \mathbf{Y}'_{i's}) \\
&\quad E(\mathbf{T}_d \mathbf{Y}_{il} \mathbf{Y}'_{il}) E(\mathbf{T}_d \mathbf{Y}_{i't} \mathbf{Y}'_{i't})) \\
&= \text{Sp} (\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_i)^2
\end{aligned}$$

4.

$$\begin{aligned}
E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's})^2 (\mathbf{Y}'_{il} \mathbf{T}_d \mathbf{Y}_{i't})^2 &= E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's})^2 E(\mathbf{Y}'_{il} \mathbf{T}_d \mathbf{Y}_{i't})^2 \\
&= \text{Sp}^2 (\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_i)
\end{aligned}$$

5. Dieser Beweis ist identisch zu 1., nur mit $\text{Cov}(\mathbf{Y}_{ik} - \mathbf{Y}_{il}) = 2\boldsymbol{\Sigma}_i$ statt $\boldsymbol{\Sigma}_i$.
6. Dieser Beweis ist aus demselben Grund identisch zu 2.

□

A.3.2 Sätze für das additiv-lineare Modell

Folgendes Lemma wird für die Beweise benötigt, die das additiv-lineare Modell benutzen:

Lemma 34. Seien \mathbf{Y}_{i1} mit $i = 1, 2$ zentrierte Zufallsvektoren im additiv-linearen Modell (3) und (4) auf S. 11, insbesondere $\text{Cov}(\mathbf{Y}_{i1}) = \boldsymbol{\Sigma}_1$,

und sei \mathbf{T}_d eine symmetrische und idempotente Matrix.

1. Dann ist mit der Abkürzung $\mathbf{S}_{11} = [s_{jj'}]_{j,j'=1,\dots,r_1} = \boldsymbol{\Gamma}'_i \mathbf{T}_d \boldsymbol{\Gamma}_i$

$$E(\mathbf{Y}'_{i1} \mathbf{T}_d \mathbf{Y}_{i1})^2 = 2\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_1)^2 + \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_1) + \sum_{j=1}^{r_1} (\delta_j - 3) s_{jj}.$$

2. Mit der Abkürzung $\mathbf{S}_{12} := \boldsymbol{\Gamma}'_1 \mathbf{T}'_d \mathbf{T}_d \boldsymbol{\Gamma}_2 = [s_{jj'}]_{\substack{j=1,\dots,r_1 \\ j'=1,\dots,r_2}}$ ist ferner

$$\begin{aligned}
E(\mathbf{Y}'_{11} \mathbf{T}_d \mathbf{Y}_{21})^4 &= 6\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_1 \mathbf{T}_d \boldsymbol{\Sigma}_2)^2 + 3\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_1 \mathbf{T}_d \boldsymbol{\Sigma}_2) \\
&\quad + \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} s_{j_1 j_2}^4 (\delta_{1j_1} \delta_{2j_2} - 9)
\end{aligned}$$

$$\begin{aligned}
& + 3 \sum_{j_1=1}^{r_1} \sum_{j_2 \neq j'_2}^{r_2} s_{j_1 j_2}^2 s_{j_1 j'_2}^2 (\delta_{1j_1} - 3) \\
& + 3 \sum_{j_1 \neq j'_1}^{r_1} \sum_{j_2=1}^{r_2} s_{j_1 j_2}^2 s_{j'_1 j_2}^2 (\delta_{2j_2} - 3).
\end{aligned}$$

Beweis. Im Beweis werden die Indices von \mathbf{Y}_{i1} fortgelassen, wo sie redundant sind.

$$\begin{aligned}
E(\mathbf{Y}'\mathbf{T}_d\mathbf{Y})^2 &= E(\mathbf{Z}'\mathbf{S}_{ii}\mathbf{Z})^2 \\
&= E\left(\sum_{j,j'} s_{jj'} Z_j Z_{j'}\right)^2 \\
&= \sum_{j,j',j'',j'''} s_{jj'} s_{j''j'''} E(Z_j Z_{j'} Z_{j''} Z_{j'''}) \\
&= \sum_{j=1}^r \delta_j s_{jj}^2 + 2 \sum_{j \neq j'}^r s_{jj'}^2 + \sum_{j \neq j'}^r s_{jj} s_{j'j'} \\
&= \sum_{j=1}^r (\delta_j - 3) s_{jj}^2 + 2 \sum_{j,j'}^r s_{jj'}^2 + \sum_{j,j'}^r s_{jj} s_{j'j'} \\
&= \sum_{j=1}^r (\delta_j - 3) s_{jj}^2 + 2\text{Sp}(\mathbf{\Gamma}\mathbf{T}_d\mathbf{\Gamma}')^2 + \text{Sp}^2(\mathbf{\Gamma}\mathbf{T}_d\mathbf{\Gamma}') \\
&= \sum_{j=1}^r (\delta_j - 3) s_{jj}^2 + 2\text{Sp}(\mathbf{T}_d\mathbf{\Sigma})^2 + \text{Sp}^2(\mathbf{T}_d\mathbf{\Sigma})
\end{aligned}$$

In der folgenden Rechnung für den zweiten Teil wird man die Terme

$$\text{Sp}(\mathbf{T}_d\mathbf{\Sigma}_1\mathbf{T}_d\mathbf{\Sigma}_2)^2 = \text{Sp}(\mathbf{S}_{12}\mathbf{S}'_{12})^2 = \sum_{j_1, j'_1=1}^{r_1} \sum_{j_2, j'_2=1}^{r_2} s_{j_1 j_2} s_{j_1 j'_2} s_{j'_1 j_2} s_{j'_1 j'_2}$$

und

$$\text{Sp}^2(\mathbf{T}_d\mathbf{\Sigma}_1\mathbf{T}_d\mathbf{\Sigma}_2) = \text{Sp}^2(\mathbf{S}_{12}\mathbf{S}'_{12}) = \left(\sum_{j_1}^{r_1} \sum_{j_2}^{r_2} s_{j_1 j_2}^2\right)^2 = \sum_{j_1, j'_1}^{r_1} \sum_{j_2, j'_2}^{r_2} s_{j_1 j_2}^2 s_{j'_1 j'_2}^2$$

wiedererkennen.

$$\begin{aligned}
E(\mathbf{Y}'_1\mathbf{T}_d\mathbf{Y}_2)^4 &= E(\mathbf{Z}'_1\mathbf{S}_{12}\mathbf{Z}_2)^4 \\
&= E\left(\sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} s_{j_1 j_2} Z_{1j_1} Z_{2j_2}\right)^4 \\
&= \sum_{j_1, j'_1, j''_1, j'''_1=1}^{r_1} \sum_{j_2, j'_2, j''_2, j'''_2=1}^{r_2} s_{j_1 j_2} s_{j'_1 j'_2} s_{j''_1 j''_2} s_{j'''_1 j'''_2} \\
&\quad \cdot E(Z_{1j_1} Z_{1j'_1} Z_{1j''_1} Z_{1j'''_1}) E(Z_{2j_2} Z_{2j'_2} Z_{2j''_2} Z_{2j'''_2})
\end{aligned}$$

$$\begin{aligned}
&= \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} s_{j_1 j_2}^4 \delta_{1j_1} \delta_{2j_2} \\
&\quad + 3 \sum_{j_1=1}^{r_1} \sum_{j_2 \neq j_2'}^{r_2} s_{j_1 j_2}^2 s_{j_1 j_2'}^2 \delta_{1j_1} \\
&\quad + 3 \sum_{j_1 \neq j_1'}^{r_1} \sum_{j_2=1}^{r_2} s_{j_1 j_2}^2 s_{j_1' j_2}^2 \delta_{2j_2} \\
&\quad + 6 \sum_{j_1 \neq j_1'}^{r_1} \sum_{j_2 \neq j_2'}^{r_2} s_{j_1 j_2} s_{j_1 j_2'} s_{j_1' j_2} s_{j_1' j_2'} \\
&\quad + 3 \sum_{j_1 \neq j_1'}^{r_1} \sum_{j_2 \neq j_2'}^{r_2} s_{j_1 j_2}^2 s_{j_1' j_2'}^2 \\
&= 6\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_1 \mathbf{T}_d \boldsymbol{\Sigma}_2)^2 + 3\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_1 \mathbf{T}_d \boldsymbol{\Sigma}_2) \\
&\quad + \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} s_{j_1 j_2}^4 (\delta_{1j_1} \delta_{2j_2} - 9) \\
&\quad + 3 \sum_{j_1=1}^{r_1} \sum_{j_2 \neq j_2'}^{r_2} s_{j_1 j_2}^2 s_{j_1 j_2'}^2 (\delta_{1j_1} - 3) \\
&\quad + 3 \sum_{j_1 \neq j_1'}^{r_1} \sum_{j_2=1}^{r_2} s_{j_1 j_2}^2 s_{j_1' j_2}^2 (\delta_{2j_2} - 3)
\end{aligned}$$

□

Lemma 35 (Dimensionsstabilität im additiv-linearen Modell). *Es gelte das additiv-lineare Modell (3)-(4) und es sei abgekürzt*

$$\tilde{\delta}_i = \max_{j=1, \dots, d} \delta_{ij}.$$

Dann ist mit $A_{(kl)}^{(i)}$ und $A_{(kl)(st)}^{(i)(i')}$ definiert wie in (36) bzw. (37)

1. $\text{Var} \left(A_{(kl)}^{(i)} A_{(st)}^{(i')} \right) \leq 484 \max \left\{ \tilde{\delta}_i, \tilde{\delta}_{i'} \right\} \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i) \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_{i'})$
2. $\text{Var} \left(\left(A_{(kl)(st)}^{(i)(i')} \right)^2 \right) \leq 4 \left(\max \left\{ 3, \tilde{\delta}_i, \tilde{\delta}_{i'} \right\} + 3 \right)^2 \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})^2.$

Beweis. Die Differenz $(\mathbf{Y}_{ik} - \mathbf{Y}_{il})$ kann geschrieben werden als

$$(\mathbf{Y}_{ik} - \mathbf{Y}_{il}) = \left(\sqrt{2} \boldsymbol{\Gamma}_i \right) \mathbf{Z}_{i(kl)} \quad (62)$$

mit einem Zufallsvektor

$$\mathbf{Z}_{i(kl)} := \frac{1}{\sqrt{2}} (Z_{ik1} - Z_{il1}, \dots, Z_{ikr_i} - Z_{ilr_i})'$$

aus unabhängigen Komponenten, für die $\text{Cov}(\mathbf{Z}_{i(kl)}) = \mathbf{I}_{r_i}$ ist und das vierte Moment $E(Z_{i(kl)j})^4 = \frac{1}{2}(\delta_{ij} - 3)$. Damit folgt auch $(\mathbf{Y}_{ik} - \mathbf{Y}_{il})$ einem additiv-linearen Modell mit Kovarianzmatrix $\sqrt{2}\mathbf{\Gamma}_i (\sqrt{2}\mathbf{\Gamma}_i)' = 2\mathbf{\Sigma}_i$. Es sei die Abkürzung

$$\mathbf{S}_i := \mathbf{\Gamma}'_i \mathbf{T}'_d \mathbf{T}_d \mathbf{\Gamma}_i = [s_{jj'}]_{j,j'=1,\dots,r_i}$$

vereinbart. Man erhält damit aus Lemma 34.1

$$\begin{aligned} E\left(A_{(kl)}^{(i)}\right)^2 &= 4\text{Sp}^2 \mathbf{S}_i + 8\text{Sp}(\mathbf{S}_i^2) + \sum_{j=1}^{r_i} \frac{1}{2}(\delta_{ij} - 3) s_{jj}^2 \\ &\leq 4\text{Sp}^2 \mathbf{S}_i + 8\text{Sp}(\mathbf{S}_i^2) + \frac{1}{2}(\tilde{\delta}_i - 3) \sum_{j=1}^{r_i} s_{jj}^2 \\ &\leq 4\text{Sp}^2 \mathbf{S}_i + 8\text{Sp}(\mathbf{S}_i^2) + \frac{1}{2}(\tilde{\delta}_i - 3) \sum_{j=1}^{r_i} \sum_{j'=1}^{r_i} s_{jj'}^2 \\ &= 4\text{Sp}^2 \mathbf{S}_i + \frac{1}{2}(\tilde{\delta}_i + 13) \text{Sp}(\mathbf{S}_i^2) \\ &\leq \frac{1}{2}(\tilde{\delta}_i + 21) \text{Sp}^2 \mathbf{S}_i \end{aligned}$$

So kann $\text{Var}\left(A_{(kl)}^{(i)} A_{(st)}^{(i')}\right)$ abgeschätzt werden durch

$$\begin{aligned} \text{Var}\left(A_{(kl)}^{(i)} A_{(st)}^{(i')}\right) &\leq E\left(A_{(kl)}^{(i)} A_{(st)}^{(i')}\right)^2 \\ &= E\left(A_{(kl)}^{(i)}\right)^2 E\left(A_{(st)}^{(i')}\right)^2 \\ &\leq \frac{1}{4}(\tilde{\delta}_i + 21)(\tilde{\delta}_{i'} + 21) \\ &\quad \cdot \text{Sp}^2(\mathbf{T}_d \mathbf{\Sigma}_{i'}) \text{Sp}^2(\mathbf{T}_d \mathbf{\Sigma}_i) \\ &= \frac{1}{4}(\tilde{\delta}_i \tilde{\delta}_{i'} + 21(\tilde{\delta}_i + \tilde{\delta}_{i'}) + 441) \text{Sp}^2(\mathbf{T}_d \mathbf{\Sigma}_{i'}) \text{Sp}^2(\mathbf{T}_d \mathbf{\Sigma}_i) \\ &\leq 484 \tilde{\delta}_i \tilde{\delta}_{i'} \text{Sp}^2(\mathbf{T}_d \mathbf{\Sigma}_{i'}) \text{Sp}^2(\mathbf{T}_d \mathbf{\Sigma}_i) \end{aligned}$$

Analog können die Differenzen $(\mathbf{Y}_{ik} - \mathbf{Y}_{il})$ und $(\mathbf{Y}_{i's} - \mathbf{Y}_{i't})$ in Lemma 34.2 eingesetzt werden. \square

A.3.3 Sätze für das elliptische Modell

Lemma 36. *Seien $\mathbf{Y}_{ik}, \mathbf{Y}_{il}$ unabhängig voneinander identisch, zentriert und elliptisch verteilt nach dem elliptischen Modell aus Abschnitt 1.6.4, wobei $\tilde{r}_i = \text{rg}(\mathbf{T}_d \mathbf{\Sigma}_i)$ ist. Entsprechend seien auch $\mathbf{Y}_{i's}, \mathbf{Y}_{i't}$ unabhängig identisch zentriert und elliptisch verteilt. Dann gelten*

1. $E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})^2 = \frac{\delta_i}{\tilde{r}_i(\tilde{r}_i+2)} (2\text{Sp}(\mathbf{T}_d \mathbf{\Sigma}_i)^2 + \text{Sp}^2(\mathbf{T}_d \mathbf{\Sigma}_i))$
2. $E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{i's})^4 = \frac{6\text{Sp}(\mathbf{T}_d \mathbf{\Sigma}_i \mathbf{T}_d \mathbf{\Sigma}_{i'})^2 + 2\text{Sp}^2(\mathbf{\Sigma}_i \mathbf{\Sigma}_{i'})}{\tilde{r}_i(\tilde{r}_i+2)\tilde{r}_{i'}(\tilde{r}_{i'}+2)} EW_{ik}^4 EW_{i's}^4$

Beweis. Sei $\tilde{\Gamma}_i$ eine $d \times \tilde{r}_i$ -Matrix, so daß $\tilde{\Gamma}_i \tilde{\Gamma}_i' = \mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d'$ ist und $\text{rg}(\tilde{\Gamma}_i) = \text{rg}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d') = \tilde{r}_i$ ist. Es sei daran erinnert, daß damit $E(W_{ik})^2 = \tilde{r}_i \geq 1$ ist.

1. Der Beweis ist am einfachsten ausgehend von der multivariaten Normalverteilung zu führen, die ein Spezialfall des elliptischen Modells ist, siehe den ähnlichen Ansatz bei Mathai, Provost und Hayakawa (1995), S. 109. Bei Normalverteilung von $\mathbf{T}_d \mathbf{Y}_{ik}$ ist demzufolge die Verteilung von $W_{ik} \sim \sqrt{\chi_d^2}$, also $\mathbf{T}_d \mathbf{Y}_{ik} = W_{ik} \tilde{\Gamma}_i \mathbf{U}_{ik}$. Sei deshalb $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, so daß $\mathbf{T}_d \mathbf{Y}_{ik} = \tilde{\Gamma}_i \mathbf{Z}$ geschrieben werden kann. Dann ist –siehe etwa Mathai und Provost (1992) S. 53–

$$\begin{aligned} E_{\mathcal{N}}(\mathbf{Y}_{ik}' \mathbf{T}_d \mathbf{Y}_{ik})^2 &= E_{\mathcal{N}}(\mathbf{Z}' \tilde{\Gamma}_i' \tilde{\Gamma}_i \mathbf{Z})^2 \\ &= 2\text{Sp}(\tilde{\Gamma}_i' \tilde{\Gamma}_i)^2 + \text{Sp}^2(\tilde{\Gamma}_i' \tilde{\Gamma}_i) \\ &= 2\text{Sp}(\boldsymbol{\Sigma}_i)^2 + \text{Sp}^2 \boldsymbol{\Sigma}_i \end{aligned}$$

Weil in $\tilde{\Gamma}_i \mathbf{Z} = \mathbf{T}_d \mathbf{Y}_{ik}$ die Kürzungsregel auf $\tilde{\Gamma}_i$ angewendet werden kann, gilt auch $\mathbf{Z} = W_{ik} \mathbf{U}_{ik}$. Durch die Unabhängigkeit von W_{ik} und \mathbf{U}_{ik} im elliptischen Modell und durch die Linearität der Spur kann anschließend EW_{ik}^4 als Faktor isoliert werden:

$$\begin{aligned} E(\mathbf{Y}_{ik}' \mathbf{Y}_{ik})^2 &= EW_{ik}^4 E(\mathbf{U}_{ik}' \tilde{\Gamma}_i' \tilde{\Gamma}_i \mathbf{U}_{ik})^2 \\ &= \frac{EW_{ik}^4}{\tilde{r}_i(\tilde{r}_i + 2)} E_{\mathcal{N}}(\mathbf{Z}' \tilde{\Gamma}_i' \tilde{\Gamma}_i \mathbf{Z})^2 \\ &= \frac{\delta_i}{\tilde{r}_i(\tilde{r}_i + 2)} (\text{Sp}^2 \boldsymbol{\Sigma}_i + 2\text{Sp}(\boldsymbol{\Sigma}_i^2)) \end{aligned}$$

2. Der Beweis beginnt wieder mit der Normalverteilung von \mathbf{Y}_{ik} und $\mathbf{Y}_{i's}$. Es sei also $\mathbf{Z}_i, \mathbf{Z}_{i'} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, damit $\tilde{\Gamma}_i \mathbf{Z}_i = \mathbf{T}_d \mathbf{Y}_{ik}$ und $\tilde{\Gamma}_{i'} \mathbf{Z}_{i'} = \mathbf{T}_d \mathbf{Y}_{i's}$ geschrieben werden können. Die Bilinearform kann bei Normalverteilung geschrieben werden als quadratische Form durch

$$\mathbf{Y}_{ik}' \mathbf{Y}_{i's} \stackrel{\mathcal{N}}{=} \frac{1}{2} (\mathbf{Z}_i \quad \mathbf{Z}_{i'})' \begin{pmatrix} \mathbf{0} & \tilde{\Gamma}_i' \tilde{\Gamma}_{i'} \\ \tilde{\Gamma}_{i'}' \tilde{\Gamma}_i & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{Z}_i \\ \mathbf{Z}_{i'} \end{pmatrix},$$

und damit ist –wiederum nach Mathai und Provost (1992), S. 53–,

$$E_{\mathcal{N}}(\mathbf{Z}_i' \tilde{\Gamma}_i' \tilde{\Gamma}_{i'} \mathbf{Z}_{i'})^4 = 6\text{Sp}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i'})^2 + 2\text{Sp}^2(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i'})$$

Wie im ersten Teil ist damit im Übergang zu elliptisch verteilten \mathbf{Y}_{ik} und $\mathbf{Y}_{i's}$

$$\begin{aligned} E(\mathbf{Y}_{ik}' \mathbf{Y}_{i's})^4 &= E(\mathbf{U}_{ik}' \tilde{\Gamma}_i' \tilde{\Gamma}_{i'} \mathbf{U}_{i's})^4 EW_{ik}^4 EW_{i's}^4 \\ &= \frac{EW_{ik}^4 EW_{i's}^4}{\tilde{r}_i(\tilde{r}_i + 2) \tilde{r}_{i'}(\tilde{r}_{i'} + 2)} E_{\mathcal{N}}(\mathbf{Z}_i' \tilde{\Gamma}_i' \tilde{\Gamma}_{i'} \mathbf{Z}_{i'})^4 \end{aligned}$$

$$= \frac{6\text{Sp}(\Sigma_i \Sigma_{i'})^2 + 2\text{Sp}^2(\Sigma_i \Sigma_{i'})}{\tilde{r}_i(\tilde{r}_i + 2)\tilde{r}_{i'}(\tilde{r}_{i'} + 2)} \delta_i \delta_{i'}$$

□

Lemma 37 (Dimensionsstabilität im elliptischen Modell). *Es folgen \mathbf{Y}_{ik} , \mathbf{Y}_{il} und $\mathbf{Y}_{i's}$, $\mathbf{Y}_{i't}$ dem elliptischen Modell (5)-(6). Dann ist*

1. $\text{Var}(A_{i(k,l)} A_{i'(s,t)}) \leq 144 \max\{\delta_i, 1\} \max\{\delta_{i'}, 1\} \text{Sp}^2(\Sigma_i) \cdot \text{Sp}^2(\Sigma_{i'})$
2. $\text{Var}(A_{i(k,l)i'(s,t)}^2) \leq 49 \cdot \max\{\delta_i \delta_{i'}, 1\} \text{Sp}^2(\Sigma_i \Sigma_{i'})$

Beweis. Weil nicht so leicht zugänglich ist, wie und mit welchen Parametern die Differenz zweier elliptisch verteilter Zufallsvariablen verteilt ist, müssen $E(A_{i(k,l)} A_{i'(s,t)})^2$ und $E(A_{i(k,l)i'(s,t)}^4)$ ausmultipliziert werden. Durch die Differenzenbildung $\mathbf{Y}_{ik} - \mathbf{Y}_{il}$ kann ohne Einschränkung dabei $\boldsymbol{\mu}_i = \boldsymbol{\mu}_{i'} = \mathbf{0}$ angenommen werden. Wie in den Beweisen zuvor wird die Notation von \mathbf{T}_d unterdrückt.

1. Nach oben abgeschätzt wird $\text{Var}(A_{i(k,l)} A_{i'(s,t)})$ mittels der ersten vier Aussagen von Lemma 36 durch

$$\begin{aligned} E(A_{i(k,l)})^2 &= E(\mathbf{Y}'_{ik} \mathbf{Y}_{ik} - 2\mathbf{Y}'_{ik} \mathbf{Y}_{il} + \mathbf{Y}'_{il} \mathbf{Y}_{il})^2 \\ &= 2E(\mathbf{Y}'_{ik} \mathbf{Y}_{ik})^2 + 2E(\mathbf{Y}'_{ik} \mathbf{Y}_{ik} \mathbf{Y}'_{il} \mathbf{Y}_{il}) \\ &\quad + 4E(\mathbf{Y}'_{ik} \mathbf{Y}_{il})^2 - 8E(\mathbf{Y}'_{ik} \mathbf{Y}_{ik} \mathbf{Y}'_{ik} \mathbf{Y}_{il}) \\ &\leq 2\delta_i (2\text{Sp}\Sigma_i^2 + \text{Sp}^2\Sigma_i) + 2\text{Sp}^2\Sigma_i \\ &\quad + 4\text{Sp}(\Sigma_i^2) \\ &\leq 12 \max\{\delta_i, 1\} \text{Sp}^2\Sigma_i \end{aligned}$$

Wegen der Unabhängigkeit von $A_{i(k,l)}$ und $A_{i'(s,t)}$ ist

$$E(A_{i(k,l)} A_{i'(s,t)})^2 \leq 144 \max\{\delta_i, 1\} \max\{\delta_{i'}, 1\} \text{Sp}^2\Sigma_i \text{Sp}^2\Sigma_{i'}$$

2. Nach oben abgeschätzt wird $\text{Var}(A_{i(k,l)i'(s,t)}^2)$ durch

$$E(A_{i(k,l)i'(s,t)}^4) = E(\mathbf{Y}'_{ik} \mathbf{Y}_{i's} - \mathbf{Y}'_{ik} \mathbf{Y}_{i't} - \mathbf{Y}'_{il} \mathbf{Y}_{i's} - \mathbf{Y}'_{il} \mathbf{Y}_{i't})^4$$

$$\begin{aligned}
&= 4E(\mathbf{Y}'_{ik}\mathbf{Y}_{i's})^4 \\
&\quad - 12E(\mathbf{Y}'_{ik}\mathbf{Y}_{i's})^2(\mathbf{Y}'_{ik}\mathbf{Y}_{i't})^2 \\
&\quad - 12E(\mathbf{Y}'_{ik}\mathbf{Y}_{i's})^2(\mathbf{Y}'_{il}\mathbf{Y}_{i's})^2 \\
&\quad + 12E(\mathbf{Y}'_{ik}\mathbf{Y}_{i's})^2 E(\mathbf{Y}'_{il}\mathbf{Y}_{i't})^2 \\
&\quad + 24E(\mathbf{Y}'_{ik}\mathbf{Y}_{i's}\mathbf{Y}'_{il}\mathbf{Y}_{i's}\mathbf{Y}'_{il}\mathbf{Y}_{i't}\mathbf{Y}'_{ik}\mathbf{Y}_{i't}) \\
&\quad + 144E(\mathbf{Y}'_{ik}\mathbf{Y}_{i's})^2(\mathbf{Y}'_{ik}\mathbf{Y}_{i't})(\mathbf{Y}'_{il}\mathbf{Y}_{i's}) \\
&\quad + 48E(\mathbf{Y}'_{ik}\mathbf{Y}_{i's})^3(\mathbf{Y}'_{il}\mathbf{Y}_{i't}) \\
&\leq 4\delta_i\delta_{i'}(6\text{Sp}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'})^2 + 3\text{Sp}^2(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'})) \\
&\quad + 12\text{Sp}^2(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'}) + \text{Sp}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'})^2 \\
&\leq 49 \cdot \max\{\delta_i\delta_{i'}, 1\} \text{Sp}^2(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_{i'})
\end{aligned}$$

□

A.3.4 Zusammensetzen der Dimensionsstabilitätsbeweise

Statt folgendem Lemma für die Varianzabschätzung von U -Statistiken könnte zwar auch die Literatur zitiert werden, siehe etwa Lemma A aus Abschnitt 5.2.1 in Serfling (1980). Allerdings wird dort von U -Statistiken mit symmetrischen Kernen ausgegangen, wenn also die Kerne invariant unter Permutationen ihrer Argumente sind. Die Schätzer in der vorliegenden Arbeit sind aber nicht unter allen Permutationen invariant. Eine Symmetrisierung könnte z. B. bei $B_2^{(i)}$ durch

$$A_{i(k,l,r,s)}^2 := A_{i(kl)i(st)}^2 + A_{i(ks)i(lt)}^2 + A_{i(kt)i(sl)}^2$$

vorgenommen werden, aber der Aufwand, hiervon die Varianz abzuschätzen, wäre größer als der Beweis für folgendes Lemma.

Lemma 38. (*Varianz von U -Statistiken*)

1. Seien Y_1, \dots, Y_n iiv. Zufallsvariablen und U_n eine U -Statistik mit einem nicht-symmetrischen Kern $h = h(y_1, \dots, y_{2m})$, $2m \leq n$, also

$$U_n = \frac{(n-m)!}{n!} \sum_{\neq k_1, \dots, k_m} h(Y_{k_1}, \dots, Y_{k_m}),$$

dessen zweites Moment existiert. Dann ist

$$\text{Var}(U_n) = \text{Var}(h(Y_1, \dots, Y_m)) \cdot o(1)$$

2. Seien $\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_n}, \mathbf{Y}_{i'_1}, \dots, \mathbf{Y}_{i'_n}$ unabhängige Beobachtungen aus zwei ver-

schiedenen Stichproben und

$$U_{ii'} = \begin{cases} \frac{(n_i - 2m)!}{n_i!} \sum_{\neq k_1, \dots, k_{2m}} h(\mathbf{Y}_{ik_1}, \dots, \mathbf{Y}_{ik_{2m}}) & i = i' \\ \frac{(n_i - m)!(n_{i'} - m)!}{n_i!n_{i'}!} \sum_{\neq k_1, \dots, k_m} \sum_{\neq l_1, \dots, l_m} h(\mathbf{Y}_{ik_1}, \dots, \mathbf{Y}_{ik_m}, \mathbf{Y}_{i'l_1}, \dots, \mathbf{Y}_{i'l_m}) & i \neq i'. \end{cases}$$

Dann gilt bei $i \neq i'$ jeweils

$$\begin{aligned} \text{Cov}(U_{ii}, U_{ii'}) &= \sqrt{\text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{i(2m)}))} \\ &\quad \sqrt{\text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{im}, \mathbf{Y}_{i'1}, \dots, \mathbf{Y}_{i'm}))} \cdot o(1) \end{aligned}$$

und

$$\begin{aligned} \text{Cov}(U_{ii'}, U_{ii''}) &= \sqrt{\text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{im}, \mathbf{Y}_{i'1}, \dots, \mathbf{Y}_{i'm}))} \\ &\quad \sqrt{\text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{im}, \mathbf{Y}_{i''1}, \dots, \mathbf{Y}_{i''m}))} \cdot o(1). \end{aligned}$$

Beweis. Unabhängig sind $h(X_{k_1}, \dots, X_{k_{2m}})$ und $h(X_{l_1}, \dots, X_{l_{2m}})$, sofern nur die Indexmengen $\{k_1, \dots, k_{2m}\}$ und $\{l_1, \dots, l_{2m}\}$ disjunkt sind. Daher verschwinden in $n(n-1) \cdot \dots \cdot (n-2m+1)$ Fällen die Kovarianzen zwischen beiden Ausdrücken. Die nicht verschwindenden Kovarianzen werden durch die Cauchy-Schwarz'sche Ungleichung

$$\text{Cov}(X_1, X_2) \leq \sqrt{\text{Var}(X_1) \text{Var}(X_2)}$$

abgeschätzt. Deswegen ist mit

$$\begin{aligned} \text{Var}(U_n) &= \frac{((n-2m)!)^2}{(n!)^2} \sum_{\neq k_1, \dots, k_{2m}} \sum_{\neq l_1, \dots, l_{2m}} \text{Cov}(h(\mathbf{Y}_{k_1}, \dots, \mathbf{Y}_{k_{2m}}), h(\mathbf{Y}_{l_1}, \dots, \mathbf{Y}_{l_{2m}})) \\ &\leq \frac{n(n-1) \cdot \dots \cdot (n-m+1) - (n-m) \cdot \dots \cdot (n-2m+1)}{n(n-1) \cdot \dots \cdot (n-m+1)} \\ &\quad \text{Var}(h(\mathbf{Y}_1, \dots, \mathbf{Y}_m)) \\ &= o(1) \cdot \text{Var}(h(\mathbf{Y}_1, \dots, \mathbf{Y}_{2m})) \end{aligned}$$

die erste Aussage bewiesen.

Der Beweis der zweiten Aussage geht genauso, da sich die zusätzlichen Faktoren mit $n_{i'}$ und $n_{i''}$ kürzen:

$$\begin{aligned} \text{Cov}(U_{ii}, U_{ii'}) &= \frac{(n_i - 2m)!(n_i - m)!(n_{i'} - m)!}{(n_i!)^2 n_{i'}!} \sum_{\neq k_1, \dots, k_m} \sum_{\neq k'_1, \dots, k'_m} \sum_{\neq k''_1, \dots, k''_m} \\ &\quad \text{Cov}(h(\mathbf{Y}_{ik_1}, \dots, \mathbf{Y}_{ik_{2m}}), h(\mathbf{Y}_{ik_1}, \dots, \mathbf{Y}_{ik_m}, \mathbf{Y}_{i'k'_1}, \dots, \mathbf{Y}_{i'k'_m})) \end{aligned}$$

$$\leq \left(1 - \frac{(n_i - 2m)(n_i - 3m + 1)}{n_i(n_i - m + 1)}\right) \cdot \sqrt{\text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{i2m})) \cdot \text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{im}, \mathbf{Y}_{i'1}, \dots, \mathbf{Y}_{i'm}))}$$

und

$$\begin{aligned} \text{Cov}(U_{ii'}, U_{ii''}) &= \frac{((n_i - m)!)^2 (n_{i'} - m)! (n_{i''} - m)!}{(n_i!)^2 n_{i'}! n_{i''}!} \sum_{\neq k_1, \dots, k_m}^{n_i} \sum_{\neq k'_1, \dots, k'_m}^{n_{i'}} \sum_{\neq k''_1, \dots, k''_m}^{n_{i''}} \sum_{\neq k'''_1, \dots, k'''_m}^{n_i} \\ &\text{Cov}(h(\mathbf{Y}_{ik_1}, \dots, \mathbf{Y}_{ik_m}, \mathbf{Y}_{i'k'_1}, \dots, \mathbf{Y}_{i'k'_m}), h(\mathbf{Y}_{ik'''_1}, \dots, \mathbf{Y}_{ik'''_m}, \mathbf{Y}_{i''k''_1}, \dots, \mathbf{Y}_{i''k''_m})) \\ &\leq \left(1 - \frac{(n_i - m) \cdot \dots \cdot (n_i - 2m + 1)}{n_i \cdot \dots \cdot (n_i - m + 1)}\right) \cdot \\ &\quad \sqrt{\text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{im}, \mathbf{Y}_{i'1}, \dots, \mathbf{Y}_{i'm}))} \cdot \\ &\quad \sqrt{\text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{im}, \mathbf{Y}_{i''1}, \dots, \mathbf{Y}_{i''m}))} \\ &= o(1) \cdot \sqrt{\text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{im}, \mathbf{Y}_{i'1}, \dots, \mathbf{Y}_{i'm}))} \cdot \\ &\quad \sqrt{\text{Var}(h(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{im}, \mathbf{Y}_{i''1}, \dots, \mathbf{Y}_{i''m}))} \end{aligned}$$

□

Beweis von Satz 14 auf Seite 49. $B_1^{(i)}$, $B_2^{(i)}$, $B_1^{(i,i')}$ und $B_2^{(i,i')}$ sind U -Statistiken. Wegen Lemma (38) sind sie also dimensionsstabil und konsistent, sofern die Varianz der Kerne nicht von d abhängen. Die Varianzabschätzungen aus Lemma (35) für das additiv-lineare Modell und Lemma 37 für das elliptische Modell zeigen, daß das tatsächlich der Fall ist.

Bleibt zu zeigen, daß B_1 , B_n und B_e dimensionsstabil sind. Weil B_1 in der Darstellung 42 eine Summe aus $B_1^{(i)}$ - und $B_1^{(i)(i')}$ -Termen ist, ist $\text{Var}(B_1)$ die Summe aus den (Ko-)Varianzen dieser Summanden. Diese können nach Lemma 38 gliedweise abgeschätzt werden. Anschließend kommt die erste Aussage von Lemma 35 bzw. Lemma 36 zur Anwendung. Zum Schluß werden positive gemischte Terme ergänzt.

$$\begin{aligned} \text{Var}(B_1) &= \sum_{i=1}^a \frac{t_{ii}^4}{n_i^4} \text{Var}(B_1^{(i)}) + 2 \sum_{i \neq i'} \frac{t_{ii}^2 t_{i'i'} t_{i'i''}}{n_i^2 n_i n_{i'}} \text{Cov}(B_1^{(i)}, B_1^{(i)(i')}) \\ &\quad + 2 \sum_{\neq i, i', i''} \frac{t_{ii}^2 t_{i'i'} t_{i''i''}}{n_i^2 n_{i'} n_{i''}} \text{Cov}(B_1^{(i)(i')}, B_1^{(i)(i'')}) \end{aligned}$$

$$\begin{aligned}
&= o(1) \left(\sum_{i=1}^a \frac{t_{ii}^4}{n_i^4} \text{Var} \left(A_{(12)}^{(i)} A_{(34)}^{(i)} \right) \right. \\
&\quad + 2 \sum_{i \neq i'}^a \frac{t_{ii}^2}{n_i^2} \frac{t_{ii'} t_{i'i'}}{n_i n_{i'}} \sqrt{\text{Var} \left(A_{(12)}^{(i)} A_{(34)}^{(i)} \right)} \sqrt{\text{Var} \left(A_{12}^{(i)} A_{12}^{(i')} \right)} \\
&\quad \left. + 2 \sum_{\substack{i, i', i'' \\ \neq i, i', i''}}^a \frac{t_{ii}^2 t_{i'i'} t_{i''i''}}{n_i^2 n_{i'} n_{i''}} \sqrt{\text{Var} \left(A_{12}^{(i)} A_{12}^{(i')} \right)} \sqrt{\text{Var} \left(A_{12}^{(i)} A_{12}^{(i'')} \right)} \right) \\
&= o(1) \tilde{\delta} \left(\sum_{i=1}^a \frac{t_{ii}^4}{n_i^4} \text{Sp}^4 \left(\mathbf{T}_d \boldsymbol{\Sigma}_i \right) \right. \\
&\quad + 2 \sum_{i \neq i'}^a \frac{t_{ii}^2}{n_i^2} \frac{t_{ii'} t_{i'i'}}{n_i n_{i'}} \text{Sp}^3 \left(\mathbf{T}_d \boldsymbol{\Sigma}_i \right) \text{Sp} \left(\mathbf{T}_d \boldsymbol{\Sigma}_{i'} \right) \\
&\quad \left. + 2 \sum_{\substack{i, i', i'' \\ \neq i, i', i''}}^a \frac{t_{ii}^2 t_{i'i'} t_{i''i''}}{n_i^2 n_{i'} n_{i''}} \text{Sp}^2 \left(\mathbf{T}_d \boldsymbol{\Sigma}_i \right) \text{Sp} \left(\mathbf{T}_d \boldsymbol{\Sigma}_{i'} \right) \text{Sp} \left(\mathbf{T}_d \boldsymbol{\Sigma}_{i''} \right) \right) \\
&\leq o(1) \tilde{\delta} \sum_{i, i', i'', i'''}^a \prod_{j \in \{i, i', i'', i'''\}} \frac{t_{jj}}{n_j} \text{Sp} \left(\mathbf{T}_d \boldsymbol{\Sigma}_j \right) \\
&= o(1) \cdot E(B_1)^2
\end{aligned}$$

Die Abschätzungen für die Varianzen von B_n , B_e und B_2 benutzen die zweite Aussage von Lemma 35, sind ansonsten aber identisch. Exemplarisch:

$$\begin{aligned}
\text{Var} (B_n) &= \sum_{i=1}^a \frac{t_{ii}^4}{n_i^4} \text{Var} \left(B_2^{(i)} \right) + 2 \sum_{i \neq i'}^a \frac{t_{ii}^2}{n_i^2} \frac{t_{ii'}}{n_i n_{i'}} \text{Cov} \left(B_2^{(i)}, B_2^{(i')} \right) \\
&\quad + 2 \sum_{\substack{i, i', i'' \\ \neq i, i', i''}}^a \frac{t_{ii'}^2 t_{i''i''}}{n_i^2 n_{i'} n_{i''}} \text{Cov} \left(B_2^{(i')} , B_2^{(i'')} \right) \\
&= o(1) \left(\sum_{i=1}^a \frac{t_{ii}^4}{n_i^4} \text{Var} \left(A_{(12)(34)}^{(i)} \right) \right. \\
&\quad + 2 \sum_{i \neq i'}^a \frac{t_{ii}^2 t_{ii'}}{n_i^3 n_{i'}} \sqrt{\text{Var} \left(A_{(12)(34)}^{(i)} \right)} \sqrt{\text{Var} \left(A_{(12)(12)}^{(i')} \right)} \\
&\quad \left. + 2 \sum_{\substack{i, i', i'' \\ \neq i, i', i''}}^a \frac{t_{ii'}^2 t_{i''i''}}{n_i^2 n_{i'} n_{i''}} \sqrt{\text{Var} \left(A_{(12)(12)}^{(i')} \right)} \sqrt{\text{Var} \left(A_{(12)(12)}^{(i'')} \right)} \right) \\
&= o(1) \tilde{\delta} \left(\sum_{i=1}^a \frac{t_{ii}^4}{n_i^4} \text{Sp}^2 \left(\mathbf{T}_d \boldsymbol{\Sigma}_i \right)^2 \right. \\
&\quad + 2 \sum_{i \neq i'}^a \frac{t_{ii}^2 t_{ii'}}{n_i^3 n_{i'}} \text{Sp} \left(\mathbf{T}_d \boldsymbol{\Sigma}_i \right)^2 \text{Sp} \left(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'} \right) \\
&\quad \left. + 2 \sum_{\substack{i, i', i'' \\ \neq i, i', i''}}^a \frac{t_{ii'}^2 t_{i''i''}}{n_i^2 n_{i'} n_{i''}} \text{Sp} \left(\mathbf{T}_d \boldsymbol{\Sigma}_i \right)^2 \text{Sp} \left(\mathbf{T}_d \boldsymbol{\Sigma}_{i'} \mathbf{T}_d \boldsymbol{\Sigma}_{i''} \right) \right)
\end{aligned}$$

$$\begin{aligned}
&\leq o(1) \tilde{\delta} \sum_{i,i',i'',i'''}^a \frac{t_{ii'}^2 t_{i''i'''}^2}{n_i n_{i'} n_{i''} n_{i'''}} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'}) \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_{i''} \mathbf{T}_d \boldsymbol{\Sigma}_{i'''}) \\
&= o(1) \cdot E(B_n)^2
\end{aligned}$$

□

Beweis von Satz 15 auf Seite 49. Mit den Erwartungswerten $E(B_1) = b_1$ und $E(B_n) = b_n$ wird von der Taylorapproximation für die Varianz ausgegangen, analog zu Korollar 31:

$$\text{Var}\left(\frac{B_1}{B_n}\right) \approx \frac{b_1^2}{b_n^2} \left(\frac{\text{Var}(B_1)}{b_1^2} + \frac{\text{Var}(B_n)}{b_n^2} - \frac{2\text{Cov}(B_1, B_n)}{b_1 b_n} \right)$$

Die Ausdrücke $\frac{\text{Var}(B_1)}{b_1^2}$ und $\frac{\text{Var}(B_n)}{b_n^2}$ hängen wegen der Dimensionsstabilität von B_1 und B_n nicht von d ab. Dasselbe gilt für

$$\frac{\text{Cov}(B_1, B_n)}{b_1 b_n} \leq \sqrt{\frac{\text{Var}(B_1)}{b_1} \frac{\text{Var}(B_n)}{b_n}}.$$

Die Konsistenz folgt aus der Konsistenz von B_n und B_1 und der Stetigkeit der Quotientenbildung. □

A.3.5 Modifizierte Huynh-Feldt- und Plug-in-Schätzer

Beweis von Satz 16 auf Seite 51. Der Beweis beruht auf dem Satz von Lehmann-Scheffé (Theorem 5.1 in Lehmann und Scheffé (1950)). Die empirische Kovarianzmatrix ist eine vollständige und suffiziente Statistik für unbekanntes $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ (siehe Anderson (1984), S. 77f). Die Huynh-Feldt-Schätzer sind nach Konstruktion im normalverteilten Modell unverzerrt und Funktionen der empirischen Kovarianzmatrix. Daher haben sie nach Lehmann-Scheffé unter allen unverzerrten Schätzern im normalverteilten Modell die kleinste Varianz, welche freilich noch nicht endlich sein muß.

Die Existenz der dimensionsstabilen Spurschätzer aus Abschnitt 3.3 beweist, daß diese kleinste Varianz tatsächlich endlich ist und zugleich nicht von d abhängt. Damit müssen die Huynh-Feldt-Schätzer $H_1^{(i)}$ und $H_2^{(i)}$ also ebenfalls dimensionsstabil sein.

Die näherungsweise Dimensionsstabilität in dem Sinne, daß die Taylor-Approximation der Varianz nicht vom Freiheitsgradschätzer abhängt, beweist man analog zu Satz 15. □

Beweis von Satz 17. Wenn $\mathbf{Y}_{ik}^{\mathcal{N}}$ ein Zufallsvektor mit $\text{Cov}(\mathbf{Y}_{ik}) = \text{Cov}(\mathbf{Y}_{ik}^{\mathcal{N}})$ ist, dann weisen die Lemmata 34.1 und 36.1 ζ_i als

$$\zeta_i = E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik}) - E\left(\mathbf{Y}'_{ik}{}^{\mathcal{N}} \mathbf{T}_d \mathbf{Y}_{ik}^{\mathcal{N}}\right)$$

aus. Dies und die Nummern 1,2 und 4 aus Lemma 33 werden im Folgenden benutzt.

Ausgegangen wird von der bekannten Tatsache, daß die empirische Kovarianzmatrix auch als U -Statistik geschrieben werden kann durch

$$\hat{\Sigma}_i = \frac{1}{2n_i(n_i-1)} \sum_{k,k'=1}^{n_i} (\mathbf{Y}_{ik} - \mathbf{Y}_{ik'}) (\mathbf{Y}_{ik} - \mathbf{Y}_{ik'})'.$$

Ferner kann ohne Einschränkung unterstellt werden, daß $\mathbf{T}_d \mathbf{Y}_{ik} = \mathbf{0}$ ist, da die Kovarianzmatrix lokationsinvariant ist. Wegen der Invarianz der Spur unter zyklischen Vertauschungen kann durch die Umformung

$$\text{Sp}(\mathbf{Y}_k \mathbf{Y}'_k \mathbf{Y}_{k'} \mathbf{Y}'_{k'}) = \text{Sp}(\mathbf{Y}'_{k'} \mathbf{Y}_k \mathbf{Y}'_k \mathbf{Y}_{k'}) = (\mathbf{Y}'_{k'} \mathbf{Y}_k)^2$$

die Abkürzung $A_{i(k,k')} = (\mathbf{Y}_{ik} - \mathbf{Y}_{ik'}) \mathbf{T}_d (\mathbf{Y}_{ik} - \mathbf{Y}_{ik'})$ aus Gleichung (36) verwendet werden. Damit ergibt sich

$$\begin{aligned} E\text{Sp}^2(\mathbf{T}_d \hat{\Sigma}_i) &= \frac{1}{4n_i^2(n_i-1)^2} \sum_{\neq k,k',k'',k'''} E(A_{i(k,k')} A_{i(k'',k''')})^2 \\ &= \frac{1}{n_i(n_i-1)} \left(EA_{i(k,k')}^2 \right. \\ &\quad \left. + 2(n_i-2) E(A_{i(k,k')} A_{i(k,k'')}) \right. \\ &\quad \left. + \frac{1}{2}(n_i-2)(n_i-3) E(A_{i(k,k')} A_{i(k'',k''')}) \right) \\ &= \frac{1}{2n_i(n_i-1)} \left(E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})^2 + 4E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik'})^2 + E^2(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik}) \right. \\ &\quad \left. + 2(n_i-2) \left(E(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik})^2 + 3E^2(\mathbf{Y}'_{ik} \mathbf{T}_d \mathbf{Y}_{ik}) \right) \right. \\ &\quad \left. + 2(n_i-2)(n_i-3) \text{Sp}^2(\mathbf{T}_d \Sigma_i) \right) \\ &= \frac{1}{2n_i(n_i-1)} \left(2(2\text{Sp}(\mathbf{T}_d \Sigma_i))^2 + \text{Sp}^2(\mathbf{T}_d \Sigma_i) + \zeta_i \right) + 4\text{Sp}(\mathbf{T}_d \Sigma_i)^2 + 2\text{Sp}^2(\mathbf{T}_d \Sigma_i) \\ &\quad + 2(n_i-2) \left(2\text{Sp}(\mathbf{T}_d \Sigma_i)^2 + 4\text{Sp}^2(\mathbf{T}_d \Sigma_i) + \zeta_i \right) \\ &\quad + 2(n_i-2)(n_i-3) \text{Sp}^2(\mathbf{T}_d \Sigma_i) \\ &= \frac{1}{n_i(n_i-1)} \left(n_i \zeta_i + n_i(n_i-1) \text{Sp}^2(\mathbf{T}_d \Sigma_i) + 2n_i \text{Sp}(\mathbf{T}_d \Sigma_i)^2 \right) \\ &= \frac{\zeta_i}{n_i-1} + \frac{2}{n_i-1} \text{Sp}(\mathbf{T}_d \Sigma_i)^2 + \text{Sp}^2(\mathbf{T}_d \Sigma_i) \end{aligned}$$

Der Erwartungswert des anderen Plug-in-Schätzers ist, wenn mit denselben Rechen-schritten und der Abkürzung

$$A_{i(k,k')i(k'',k''')} = (\mathbf{Y}_{ik} - \mathbf{Y}_{ik'})' \mathbf{T}_d (\mathbf{Y}_{ik''} - \mathbf{Y}_{ik'''})$$

aus Gleichung (37) gearbeitet wird,

$$\begin{aligned}
E\text{Sp}\left(\mathbf{T}_d\hat{\Sigma}_i\right)^2 &= \frac{1}{4n_i^2(n_i-1)^2} \sum_{k,k',k'',k'''=1}^{n_i} EA_{i(k,k')i(k'',k''')}^2 \\
&= \frac{1}{4n_i^2(n_i-1)^2} \left(2n_i(n_i-1)EA_{i(1,2)i(1,2)}^2\right. \\
&\quad + 4n_i(n_i-1)(n_i-2)EA_{i(1,2)i(1,3)}^2 \\
&\quad \left.+ n_i(n_i-1)(n_i-2)(n_i-3)EA_{i(1,2)i(3,4)}^2\right) \\
&= \frac{1}{2n_i(n_i-1)} \left(\left(2E(\mathbf{Y}'_{ik}\mathbf{Y}_{ik})^2 + 2E^2(\mathbf{Y}'_{ik}\mathbf{Y}_{ik}) + 4E(\mathbf{Y}'_{ik}\mathbf{Y}_{ik'})^2\right)\right. \\
&\quad \left.+ 2(n_i-2)\left(3E(\mathbf{Y}'_{ik}\mathbf{Y}_{ik'})^2 + E(\mathbf{Y}'_{ik}\mathbf{Y}_{ik})^2\right)\right. \\
&\quad \left.+ 2(n_i-2)(n_i-3)\text{Sp}(\mathbf{T}_d\boldsymbol{\Sigma}_i)^2\right) \\
&= \frac{1}{2n_i(n_i-1)} \left(2(2\text{Sp}(\mathbf{T}_d\boldsymbol{\Sigma}_i)^2 + \text{Sp}^2(\mathbf{T}_d\boldsymbol{\Sigma}_i) + \zeta_i)\right. \\
&\quad \left.+ 2\text{Sp}^2(\mathbf{T}_d\boldsymbol{\Sigma}_i) + 4\text{Sp}(\mathbf{T}_d\boldsymbol{\Sigma}_i)^2\right. \\
&\quad \left.+ 2(n_i-2)(3\text{Sp}(\mathbf{T}_d\boldsymbol{\Sigma}_i)^2 + 2\text{Sp}(\mathbf{T}_d\boldsymbol{\Sigma}_i)^2 + \text{Sp}^2(\mathbf{T}_d\boldsymbol{\Sigma}_i) + \zeta_i)\right. \\
&\quad \left.+ 2(n_i-2)(n_i-3)\text{Sp}(\mathbf{T}_d\boldsymbol{\Sigma}_i)^2\right) \\
&= \frac{\zeta_i}{n_i} + \frac{1}{n_i-1}\text{Sp}^2(\mathbf{T}_d\boldsymbol{\Sigma}_i) + \frac{n_i}{n_i-1}\text{Sp}(\mathbf{T}_d\boldsymbol{\Sigma}_i)^2.
\end{aligned}$$

□

Beweis von Satz 18. Es sei mit den Bezeichnungen aus der 2 zum additiv-linearen Modell $\tilde{\delta}_i := \max_{j=1,\dots,r_i} \{\delta_{ij}\}$. Dann ist in diesem Modell ausgehend von Abkürzung (48)

$$\begin{aligned}
\zeta_i &= \sum_{j=1}^{r_i} (\delta_{ij} - 3) a_{jj}^2 \\
&\leq (\tilde{\delta}_i - 3) \sum_{j=1}^{r_i} a_{jj}^2 \\
&\leq (\tilde{\delta}_i - 3) \sum_{j,j'=1}^{r_i} a_{jj'}^2 \\
&= (\tilde{\delta}_i - 3) \text{Sp}(\mathbf{A}\mathbf{A}') = (\tilde{\delta}_i - 3) \text{Sp}(\mathbf{T}_d\boldsymbol{\Sigma}_i)^2 \\
&\leq (\tilde{\delta}_i - 3) \text{Sp}^2(\mathbf{T}_d\boldsymbol{\Sigma}_i)
\end{aligned} \tag{63}$$

Die Folge c_n ist damit bei $H_1^{(i)}$ wegen Gleichung (50)

$$c_n = \frac{(n_i^2 - 2n_i + 2)(\tilde{\delta}_i - 3)}{n_i(n_i + 1)(n_i - 2)} \xrightarrow{n_i \rightarrow \infty} 0.$$

Für $H_2^{(i)}$ ist die Nullfolge entsprechend

$$c_n = \frac{(n_i^2 - n_i + 1) (\tilde{\delta}_i - 3)}{n_i (n_i + 1) (n_i - 2)}.$$

Im elliptischen Modell ist die Behauptung analog für $H_1^{(i)}$ aus der auf ζ_i angewendeten Abschätzung $\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 \leq \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)$, zu gewinnen:

$$\begin{aligned} \frac{EH_2^{(i)} - \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2}{\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2} &= (\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2)^{-1} \frac{(n_i^2 - n_i + 1)}{n_i (n_i + 1) (n_i - 2)} \zeta_i \\ &\leq 3 \frac{(n_i^2 - n_i + 1)}{n_i (n_i + 1) (n_i - 2)} \frac{\delta_i - 3}{r_i (r_i + 2)} \\ &\leq 3 \frac{(n_i^2 - n_i + 1) (\delta_i - 3)}{n_i (n_i + 1) (n_i - 2)} \end{aligned}$$

Bei $H_2^{(i)}$ hilft die Tatsache, daß nach der Cauchy-Schwarz'schen Ungleichung

$$\frac{\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)}{\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2} \leq r_i$$

ist, da r_i in Definition 4 $r_i = \text{rg}(\boldsymbol{\Sigma}_i)$ ist:

$$\begin{aligned} \frac{EH_2^{(i)} - \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2}{\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2} &\leq \frac{(n_i^2 - n_i + 1)}{n_i (n_i + 1) (n_i - 2)} \frac{\delta_i - 3}{r_i (r_i + 2)} (2\text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 + \text{Sp}^2 \mathbf{T}_d \boldsymbol{\Sigma}_i) \\ &\leq \frac{(n_i^2 - n_i + 1)}{n_i (n_i + 1) (n_i - 2)} \frac{\delta_i - 3}{r_i (r_i + 2)} (2 + r_i) \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 \\ &= \frac{(n_i^2 - n_i + 1)}{n_i (n_i + 1) (n_i - 2) r_i} \end{aligned}$$

□

A.4 Beweise zum Kapitel 4 Korrektur

A.4.1 Exakte Berechnung von $\text{Var}_{\mathcal{N}}(B_n)$

Lemma 39. Seien $Q_1 = \mathbf{X}' \mathbf{A}_1 \mathbf{X}$ und $Q_2 = \mathbf{X}' \mathbf{A}_2 \mathbf{X}$ quadratische Formen in normalverteilten Zufallsvektoren $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit symmetrischen Matrizen \mathbf{A}_1 und \mathbf{A}_2 , durch die $\text{Sp}(\mathbf{A}_1 \boldsymbol{\Sigma}) = \text{Sp}(\mathbf{A}_2 \boldsymbol{\Sigma}) = 0$ und $E(\mathbf{A}_1 \mathbf{X}) = E(\mathbf{A}_2 \mathbf{X}) = \mathbf{0}$ ist. Dann ist

$$\text{Cov}_{\mathcal{N}}(Q_1^2, Q_2^2) = 32\text{Sp}((\mathbf{A}_1 \boldsymbol{\Sigma})^2 (\mathbf{A}_2 \boldsymbol{\Sigma})^2) + 16\text{Sp}(\mathbf{A}_1 \boldsymbol{\Sigma} \mathbf{A}_2 \boldsymbol{\Sigma})^2 + 8\text{Sp}^2(\mathbf{A}_1 \boldsymbol{\Sigma} \mathbf{A}_2 \boldsymbol{\Sigma})$$

Beweis. Der Beweis soll über die kumulantenenerzeugende Funktion geführt werden. Dafür sei $M(t_1, t_2)$ die produktmomenterzeugende Funktion von Q_1 und Q_2 , so

daß $\ln(M(t_1, t_2))$ die gemeinsame kumulantenenerzeugende Funktion ist. Die zweite Kumulante in jeweils Q_1 und Q_2 kann durch

$$\begin{aligned}
\kappa_{2,2} &= \left. \frac{\partial^2 \ln M}{\partial t_1^2 \partial t_2^2} \right|_{t_1, t_2=0} \\
&= EQ_1^2 Q_2^2 - E^2 Q_1 Q_2 \\
&\quad - 2(E^2 Q_1 E Q_2^2 + E Q_1^2 E^2 Q_2 - E Q_1^2 Q_2 E Q_2 - E Q_1 E Q_1 Q_2^2) \\
&\quad - 6E^2 Q_1 E^2 Q_2 + 8E Q_1 E Q_1 Q_2 E Q_2 \\
&= EQ_1^2 Q_2^2 - 2E^2 Q_1 Q_2 - E Q_1^2 E Q_2^2 \\
&= \text{Cov}(Q_1^2, Q_2^2) - 2\text{Cov}^2(Q_1, Q_2)
\end{aligned}$$

berechnet werden. Bei Normalverteilung ist

$$\begin{aligned}
\kappa_{2,2} &= 32\text{Sp}\left((\Gamma' \mathbf{A}_1 \Gamma)^2 (\Gamma' \mathbf{A}_2 \Gamma)^2\right) + 16\text{Sp}(\Gamma' \mathbf{A}_1 \Gamma \Gamma' \mathbf{A}_2 \Gamma)^2 \\
&= 32\text{Sp}\left((\Sigma \mathbf{A}_1)^2 (\Sigma \mathbf{A}_2)^2\right) + 16\text{Sp}(\mathbf{A}_1 \Sigma \mathbf{A}_2 \Sigma)^2,
\end{aligned}$$

siehe Mathai, Provost und Hayakawa (1995), Theorem 2.4.1 und insbesondere Gleichung (2.4.9) auf S. 31. Ferner ist $\text{Cov}_{\mathcal{N}}(Q_1, Q_2) = 2\text{Sp}(\mathbf{A}_1 \Sigma \mathbf{A}_2 \Sigma)$, siehe hierzu bei Mathai und Provost (1992) das Theorem 3.2d.3 auf S. 75. Also ergibt sich

$$\text{Cov}(Q_1^2, Q_2^2) = \kappa_{2,2} + 8\text{Sp}^2(\mathbf{A}_1 \Sigma \mathbf{A}_2 \Sigma)$$

□

Lemma 40. *Es seien $\mathbf{Y}_1, \dots, \mathbf{Y}_6 \sim \mathcal{N}(\boldsymbol{\mu}_i, \Sigma_i)$ und $\mathbf{Y}_7, \dots, \mathbf{Y}_{10} \sim \mathcal{N}(\boldsymbol{\mu}_{i'}, \Sigma_{i'})$ sowie $A_{(k,l)(s,t)} = (\mathbf{Y}_k - \mathbf{Y}_l)' (\mathbf{Y}_s - \mathbf{Y}_t)$.*

1. $\text{Var}_{\mathcal{N}}\left(A_{(1,2)(7,8)}^2\right) = 96\text{Sp}(\Sigma_i \Sigma_{i'})^2 + 32\text{Sp}^2(\Sigma_i \Sigma_{i'})$
2. $\text{Cov}_{\mathcal{N}}\left(A_{(1,2)(3,4)}^2, A_{(1,3)(2,4)}^2\right) = 48\text{Sp}(\Sigma_i^4) + 8\text{Sp}^2(\Sigma_i^2)$
3. $\text{Cov}_{\mathcal{N}}\left(A_{(1,2)(7,8)}^2, A_{(1,2)(7,9)}^2\right) = 48\text{Sp}(\Sigma_i \Sigma_{i'})^2 + 8\text{Sp}^2(\Sigma_i \Sigma_{i'})$
4. $\text{Cov}_{\mathcal{N}}\left(A_{(1,2)(3,4)}^2, A_{(1,3)(2,5)}^2\right) = 26\text{Sp}(\Sigma_i^4) + 4\text{Sp}^2(\Sigma_i^2)$
5. $\text{Cov}_{\mathcal{N}}\left(A_{(1,2)(3,4)}^2, A_{(1,4)(3,5)}^2\right) = 26\text{Sp}(\Sigma_i^4) + 4\text{Sp}^2(\Sigma_i^2)$
6. $\text{Cov}_{\mathcal{N}}\left(A_{(1,2)(7,8)}^2, A_{(1,2)(9,10)}^2\right) = 32\text{Sp}(\Sigma_i \Sigma_{i'})^2$
7. $\text{Cov}_{\mathcal{N}}\left(A_{(1,2)(3,4)}^2, A_{(1,5)(2,6)}^2\right) = 16\text{Sp}(\Sigma_i^4)$
8. $\text{Cov}_{\mathcal{N}}\left(A_{(1,2)(7,8)}^2, A_{(1,5)(7,9)}^2\right) = 18\text{Sp}(\Sigma_i \Sigma_{i'})^2 + 2\text{Sp}^2(\Sigma_i \Sigma_{i'})$
9. $\text{Cov}_{\mathcal{N}}\left(A_{(1,2)(6,7)}^2, A_{(1,5)(8,9)}^2\right) = 8\text{Sp}(\Sigma_i \Sigma_{i'})^2$

Beweis. Der Beweis wird für $\text{Cov}_{\mathcal{N}}(A^2_{(12)(78)}, A^2_{(12)(79)})$ geführt. Die anderen Nummern werden analog gerechnet. Um Lemma 39 anwenden zu können, werden die Matrizen

$$\mathbf{A}_1 := \frac{1}{2} \begin{pmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 & \\ 1 & -1 & & & \\ -1 & 1 & & 0 & \\ 0 & & & & \end{pmatrix} \quad \text{und} \quad \mathbf{A}_2 := \frac{1}{2} \begin{pmatrix} 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 & 1 \\ 1 & -1 & & & \\ 0 & 0 & & 0 & \\ -1 & 1 & & & \end{pmatrix}$$

definiert. Mit dem Zufallsvektor $\mathbf{Y} := (\mathbf{Y}'_1, \mathbf{Y}'_2, \mathbf{Y}'_7, \mathbf{Y}'_8, \mathbf{Y}'_9)'$ kann nun geschrieben werden

$$A_{(12)(78)} = \mathbf{Y}' \mathbf{A}_1 \mathbf{Y} \quad \text{und} \quad A_{(12)(79)} = \mathbf{Y}' \mathbf{A}_2 \mathbf{Y}.$$

Es ist $\text{Cov}_{\mathcal{N}}(\mathbf{Y}) = \boldsymbol{\Sigma}_i \oplus \boldsymbol{\Sigma}_i \oplus \boldsymbol{\Sigma}_{i'} \oplus \boldsymbol{\Sigma}_{i'} \oplus \boldsymbol{\Sigma}_{i'} =: \boldsymbol{\Sigma}$. Die übrigen Voraussetzungen an Lemma 39 sind mit $\mathbf{A}_1 \mathbf{E} \mathbf{Y} = \mathbf{A}_2 \mathbf{E} \mathbf{Y} = \mathbf{0}$ und $\text{Sp}(\boldsymbol{\Sigma} \mathbf{A}_1) = \text{Sp}(\boldsymbol{\Sigma} \mathbf{A}_2) = 0$ erfüllt. Der Rest ist Matrixrechnung. \square

Beweis von Satz 19 auf Seite 59. Die erste Summe ist $\sum_{i=1}^a \text{Var}_{\mathcal{N}}\left(\frac{t_{ii}^2}{n_i} B_i\right)$. Es ist

$$\begin{aligned} \text{Var}_{\mathcal{N}}(B_i) &= (4n_i^3 (n_i - 1) (n_i - 2) (n_i - 3))^{-2} \sum_{k,l,s,t}^* \sum_{k',l',s',t'}^* \text{Cov}_{\mathcal{N}}(A^2_{(k,l)(s,t)}, A^2_{(k',l')(s',t')}) \\ &= (4n_i^3 (n_i - 1) (n_i - 2) (n_i - 3))^{-2} \left(\binom{n_i}{4} 3 \cdot 2^6 \text{Var}_{\mathcal{N}}(A^2_{(1,2)(3,4)}) \right. \\ &\quad + \binom{n_i}{4} 3 \cdot 2^7 \text{Cov}_{\mathcal{N}}(A^2_{(1,2)(3,4)}, A^2_{(1,3)(2,4)}) \\ &\quad + \binom{5}{2} \binom{n_i}{5} 3 \cdot 2^7 \text{Cov}_{\mathcal{N}}(A^2_{(1,2)(3,4)}, A^2_{(1,2)(3,5)}) \\ &\quad + \binom{5}{2} \binom{n_i}{5} 3 \cdot 2^8 \text{Cov}_{\mathcal{N}}(A^2_{(1,2)(3,4)}, A^2_{(1,3)(2,5)}) \\ &\quad + \binom{6}{2} \binom{n_i}{6} 3 \cdot 2^7 \text{Cov}_{\mathcal{N}}(A^2_{(1,2)(3,4)}, A^2_{(1,2)(5,6)}) \\ &\quad + \binom{6}{2} \binom{n_i}{6} 3 \cdot 2^9 \text{Cov}_{\mathcal{N}}(A^2_{(1,2)(3,4)}, A^2_{(1,5)(2,6)}) \\ &\quad + \binom{6}{2} \binom{n_i}{6} 3 \cdot 2^9 \text{Cov}_{\mathcal{N}}(A^2_{(1,2)(3,4)}, A^2_{(1,5)(3,6)}) \\ &\quad \left. + \binom{n_i}{7} 7! \cdot 2^4 \text{Cov}_{\mathcal{N}}(A^2_{(1,2)(3,4)}, A^2_{(1,5)(6,7)}) \right) \\ &= \frac{4(2n_i^3 - 9n_i^2 + 9n_i + 4)}{n_i^5 (n_i - 1) (n_i - 2) (n_i - 3)} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i)^4 + \frac{4}{n_i^5 (n_i - 3)} \text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i)^2 \end{aligned}$$

Die Faktoren $\binom{n_i}{4} 3 \cdot 2^6$, $\binom{n_i}{4} 3 \cdot 2^7$ usw. kommen von den Wiederholungen, die durch

Symmetrien, Erwartungswertgleichheiten und der Kombinatorik der Indices zustandekommen.

Erwartungswertgleichheit zu $\text{Var}_{\mathcal{N}} \left(A_{(1,2)(3,4)}^2 \right)$ gibt es für die $\binom{n_i}{4}$ Auswahlen der 4 Indices. Sind 4 Indices fixiert, existieren in $\text{Cov} \left(A_{(1,2)(3,4)}^2, A_{(1,2)(3,4)}^2 \right)$ noch 6 Symmetrien der Art $A_{(1,2)(3,4)}^2 = A_{(2,1)(3,4)}^2$ und $A_{(1,2)(3,4)}^2 = A_{(3,4)(1,2)}^2$, also 4 Vertauschungen innerhalb der Klammern und 2 Vertauschungen der gesamten Klammern, so daß der Faktor 2^6 hinzukommt. Ferner gilt Erwartungswertgleichheit durch Permutation der Indices, die nicht durch die Ausführung der Symmetrien darstellbar ist. 4 Indices können auf $4! = 24$ Weisen angeordnet werden, aber „modulo“ der $2^3 = 8$ Symmetrien bleiben 3 Kombinationen, die erwartungswertgleich sind; namentlich

$$\text{Var} \left(A_{(1,3)(2,4)}^2 \right) = \text{Var} \left(A_{(1,4)(2,3)}^2 \right) = \text{Var} \left(A_{(1,2)(3,4)}^2 \right).$$

Zusätzlich zu denselben Überlegungen kommt vor $\text{Cov}_{\mathcal{N}} \left(A_{(1,2)(3,4)}^2, A_{(1,3)(2,4)}^2 \right)$ der Faktor 2 wegen der Symmetrie der Kovarianz hinzu. Zusätzlich steht der Faktor $\binom{5}{2}$ vor $\text{Cov}_{\mathcal{N}} \left(A_{(1,2)(3,4)}^2, A_{(1,2)(3,5)}^2 \right)$ wegen der Anzahl der Möglichkeiten, daß von den 5 Indices sich zwei in beiden Argumenten wiederholen.

Der zusätzliche Faktor 2 vor $\text{Cov}_{\mathcal{N}} \left(A_{(1,2)(3,4)}^2, A_{(1,3)(2,5)}^2 \right)$ rührt von der Gleichheit in Erwartung

$$\text{Cov} \left(A_{(1,2)(3,4)}^2, A_{(2,3)(1,5)}^2 \right) = \text{Cov}_{\mathcal{N}} \left(A_{(1,2)(3,4)}^2, A_{(1,3)(2,5)}^2 \right)$$

her. Bei $\text{Cov}_{\mathcal{N}} \left(A_{(1,2)(3,4)}^2, A_{(1,5)(2,6)}^2 \right)$ und $\text{Cov}_{\mathcal{N}} \left(A_{(1,2)(3,4)}^2, A_{(1,5)(3,6)}^2 \right)$ können die Rollen der Indices 5 und 6 sowie 3 und 4 untereinander zusätzlich tauschen.

Der Faktor vor $\text{Cov}_{\mathcal{N}} \left(A_{(1,2)(3,4)}^2, A_{(1,5)(6,7)}^2 \right)$ kann erklärt werden, da die 7 Indices permutiert werden können und der Index, der sich in beiden Argumenten wiederholt, in beiden Argumenten jeweils 4 Positionen haben kann, ohne daß sich an der Kovarianz etwas verändert. Zum Schluß werden die Ergebnisse aus Lemma 40 eingesetzt.

Die zweite Summe enthält die Varianzen der gemischten Terme $\text{Var}_{\mathcal{N}} \left(\frac{t_{ii'}^2}{n_i n_{i'}} B_{ii'} \right)$ und $\text{Cov}_{\mathcal{N}} \left(\frac{t_{ii'}^2}{n_i n_{i'}} B_{ii'}, \frac{t_{ii'}^2}{n_i n_{i'}} B_{i'i} \right)$. Es sei daran erinnert, daß

$$A_{i(kl)i'(st)} = (\mathbf{Y}_{ik} - \mathbf{Y}_{il})' (\mathbf{Y}_{i's} - \mathbf{Y}_{i't})$$

ist, also die Bilinearform aus den Beobachtungen k und l der Stichprobe i sowie s und t der Stichprobe i' . Es ist

$$\text{Var}_{\mathcal{N}} (B_{ii'}) = (4n_i^2 (n_i - 1) n_{i'}^2 (n_{i'} - 1))^{-2} \sum_{k \neq l}^{n_i} \sum_{r \neq s}^{n_i} \sum_{k' \neq l'}^{n_{i'}} \sum_{s' \neq t'}^{n_{i'}} \text{Cov}_{\mathcal{N}} \left(A_{i(kl)i'(k'l')}^2, A_{i(st)i'(s't')}^2 \right)$$

$$\begin{aligned}
&= (4n_i^2 (n_i - 1) n_{i'}^2 (n_{i'} - 1))^{-2} \left(\binom{n_i}{2} \binom{n_{i'}}{2} 2^4 \text{Var}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2) \right. \\
&+ \binom{n_i}{2} \binom{n_{i'}}{3} \cdot 3 \cdot 2^5 \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,2)i'(1,3)}^2) \\
&+ \binom{n_i}{3} \binom{n_{i'}}{2} \cdot 3 \cdot 2^5 \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,3)i'(1,2)}^2) \\
&+ \binom{n_i}{3} \binom{n_{i'}}{3} \cdot 3 \cdot 2^6 \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,3)i'(1,3)}^2) \\
&+ \binom{n_i}{2} \binom{n_{i'}}{4} \cdot 3 \cdot 2^2 \cdot 4! \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,2)i'(3,4)}^2) \\
&+ \binom{n_i}{4} \binom{n_{i'}}{2} \cdot 3 \cdot 2^2 \cdot 4! \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(3,4)i'(1,2)}^2) \\
&+ \binom{n_i}{3} \binom{n_{i'}}{4} \cdot 3 \cdot 2^2 \cdot 4! \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,3)i'(3,4)}^2) \\
&\left. \left(+ \binom{n_i}{4} \binom{n_{i'}}{3} \cdot 3 \cdot 2^2 \cdot 4! \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(3,4)i'(1,3)}^2) \right) \right) \\
&= \frac{2(n_i + n_{i'} - 1) \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})^2 + 2\text{Sp}^2(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'})}{n_i^2 n_{i'}^2 (n_i - 1) (n_{i'} - 1)}.
\end{aligned}$$

Die dritte Summe führt die Kovarianzen $\text{Cov}_{\mathcal{N}} \left(\frac{t_{ii'}^2}{n_i n_{i'}} B_{ii'}, \frac{t_{ii''}^2}{n_i n_{i''}} B_{ii''} \right)$.

$$\begin{aligned}
\text{Cov}_{\mathcal{N}} \left(B_2^{(i,i')}, B_2^{(i,i'')} \right) &= (16n_i^4 (n_i - 1)^2 n_{i'}^2 (n_{i'} - 1) n_{i''}^2 (n_{i''} - 1))^{-1} \\
&\cdot \sum_{k \neq l}^{n_i} \sum_{s \neq t}^{n_i} \sum_{k' \neq l'}^{n_{i'}} \sum_{k'' \neq l''}^{n_{i''}} \text{Cov}_{\mathcal{N}} (A_{i(k,l)i'(k',l')}^2, A_{i(s,t)i''(k'',l'')}^2) \\
&= (16n_i^4 (n_i - 1)^2 n_{i'}^2 (n_{i'} - 1) n_{i''}^2 (n_{i''} - 1))^{-1} \\
&\cdot \left(\binom{n_i}{2} \binom{n_{i'}}{2} \binom{n_{i''}}{2} \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,2)i''(1,2)}^2) \right. \\
&\quad \left. \binom{n_i}{3} \binom{n_{i'}}{2} \binom{n_{i''}}{2} \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,3)i''(1,2)}^2) \right) \\
&= \frac{2}{n_i^2 (n_i - 1) n_{i'} n_{i''}} \text{Sp}(\mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i'} \mathbf{T}_d \boldsymbol{\Sigma}_i \mathbf{T}_d \boldsymbol{\Sigma}_{i''})
\end{aligned}$$

Die vierte Summe faßt alle Terme der Art $\text{Cov}_{\mathcal{N}} \left(\frac{t_{ii}^2}{n_i^2} B_i, \frac{t_{i'i'}^2}{n_i n_{i'}} B_{i'i'} \right)$ zusammen.

$$\begin{aligned}
\text{Cov}_{\mathcal{N}} \left(B_2^{(i,i')}, B_2^{(i)} \right) &= (16n_i^5 (n_i - 1)^2 (n_i - 2) (n_i - 3) (n_{i'}^2) (n_{i'} - 1))^{-1} \\
&\cdot \sum_{k \neq l}^{n_i} \sum_{k' \neq l'}^{n_{i'}} \sum_{k'' \neq l''}^{n_i} \sum_{s'' \neq t''}^{n_i} \text{Cov}_{\mathcal{N}} (A_{i(k,l)i'(k',l')}^2, A_{i(k'',l'')i(s'',t'')}^2)
\end{aligned}$$

$$\begin{aligned}
&= (16n_i^5 (n_i - 1)^2 (n_i - 2) (n_i - 3) (n_i^2) (n_{i'} - 1))^{-1} \\
&\quad \cdot \left(\binom{n_i}{4} \binom{n_{i'}}{2} 2^6 3 \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,2)i(3,4)}^2) \right. \\
&\quad + \binom{n_i}{4} \binom{n_{i'}}{2} 2^7 3 \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,3)i(2,4)}^2) \\
&\quad \left. + \binom{n_i}{5} \binom{n_{i'}}{2} 5! \cdot 2^4 \text{Cov}_{\mathcal{N}} (A_{i(1,2)i'(1,2)}^2, A_{i(1,3)i(4,5)}^2) \right) \\
&= \frac{4}{n_i^3 (n_i - 1) n_{i'}} \text{Sp} (\boldsymbol{\Sigma}_i^3 \boldsymbol{\Sigma}_{i'})
\end{aligned}$$

□

Beweis zu Lemma 20. Seien $\lambda_{ij} \geq 0$ die Eigenwerte von $\boldsymbol{\Sigma}_i$.

1. Ohne Einschränkung kann gefordert werden $\lambda_{11} \geq \lambda_{12} \geq \dots \geq 0$. Es ist nach der Cauchy-Schwarz'schen Ungleichung

$$\frac{1}{d} \leq \frac{\text{Sp}(\boldsymbol{\Sigma}_1)^2}{\text{Sp}^2(\boldsymbol{\Sigma}_1)} = \frac{\sum \lambda_{1i}^2}{(\sum \lambda_{1i})^2} \leq 1,$$

wobei Gleichheit links bei $\lambda_{11} = \dots = \lambda_{1d}$ gilt und Gleichheit rechts bei $\lambda_{11} \neq \lambda_{12} = \dots = 0$. So ist schon bei Geisser und Greenhouse (1958) der Beweis für den Bereich des Box'schen Epsilons geführt. Die selben Extrema mit denselben Extremstellen hat man auch nach dem Übergang von $\boldsymbol{\Sigma}_1$ zu $\boldsymbol{\Sigma}_1^2$.

2. Seien $\lambda_{(12)1}, \dots, \lambda_{(12)d}$ die Eigenwerte von $\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2$. Dann ist

$$\begin{aligned}
0 &\leq \frac{\text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)}{\text{Sp}(\boldsymbol{\Sigma}_1) \text{Sp}(\boldsymbol{\Sigma}_2)} \\
&= \frac{\sum \lambda_{(12)j}}{(\sum \lambda_{1j}) (\sum \lambda_{2j})} \\
&\leq \frac{\sum_j \lambda_{1j} \lambda_{2j}}{\left(\sum_{j,j'} \lambda_{1j} \lambda_{2j'} \right)} \\
&\leq 1,
\end{aligned}$$

wobei im letzten Schritt Gleichheit gilt, wenn alle bis auf ein $\lambda_{(12)j}$ verschwinden. Im ersten Schritt gilt Gleichheit, falls $\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2 = \mathbf{0}$ ist. Damit hat man für dieselben $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ dieselben Extrema von $\frac{\text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)^2}{\text{Sp}^2(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)}$ und $\frac{\text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)}{\text{Sp}(\boldsymbol{\Sigma}_1) \text{Sp}(\boldsymbol{\Sigma}_2)}$.

3. Wie bei 2. ist

$$0 \leq \frac{\text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_3)}{\text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2) \text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_3)} \leq 1.$$

Laut Marshall und Olkin 1979 , S. 248, ist $\sum_j \lambda_{(12)j} \leq \sum_j \lambda_{1j} \sum_j \lambda_{2j}$. Damit ergibt sich

$$\begin{aligned}
0 &\leq \frac{(\text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2) \text{Sp}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_3) \text{Sp}(\boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_3) \text{Sp}(\boldsymbol{\Sigma}_1^2))^{\frac{1}{2}}}{\text{Sp}^2(\boldsymbol{\Sigma}_1) \text{Sp}(\boldsymbol{\Sigma}_2) \text{Sp}(\boldsymbol{\Sigma}_3)} \\
&= \frac{(\sum \lambda_{(12)j} \sum \lambda_{(13)j} \sum \lambda_{(23)j} \sum \lambda_{1j}^2)^{\frac{1}{2}}}{(\sum \lambda_{1j})^2 \sum \lambda_{2j} \sum \lambda_{3j}} \\
&\leq \frac{(\sum \lambda_{1j} \lambda_{2j} \sum \lambda_{1j} \lambda_{3j} \sum \lambda_{2j} \lambda_{3j} \sum \lambda_{1j}^2)^{\frac{1}{2}}}{(\sum \lambda_{1j})^2 \sum \lambda_{2j} \sum \lambda_{3j}} \\
&\leq \frac{(\sum \lambda_{1j} \lambda_{2j} \sum \lambda_{1j} \lambda_{3j} \sum \lambda_{2j} \lambda_{3j})^{\frac{1}{2}}}{\sum \lambda_{1j} \sum \lambda_{2j} \sum \lambda_{3j}} \\
&\leq \frac{(\sum \lambda_{1j} \lambda_{3j} \sum \lambda_{2j} \lambda_{3j})^{\frac{1}{2}}}{\sqrt{\sum \lambda_{1j} \lambda_{2j}} \sum \lambda_{3j}} \leq \frac{(\sum \lambda_{1j} \lambda_{3j})^{\frac{1}{2}}}{\sqrt{\sum \lambda_{1j}} \sum \lambda_{3j}} \\
&\leq 1.
\end{aligned}$$

4. Dies folgt aus 3. □

A.4.2 Ableitungen der F_F -Quantilfunktion und Integrierbarkeit

Beweis von Satz 21 auf Seite 62. Dieser Satz benötigt folgendes Lemma:

Lemma 41. *Die Funktion*

$$\begin{aligned}
\tau_1(t, \nu_1, \nu_2) &= t^{\frac{\nu_1}{2}-1} \left(\frac{\nu_1 t + \nu_2}{\nu_2} \right)^{-\frac{\nu_1 + \nu_2 + 2}{2}} \left(\nu_2(t-1) + (\nu_1 t + \nu_2) \cdot \right. \\
&\quad \left. \left[\log \left(\frac{\nu_1 t + \nu_2}{\nu_1 t} \right) - \psi_0 \left(\frac{\nu_1 + \nu_2}{2} \right) + \psi_0 \left(\frac{\nu_1}{2} \right) \right] \right)
\end{aligned}$$

ist für finite $\nu_1 < \nu_2$ und finite obere Grenze uneigentlich integrierbar, d. h.

$$\lim_{x \searrow 0} \int_x^{x_0} \tau_1(t, \nu_1, \nu_2) dt$$

existiert für jedes endliche x_0 . Dasselbe gilt für

$$\begin{aligned}
\frac{\partial \tau_1(t, \nu_1, \nu_2)}{\partial \nu_1} &= t^{\frac{\nu_1}{2}-1} \left(\frac{\nu_1 t + \nu_2}{\nu_2} \right)^{-\frac{\nu_1 + \nu_2 + 2}{2}} \cdot \\
&\left(\frac{1}{2} \left(\log(t) - \frac{t(\nu_1 + \nu_2 + 2)}{\nu_1 t + \nu_2} - \log \left(\frac{\nu_1 t + \nu_2}{\nu_2} \right) \right) \left((\nu_1 t + \nu_2) \left[\dots \right] + \nu_2(t-1) \right) \right. \\
&\quad \left. + \left(t \left[\dots \right] + t - \frac{1}{2}(\nu_1 t + \nu_2) \left(\psi' \left(\frac{\nu_1 + \nu_2}{2} \right) + \frac{2}{\nu_1} - \psi' \left(\frac{\nu_1}{2} \right) \right) \right) \right),
\end{aligned}$$

wobei [...] die eckige Klammer in der Definition von τ_1 abkürzt.

Beweis. Den Beweis kann man führen, indem man die rechten Seiten ausmultipliziert und die entstehenden Summanden separat auf Integrierbarkeit im interessierenden Intervall untersucht. In beiden Funktionen ist dabei eine Polstelle bei $t = 0$ zu beachten, wenn $1 \leq \nu_1 \leq 2$ ist.

Der ausmultiplizierte Funktionsterm von τ_1 besteht aus acht Summanden, nach Majorante und Polstelle tabelliert in Tabelle 6.

	Produkt aus ...	Majorante	Polstelle
	$-\nu_2$	$t^{-\frac{1}{2}}$	Ja, $t = 0$
	$\nu_2 t$	\sqrt{t}	Nein
$t^{\frac{\nu_1}{2}-1} \underbrace{(\nu_1 t + \nu_2)^{-\frac{\nu_1+\nu_2+2}{2}}}_{<1}$	$\nu_1 t \log\left(\frac{\nu_1 t + \nu_2}{\nu_1 t}\right)$	\sqrt{t}	Nein
	$-\psi\left(\frac{\nu_1 + \nu_2}{2}\right)$	\sqrt{t}	Nein
	$\psi_0\left(\frac{\nu_1}{2}\right)$	\sqrt{t}	Nein
	$\log\left(\frac{\nu_1 t + \nu_2}{\nu_1 t}\right)$	s. u.	Ja, $t = 0$
ν_2	$-\psi\left(\frac{\nu_1 + \nu_2}{2}\right)$	$t^{-\frac{1}{2}}$	Ja, $t = 0$
	$\psi_0\left(\frac{\nu_1}{2}\right)$	$t^{-\frac{1}{2}}$	Ja, $t = 0$

Tabelle 6: Majoranten und Polstellen der Summanden des Integranden

Von denen haben vier nie eine Polstelle. Von den übrigen vier Summanden lassen sich bei dreien die Beträge gegen $ct^{\frac{\nu_1}{2}-1}$ mit $c = \nu_2 \max\left\{\psi_0\left(\frac{\nu_1}{2}\right), \psi_0\left(\frac{\nu_1+\nu_2}{2}\right), 1\right\}$ abschätzen. Auf Integrierbarkeit an der Polstelle $t = 0$ zu untersuchen bleibt daher nur noch $t^{\frac{\nu_1}{2}-1} (\nu_1 t + \nu_2)^{-\frac{\nu_1+\nu_2+2}{2}} \log\left(\frac{\nu_1 t + \nu_2}{\nu_1 t}\right)$. Hier genügt eine vereinfachte Betrachtung für den Fall $\nu_1 = 1$ wegen der Abschätzungen

$$\begin{aligned} t^{\frac{\nu_1}{2}-1} (\nu_1 t + \nu_2)^{-\frac{\nu_1+\nu_2+2}{2}} \log\left(\frac{\nu_1 t + \nu_2}{\nu_1 t}\right) &\leq t^{\frac{\nu_1}{2}-1} \log\left(\frac{\nu_1 t + \nu_2}{\nu_1 t}\right) \\ &\leq t^{-\frac{1}{2}} \log\left(1 + \frac{\nu_2}{t}\right). \end{aligned}$$

Letzterer Ausdruck hat als Stammfunktion $2\sqrt{t} \log\left(1 + \frac{\nu_2}{t}\right) - 4\sqrt{\nu_2} \arctan\left(\sqrt{\frac{\nu_2}{t}}\right)$. Diese beiden Summanden sind beschränkt für $t \in (0, 1]$. Für den ersten Term ist dies ersichtlich durch Anwendung des Satzes von de l'Hospital

$$\begin{aligned} \lim_{t \rightarrow 0} \sqrt{t} \log\left(1 + \frac{\nu_2}{t}\right) &= \lim_{t \rightarrow 0} \frac{\log\left(1 + \frac{\nu_2}{t}\right)}{t^{-\frac{1}{2}}} \\ &= \lim_{t \rightarrow 0} \frac{\frac{\nu_2}{t(\nu_2+t)}}{\frac{1}{2}t^{-\frac{3}{2}}} \\ &= \lim_{t \rightarrow 0} \frac{2t^{\frac{1}{2}}\nu_2}{\nu_2 + t} = 0, \end{aligned}$$

für den zweiten Term durch die Beschränktheit des Arkustangens. Damit ist gezeigt, daß τ_1 integrierbar ist.

Der Funktionsterm von τ_2 besteht ebenfalls aus Summanden, die im Integrationsbereich beschränkt sind, und zehn Summanden, die eine Polstelle an $t = 0$ haben. Diese Summanden lassen sich bis auf von t unabhängige Faktoren abschätzen durch eine der Majoranten in folgender Tabelle:

Majorante	Stammfunktion
$t^{-\frac{1}{2}} \log(t) \log\left(1 + \frac{\nu_2}{t}\right)$	s. u.
$\frac{\log(t)}{\sqrt{t}}$	$\sqrt{t}(2 \log(t) - 4)$
$t^{-\frac{1}{2}} \log\left(1 + \frac{\nu_2}{t}\right)$	siehe Integrierbarkeit von τ_1
$t^{-\frac{1}{2}}$	$2\sqrt{t}$

Zu zeigen bleibt daher nur noch die Existenz des uneigentlichen Integrals

$$\lim_{x \rightarrow 0} \int_x^1 t^{-\frac{1}{2}} \log(t) \log\left(1 + \frac{\nu_2}{t}\right) dt.$$

Sie wird zunächst für den vereinfachten Integranden $t^{-\frac{1}{2}} \log(t) \log\left(\frac{\nu_2}{t}\right)$ gezeigt. Anschließend kann analytisch bewiesen werden, daß im Intervall $(0, 1]$ die Differenz zwischen dem ursprünglichen Integranden und seiner Vereinfachung beschränkt ist.

Eine Stammfunktion für $t^{-\frac{1}{2}} \log(t) \log\left(\frac{\nu_2}{t}\right)$ ist

$$2\sqrt{t} \left(\log\left(\frac{\nu_2}{t}\right) (\log(t) - 2) + 2 \log(t) - 8 \right).$$

Sie ist im Integrationsbereich beschränkt, denn es ist nach dem Satz von de l'Hospital

$$\begin{aligned} \lim_{t \rightarrow 0} \sqrt{t} \log\left(\frac{\nu_2}{t}\right) \log(t) &= \lim_{t \rightarrow 0} \frac{\log\left(\frac{\nu_2}{t}\right) \log(t)}{t^{-\frac{1}{2}}} \\ &= \lim_{t \rightarrow 0} \frac{t^{-1} \log\left(\frac{\nu_2}{t}\right) - t^{-1} \log(t)}{-\frac{1}{2} t^{-\frac{3}{2}}} \\ &= \lim_{t \rightarrow 0} \frac{\log\left(\frac{\nu_2}{t}\right) - \log(t)}{-\frac{1}{2} t^{-\frac{1}{2}}} \\ &= \lim_{t \rightarrow 0} \frac{\log(\nu_2) - 2 \log(t)}{-\frac{1}{2} t^{-\frac{1}{2}}} \\ &= \lim_{t \rightarrow 0} \frac{-2t^{-1}}{\frac{1}{4} t^{-\frac{3}{2}}} \\ &= \lim_{t \rightarrow 0} -8\sqrt{t} = 0 \end{aligned}$$

Damit existiert das unbestimmte Integral $\lim_{x \rightarrow 0} \int_x^1 t^{-\frac{1}{2}} \log(t) \log\left(\frac{\nu_2}{t}\right) dt$. Der Abstand des vereinfachten vom ursprünglichen Integranden

$$\tau_3(t) := \left| t^{-\frac{1}{2}} \log(t) \left(\log\left(\frac{\nu_2}{t}\right) - \log\left(1 + \frac{\nu_2}{t}\right) \right) \right| = \left| t^{-\frac{1}{2}} \log(t) \log\left(\frac{\nu_2}{t + \nu_2}\right) \right|$$

ist stetig auf $(0, 1]$ und verschwindet für $t \rightarrow 0$, so daß τ_3 stetig auf $[0, 1]$ fortgesetzt

werden kann. Die Fortsetzung ist offensichtlich stetig und beschränkt, so daß sie ein Maximum c besitzt. Durch

$$\begin{aligned} \lim_{x \rightarrow 0} \int_x^1 t^{-\frac{1}{2}} \log(t) \log\left(1 + \frac{\nu_2}{t}\right) dt &\leq \lim_{x \rightarrow 0} \int_x^1 \left| t^{-\frac{1}{2}} \log(t) \log\left(\frac{\nu_2}{t}\right) \right| + c dt \\ &= \lim_{x \rightarrow 0} \int_x^1 \left| t^{-\frac{1}{2}} \log(t) \log\left(\frac{\nu_2}{t}\right) \right| dt + c \end{aligned}$$

existiert auch das letzte benötigte uneigentliche Integral. \square

Wie angekündigt wird der Satz über die Ableitung implizit definierter Funktionen angewendet. Dafür ist zu zeigen, daß die partiellen Ableitungen $\frac{\partial g}{\partial \nu_1}$ und $\frac{\partial g}{\partial x}$ des Integrals in Gleichung (55) auf Seite 62 existieren und $\frac{\partial g}{\partial x}$ nicht verschwindet. Letzteres ist gerade die Dichte F'_F . Sie ist für alle $x > 0$ positiv.

Die Ableitung $\frac{\partial g}{\partial \nu_1}$ kann nach Satz 128.2 aus Heuser (2000) im Punkt (q_F, ν_1, ν_2) als Parameterintegral

$$\begin{aligned} \frac{\partial g(q_F, \nu_1, \nu_2)}{\partial \nu_1} &= \frac{\partial}{\partial \nu_1} F_F(q_F, \nu_1, \nu_2) = \frac{\partial}{\partial \nu_1} \left(\int_0^{q_F} F'(t, \nu_1, \nu_2) dt \right) \\ &= \frac{\partial}{\partial \nu_1} \int_0^{q_F} F'(t, \nu_1, \nu_2) dt \end{aligned}$$

gewonnen werden, sofern der Betrag der Ableitung $\frac{\partial}{\partial \nu_1} F'_F(x, \nu_1, \nu_2)$ für alle ν_1 durch eine nach t integrierbare Funktion majorisiert wird. Die Ableitung lautet

$$\begin{aligned} \frac{\partial F'_F}{\partial \nu_1}(t, \nu_1, \nu_2) &= -\nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} t^{\frac{\nu_1}{2}-1} \left(\nu_2(t-1) + (\nu_1 t + \nu_2) \cdot \right. \\ &\left. \left[\log\left(\frac{\nu_1 t + \nu_2}{\nu_1 t}\right) - \psi\left(\frac{\nu_1 + \nu_2}{2}\right) + \psi_0\left(\frac{\nu_1}{2}\right) \right] \right) \left(2B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) (\nu_1 t + \nu_2)^{\frac{n+m+2}{2}} \right)^{-1}. \end{aligned} \quad (64)$$

$\frac{\partial F'_F}{\partial \nu_1}$ ist im Falle $\nu_1 > 2$ beschränkt für alle nicht-negativen t . Im Falle $1 \leq \nu_1 \leq 2$ gibt es an $t = 0$ eine Polstelle, die aber (Lebesgue-)integrierbar ist nach Lemma 41 auf Seite 106

Damit ist $\frac{\partial F'_F}{\partial \nu_1}$ integrierbar und es ist nach dem Satz über die Ableitung impliziter Funktionen

$$\frac{\partial q_F(q_F(p, \nu_1, \nu_2), \nu_1, \nu_2)}{\partial \nu_1} = -\frac{\frac{\partial q_F(p, \nu_1, \nu_2)}{\partial \nu_1}}{F'_F(q_F(p, \nu_1, \nu_2), \nu_1, \nu_2)}.$$

\square

B Simulierte Konstellationen

B.1 Konstellationen mit identischen Kovarianzmatrizen.

n_1	n_2	n_3	$\Sigma_1 = \Sigma_2 = \dots$
10	20		CS
10	20		$AR(0, 6)$
10	20		$AR(0, 9)$
10	20		$harToe$
10	20		$linToe$
10	15	25	CS
10	15	25	$harToe$
10	15	25	$AR(0, 6)$
10	15	25	$AR(0, 9)$
10	15	25	$linToe$

B.2 Konstellationen aus zwei Stichproben bei unterschiedlichen Kovarianzmatrizen.

n_1	Σ_1	n_2	Σ_2
10	CS	20	$2 \cdot CS$
20	CS	10	$2 \cdot CS$
10	$AR(0, 6)$	20	$2 \cdot AR(0, 6)$
20	$AR(0, 6)$	10	$2 \cdot AR(0, 6)$
10	$AR(0, 9)$	20	$2 \cdot AR(0, 9)$
20	$AR(0, 9)$	10	$2 \cdot AR(0, 9)$
10	$linToe$	20	$2 \cdot linToe$
20	$linToe$	10	$2 \cdot linToe$
10	$harToe$	20	$2 \cdot harToe$
20	$harToe$	10	$2 \cdot harToe$
10	CS	20	$AR(0, 9)$
20	CS	10	$AR(0, 9)$
10	CS	20	$linToe$
20	CS	10	$linToe$
10	CS	20	$harToe$
20	CS	10	$harToe$
10	$linToe$	20	$AR(0, 9)$
20	$linToe$	10	$AR(0, 9)$
10	$harToe$	20	$AR(0, 9)$
20	$harToe$	10	$AR(0, 9)$

B.3 Konstellationen aus drei und vier Stichproben bei unterschiedlichen Kovarianzmatrizen.

n_1	Σ_1	n_2	Σ_2	n_3	Σ_3	n_4	Σ_4
30	$AR(0,6)$	15	$2 \cdot AR(0,6)$	30	CS		
15	$AR(0,6)$	30	$2 \cdot AR(0,6)$	15	CS		
30	$AR(0,9)$	15	$2 \cdot AR(0,9)$	30	$linToe$		
15	$AR(0,9)$	30	$2 \cdot AR(0,9)$	15	$linToe$		
30	$AR(0,9)$	15	$2 \cdot AR(0,9)$	30	$harToe$		
15	$AR(0,9)$	30	$2 \cdot AR(0,9)$	15	$harToe$		
10	$2 \cdot AR(0,6)$	15	$3 \cdot AR(0,6)$	20	$4 \cdot AR(0,6)$	30	$5 \cdot AR(0,6)$
30	$2 \cdot AR(0,6)$	20	$3 \cdot AR(0,6)$	15	$4 \cdot AR(0,6)$	10	$5 \cdot AR(0,6)$
10	$2 \cdot AR(0,9)$	15	$3 \cdot AR(0,9)$	20	$4 \cdot AR(0,9)$	30	$5 \cdot AR(0,9)$
30	$2 \cdot AR(0,9)$	20	$3 \cdot AR(0,9)$	15	$4 \cdot AR(0,9)$	10	$5 \cdot AR(0,9)$

C Literatur

Literatur

- Abramowitz, M. und I. A. Stegun (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. **55**. Washington DC: National Bureau of Standards (siehe S. 63, 84).
- Akhiezer, N. I. (1965). *The classical moment problem and some related questions in analysis*. Hafner Publishing Co., New York (siehe S. 75).
- Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. Second. John Wiley & Sons, Inc., New York (siehe S. 97).
- Atiqullah, M. (1962). „The estimation of residual variance in quadratically balanced least-squares problems and the robustness of the F -test“. *Biometrika* **49**, S. 83–91 (siehe S. 15, 18, 20, 27, 74).
- Bai, Z. und H. Saranadasa (1996). „Effect of high dimension: by an example of a two sample problem“. *Statist. Sinica* **6** (2), S. 311–329 (siehe S. 22).
- Bathke, A. C. und S. W. Harrar (2008). „Nonparametric methods in multivariate factorial designs for large number of factor levels“. *Journal of Statistical Planning and Inference* **138** (3), S. 588–610. DOI: DOI : 10 . 1016 / j . jspi . 2006 . 11 . 004 (siehe S. 30, 77).
- Bathke, A. C., O. Schabenberger u. a. (2009). „Greenhouse-Geisser adjustment and the ANOVA-type statistic: cousins or twins?“ *Amer. Statist.* **63** (3), S. 239–246. DOI: 10.1198/tast.2009.08187 (siehe S. 18, 37).
- Becker, B. (2010). „Test für hochdimensionale Messwiederholungen mit unbekanntem Kovarianzmatrizen“. Diplomarbeit. Inst. für Mathematische Stochastik, Universität Göttingen (siehe S. 16, 30, 44, 47–49, 68, 74).
- Bickel, P. J. und E. L. Lehmann (1969). „Unbiased estimation in convex families“. *Ann. Math. Statist.* **40**, S. 1523–1535 (siehe S. 58).
- Bono, R. u. a. (Dez. 2016). „Sphericity estimation bias for repeated measures designs in simulation studies“. *Behavior Research Methods* **48** (4), S. 1621–1630. DOI: 10.3758/s13428-015-0673-1 (siehe S. 45).
- Box, G. E. P. (1954a). „Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification“. *Ann. Math. Statistics* **25**, S. 290–302 (siehe S. 14–16).
- (1954b). „Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification“. *Ann. Math. Statistics* **25**, S. 484–498 (siehe S. 14–16).
- Bradley, J. V. (1978). „Robustness?“ *British Journal of Mathematical and Statistical Psychology* **31**, S. 144–152 (siehe S. 25, 61).
- Brunner, E. und U. Munzel (2013). *Nichtparametrische Datenanalyse: Unverbundene Stichproben*. Springer Berlin Heidelberg (siehe S. 77).

- Brunner, E., A. C. Bathke und M. Placzek (2012). „Estimation of Box’s ϵ for low- and high-dimensional repeated measures designs with unequal covariance matrices“. *Biom. J.* **54** (3), S. 301–316. DOI: 10.1002/bimj.201100160 (siehe S. 15, 16, 30, 44, 45, 50, 68, 74).
- Brunner, E., H. Dette und A. Munk (1997). „Box-type approximations in nonparametric factorial designs“. *J. Amer. Statist. Assoc.* **92** (440), S. 1494–1502 (siehe S. 15, 75).
- Brunner, E., U. Munzel und M. L. Puri (2002). „The multivariate nonparametric Behrens-Fisher problem“. *J. Statist. Plann. Inference* **108** (1-2). C. R. Rao 80th birthday felicitation volume, Part II, S. 37–53 (siehe S. 75).
- Cambanis, S., S. Huang und G. Simons (1981). „On the theory of elliptically contoured distributions“. *J. Multivariate Anal.* **11** (3), S. 368–385. DOI: 10.1016/0047-259X(81)90082-8 (siehe S. 12).
- Chen, L. H. und Q.-M. Shao (2001). „A non-uniform Berry–Esseen bound via Stein’s method“. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **120**, S. 236–254 (siehe S. 75).
- Chen, S. X. und Y.-L. Qin (2010). „A two-sample test for high-dimensional data with applications to gene-set testing“. *Ann. Statist.* **38** (2), S. 808–835. DOI: 10.1214/09-AOS716 (siehe S. 22).
- Chiang, C.-Y. (2005). „Multidimensional scale types and invariant multivariate statistics“. *British Journal of Mathematical and Statistical Psychology* **58** (1), S. 83–95. DOI: 10.1348/000711005X38717 (siehe S. 22).
- Collier, R. O. u. a. (1967). „Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design“. *Psychometrika* **32** (3), S. 339–353. DOI: 10.1007/BF02289596 (siehe S. 43).
- Feng, L. u. a. (2015). „Two-sample Behrens-Fisher problem for high-dimensional data“. *Statist. Sinica* **25** (4), S. 1297–1312 (siehe S. 22).
- Friedrich, S. und M. Pauly (2018). „MATS: Inference for potentially singular and heteroscedastic MANOVA“. *J. Multivariate Analysis* **165**, S. 166–179. DOI: 10.1016/j.jmva.2017.12.008 (siehe S. 22, 76).
- Galassi, M. u. a. (2011). *GNU Scientific Library Reference Manual*. The GSL Team (siehe S. 63).
- Geisser, S. und S. W. Greenhouse (1958). „An extension of Box’s results on the use of the F distribution in multivariate analysis“. *Ann. Math. Statist.* **29**, S. 885–891 (siehe S. 13, 43, 105).
- Greenhouse, S. W. und S. Geisser (1959). „On methods in the analysis of profile data“. *Psychometrika* **24**, S. 95–112 (siehe S. 13, 43).
- Halmos, P. R. (1946). „The theory of unbiased estimation“. *Ann. Math. Statistics* **17**, S. 34–43 (siehe S. 79).
- Happ, M., S. W. Harrar und A. C. Bathke (2016). „Inference for low- and high-dimensional multigroup repeated measures designs with unequal covariance matrices“. *Biometrical Journal* **58** (4), S. 810–830. DOI: 10.1002/bimj.201500064 (siehe S. 5, 16, 22, 44, 45, 50, 68, 69, 74).

- Harden, M. (2013). „Das Behrens-Fisher-Problem für hochdimensionale Split-Plot-Designs“. Magisterarb. Inst. für Mathematische Stochastik, Universität Göttingen (siehe S. 6, 16, 17, 30, 44, 47–49, 68, 71, 74).
- Helms, H.-J. (2010). „Robuste Verfahren für strukturierte hochdimensionale Repeated-Measures-Designs unter Nicht-Normalverteilung“. Diplomarbeit. Inst. für Mathematische Stochastik, Universität Göttingen (siehe S. 18, 20, 47).
- Heuser, H. (2000). *Lehrbuch der Analysis, Teil 2*. Stuttgart: Teubner (siehe S. 62, 109).
- Hoekstra, R. M. und T. C. Chenier (1994). „A note on the box correction factor for degrees of freedom in certain F tests“. *Comm. Statist. Theory Methods* **23** (5), S. 1483–1505. DOI: 10.1080/03610929408831335 (siehe S. 43, 56).
- Hofhuis, J. (2013). „Dysferlin and its role in the pathogenesis of muscular dystrophy“. Diss. Georg-August-Universität Göttingen (siehe S. 6).
- Huber, P. J. (1981). *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York (siehe S. 8, 10, 47).
- Huynh, H. (1978). „Some approximate tests for repeated measurement designs“. *Psychometrika* **43** (2), S. 161–175. DOI: 10.1007/BF02293860 (siehe S. 44).
- Huynh, H. und L. S. Feldt (1970). „Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F -Distributions“. *Journal of the American Statistical Association* **65** (332), S. 1582–1589 (siehe S. 14, 18).
- (1976). „Estimation of the Box Correction for Degrees of Freedom from Sample Data in Randomized Block and Split-Plot Designs“. *Journal of Educational Statistics* **1** (1), S. 69–82 (siehe S. 43, 44).
- IBM SPSS Statistics for Windows, Version 25.0* (2017). IBM Corp. (siehe S. 13, 15).
- Johnson, M. E. (1987). *Multivariate Statistical Simulation*. John Wiley & sons (siehe S. 13, 37).
- Kariya, T. (1981). „A Robustness Property of Hotelling’s T^2 -Test“. *The Annals of Statistics* **9** (1), S. 211–214 (siehe S. 8).
- Kariya, T. und B. K. Sinha (1989). *Robustness of statistical tests*. Boston, MA: Academic Press Inc. (siehe S. 8, 12).
- Katayama, S., Y. Kano und M. S. Srivastava (2013). „Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension“. *J. Multivariate Analysis* **116**, S. 410–421 (siehe S. 22).
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences*. 3rd ed. Brooks/Cole Publishing Co. Wadsworth Publishing Co., Inc., Belmont, Calif. (siehe S. 44).
- Krishnamoorthy, K., F. Lu und T. Mathew (2007). „A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models“. *Computational Statistics & Data Analysis* **51**, S. 5731–5742 (siehe S. 23).
- Lane, D. (2016). „The assumption of sphericity in repeated-measures designs: What it means and what to do when it is violated“. *The Quantitative Methods for Psychology* **12**, S. 114–122 (siehe S. 44, 57, 58, 66).

- Lecoutre, B. (1991). „A correction for the approximate test in repeated measures designs with two or more independent groups“. *Journal of Educational Statistics* **16**, S. 371–372 (siehe S. 44).
- Lehmann, E. L. und J. P. Romano (2005). *Testing statistical hypotheses*. Third. Springer, New York (siehe S. 64).
- Lehmann, E. L. und H. Scheffé (1950). „Completeness, similar regions, and unbiased estimation. I“. *Sankhyā* **10**, S. 305–340 (siehe S. 97).
- Lehmann, E. L. (2004). *Elements of Large-Sample Theory*. New York: Springer-Verlag (siehe S. 48, 77).
- Lindsay, B. G. und P. Basak (2000). „Moments determine the tail of a distribution (but not much else)“. *Amer. Statist.* **54** (4), S. 248–251. DOI: 10.2307/2685775 (siehe S. 75, 76).
- Linnik, J. W. (1968). *Statistical problems with nuisance parameters*. American Mathematical Soc. (siehe S. 9).
- Marshall, A. W. und I. Olkin (1979). *Inequalities: theory of majorization and its applications*. New York: Acad. Press (siehe S. 106).
- Mathai, A. M. und S. B. Provost (1992). *Quadratic forms in random variables*. **126**. Theory and applications. New York: Marcel Dekker Inc. (siehe S. 30, 77, 91, 101).
- Mathai, A. M., S. B. Provost und T. Hayakawa (1995). *Bilinear forms and zonal polynomials*. **102**. New York: Springer-Verlag. DOI: 10.1007/978-1-4612-4242-0 (siehe S. 12, 91, 101).
- Maxwell, S. E. und R. D. Arvey (1982). „Small sample profile analysis with many variables.“ *Psychological Bulletin* **92** (3), S. 778–785 (siehe S. 44).
- Neammanee, K. (2005). „On the constant in the nonuniform version of the Berry-Esseen theorem“. *International Journal of Mathematics and Mathematical Sciences* **2005**, S. 1951–1967 (siehe S. 75).
- Nelsen, R. B. (1999). *An introduction to copulas*. **139**. Springer-Verlag, New York. DOI: 10.1007/978-1-4757-3076-0 (siehe S. 75).
- Oberfeld, D. und T. Franke (2013). „Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data“. *Behavior Research Methods* **45** (3), S. 792–812. DOI: 10.3758/s13428-012-0281-2 (siehe S. 15, 41, 45).
- Park, J. und D. N. Ayyala (2013). „A test for the mean vector in large dimension and small samples“. *Journal of Statistical Planning and Inference* **143** (5), S. 929–943. DOI: <http://dx.doi.org/10.1016/j.jspi.2012.11.001> (siehe S. 22).
- Patnaik, P. B. (1949). „The non-central χ^2 and F -distributions and their applications“. *Biometrika* **36**, S. 202–232 (siehe S. 14–16).
- Pauly, M., D. Ellenberger und E. Brunner (2015). „Analysis of high-dimensional one group repeated measures designs“. *Statistics* **49** (6), S. 1243–1261. DOI: 10.1080/02331888.2015.1050022 (siehe S. 22, 23).
- Portnoy, S. (1986). „On the central limit theorem in \mathbf{R}^p when $p \rightarrow \infty$ “. *Probab. Theory Related Fields* **73** (4), S. 571–583. DOI: 10.1007/BF00324853 (siehe S. 22).

- Quintana, S. M. und S. E. Maxwell (1994). „A Monte Carlo Comparison of Seven ϵ -Adjustment Procedures in Repeated Measures Designs with Small Sample Sizes“. *Journal of Educational Statistics* **19** (1), S. 57–71 (siehe S. 44, 57, 66).
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria (siehe S. 36, 62).
- Rogan, J. C., H. J. Keselman und J. L. Mendoza (1979). „Analysis of repeated measurements“. *British Journal of Mathematical and Statistical Psychology* **32** (2), S. 269–286. DOI: 10.1111/j.2044-8317.1979.tb00598.x (siehe S. 44).
- SAS/STAT [®] 9.3 User's Guide (2011). SAS Institute Inc. (siehe S. 13, 15).
- Sattler, P. und M. Pauly (2018). „Inference for high-dimensional split-plot-designs: A unified approach for small to large numbers of factor levels“. *Electron. J. Statist.* **12** (2), S. 2743–2805. DOI: 10.1214/18-EJS1465 (siehe S. 4, 23).
- Searle, S. R., G. Casella und C. E. McCulloch (1992). *Variance components*. A Wiley-Interscience Publication. New York: John Wiley & Sons Inc. (siehe S. 14, 34).
- Searle, S. R. und M. H. J. Gruber (2017). *Linear models*. Second edition [of MR1461543]. John Wiley & Sons, Inc., Hoboken, NJ (siehe S. 5, 78).
- Secchi, P., A. Stamm und S. Vantini (2013). „Inference for the mean of large p small n data: A finite-sample high-dimensional generalization of Hotelling's theorem“. *Electron. J. Statist.* **7**, S. 2005–2031. DOI: 10.1214/13-EJS833 (siehe S. 20, 21).
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons Inc. (siehe S. 46, 93).
- Srivastava, M. S. und M. Du (2008). „A test for the mean vector with fewer observations than the dimension“. *J. Multivariate Anal.* **99** (3), S. 386–402 (siehe S. 22).
- Srivastava, M. S. und Y. Fujikoshi (2006). „Multivariate analysis of variance with fewer observations than the dimension“. *J. Multivariate Anal.* **97** (9), S. 1927–1940 (siehe S. 20, 22).
- Srivastava, M. S., S. Katayama und Y. Kano (2013a). „A two sample test in high dimensional data“. *J. Multivariate Analysis* **114**, S. 349–358 (siehe S. 22, 23).
- (2013b). „Corrigendum to „A two sample test in high dimensional data“ [J. Multivariate Analysis 114 (2013) 349–358]“. *J. Multivariate Analysis* **119**, S. 209 (siehe S. 22).
- Statistica* 13.3.1 (2018) (siehe S. 13).
- Stoloff, P. H. (1970). „Correcting for Heterogeneity of Covariance for Repeated Measures Designs of the Analysis of Variance“. *Educational and Psychological Measurement* **30** (4), S. 909–924. DOI: 10.1177/001316447003000411 (siehe S. 43).
- Tashiro, Y. (1977). „On methods for generating uniform random points on the surface of a sphere“. *Ann. Inst. Statist. Math.* **29** (2), S. 295–300. DOI: 10.1007/BF02532791 (siehe S. 37).
- Wang, H. und M. G. Akritas (2010). „Inference from heteroscedastic functional data“. *Journal of Nonparametric Statistics* **22** (2), S. 149–168. DOI: 10.1080/10485250903171621 (siehe S. 23).

Werner, C. (2002). „Dimensionsstabile Approximation für Verteilungen von zufälligen quadratischen Formen im Repeated-Measures-Design“. Diplomarbeit. Inst. für Mathematische Stochastik, Universität Göttingen (siehe S. 45, 46).