

# **Nichtparametrische Cross-Over-Verfahren**

Dissertation  
zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultäten  
der Georg-August-Universität zu Göttingen

vorgelegt von  
Bettina Kulle  
aus Stuttgart

Göttingen 2001

D 7

Referent: Prof. Dr. Edgar Brunner

Koreferent: Prof. Dr. Manfred Denker

Tag der mündlichen Prüfung: 30. Januar 2002

## Danksagung

Mein besonderer Dank gilt Herrn Prof. Dr. Edgar Brunner, der mich während der Entstehung dieser Arbeit durch seine engagierte Betreuung maßgeblich unterstützte. Ferner bedanke ich mich bei ihm für die hervorragenden Arbeitsmöglichkeiten in der Abteilung Medizinische Statistik der Universität Göttingen.

Für die Erstellung der Gutachten bin ich Herrn Prof. Brunner und Herrn Prof. Manfred Denker sehr dankbar.

Ein besonderes Dankeschön richte ich an Dr. Sebastian Domhof, der stets ein kompetenter und hilfsbereiter Ansprechpartner für mich war.

Ebenfalls danken möchte ich Anke Görlitz, Markus und Björn Andreassen sowie Jan-Frederik Engelhardt, die beim Korrekturlesen den einen oder anderen wertvollen Tip für mich hatten.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Beispiele</b>	<b>5</b>
2.1	Asthma-Studie I . . . . .	5
2.2	Asthma-Studie II . . . . .	7
<b>3</b>	<b>Der Cross-Over-Plan</b>	<b>9</b>
3.1	Motivation . . . . .	9
3.2	Cross-Over-Versuchsanlage . . . . .	9
3.3	Elementare grafische Darstellungen des Datenmaterials . . . . .	10
3.4	Effekte im Cross-Over-Plan . . . . .	13
3.5	Modelle . . . . .	17
3.5.1	Lineares Modell . . . . .	17
3.5.2	Nichtparametrisches Modell . . . . .	19
<b>4</b>	<b>Statistiken für den Cross-Over-Plan</b>	<b>21</b>
4.1	Lineares Modell . . . . .	21
4.1.1	Schätzer der Effekte . . . . .	22
4.1.1.1	Schätzer für den Residualunterschied . . . . .	22
4.1.1.2	Schätzer für den Periodenunterschied . . . . .	22
4.1.1.3	Schätzer für den Behandlungsunterschied . . . . .	23
4.1.2	Hypothesen und Tests . . . . .	24
4.1.2.1	Hypothesen und Tests für einen Residualunterschied . . . . .	24
4.1.2.2	Hypothesen und Tests auf einen Periodenunterschied . . . . .	25
4.1.2.3	Hypothesen und Tests auf einen Behandlungsunterschied (1) . . . . .	26
4.1.2.4	Hypothesen und Tests auf einen Behandlungsunterschied (2) . . . . .	28
4.2	Nichtparametrisches Modell . . . . .	28
4.2.1	Motivation und Definition . . . . .	28
4.2.2	Hypothesen . . . . .	30
4.2.3	Schätzer . . . . .	34
4.2.4	Harmonische Ränge . . . . .	35
4.2.5	Asymptotik und Grenzverteilungen . . . . .	37
4.2.6	Teststatistiken für $H_0^F$ . . . . .	39
4.2.7	Teststatistiken für $H_0^P$ . . . . .	41

4.2.8	Eigenschaften der Teststatistiken . . . . .	42
4.3	Auswertung der Beispiele . . . . .	43
4.3.1	Asthma-Studie II . . . . .	43
4.3.2	Asthma-Studie I . . . . .	45
<b>5</b>	<b>Der Nacheffekt</b>	<b>47</b>
5.1	Allgemeines und Historisches . . . . .	47
5.2	Grafische Bearbeitung des Problems . . . . .	49
5.3	Die Two-Stage-Analyse . . . . .	52
5.4	Konfidenzintervallverfahren . . . . .	53
5.5	Äquivalenztests . . . . .	55
<b>6</b>	<b>Baseline-Werte</b>	<b>59</b>
6.1	Einführung und Motivation . . . . .	59
6.2	Lineares Modell . . . . .	60
6.3	Nichtparametrisches Modell . . . . .	64
6.3.1	Notation und Modell . . . . .	64
6.3.2	Regressionsmodell . . . . .	67
6.3.3	Asymptotik und Grenzverteilungen . . . . .	68
6.3.4	Teststatistiken für $H_0^F$ . . . . .	72
6.3.5	Teststatistiken für $H_0^{preg}$ . . . . .	74
6.4	Auswertung der Asthma-Studie II . . . . .	76
6.4.1	Lineares Modell . . . . .	76
6.4.2	Nichtparametrisches Modell . . . . .	76
<b>7</b>	<b>Simulation</b>	<b>79</b>
7.1	Niveausimulation . . . . .	79
7.1.1	Simulation ohne Baseline-Werte . . . . .	80
7.1.2	Simulationsvergleiche mit den Verfahren des linearen Modells . . . . .	85
7.1.3	Simulation mit Baseline-Werten . . . . .	88
7.1.4	Simulation des Äquivalenzverfahrens . . . . .	91
7.1.5	Simulation des Konfidenzintervallverfahrens . . . . .	91
7.2	Powersimulationen . . . . .	92
7.2.1	Simulation ohne Baseline-Werte . . . . .	94
7.2.2	Simulationsvergleiche mit den Verfahren des linearen Modells . . . . .	97
7.2.3	Simulation mit Baseline-Werten . . . . .	99
7.2.4	Simulation des Äquivalenzverfahrens . . . . .	104
<b>8</b>	<b>Zusammenfassung und Ausblick</b>	<b>107</b>
<b>A</b>	<b>Theorie</b>	<b>111</b>
A.1	Notation . . . . .	111
A.1.1	Vektoren und Matrizen . . . . .	111
A.1.2	Konvergenz . . . . .	112
A.2	Grundlegende Lemmata . . . . .	113
A.2.1	Grenzwertsätze . . . . .	113
A.3	Definitionen, Sätze und Beweise . . . . .	114

<b>Literatur</b>	<b>131</b>
<b>Lebenslauf</b>	<b>137</b>





# Kapitel 1

## Einleitung

In mehreren Phasen ihrer Entwicklung werden Medikamente oder Behandlungen für verschiedene Krankheiten im Rahmen klinischer Studien miteinander verglichen. In der Studienform, die in dieser Arbeit betrachtet wird, werden die Medikamente bzw. Behandlungen beurteilt, indem ihre Effekte in den Versuchseinheiten verglichen werden. Bei einem **Cross-Over-Verfahren** wird jede Versuchseinheit zwei oder mehr unterschiedlichen Medikamenten bzw. Behandlungen unterzogen. Die Reihenfolge, in der diese verabreicht werden, hängt vom gewählten Design ab. Der einfachste und verbreitetste Versuchsaufbau eines Cross-Over-Plans ist das  $2 \times 2$ -Design, welches auch in dieser Arbeit betrachtet wird. Jede Versuchseinheit erhält zwei verschiedene Behandlungen  $A$  und  $B$ . Ferner werden die Patienten den Gruppen 1 und 2 zufällig zugeteilt. Die erste Gruppe erhält zunächst einmal  $A$ , und dann, nach einer zuvor festgelegten Zeitspanne, Behandlung  $B$ . Die verbleibenden Versuchseinheiten aus Gruppe 2 erhalten die Behandlungen oder Medikamente genau in umgekehrter Reihenfolge. Der Unterschied zwischen den beiden Gruppen liegt demnach in der Reihenfolge der Behandlungen. Das Design dieses Standard-Cross-Over-Plans wird daher häufig auch synonym mit  $AB/BA$ -Design bezeichnet. Das Ziel, das mit diesem Design verfolgt wird, ist die Verringerung der Zahl der Versuchseinheiten bei gleichbleibender Präzision oder entsprechender Erhöhung der Präzision bei gleichbleibender Zahl der Versuchseinheiten.

Erstmals erwähnt wurde das Cross-Over-Design Mitte des 19. Jahrhunderts in der Landwirtschaft, als Bennett Lawes dieses Verfahren genutzt hat, um seine Ansichten bzgl. der Düngung von Feldern gegenüber Justus von Liebig zu rechtfertigen. LIEBIG (1847) meinte, dass die Zugabe von Mineralien nötig sei. In seinem Versuch konnte Lawes jedoch nachweisen, dass das Feld mit der Reihenfolge Mineralien-Ammoniak zu guter Ernte führte, während das Feld mit umgekehrter Düngungsreihenfolge am Ende eine sehr viel schlechtere Ernte hervorbrachte. Damit versuchte er seine These zu untermauern, dass Ammoniak der bessere Dünger sei. Die erste bekannte *klinische* Studie mit dem hier besprochenen Design wurde von CUSHNY UND PEEBLES (1905) durchgeführt. Diese Studie konnte zwar noch nicht formal ausgewertet werden, war aber Erklärungsgrundlage vieler später entwickelter Verfahren, siehe z.B. STUDENT (1908) oder FISHER (1925). COCHRAN (1939) trennte erstmals die Behandlungseffekte in den *reinen Behandlungseffekt*, der lediglich für den Unterschied der beiden Behandlungen steht, und den *Überhangseffekt*, d.h. die Nachwirkungen der Behandlungen aus der ersten Periode. Die Begriffe Überhangs-, Nach- und Residualeffekt werden synonym verwendet. GRIZZ-

LEs Arbeit von 1965 ist zusammen mit der Arbeit von HILLS UND ARMITAGE (1979) der Meilenstein in der Geschichte der Tests für den Cross-Over-Plan. Die in diesen Arbeiten vorgestellten parametrischen Verfahren werden mit Ausnahme von einer Handvoll Verfeinerungen bis zum heutigen Tag eingesetzt.

Aufgrund der speziellen Problematiken, die dieses Design mit sich bringt, ist das Cross-Over-Design weder universell einsetzbar, noch bringt es unter allen Umständen eine Verbesserung der Effizienz. Die Einsetzbarkeit beschränkt sich aufgrund des Designs auf Behandlungsmethoden, deren Wirkung abschätzbar lange anhält. Ebenso werden Cross-Over-Verfahren im klinischen Bereich nur bei Krankheiten angewandt, die einen chronischen Verlauf haben. Das größte Problem jedoch stellt ein eventueller Nacheffekt dar, der eintreten kann, falls die Wirkung der Behandlung aus der ersten Periode zu Beginn der zweiten Behandlung noch nicht völlig abgeklungen ist. Insbesondere BROWN (1980) hat auf die damit verbundenen Schwierigkeiten des betrachteten Designs aufmerksam gemacht:

- Falls ein Residualeffekt nicht a priori ausgeschlossen werden kann, ist unklar, wie der Test auf den Behandlungseffekt im Weiteren zu handhaben ist, bzw. wie der (verallgemeinerte) Behandlungseffekt im Folgenden zu interpretieren ist.
- Die zuerst von GRIZZLE (1965) vorgeschlagene Two-Stage-Methode ist ein zweistufiges Verfahren, dass zuerst auf einen Residualeffekt testet. Bei Nichtablehnung des Tests wird dieser als nicht vorhanden vorausgesetzt. Diese Konklusion widerspricht der Aussagekraft eines statistischen Tests, denn man schließt aus der Nichtablehnung der Hypothese fälschlicherweise auf deren Gültigkeit.

Ein restriktives Modell für die Cross-Over-Versuchsanlage stellt das weit verbreitete lineare Modell dar, in dem vorausgesetzt wird, dass sich die einzelnen möglicherweise auftretenden Effekte additiv zusammensetzen. Im linearen Modell sind das neben dem Erwartungswert der Stichprobe, dem individuellen Effekt und dem Versuchsfehler insbesondere folgende Effekte: der Behandlungs-, Perioden- und Nacheffekt. In GRIZZLES Modell wird zudem auch die Normalverteilung der Daten vorausgesetzt. Diese Voraussetzungen, sowohl die Linearität als auch die Verteilungsannahmen, sollen mit dem in dieser Arbeit diskutierten Ansatz fallengelassen werden. Für die Zielvariablen wird hier lediglich mindestens ein ordinales Niveau erwartet, d.h. es muss eine Anordnung in den Daten vorhanden sein. Ein Beispiel stellt die häufige Verwendung von Scores dar, wie etwa Schmerz- oder Angst-Scores.

Der erste nichtparametrische Ansatz wurde für das klassische lineare Modell des Cross-Over-Designs von KOCH (1972) vorgeschlagen. Diese Verfahren setzten immer noch stetige Verteilungsfunktionen voraus und waren somit ungeeignet für ordinale Daten und immer noch auf das restriktive additive Modell beschränkt. Der Unterschied zu den zuvor erklärten Verfahren von GRIZZLE bestand lediglich darin, dass die Normalverteilungsannahme fallen gelassen wurde und demzufolge statt mit dem unverbundenen  $t$ -Test mit dem Wilcoxon-Mann-Whitney-Test gearbeitet wurde. In allgemeinen Versuchdesigns wurden relative Effekte eingesetzt, wofür kein lineares Modell oder Verteilungsannahmen zugrunde gelegt werden mussten. AKRITAS UND ARNOLD (1994) setzten diese Effekte ein und entwickelten Rangtests, die nicht auf Summen oder Differenzen der Daten beruhten. Damit erreichte man die Invarianz des Testergebnisses unter monotonen Transformationen der Daten und konnte mit ordinalen Daten sinnvoll umgehen. Die Hypothesen,

unter denen die Teststatistiken hergeleitet wurden, wurden dabei stets über die Verteilungsfunktionen gestellt. Diese Tests basieren alle auf den gewichteten relativen Effekten, die die Eigenschaft besitzen, von den Stichprobenumfängen abhängig zu sein, weshalb in KULLE (1999) für unverbundene Stichproben und in SIEMER (1999) im Rahmen der Kovarianzanalyse erstmals die ungewichteten relativen Effekte eingesetzt wurden. Diese sind zum Einen selbst unabhängig von den Stichprobenumfängen, und zum Anderen ist auch bei Hypothesenstellung über diese Art von Effekten die Hypothese unabhängig von den Stichprobenumfängen, was sonst offensichtlich zu Interpretationsproblemen führt. In dieser Arbeit werden im später eingeführten nichtparametrischen Modell für das Cross-Over-Design die ungewichteten relativen Effekte benutzt, um Teststatistiken sowohl unter  $H_0^F : \mathbf{cF} = 0$  als auch unter  $H_0^p : \mathbf{cp} = 0$  herzuleiten.

Die Meinungen über die Anwendbarkeit des Cross-Over-Designs liegen weit auseinander. TRAMPISCH (1995) bemerkte treffend, was charakteristisch für das Meinungsfeld zu diesem Design ist: Seit der Veröffentlichung von GRIZZLE (1965) variierten die Ansichten über die Anwendbarkeit des Cross-Over-Verfahrens. Von jeglichem Absehen der Durchführung einer Studie mit Hilfe dieses Designs bis hin zur universellen Einsetzbarkeit, auch bei nicht chronischen Krankheiten, ginge dabei die Skala. Heute hingegen sei man sich weitestgehend einig, dass dieses Verfahren in bestimmten Fällen inklusive pharmakokinetischen und Bioäquivalenz-Studien sehr nützlich ist. Hinzu käme das Feld der chronischen Krankheiten, in dem Cross-Over-Verfahren allgemein benutzt und akzeptiert werden.

Der Aufbau dieser Arbeit gestaltet sich wie folgt: Im Anschluss an diese Einleitung werden im 2. Kapitel zwei Beispiele vorgestellt, die durch die gesamte Arbeit hindurch die Vorgehensweise und einzelne Aspekte der betrachteten Verfahren veranschaulichen sollen. Anschließend wird der Cross-Over-Plan in Kapitel 3 vorgestellt. Dabei werden neben der Motivation, der Vorstellung der Versuchsanlage und des Aufzeigens, wie das Datenmaterial sinnvoll grafisch aufbereitet werden kann, auch die einzelnen im Cross-Over-Plan auftretenden Effekte beschrieben. Das Kapitel schließt mit der Vorstellung des linearen und nichtparametrischen Modells. In Kapitel 4 werden zunächst die bereits bekannten Tests auf die entsprechenden Effekte im linearen Modell für das Cross-Over-Design vorgestellt. Danach werden die Statistiken für das dieser Arbeit zugrunde gelegte nichtparametrische Modell hergeleitet. Dabei wird als Effekt der ungewichtete relative Effekt gewählt. Es werden Tests für die in diesem Modell von AKRITAS UND ARNOLD (1994) eingeführten Hypothesen über die Verteilungsfunktionen ( $H_0^F : \mathbf{cF} = 0$ ) hergeleitet sowie auch Tests für Hypothesen, die über die zuvor erwähnten relativen Effekte gestellt werden ( $H_0^p : \mathbf{cp} = 0$ ). Die Ergebnisse werden mit den Ergebnissen auf Basis des linearen Modells verglichen und diskutiert. Mathematische Beweise und detaillierte oder notationsintensive Rechnungen werden dabei zugunsten der Lesbarkeit in den Anhang ausgegliedert. Zum Abschluss dieses Kapitels werden die Beispiele ausgewertet. Kapitel 5 erläutert die Problematik, die auftritt, wenn ein Nacheffekt nicht von vornherein ausgeschlossen werden kann. Dabei wird neben der grundsätzlichen Klärung des Sachverhalts auch auf verschiedene mögliche Vorgehensweisen für diesen Fall eingegangen. Ein besonderes Augenmerk liegt auf der Einsetzbarkeit von Äquivalenztests für eine modifizierte Two-Stage-Analyse. Kapitel 6 widmet sich ganz der Einbeziehung von Baseline-Werten in die Analyse eines Cross-Over-Problems. Zunächst wird dabei auf eine Lösungsmöglichkeit im linearen Modell eingegangen, die Adjustierung der Mess-

werte mit den entsprechenden Baseline-Werten. Anschließend wird die Anwendung der nichtparametrischen Kovarianzanalyse nach SIEMER (1999) auf den um Baseline-Werte ergänzten Cross-Over-Plan erläutert. In Kapitel 7 werden die neuen Statistiken einer genauen Überprüfung bezüglich der Einhaltung des Niveaus unterzogen und ihre Power untersucht. Dabei werden die Teststatistiken für beide Arten von nichtparametrischen Hypothesen ( $H_0^F$  und  $H_0^p$ ) überprüft. Ferner wird ein Vergleich zwischen den Verfahren im linearen Modell und den in dieser Arbeit entwickelten neuen Verfahren gezogen. Dazu kommen noch Untersuchungen des Konfidenzintervallverfahrens und Äquivalenztests. Dieses Kapitel soll einen Einblick geben, wie gut die verschiedenen Tests unter Berücksichtigung verschiedener Effekte oder mit bzw. ohne Baseline-Werte sind. Dies alles geschieht mit der Intention, dass der Anwender eine Vorstellung davon bekommt, ob für seinen eigenen zugrunde liegenden Versuch die vorgestellten Verfahren in Frage kommen und welche Eigenschaften sie haben. Im Kapitel 8 werden die Ergebnisse der vorliegenden Arbeit zusammengefasst, diskutiert und Ansätze für weitere mögliche Untersuchungen bzw. Weiterentwicklungen gegeben.

Ziel dieser Arbeit ist es, nichtparametrische Tests unter  $H_0^F$  und  $H_0^p$  auf das Cross-Over-Design zu übertragen und dabei Interpretationen, Schwierigkeiten und Anwendbarkeit in diesem speziellen Design aufzuzeigen. Ferner geht es darum, dem Anwender eine theoretische Grundlage zu geben, wie er ohne jede Verteilungs- oder Modell-Annahme, also beispielsweise auch für ordinale Daten, Cross-Over-Pläne auswerten kann. Dabei werden auch ausführlich die Problematik mit dem Nacheffekt und mögliche Baseline-Werte berücksichtigt.

# Kapitel 2

## Beispiele

In diesem Kapitel werden zwei Beispiele vorgestellt. Zunächst ein Beispiel im klassischen  $AB/BA$ -Cross-Over-Design mit ordinalen Daten: die Asthma-Studie I. Anschließend wird die Asthma-Studie II vorgestellt mit stetigen Daten und zusätzlicher Berücksichtigung von Baseline-Werten.

### 2.1 Asthma-Studie I

In einer Doppel-Blind Studie aus SENN (1993) werden 7-13 Jahre alte asthmakranke Kinder zufällig einer der beiden Gruppen 1 bzw. 2 zugeteilt. Die Studie ist als Cross-Over-Plan angelegt, und zwar sei Behandlung  $A$ :  $12\mu g$  Formoterol Spraylösung und Behandlung  $B$ :  $200\mu g$  Salbutamol Spraylösung. Zum Messzeitpunkt wurde ein Reiz in der Lunge herbeigeführt, was sich in einem reduzierten  $FEV_1$ -Wert (forced expired volume in one second) niederschlägt. Jedes Kind wurde am Versuchstag derselben Versuchsbedingung ausgesetzt, und es wurden jeweils 2 und 8 Stunden nach der Behandlungen  $A$  oder  $B$  verschiedene Lungenfunktionsparameter gemessen. Die Ergebnisse wurden mit Hilfe einer Vierpunkte-Skala zusammengefasst:

**1:** schlecht

**2:** mäßig

**3:** mittel

**4:** gut

Die Einstufung nahm stets der Untersuchungsleiter vor.

Die Kinder aus Gruppe 1 erhielten zunächst  $12\mu g$  Formoterol Spraylösung (Behandlung  $A$ ) und in der zweiten Periode  $200\mu g$  Salbutamol Spraylösung (Behandlung  $B$ ). In Gruppe 2 wurden diese Behandlungen in entgegengesetzter Reihenfolge vorgenommen. Die resultierenden Daten dieser Studie befinden sich in Tabelle 2.1.

In diesem Beispiel haben wir das Versuchsdesign aus Tabelle 2.2. In dieser Tabelle sind die Behandlungen der beiden Gruppen zu den verschiedenen Zeitpunkten festgehalten. An dieser Stelle sieht man deutlich, woher der Name *Cross-Over-Design* kommt, denn bildlich gesprochen werden die Behandlungen über Kreuz vergeben.

Tabelle 2.1: Untersuchungsergebnisse Asthma-Studie I

Gruppe	Indiv.	Periode 1	Periode 2
1	3	4	4
	4	3	1
	7	4	1
	8	4	3
	9	4	4
	11	4	3
	15	4	3
	16	4	1
	19	4	3
	20	4	1
	22	4	3
23	4	2	
2	1	4	2
	2	4	3
	5	4	4
	6	4	4
	10	4	4
	12	4	4
	13	4	4
	14	3	4
	17	4	3
	18	4	2
	21	4	2
	24	4	3

Tabelle 2.2: Behandlungsabfolge in der Asthma-Studie I

Gruppe	Zeitpunkt	
	Periode 1	Periode 2
1	A	B
2	B	A

Tabelle 2.3:  $FEV_1$ -Werte der Asthma-Studie II

Gruppe	Indiv.	Run-In	Periode 1	Wash-Out	Periode 2
1	1	1.09	1.28	1.24	1.33
	2	1.38	1.60	1.90	2.21
	3	2.27	2.46	2.19	2.43
	4	1.34	1.41	1.47	1.81
	5	1.31	1.40	0.85	0.85
	6	0.96	1.12	1.12	1.20
	7	0.66	0.90	0.78	0.90
	8	1.69	2.41	1.90	2.79
2	1	1.74	3.06	1.54	1.38
	2	2.41	2.68	2.13	2.10
	3	3.05	2.60	2.18	2.32
	4	1.20	1.48	1.41	1.30
	5	1.70	2.08	2.21	2.34
	6	1.89	2.72	2.05	2.48
	7	0.89	1.94	0.72	1.11
	8	2.41	3.35	2.83	3.23
	9	0.96	1.16	1.01	1.25

Dieser Datensatz konnte mit bisher vorhandenen Verfahren unter Berücksichtigung von allen Informationen nicht ausgewertet werden, da bisher kein Verfahren in der Lage war, ordinale Daten in einem Cross-Over-Plan auf einen Behandlungsunterschied zu untersuchen. In SENN (1993) wurde der Datensatz folgendermaßen gehandhabt. Die vier Scores (1-4) wurden auf zwei Scores reduziert. Das Zeichen - entsprach dem Score 1 bis 3 und + dem Score 4. Daraufhin wurde ein Verfahren für binäre Daten angewandt.

## 2.2 Asthma-Studie II

Die in der Studie beobachteten Patienten haben leichtes bis akutes bronchiales Asthma und wurden den beiden Behandlungsgruppen *A* und *B* zufällig zugeteilt. Die Behandlungsgruppen unterscheiden sich in der Reihenfolge der Vergabe einer einmaligen Dosis von zwei Medikamenten, die im Folgenden mit *A* und *B* bezeichnet seien. Zielvariable ist die Ein-Sekunden-Kapazität  $FEV_1$  in Litern (forced expired volume in one second). Behandlungsgruppe 1 bekommt die Behandlungen in der Reihenfolge *AB* und Behandlungsgruppe 2 in umgekehrter Reihenfolge. Der erste Baseline-Wert wurde unmittelbar vor der Behandlung zum 1. Zeitpunkt erhoben und in der Spalte mit der Überschrift *Run-In* abgetragen. 2-3 Stunden nach Verabreichung des ersten Medikaments wurden  $FEV_1$ -Messungen vorgenommen. Der Mittelwert dieser Messungen befindet sich in der Spalte mit der Überschrift *Periode 1*. Nach angemessener Zeit wurde unmittelbar vor der Verabreichung vom jeweils anderen Medikament der zweite Baseline-Wert (*Wash-Out*) erhoben. Wiederum nach ca. 2-3 Stunden wurden dann erneut  $FEV_1$ -Messungen durchgeführt. Der Mittelwert befindet sich in der Spalte mit der Überschrift *Periode 2*. Der resultierende Datensatz von PATEL (1983) ist in Tabelle 2.3 angegeben.

Tabelle 2.4: Behandlungsabfolge in der Asthma-Studie II

Gruppe	Zeitpunkt			
	Run-In	Periode 1	Wash-Out	Periode 2
1	0	A	0	B
2	0	B	0	A

Damit erhält man das folgende Versuchsdesign aus Tabelle 2.4. In dieser Tabelle sind die Behandlungen der beiden Gruppen zu den verschiedenen Zeitpunkten festgehalten. Eine 0 bezeichnet keine Behandlung. Diese beiden Werte nennt man auch Baseline-Werte, weil Messungen zu diesen Zeitpunkten unter der Annahme erhoben werden, dass *kein* Effekt einer untersuchten Behandlung vorliegt.

Es soll hier nicht unerwähnt bleiben, dass an diesem Beispiel auch Kritik geübt werden kann. PATEL (1983) macht in seiner Arbeit keine Angabe darüber, ob die Patienten randomisiert wurden, so dass an dieser Stelle die Annahme getroffen werden muss, dass die Patienten den beiden Gruppen randomisiert zugeteilt wurden. GRIEVE UND SENN (1998) kritisieren einige weitere Punkte, z.B. dass der Run-In-Wert nur aus einer Messung besteht und der Wert von Periode 1 das Mittel von zwei Versuchen ist, was unterschiedliche Varianzen zur Folge hat. Auch offensichtliche Tippfehler in der Originalarbeit werden angemahnt.

Der Unterschied zum vorigen Beispiel besteht darin, dass hier zusätzlich sogenannte Baseline-Werte erhoben wurden. Dabei handelt es sich um Werte, die nicht unter Behandlung erhoben wurden. Den ersten Baseline-Wert bezeichnet man auch als *Run-In*-Wert. Dieser wird vor Beginn der Behandlung in der ersten Periode erhoben und ist damit in jedem Fall unbeeinflusst von den darauffolgenden Behandlungen. Der zweite Baseline-Wert (*Wash-Out*) wird nach der Wash-Out-Phase erhoben. Als Wash-Out-Phase wird die Erholungsphase nach der ersten Behandlung bezeichnet, in der möglichst die Wirkung der vergangenen Behandlungen abklingt. Vernachlässigt man die Baseline-Werte, so hat man wieder den Standard-Cross-Over-Plan aus dem vorigen Beispiel.



# Kapitel 3

## Der Cross-Over-Plan

### 3.1 Motivation

Cross-Over-Versuchspläne haben die Eigenschaft, dass jede Versuchseinheit sowohl in der ersten Periode als auch in der zweiten Periode beurteilt wird. Damit verfolgt man das Ziel, die Streuung der Kenngrößen im Vergleich zu Zwei-Stichproben-Designs zu reduzieren. In der Praxis sollten demnach weniger Versuchseinheiten bei gleichbleibender Präzision benötigt werden, oder die Präzision des Tests wird bei gleichbleibender Anzahl an Versuchseinheiten verbessert. In diesem Versuchsdesign wird aber auch der Tatsache Rechnung getragen, dass jede Versuchseinheit auf eine Behandlung anders reagiert. Es ist dabei berücksichtigt, dass aufgrund eines möglichen Zeiteffekts die Verteilungen in beiden Perioden verschieden sein können, auch wenn kein Behandlungsunterschied vorliegt. Ferner kann eine Behandlung, die in der ersten Periode angewandt wurde, in der zweiten Periode einen Nacheffekt haben, der dann den Behandlungseffekt der nachfolgenden Behandlung überlagert.

Wenn es um den Vergleich zweier Behandlungen geht, können Cross-Over-Pläne nicht immer sinnvoll eingesetzt werden. In klinischen Studien sind sie beispielsweise dann ungeeignet, wenn Patienten nach der ersten Periode bereits ein stark abgeschwächtes Krankheitsbild aufweisen oder sogar geheilt sind. Dann wäre die Untersuchung des anderen Medikaments in der zweiten Periode selbstverständlich überflüssig. Die Anwendung von Cross-Over-Plänen ist demnach lediglich dann ratsam, wenn der Behandlungseffekt nicht dauerhaft ist, damit die Behandlung in der zweiten Periode überhaupt noch sinnvoll ist. Klassische Einsatzbereiche sind Ernährungsstudien, Therapiestudien bei chronischen Erkrankungen sowie Bioverfügbarkeitsstudien.

### 3.2 Cross-Over-Versuchsanlage

Die Versuchsanlage eines Cross-Over-Plans, im Folgenden kurz mit COP bezeichnet, ist ein Spezialfall der Versuchsanlagen mit longitudinalen Daten, wie sie in BRUNNER UND LANGER (1999) vorgestellt wurden. Man hat hier einen Whole-Plot-Faktor Gruppe (Faktor  $G$ ) und einen Sub-Plot-Faktor Zeit (Faktor  $T$ ), da jedes Individuum zu zwei Zeitpunkten gemessen wird. Die einzelnen Stufen lassen sich als Überlagerung verschiedener Effekte (Nach-, Behandlungs- und Periodeneffekt) darstellen.

Tabelle 3.1: Allgemeine Versuchsanlage des Standard-COPs

		Faktor $T$	
		$s = 1$	$s = 2$
$i = 1$	$k = 1$	$X_{111}$	$X_{112}$
	$\vdots$	$\vdots$	$\vdots$
	$k = n_1$	$X_{1n_11}$	$X_{1n_12}$
$i = 2$	$k = 1$	$X_{211}$	$X_{212}$
	$\vdots$	$\vdots$	$\vdots$
	$k = n_2$	$X_{2n_21}$	$X_{2n_22}$

Aus Tabelle 3.1 ist ersichtlich, dass  $X_{iks}$  den Wert der  $k$ -ten Versuchseinheit der  $i$ -ten Gruppe zum Zeitpunkt  $s$  mit  $i = 1, 2$ ,  $k = 1, \dots, n_i$  und  $s = 1, 2$  beschreibt. Faktor  $G$  beschreibt die entsprechende Gruppe  $i = 1, 2$  und Faktor  $T$  den Zeitpunkt bzw. die Periode  $s = 1, 2$ .

Ein Beispiel für einen Standard-COP mit ordinalen Daten ist die Asthma-Studie I aus dem Datensatz von SENN (1993), der in Abschnitt 2.1 eingeführt wurde. Hier gibt es die beiden Gruppen 1 und 2, wobei in Gruppe 1 Formoterol und danach Salbutamol verabreicht wurde. In der 2. Gruppe werden die gleichen Substanzen in umgekehrter Reihenfolge gegeben. Die Werte  $X_{iks}$  entsprechen den Scores, die der Untersucher dem entsprechenden  $k$ -ten Individuum in der  $i$ -ten Gruppe zum Zeitpunkt  $s$  nach der Bestimmung der Lungenfunktionsparameter vergeben hat.

Die Standard-Versuchsanlage aus Tabelle 3.1 kann bei der Hinzunahme von Baseline-Werten (siehe Kapitel 6) erweitert werden. Handelt es sich um einen Eingangswert (Run-In-Wert), also eine Beobachtung vor Beginn der Behandlungen, so ordnet man ihn mit der Bezeichnung  $s = R$  vor dem Wert für Zeitpunkt 1 an. Der in der Wash-Out-Phase erhobene Wert liegt zeitlich zwischen dem Wert vom ersten Zeitpunkt ( $s = 1$ ) und zweiten Zeitpunkt ( $s = 2$ ). Die Indizierung ist entsprechend  $s = W$  für den Wert am Ende der Wash-Out-Phase. Ein Beispiel mit Baseline-Werten vor der ersten Periode und am Ende der Wash-Out-Phase ist die von PATEL (1983) beschriebene Asthma-Studie II. Hier wurde vor und nach den beiden Behandlungen bei jedem Patienten die Ein-Sekunden-Kapazität gemessen. In Tabelle 3.2 ist die allgemeine Versuchsanlage mit Baseline-Werten dargestellt. Sie entspricht für den Fall  $n_1 = 8$  und  $n_2 = 9$  gerade dem Versuchsplan der Asthma-Studie II.  $X_{iks}$  bezeichnet den  $FEV_1$ -Wert des  $k$ -ten Patienten der  $i$ -ten Gruppe zum Zeitpunkt  $s = R, 1, W, 2$  mit  $s = R$  bzw.  $s = W$  für die beiden Baseline-Werte.

### 3.3 Elementare grafische Darstellungen des Datenmaterials

Ergänzend zu den Verfahren der schließenden Statistik können Grafiken zur Veranschaulichung von zugrundeliegenden Trends betrachtet werden. KENWARD UND JONES (1989) empfehlen, sich bei jeder Studie mit Hilfe von geeigneten Grafiken einen Überblick über das vorhandene Datenmaterial zu verschaffen. Grafiken stellen stets in unverfälschter Form die Daten mit ihren Eigenarten und Trends dar. Dazu gibt es in erster

Tabelle 3.2: Allgemeine Versuchsanlage des COPs mit Baseline-Werten

Gruppe		Zeitpunkt			
		Run-In	Periode 1	Wash-Out	Periode 2
$i = 1$	$k = 1$	$X_{11R}$	$X_{111}$	$X_{11W}$	$X_{112}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$k = n_1$	$X_{1n_1R}$	$X_{1n_11}$	$X_{1n_1W}$	$X_{1n_12}$
$i = 2$	$k = 1$	$X_{21R}$	$X_{211}$	$X_{21W}$	$X_{212}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$k = n_2$	$X_{2n_2R}$	$X_{2n_21}$	$X_{2n_2W}$	$X_{2n_22}$

Linie drei Arten:

- Für beide Gruppen werden die Beobachtungen nach der Zeit in jeweils einen Grafen aufgetragen und pro Versuchseinheit miteinander verbunden. Mit Hilfe dieser Darstellung wird der behandlungsabhängige Trend der Daten über die beiden Perioden offengelegt. Welcher Trend aus welchen Anteilen besteht, ist ohne weitere Voraussetzungen nicht auszumachen. Diese Form der Darstellung ist lediglich für kleine Stichprobenzahlen geeignet (siehe Abbildung 3.1), da sonst die schlechte Übersichtlichkeit die Vorteile dieser grafischen Darstellung kompensiert.
- Es werden die Mittelwerte beider Gruppen gegen die Zeit aufgetragen und pro Behandlung die Punkte verbunden. Hier wird die mittlere Tendenz der Daten im Hinblick auf die Veränderung der jeweiligen Behandlung deutlich (siehe Abbildung 3.2), demnach die Frage beantwortet: Wie ändern sich die Werte der jeweiligen Behandlungen beider Gruppen in den Perioden 1 und 2?
- Es werden die Mittelwerte beider Gruppen gegen die Zeit aufgetragen und pro Gruppe die Punkte verbunden. Hier werden die mittleren Behandlungswerte dargestellt. Man sieht den Verlauf der Werte innerhalb der Gruppe (siehe Abbildung 3.3), also die Behandlungsunterschiede innerhalb einer Gruppe und die Verlaufsunterschiede zwischen den Gruppen.

Die letzten beiden Darstellungen implizieren streng genommen bereits die metrische Skalierung der Daten, da Mittelwerte der Beobachtungen gebildet werden. Im Falle ordinaler Daten kann die Kenngröße aber auch abgewandelt werden und durch den Median bzw. relativen Effekt (siehe nachfolgendes Kapitel) ersetzt werden. Der Grafiktyp ändert sich dadurch nicht. Es handelt sich weiterhin um Grafiken, die den Trend innerhalb der Behandlungen (Abbildung 3.2) bzw. der Gruppen (Abbildung 3.3) über die Zeit betrachten. In den Grafiken werden mit  $A$  und  $B$  die beiden Behandlungen, die den beiden Gruppen in umgekehrter Reihenfolge zugute kommen, bezeichnet. Die Gruppen werden lediglich mit 1 bzw. 2 beziffert. Die vertikale Achse repräsentiert die Zielvariable, wohingegen auf der horizontalen Achse die Zeit (also die erste und zweite Periode) aufgetragen ist.

Für die Grafiken wurde der Datensatz der Asthma-Studie II (siehe S. 7) von PATEL (1983) mit den Werten unter Behandlung von Periode 1 und 2 benutzt. Abbildung 3.1 zeigt für Gruppe 1 einen Trend von  $FEV_1$ -Werten unter Behandlung  $B$  größer zu sein als unter

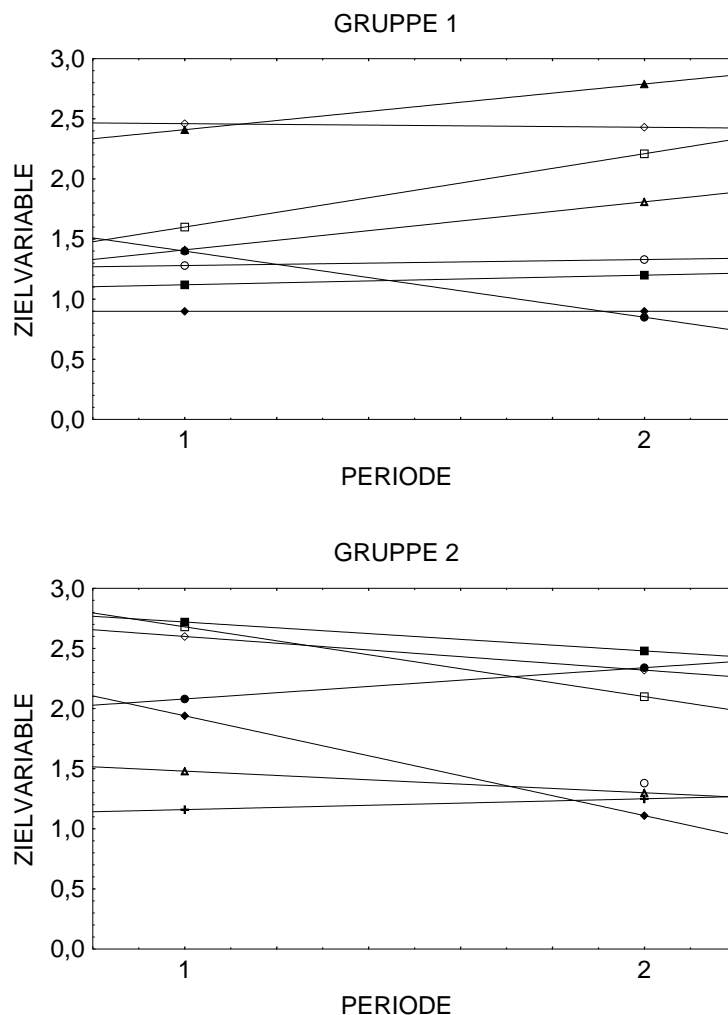


Abbildung 3.1: Die Beobachtungen pro Gruppe nach der Zeit

Behandlung *A*. In Gruppe 2 wirkt dieser beobachtete Trend noch überzeugender. Es gibt zwar mehr gegenläufige Trends, aber dafür auch deutlichere Unterschiede von Behandlung *B* zu *A* im erwarteten Trend. Das bedeutet, dass in beiden Gruppen die  $FEV_1$ -Werte unter Behandlung *B* tendenziell größer scheinen als unter Behandlung *A*. In [Abbildung 3.2](#) zeigt sich, dass sich unter Behandlung *B* im Mittel deutlich niedrigere  $FEV_1$ -Werte in der zweiten Periode ergeben (wenn zuvor mit *A* behandelt wurde) als ohne vorige Behandlung. Der Unterschied innerhalb von Behandlung *A* ist nicht so groß über die zwei Perioden betrachtet. Jedoch ist hier eine wachsende Tendenz zu beobachten, d.h. Behandlung *A* führt im Mittel in der ersten Periode zu niedrigeren  $FEV_1$ -Werten als in der zweiten Periode (wenn zuvor Behandlung *B* verabreicht wurde). In [Abbildung 3.3](#) zeigt sich der in [Abbildung 3.1](#) beobachtete Trend bestätigt: Die  $FEV_1$  Werte der ersten Gruppe verlaufen deutlich niedriger als Werte der zweiten Gruppe, wobei dieser Unterschied in Periode 1 größer ist als in Periode 2. Ferner ist zu erkennen, dass Werte unter Behandlung *B* im Mittel bei beiden Gruppen über denen von Behandlung *A* liegen, wobei dies bei der zweiten Gruppe noch deutlicher zu sehen ist. Diese Tatsache kann man anhand der

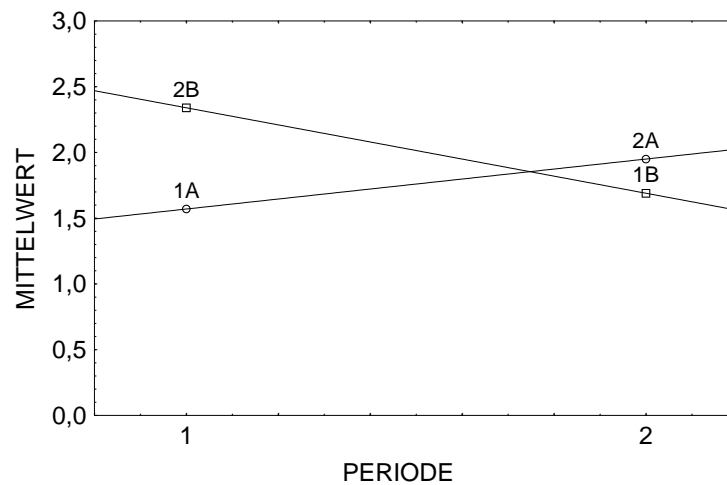


Abbildung 3.2: Die Mittelwerte der Beobachtungen pro Behandlungen nach der Zeit

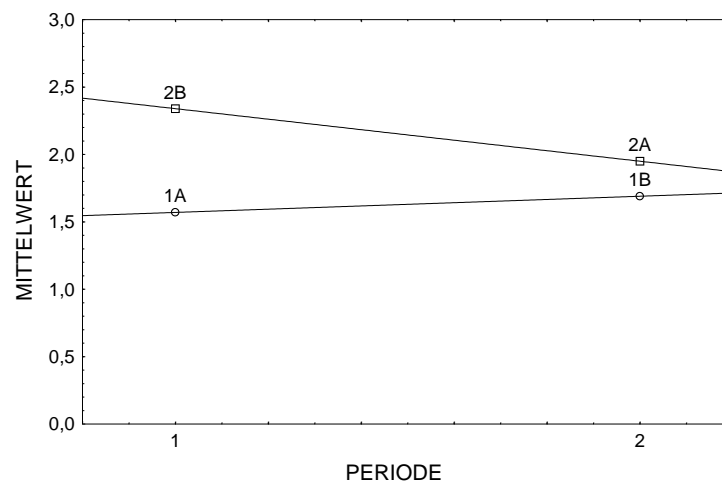


Abbildung 3.3: Die Mittelwerte der Beobachtungen pro Gruppe nach der Zeit

unterschiedlichen Steigungen der Geraden ablesen.

Grafiken kommen ebenfalls bei der Analyse der Daten auf einen Behandlungseffekt, bei vorhandenem Residualeffekt, zum Tragen (siehe dazu Abschnitt 5.2 auf Seite 49).

### 3.4 Effekte im Cross-Over-Plan

Im Folgenden werden die einzelnen Effekte im Cross-Over-Plan vorgestellt und anhand einiger Grafiken veranschaulicht. Hierbei wurde zur Veranschaulichung die Darstellungsform der Gruppenverläufe (siehe Abbildung 3.3) gewählt. Das heißt, für die Veranschaulichung werden der einfachen Verständlichkeit halber stetige Daten vorausgesetzt.

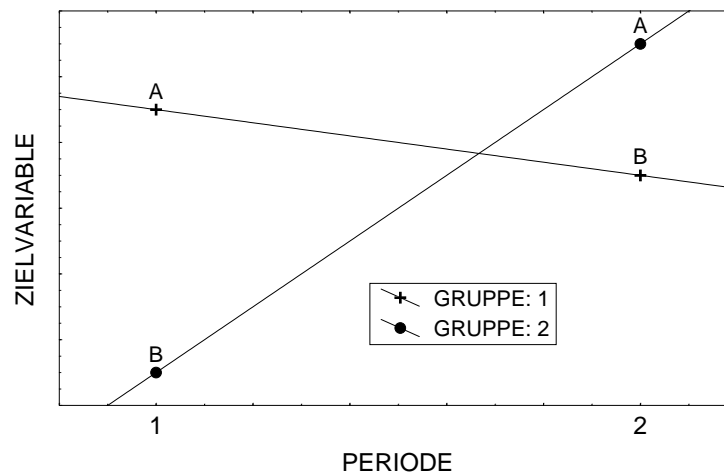


Abbildung 3.4: Positiver Residualeffekt

### Residualeffekt

Der Residualeffekt hat die Synonyme Nacheffekt und Carry-Over-Effekt, die in dieser Arbeit alle gleichberechtigt benutzt werden.

Ein vorhandener Residualeffekt ist ein Indiz für die Nachwirkung einer Behandlung aus der ersten Periode. Diese Nachwirkung, z.B. einer Medikation, hat bis zur Behandlung in der nachfolgenden Periode Auswirkungen. Genau dann, wenn eine solche Nachwirkung besteht, ist der Unterschied zwischen den Behandlungen *A* und *B* in der ersten Periode anders als in der zweiten Periode. Eine Möglichkeit der Quantifizierung des Residualunterschieds ist die Veränderung des Behandlungsunterschiedes von der ersten zur zweiten Periode. Man unterscheidet hier im Wesentlichen zwei Fälle:

- *Verkleinerung* des Behandlungsunterschieds: möglicherweise Überhangseffekt in erster Gruppe von Behandlung *A* in die zweite Periode, z.B. wenn die Wash-Out-Phase zu kurz war. Diesen Sachverhalt veranschaulicht Abbildung 3.4.
- *Vergrößerung* des Behandlungsunterschieds: möglicherweise Entzugseffekt, der sich nach Absetzen von Behandlung *A* in der zweiten Periode bemerkbar macht. Abbildung 3.5 ist ein mögliches Beispiel.

Der Residualeffekt kann wichtige Hinweise geben, z.B. auf pharmakologische Wirkungsmechanismen wie die Dauer der Wirkung eines Medikaments. Häufig erhält man auch zusätzliche Informationen bei Wechselwirkungen zweier Medikationen, da möglicherweise andere Wirkungsmechanismen auftreten als unter dem separaten Einsatz zweier Medikamente.

Ist ein Residualeffekt in einer Studie auf Basis eines Cross-Over-Plans nicht auszuschließen, ergeben sich große Schwierigkeiten in Bezug auf die Auswertung der Studie. Welcher Art die Schwierigkeiten sind und mögliche Lösungsansätze stellt das Kapitel 5 bereit.

### Behandlungseffekt

Ein Behandlungseffekt liegt vor, wenn die Wirkung der beiden Behandlungen *A* und *B*

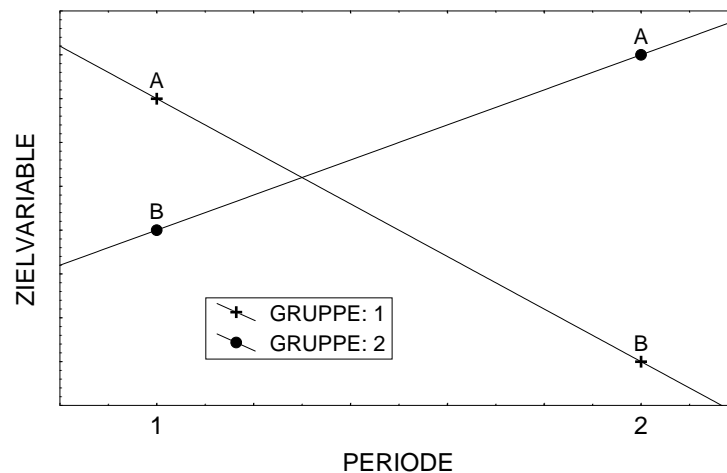


Abbildung 3.5: Negativer Residualeffekt

verschieden sind. Liegt kein Residualunterschied vor, so tritt der Behandlungsunterschied in beiden Gruppen gleich stark auf, und dieser Unterschied kann effizient aus den Daten der beiden Perioden analysiert werden. Kann ein Residualunterschied nicht ausgeschlossen werden, so kann lediglich ein verallgemeinerter Behandlungsunterschied untersucht werden, der sogenannte Cross-Over-Effekt: eine Überlagerung des reinen Behandlungseffekts und des Residualeffekts. Als reinen Behandlungseffekt bezeichnet man den Effekt, der ausschließlich durch den Unterschied der beiden Behandlungen verursacht wird. Nach LEHMACHER (1987) können sogar im Falle eines vorhandenen Residualunterschieds nur Daten der ersten Periode benutzt werden, um einen (reinen) Behandlungsunterschied aufzudecken.

In Abbildung 3.6 ist ein reiner Behandlungseffekt dargestellt, d.h. hier wird davon ausgegangen, dass es, erstens, keinen Einfluss des Zeitpunkts der Behandlungen und, zweitens, keinen Einfluss der in Periode 1 verabreichten Medikamente in der Nachfolgeperiode gibt.

### Periodeneffekt

Bei einem Periodeneffekt liegt das Mittel der beiden Behandlungswirkungen in der zweiten Periode auf einem anderen Niveau als das der ersten Periode. Die Differenz dieser Mittel ist der Periodenunterschied, wenn die Medikation der ersten Periode keinen Einfluss auf die Daten der zweiten Periode hat. Demnach ist der reine Periodeneffekt dann gegeben, wenn dieser Residualeffekt ausgeschlossen werden kann (analog zum reinen Behandlungseffekt). Es geht hier nicht um die Änderung der Zielvariable über die Zeit abhängig von der Behandlung, sondern um die Änderung, die auch ohne die Behandlung eingetreten wäre. Abbildung 3.7 veranschaulicht einen reinen Periodeneffekt (zusammen mit einem Behandlungseffekt).

Mögliche Ursachen für einen Periodeneffekt:

- Beeinträchtigung der Behandlungswirkung durch äußere Einflüsse, die sich über die Zeit ändern.

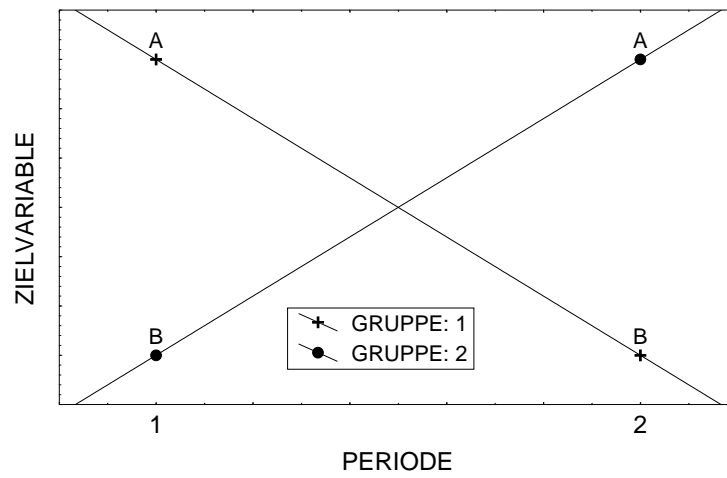


Abbildung 3.6: Reiner Behandlungseffekt

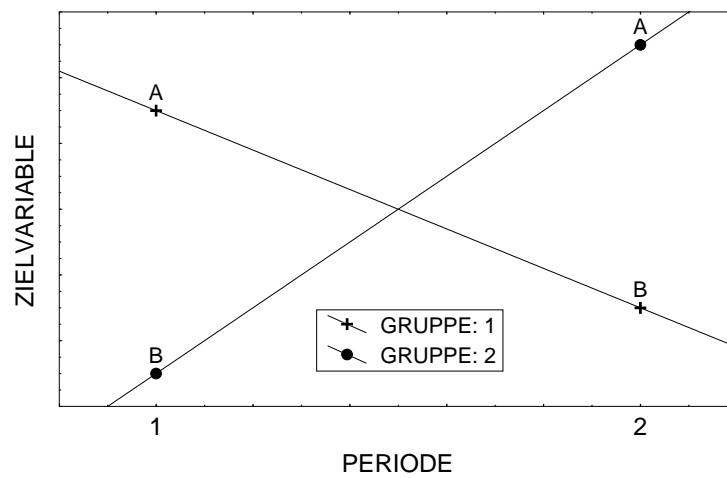


Abbildung 3.7: Nur Behandlungs- und Periodeneffekt



- Lernzuwachs oder Gewöhnung der Versuchseinheiten an die Studie, wodurch der placebo-artige Anfangsvorteil beider Behandlungen nachlässt.
- Veränderung des Krankheitsstatus und dadurch andere Ausgangsbedingung in der zweiten Periode als zu Beginn. Extremes Beispiel: Heilung eines Patienten durch die Behandlung in der ersten Periode.

Ohne Berücksichtigung eines möglichen Nacheffekts der Medikation in der vorhergehenden Periode, wirken sich diese Ursachen als Periodeneffekt aus. Der Basis-COP kann nicht zwischen behandlungsbedingten Nacheinflüssen und extern sich ändernden Bedingungen unterscheiden. Solche Perioden-Effekte sind unproblematisch und beeinflussen die statistische Analyse nicht.

Nach SENN (1993) sollte die (Nicht-)Berücksichtigung des Periodeneffekts aus apriori Gründen geschehen, da bei entsprechender Nachweisführung mit Hilfe der schließenden Statistik verschiedene Probleme auftreten. Dazu mehr im folgenden Kapitel.

## 3.5 Modelle

In diesem Abschnitt wird zunächst das klassische lineare Modell von GRIZZLE (1965) angegeben, wie es von KOCH (1972) erstmals ohne Normalverteilungsannahme des Fehlerterms benutzt wurde. Anschließend wird in Abschnitt 3.5.2 das weniger restriktive nichtparametrische Modell vorgestellt, wofür dann im folgenden Kapitel aufgezeigt wird, wie mit geeigneten Verfahren auf Basis dieses Modells Daten auf einen Behandlungs-, Residual- oder Periodeneffekt untersucht werden können. In diesem nichtparametrischen Modell wird keine Additivität der einzelnen Erwartungswerte mehr vorausgesetzt. Daraus folgt, dass keine restriktiven Voraussetzungen mehr an die Verteilungsfunktionen gestellt werden und die Zielvariablen lediglich mindestens ordinales Niveau haben brauchen.

### 3.5.1 Lineares Modell

Im linearen Modell nach GRIZZLE (1965) setzen sich die einzelnen auftretenden Effekte additiv zusammen. Das bedeutet folgendes: Jede Zufallsvariable besteht aus dem mittleren Erwartungswert der Verteilung der Gesamtstichprobe zuzüglich dem entsprechenden Behandlungseffekt für Behandlung  $A$  oder  $B$ . Als nächstes wird der Zeiteffekt dazu addiert, also der entsprechende Effekt für Periode 1 oder Periode 2, je nachdem wann der Wert erhoben wurde. Für Werte, die in der zweiten Periode erhoben wurden, gibt es zusätzlich noch die Möglichkeit eines Nacheffekts der Behandlungen aus der vorhergehenden Methode. Letztendlich gibt es noch den individuellen Effekt, der nur von der entsprechenden Versuchseinheit abhängt und den Versuchsfehler. Man erhält demnach folgendes Modell:

$$X_{iks} = \mu + \Phi_{2-\delta_{is}} + \pi_s + \delta_{2s}\lambda_i + S_{ik} + \epsilon_{iks}, \quad (3.1)$$

wobei  $i = 1, 2, k = 1, \dots, n_i$  und  $s = 1, 2$ .

$\delta_{is}$  bezeichnet das Kronecker-Delta, d.h.

$$\delta_{is} = \begin{cases} 1, & \text{wenn } i=s \\ 0, & \text{sonst} \end{cases}$$

Dabei gelten folgende Bezeichnungen:

- $\mu$  : mittlerer Erwartungswert
- $\Phi_{2-\delta_{is}}$  : Behandlungseffekt
- $\pi_s$  : Effekt der  $s$ -ten Periode
- $\lambda_i$  : Residualeffekt der Behandlung in Gruppe  $i$  in der zweiten Periode
- $S_{ik}$  : unabhängig, identisch verteilte Zufallsvariablen, die den Effekt des  $k$ -ten Individuums beschreiben mit  $E(S_{ik}) = 0$ .
- $\epsilon_{iks}$  : unabhängiger Versuchsfehler mit  $E(\epsilon_{iks}) = 0$  und unabhängig von  $S_{ik}$ .

Man reparametrisiert mit den Bedingungen

$$\begin{aligned}\Phi &= \Phi_1 = -\Phi_2 \\ \lambda &= \lambda_1 = -\lambda_2 \\ \pi &= \pi_1 = -\pi_2\end{aligned}$$

und kann das lineare Modell in vereinfachter Form schreiben als

$$\begin{aligned}X_{1k1} &= \mu_{11} + S_{1k} + \epsilon_{1k1}, \quad \mu_{11} = \mu + \Phi + \pi \\ X_{1k2} &= \mu_{12} + S_{1k} + \epsilon_{1k2}, \quad \mu_{12} = \mu - \Phi - \pi + \lambda \\ X_{2k1} &= \mu_{21} + S_{2k} + \epsilon_{2k1}, \quad \mu_{21} = \mu - \Phi + \pi \\ X_{2k2} &= \mu_{22} + S_{2k} + \epsilon_{2k2}, \quad \mu_{22} = \mu + \Phi - \pi - \lambda.\end{aligned}\tag{3.2}$$

Die einzelnen Erwartungswerte können in einem Vektor

$$\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})^T$$

zusammengefasst werden.

**Notation:** Auch zukünftig werden Vektoren mit fett gedruckten Buchstaben bezeichnet, ebenfalls Matrizen. Aus dem Zusammenhang ergibt sich jeweils, um was es sich handelt. Zur genaueren Beschreibung sei auf den Abschnitt A.1 auf Seite 111 im Anhang verwiesen.

Die Hypothesen werden in diesem Modell in der Regel direkt über die Effekte  $\Phi$ ,  $\lambda$  und  $\pi$  gestellt. Dies ist äquivalent zur Hypothesenstellung über den Erwartungswertvektor mit Hilfe von Kontrastvektoren. Dazu jedoch näheres in Abschnitt 4.1.2 auf Seite 24.

Um Tests für dieses Modell herzuleiten, wurde zunächst von GRIZZLE (1965) das univariate Modell angesetzt, wobei der Individualeffekt ( $S_{ik}$ ) und der Fehlerterm ( $\epsilon_{iks}$ ) normalverteilt sind mit Erwartungswert 0 und Varianz  $\sigma_s^2$  bzw.  $\sigma_e^2$ . D.h. es werden nicht nur für jede Versuchseinheit und jeden Fehlerterm gleiche Varianzen angenommen, sondern auch für alle Beobachtungen identische Varianzen  $\sigma_s^2 + \sigma_e^2$ . Außerdem folgt aus dem zugrundegelegten Modell, dass die Kovarianz zwischen Beobachtungen der ersten und zweiten Periode  $\text{Cov}(X_{ik1}, X_{ik2}) = \sigma_s^2$  stets positiv ist.

LEHMACHER (1987) schwächte diese Voraussetzungen ein wenig ab, indem er das Modell von ZIMMERMANN UND RAHLFS (1980) sowie SCHNEIDER (1983) zugrunde legte. Der Individualeffekt  $S_{ik}$  geht dabei in dem Modell (3.1) in den Versuchsfehler mit ein. Für die zweidimensionalen Vektoren  $\boldsymbol{\epsilon}_{ik} = (\epsilon_{ik1}, \epsilon_{ik2})$  wird vorausgesetzt, dass sie unabhängig und identisch multivariat normalverteilt sind mit Erwartungswertvektor  $\mathbf{0}$  und

beliebiger (nichtsingulärer) Kovarianzmatrix. Im Folgenden wird stets Bezug auf dieses Modell genommen, was als Spezialfall das Modell von GRIZZLE (1965) enthält.

KOCH (1972) erweiterte das Modell ein weiteres Mal und prägte den ersten semiparametrischen Ansatz, indem er aus dem zuvor eingeführten linearen Modell zwar die Additivität der einzelnen Effekte übernahm, jedoch keine Normalverteilungsannahme mehr für den Fehlerterm machte.

Ob parametrischer Ansatz oder nicht: Voraussetzung für alle Varianten ist nach wie vor die Annahme der Additivität der einzelnen Effekte, sowie die sinnvolle Interpretation von Summen und Differenzen. Damit können Daten, die die entsprechenden Voraussetzungen an die Verteilungsfunktion nicht erfüllen, nicht behandelt werden, z.B. ordinale Daten wie im Beispiel der Asthma-Studie I. Ebenfalls ist die Annahme der Linearität der Effekte eine Voraussetzung, die nicht überprüft werden kann. Zudem gibt es Schwierigkeit bei *floor and ceiling* Effekten, also beschränkten Daten. Ein Beispiel wären Messungen des Puls unter zwei verschiedenen Behandlungen. Der Unterschied zwischen einem Puls von 70 und 80 Schlägen pro Minute beträgt wie bei einem Unterschied von 0 zu 10 jeweils 10 Schläge pro Minute. Die Bedeutung dieser Differenz ist jedoch im unteren Bereich (floor) deutlich höher als im normalen Bereich. Allgemein kann sich eben an den Ober- bzw. Untergrenzen der Daten eine Veränderung der Daten erheblich mehr auswirken als im Mittelfeld. SALSBURG (1999) schlägt dafür bei parametrischen Verfahren, wie beispielsweise der ANOVA, eine monotone Transformation vor, die an den Ober- bzw. Untergrenzen die Daten *auseinanderzieht*. Das Problem dabei ist, dass sich im linearen Modell die Effekte unter solchen Transformationen verändern und möglicherweise auch das ganze Modell (3.1) die Gültigkeit verliert.

### 3.5.2 Nichtparametrisches Modell

**Definition 3.5.1 (Nichtparametrisches gemischtes Modell (GM))** *Das nichtparametrische gemischte Modell hat die folgende Form. Gegeben seien für  $i = 1, 2$  und  $k = 1, \dots, n_i$  unabhängige Zufallsvektoren*

$$\mathbf{X}_{ik} = (X_{ik1}, X_{ik2})^T \sim \mathbf{F}_i.$$

Für die Einträge gelte:

$$X_{iks} \sim F_{is} \quad (3.3)$$

mit  $i = 1, 2$ ,  $k = 1, \dots, n_i$  und  $s = 1, 2$ .

Zwischen den Einträgen des Vektors  $\mathbf{X}_{ik}$  seien beliebige Abhängigkeiten zugelassen. Das bedeutet, dass berücksichtigt wird, dass  $X_{ik1}$  und  $X_{ik2}$  zwei Messungen an derselben Versuchseinheit sind. Im Folgenden sei lediglich vom nichtparametrischen Modell gesprochen oder einfach als Abkürzung (GM) mit GM = gemischtes Modell. Der Zusatz *gemischt* verdeutlicht lediglich, dass in diesem Modell sowohl unabhängige (Gruppe) als auch abhängige Faktoren (Zeitpunkt) auftreten.

Im linearen Modell aus dem vorangegangenen Abschnitt standen in natürlicher Weise leicht interpretierbare Parameter für die Bildung von Effekten und Hypothesen zur Verfügung. Im nichtparametrischen Modell kann man die Hypothesen über die entsprechenden Verteilungsfunktionen oder über Funktionale dieser Verteilungsfunktionen, die

relativen Effekte, stellen. Letztere sind in ihrer Interpretation möglicherweise schwieriger, aber dafür realistischer.

Der Vorteil des nichtparametrischen Modells gegenüber dem linearen Modell ist vor allem in der Anwendbarkeit in der Praxis zu sehen. Dort sind beispielsweise die Additivität der Effekte und positive Kovarianzen (Voraussetzungen für GRIZZLEs Modell) nicht überprüfbar, oder die Annahme der Normalverteilung ist unter Umständen nicht angebracht. Im nichtparametrischen Modell gibt es keine Restriktionen mehr, weshalb dieses Modell die Realität in der Regel besser widerspiegelt.

# Kapitel 4

## Statistiken für den Cross-Over-Plan

Dieses Kapitel stellt bereits bekannte Verfahren im linearen Modell für den Cross-Over-Plan vor und entwickelt neue Teststatistiken für den nichtparametrischen Ansatz. Dies alles geschieht mit der Intention, dem Leser ohne viel Notation und theoretische Ausführung, z.B. in der Form von Beweisen, die Grundlagen und Vorgehensweisen zu vermitteln. Der mehr theoretisch interessierte Leser findet Notation, exakte Definitionen und Sätze mit Beweisen im Anhang ab Seite 111.

In Abschnitt 4.1 werden die Verfahren für das lineare Modell beschrieben. Abschnitt 4.2 auf Seite 28 widmet sich der Herleitung von Testverfahren für nichtparametrische Verfahren unter den Hypothesen  $H_0^F : \mathbf{cF} = 0$  und  $H_0^p : \mathbf{cp} = 0$ . Dabei wird im Teil 4.2.8 auf Seite 42 gezeigt, dass die in diesem Abschnitt entwickelten neuen Verfahren eine sinnvolle Erweiterung der Verfahren im linearen Modell darstellen. An den beiden Asthma-Studien werden diese Ergebnisse abschließend in Abschnitt 4.3 veranschaulicht.

### 4.1 Lineares Modell

Dieser Abschnitt hat in erster Linie historischen und ergänzenden Charakter. Hier werden die Verfahren vorgestellt, die in der Vergangenheit für das lineare Modell entwickelt wurden. Dabei wird von dem Modell von LEHMACHER (1987) ausgegangen, dass in Abschnitt 3.5.1 auf Seite 17 eingeführt wurde:

$$X_{iks} = \mu + \Phi_{2-\delta_{is}} + \pi_s + \delta_{2s}\lambda_i + \epsilon_{iks},$$

wobei  $i, s = 1, 2$  und  $k = 1, \dots, n_i$ . Die auftretenden Parameter  $\mu$ ,  $\Phi_{2-\delta_{is}}$ ,  $\pi_s$ ,  $\lambda_i$  und  $\epsilon_{iks}$  wurden bereits auf Seite 17 definiert. Die Beobachtungen sind normalverteilt und der Individualeffekt  $S_{ik}$  ist in  $\epsilon_{iks}$  enthalten. Die Resultate der folgenden Abschnitte sind uneingeschränkt auch für das Modell von GRIZZLE (1965) gültig. Wenn die Normalverteilungsannahme nicht gerechtfertigt ist, kann immer noch auf den semiparametrischen Ansatz von KOCH (1972) zurückgegriffen werden. Dieser beruht auf dem gleichen Modell mit einer Ausnahme: die Voraussetzung der Normalverteilung für die Fehlerterme wird fallengelassen.

Bevor Hypothesen und Tests für die einzelnen Effekte vorgestellt werden, seien noch ein paar Aspekte der Schätz- und Testbarkeit derselben erwähnt.

Unter den entsprechenden Voraussetzungen des Modells ergeben sich die Erwartungswerte aus Tabelle 4.1.

Tabelle 4.1: Erwartungswerte

		Periode	
		1	2
Gruppe	1	$\mu + \Phi_1 + \pi_1$	$\mu + \Phi_2 + \pi_2 + \lambda_1$
	2	$\mu + \Phi_2 + \pi_1$	$\mu + \Phi_1 + \pi_2 + \lambda_2$

### 4.1.1 Schätzer der Effekte

Um ein Gefühl für einzelne Effekte zu bekommen, ist es oftmals hilfreich, die Konstruktion der Schätzer zu betrachten. Ferner werden die verschiedenen Effektschätzer für die im anschließenden Abschnitt angegebenen Tests benötigt.

#### 4.1.1.1 Schätzer für den Residualunterschied

Der Residualunterschied  $d_\lambda$  sei definiert als die Differenz des Residualeffekts der 1. Gruppe  $\lambda_1$  minus den Residualeffekt der 2. Gruppe  $\lambda_2$ :

$$d_\lambda := \lambda_1 - \lambda_2. \quad (4.1)$$

Beim Unterschied des Residualeffekts handelt es sich um den Betrag, um den sich der Behandlungsunterschied vom Übergang der ersten zur zweiten Periode ändert. Und so wird auch der entsprechende Schätzer konstruiert:

$$\begin{aligned} \hat{d}_\lambda &= (\bar{X}_{1.1} - \bar{X}_{2.1}) - (\bar{X}_{2.2} - \bar{X}_{1.2}) \\ &= (\bar{X}_{1.1} + \bar{X}_{1.2}) - (\bar{X}_{2.1} + \bar{X}_{2.2}). \end{aligned} \quad (4.2)$$

Dieser Schätzer ist für alle  $i, s = 1, 2$  und  $k = 1, \dots, n_i$  erwartungstreu, denn:

$$\begin{aligned} E(\hat{d}_\lambda) &= E((\bar{X}_{1.1} - \bar{X}_{2.1}) - (\bar{X}_{2.2} - \bar{X}_{1.2})) \\ &= (\Phi_1 + \pi_1 - \Phi_2 - \pi_1) - (\Phi_1 + \pi_2 + \lambda_2 - \Phi_2 - \pi_2 - \lambda_1) \\ &= \lambda_1 - \lambda_2. \end{aligned}$$

#### 4.1.1.2 Schätzer für den Periodenunterschied

Der Periodenunterschied  $d_\pi$  sei definiert als die Differenz zwischen der ersten Periode und der zweiten Periode:

$$d_\pi = \pi_1 - \pi_2.$$

Der Periodenunterschied ist als der Betrag zu verstehen, um den sich das Mittel der ersten Periode gegenüber dem Mittel der zweiten Periode verändert. Dementsprechend wird der Schätzer  $\hat{d}_\pi$  konstruiert:

$$\begin{aligned} \hat{d}_\pi &= \frac{1}{2} (\bar{X}_{1.1} + \bar{X}_{2.1}) - \frac{1}{2} (\bar{X}_{1.2} + \bar{X}_{2.2}) \\ &= \frac{1}{2} ((\bar{X}_{1.1} - \bar{X}_{1.2}) - (\bar{X}_{2.2} - \bar{X}_{2.1})). \end{aligned} \quad (4.3)$$

Dieser Schätzer ist erwartungstreu, denn es gilt mit Reparametrisierungsbedingung für den Residualeffekt  $\lambda = \lambda_1 = -\lambda_2$  folgendes:

$$\begin{aligned} E(\widehat{d}_\pi) &= \frac{1}{2}(\Phi_1 + \pi_1 + \Phi_2 + \pi_1) - \frac{1}{2}(\Phi_2 + \pi_2 + \lambda_1 + \Phi_1 + \pi_2 + \lambda_2) \\ &= \frac{1}{2}(2\pi_1 - 2\pi_2) + \frac{1}{2}\underbrace{(\lambda_1 + \lambda_2)}_{=0} \\ &= \pi_1 - \pi_2. \end{aligned}$$

Ohne die Reparametrisierungsbedingung  $\lambda = \lambda_1 = -\lambda_2$  wäre eine erwartungstreue Schätzung des Periodenunterschieds offensichtlich nicht möglich. Diese Bedingung bedeutet, dass der mittlere Residualunterschied  $(\lambda_1 + \lambda_2)/2$  per definitionem im Periodenunterschied enthalten ist.

Ein reiner Periodeneffekt, der nicht auf irgend eine Art und Weise den Residualeffekt enthält, ist nicht schätzbar.

#### 4.1.1.3 Schätzer für den Behandlungsunterschied

Der Behandlungsunterschied  $d_\Phi$  sei definiert als die Differenz zwischen dem Effekt von Behandlung  $A$  (bezeichnet als  $\Phi_1$ ) und Behandlung  $B$  (bezeichnet als  $\Phi_2$ )

$$d_\Phi = \Phi_1 - \Phi_2. \quad (4.4)$$

Demnach ist dies stets der Behandlungsunterschied zwischen Behandlung  $A$  und  $B$  in der ersten Periode und auch der Behandlungsunterschied zwischen  $A$  und  $B$  in der zweiten Periode, wenn kein Residualeffekt vorliegt.

Man kann den Behandlungseffekt mit dem Mittel aus dem Behandlungsunterschied der ersten und der zweiten Periode schätzen:

$$\begin{aligned} \widehat{d}_\Phi &= \frac{1}{2}((\bar{X}_{1.1} - \bar{X}_{2.1}) + (\bar{X}_{2.2} - \bar{X}_{1.2})) \\ &= \frac{1}{2}((\bar{X}_{1.1} - \bar{X}_{1.2}) - (\bar{X}_{2.1} - \bar{X}_{2.2})). \end{aligned} \quad (4.5)$$

Betrachtet man den Erwartungswert dieses Schätzers, so ergibt sich folgendes:

$$\begin{aligned} E(\widehat{d}_\Phi) &= \frac{1}{2}(\Phi_1 + \pi_1 - \Phi_2 - \pi_1 + \Phi_1 + \pi_2 + \lambda_2 - \Phi_2 - \pi_2 - \lambda_1) \\ &= \frac{1}{2}(2\Phi_1 - 2\Phi_2 - (\lambda_1 - \lambda_2)) \\ &= d_\Phi - \frac{1}{2}d_\lambda. \end{aligned}$$

Dieser Schätzer ist offensichtlich nur dann erwartungstreu, wenn apriori ein Residualeffekt ausgeschlossen werden kann, demnach gilt:  $\lambda = d_\lambda = 0$ . Trifft man jedoch diese Annahme zu unrecht, so ergibt sich ein verzerrter Schätzer mit einem Bias von  $d_\lambda/2$ . Die Richtung der Verfälschung ergibt eine Über- bzw. Unterschätzung des wahren Werts:

$$\begin{aligned} d_\lambda < 0 &\Rightarrow E(\widehat{d}_\Phi) > d_\Phi && \text{Überschätzung und} \\ d_\lambda > 0 &\Rightarrow E(\widehat{d}_\Phi) < d_\Phi && \text{Unterschätzung des Behandlungseffekts.} \end{aligned}$$

Eine andere Möglichkeit kann vorgeschlagen werden. Wenn einfach nur die Daten der ersten Periode benutzt werden, spielt es überhaupt keine Rolle, ob ein Nacheffekt bei der Behandlung in der Folgeperiode zum Tragen kommt oder nicht. Dafür lautet der Schätzer

$$\hat{d}_{\Phi,1} = \bar{X}_{1,1} - \bar{X}_{2,1}.$$

Dieser Schätzer ist unter allen Umständen, ob mit oder ohne Residualeffekt, erwartungstreu, da

$$\begin{aligned} E(\hat{d}_{\Phi,1}) &= \Phi_1 + \pi_1 - \Phi_2 - \pi_1 \\ &= \Phi_1 - \Phi_2. \end{aligned}$$

Dafür benutzt man nur die Hälfte der erhobenen Daten, weil auf die Daten der zweiten Periode vollständig verzichtet wird. Nach LEHMACHER (1987) verzichtet man dadurch auf die durch das Cross-Over-Design erhaltenen Vorteile wie beispielsweise eine geringere Varianz des Schätzers. Häufig bezeichnet man den ersten Schätzer  $\hat{d}_{\Phi}$  auch als Schätzer für einen verallgemeinerten Behandlungsunterschied.

## 4.1.2 Hypothesen und Tests

Zunächst werden die Hypothesen vorgestellt, erklärt und die verschiedenen Schreibweisen aufgezeigt. Anschließend werden für die zuvor erläuterten Effekte Hypothesen und Teststatistiken beschrieben. Dabei wird neben der unter Normalverteilungsannahme auch auf die hauptsächlich von KOCH (1972) geprägte semiparametrische Variante eingegangen.

### 4.1.2.1 Hypothesen und Tests für einen Residualunterschied

Im linearen Modell formuliert man die Hypothesen über die Effekte des zugrundeliegenden Modells (3.1) in der folgenden natürlichen Form

$$H_0(\lambda) : \lambda = 0 \text{ kein Residual-Effekt}$$

Mit Hilfe des Kontrastvektors  $\mathbf{c}_{\lambda} = (1, 1, -1, -1)$  lässt sich die o.g. Hypothese auch in der Form

$$H_0^{\mu}(\lambda) : \mathbf{c}_{\lambda}\boldsymbol{\mu} = 0$$

mit  $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})^T$  schreiben, was äquivalent ist zur oben aufgezeigten Hypothesenstellung über die Modellparameter, denn es gilt:

$$\begin{aligned} & \mathbf{c}_{\lambda}\boldsymbol{\mu} = 0 \\ \iff & \mu_{11} + \mu_{12} = \mu_{21} + \mu_{22} \\ \iff & \mu + \Phi + \pi + \mu - \Phi - \pi + \lambda = \mu - \Phi + \pi + \mu + \Phi - \pi - \lambda \\ \iff & \lambda = -\lambda \\ \iff & \lambda = 0. \end{aligned}$$



Gemäß dem in (4.2) angegebenen Schätzer für einen Residualunterschied, bildet man pro Versuchseinheit die Summe der Beobachtungen für die beiden Behandlungen, d.h.

$$s_{ik} = X_{ik1} + X_{ik2} \quad \text{für } i = 1, 2, k = 1, \dots, n_i$$

und beschreibt diesen Schätzer demnach durch  $\hat{d}_\lambda = \bar{s}_{1.} - \bar{s}_{2.}$ . Unter Voraussetzung der Normalverteilung kann damit der unverbundene  $t$ -Test auf diese Individuumssummen  $s_{ik}$  angewendet werden. Man erhält die Teststatistik

$$U_{LEH}^\lambda = \frac{\bar{s}_{1.} - \bar{s}_{2.}}{\hat{\sigma}_\lambda} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

mit

$$\hat{\sigma}_\lambda^2 = \frac{(n_1 - 1)\hat{\sigma}_{\lambda,1}^2 + (n_2 - 1)\hat{\sigma}_{\lambda,2}^2}{n_1 + n_2 - 2}, \quad (4.6)$$

wobei  $\hat{\sigma}_{\lambda,i}$  für  $i = 1, 2$  die empirische Standardabweichung von  $s_{ik}$

$$\hat{\sigma}_{\lambda,i}^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (s_{ik} - \bar{s}_{i.})^2$$

für die Gruppe  $i$  bezeichnet. Ohne Verteilungsannahme kann asymptotisch auch der semiparametrische Ansatz von KOCH (1972) verwendet werden. Bei kleinen Stichprobenumfängen kann der exakte Wilcoxon-Mann-Whitney Test, auf die Ränge der Individuumssummen  $s_{ik}$  angewendet, benutzt werden. Ebenso bietet sich bei größeren Stichproben auch der asymptotische Wilcoxon-Mann-Whitney Test an.

Zur Interpretation dieser Vorgehensweise sei gesagt, dass sich ein Behandlungseffekt auf die Differenzen der Beobachtungssummen nicht auswirkt. Auch ein reiner Periodeneffekt wird dann nicht sichtbar. Wohl aber erkennt man unterschiedliche Nacheffekte.

Die Problematik, die später vertieft aufgegriffen wird, besteht darin, dass durch Nichtverwerfen der Hypothese nicht auf deren Gültigkeit geschlossen werden kann. Man kann lediglich mit der zugrunde gelegten Irrtumswahrscheinlichkeit auf ein Vorhandensein des Nacheffekts schließen. Es ist jedoch eigentlich die Intention, diesen Residualeffekt auszuschließen. SENN (1993) fügt hinzu, dass sich im vorliegenden Design, streng genommen, der Nacheffekt und die Behandlungs-Periode-Wechselwirkung nicht trennen lassen. Er meint, dass dieser Test nicht angewandt werden sollte, sondern eine adäquate Wash-Out-Phase im Vorfeld einen Residualeffekt ausschließen müsse.

#### 4.1.2.2 Hypothesen und Tests auf einen Periodenunterschied

Es gelte das Modell (3.1). In natürlicher Form kann man die Hypothese über den Parameter  $\pi$  formulieren:

$$H_0(\pi) : \pi = 0 \quad \text{kein Periodenunterschied.}$$

Unter Verwendung des entsprechenden Kontrastvektors  $\mathbf{c}_\pi := (1, -1, 1, -1)$  kann sich der äquivalenten Schreibweise

$$H_0^\mu(\pi) : \mathbf{c}_\pi \boldsymbol{\mu} = 0$$

bedient werden. Gemäß dem in (4.3) angegebenen Schätzer für den Periodenunterschied, werden von LEHMACHER (1987) und HILLS UND ARMITAGE (1979) die sog. Cross-Over-Differenzen, d.h. Wert von Behandlung  $A$  minus Wert von Behandlung  $B$  in jeder Gruppe verwendet:

$$\begin{aligned} c_{1k} &= X_{1k1} - X_{1k2} \\ c_{2k} &= X_{2k2} - X_{2k1} \end{aligned}$$

mit  $k = 1, \dots, n_i$ . Auf diese  $n = n_1 + n_2$  Cross-Over-Differenzen wird unter Normalverteilungsannahme der unverbundene  $t$ -Test, bzw. ohne Verteilungsannahme, nach KOCH (1972) der exakte oder asymptotische Wilcoxon-Mann-Whitney Test angewendet. Man erhält demnach für den parametrischen Fall die Teststatistik

$$U_{LEH}^{\pi} = \frac{\bar{c}_{1.} - \bar{c}_{2.}}{\hat{\sigma}_{\pi}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}.$$

Der Varianzschätzer ergibt sich wie folgt durch

$$\hat{\sigma}_{\pi}^2 = \frac{(n_1 - 1)\hat{\sigma}_{\pi,1}^2 + (n_2 - 1)\hat{\sigma}_{\pi,2}^2}{n_1 + n_2 - 2},$$

wobei  $\hat{\sigma}_{\pi,i}$  für  $i = 1, 2$  die empirische Standardabweichung der Cross-Over-Differenzen  $c_{ik}$  ist:

$$\hat{\sigma}_{\pi,i}^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (c_{ik} - \bar{c}_{i.})^2.$$

Es gilt offensichtlich  $\hat{d}_{\pi} = (\bar{c}_{1.} - \bar{c}_{2.})/2$ . Der Vorfaktor  $1/2$  kann weggelassen werden, da auch beim Varianzschätzer auf ihn verzichtet wird, so dass man die Teststatistik  $U^{\pi}$  erhält.

Einen wesentlichen Nachteil erwähnt SENN (1993): Wie beim Residualeffekt kann man nicht aus Nichtablehnung der Hypothese auf deren Gültigkeit schließen, d.h. auf das Nichtvorhandensein eines Periodenunterschieds. Die Entscheidung über einen eventuellen Periodenunterschied sollte demnach a priori aus sachlichen Überlegungen geschehen.

#### 4.1.2.3 Hypothesen und Tests auf einen Behandlungsunterschied (1)

Die Hypothese auf einen Behandlungsunterschied kann folgendermaßen formuliert werden:

$$H_0(\Phi) : \Phi = 0.$$

Mit Hilfe des Kontrastvektors  $\mathbf{c}_{\Phi} := (1, -1, -1, 1)$  ausgedrückt erhält man die äquivalente Form

$$H_0^{\mu}(\Phi) : \mathbf{c}_{\Phi} \boldsymbol{\mu} = 0.$$

LEHMACHER (1987) schlägt wie HILLS UND ARMITAGE (1979) hier Tests auf Basis der Periodenunterschiede  $p_{ik}$  vor mit

$$p_{ik} = X_{ik1} - X_{ik2} \quad \text{für } i = 1, 2, k = 1, \dots, n_i$$

Periodenunterschied der  $k$ -ten Versuchseinheit der  $i$ -ten Gruppe. Dieser Periodenunterschied darf nicht verwechselt werden mit dem notationsgleichen im nächsten Kapitel eingeführten relativen Effekt. Der Schätzer für den Behandlungseffekt aus (4.5) kann somit auch dargestellt werden als  $\hat{d}_\Phi = (\bar{p}_{1.} - \bar{p}_{2.})/2$ . Die resultierende Teststatistik ist

$$U_{LEH}^\Phi = \frac{\bar{p}_{1.} - \bar{p}_{2.}}{\hat{\sigma}_\Phi} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (4.7)$$

mit  $\hat{\sigma}_\Phi$  die empirische Standardabweichung gemäß

$$\hat{\sigma}_\Phi^2 = \frac{(n_1 - 1)\hat{\sigma}_{\Phi,1}^2 + (n_2 - 1)\hat{\sigma}_{\Phi,2}^2}{n_1 + n_2 - 2},$$

wobei  $\hat{\sigma}_{\Phi,i}$  die empirische Standardabweichung ist von  $p_{ik}$  für die Gruppe  $i$

$$\hat{\sigma}_{\Phi,i}^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (p_{ik} - \bar{p}_{i.})^2. \quad (4.8)$$

Wenn man keine Verteilungsannahme im linearen Modell treffen kann, sollte auch hier auf den exakten oder asymptotischen Wilcoxon-Mann-Whitney Test, angewandt auf die Ränge der Periodenunterschiede  $p_{ik}$ , zurückgegriffen werden. Ansonsten ist die Normalverteilung dieser Statistik unter Hypothese gewährleistet.

Mit diesem Test prüft man lediglich auf einen reinen Behandlungsunterschied, wenn kein Residualeffekt vorhanden ist, denn nur dann ist  $\hat{d}_\Phi = (\bar{p}_{1.} - \bar{p}_{2.})/2$  ein erwartungstreuer Schätzer für den Periodenunterschied  $d_\Phi$ .

Wenn man diesen Test anwendet, ohne die Annahme zu treffen, dass kein Residualeffekt vorhanden ist, so bekommt man einen Test auf diesen verallgemeinerten Behandlungseffekt

$$\begin{aligned} \delta &:= d_\Phi - \frac{1}{2}d_\lambda \\ &= 2 \cdot \Phi - \lambda, \end{aligned} \quad (4.9)$$

den Test auf einen Cross-Over-Effekt. Der Cross-Over-Effekt ist derjenige Effekt, der nach Definition erwartungstreu aus  $\hat{d}_\Phi$  geschätzt wird. Die Überlagerung des (reinen) Behandlungseffekts und des Residualeffekts macht die Problematik des Cross-Over-Plans aus. Siehe dazu auch LEHMACHER (1987), SENN (1993) und Kapitel 5.

Die Teststatistik  $U_{LEH}^\Phi$  ist somit für den Test auf einen Behandlungseffekt und einen Cross-Over-Effekt gleich, jedoch müssen die Unterschiede in Anwendung und Interpretation der resultierenden Tests beachtet werden. Diese Teststatistik testet, wenn ein möglicher Residualeffekt berücksichtigt wird, auf einen verallgemeinerten Behandlungseffekt, den Cross-Over-Effekt, und sonst auf einen reinen Behandlungseffekt.

Die Annahme, dass kein Residualunterschied vorhanden ist, kann man jedoch nicht immer treffen. Entweder ist man aus theoretischen Vorüberlegungen zur Versuchsanlage zu dem Entschluss gekommen, dass eine Nachwirkung einer Behandlung der ersten Periode nicht mit absoluter Sicherheit auszuschließen ist, oder ein entsprechender Test (siehe Abschnitt 5.5 auf Seite 55) kam nicht zu dem Ergebnis, dass mit einer beliebig kleinen Irrtumswahrscheinlichkeit ein zu definierender kleiner Residualunterschied vorliegt. Daher stellt sich die Frage, was mit diesem Test passiert, wenn  $\lambda = 0$  nicht angenommen werden darf, und wie der (reine) Behandlungseffekt dennoch getestet werden kann.

#### 4.1.2.4 Hypothesen und Tests auf einen Behandlungsunterschied (2)

Wenn ein Residualunterschied vorhanden ist, erhält man nach Abschnitt 4.1.1.3 nur einen sicher erwartungstreuen Schätzer, wenn die Daten der zweiten Periode vernachlässigt werden. Auf die Daten der ersten Periode wird dann ein unverbundener  $t$ -Test bzw. ohne Verteilungsannahme ein Wilcoxon-Mann-Whitney Test im klassischen Zwei-Stichproben-Design angewandt. In einem solchen Fall wird auf sämtliche in der zweiten Periode erhobenen Daten verzichtet und damit auf die komplette Cross-Over-Versuchsanlage.

## 4.2 Nichtparametrisches Modell

Dieser Abschnitt widmet sich den Grundlagen, basierend auf dem in Abschnitt 3.5.2 vorgestellten Modell (GM): Für die Einträge der unabhängigen Zufallsvektoren

$$\mathbf{X}_{ik} = (X_{ik1}, X_{ik2})^T \sim \mathbf{F}_i$$

gilt:

$$X_{iks} \sim F_{is}$$

mit  $i = 1, 2$ ,  $k = 1, \dots, n_i$  und  $s = 1, 2$ . Bevor Tests für dieses Modell entwickelt werden können, müssen grundlegende Definitionen, Hypothesenstellung und asymptotische Resultate geklärt werden. Alle Beobachtungen können in dem Vektor  $\mathbf{X} = (X_{111}, \dots, X_{2n_22})$  zusammengefasst werden.

### 4.2.1 Motivation und Definition

Bei Einführung des nichtparametrischen Modells in Abschnitt 3.5.2 auf Seite 19 wurden bereits viele Vorteile des nichtparametrischen Modells gegenüber dem linearen Modell erläutert. Dazu gehörten:

- Anwendbarkeit auch bei ordinalen Daten,
- Nicht die Annahme eines Lokationsmodells,
- stets keine Verteilungsannahmen für Fehlerterme (oder ggf. Individualeffekte).

Durch die im Folgenden erläuterte Vorgehensweise mit Hilfe der relativen Effekte, ergeben sich noch Vorteile wie die Robustheit gegen Ausreißer und Invarianz unter monotonen Transformationen.

Bei dem bisher in nichtparametrischen Verfahren üblichen *gewichteten* relativen Effekt z.B. aus BRUNNER UND PURI (1996), der über das gewichtete Mittel der Verteilungsfunktionen gebildet wird

$$H(x) = \frac{1}{N} \sum_{i=1}^2 n_i (F_{i1}(x) + F_{i2}(x)),$$

hängt dieser von den einzelnen Stichprobenumfängen ab. Es ist aber schwer, eine Interpretation dafür zu finden, dass sich der gewählte Effekt  $\mathbf{p}_N = \int H_N d\mathbf{F}$  in unbalancierten

Versuchsplänen in Abhängigkeit von den Stichprobenumfängen verändert. Der gewichtete relative Effekt verliert damit die Bedeutung einer festen Modellgröße, die zu schätzen ist und für die Konfidenzintervalle bestimmt werden sollen.  $\mathbf{F} = (F_{11}, F_{12}, F_{21}, F_{22})^T$  bezeichnet den Vektor der Verteilungsfunktionen. Ferner hängt dann auch zwangsläufig die Hypothese  $H_0^p : \mathbf{c}\mathbf{p}_N = 0$  von den Stichprobenumfängen ab. Eine Hypothese erhebt jedoch für sich den Anspruch, unabhängig von den Stichprobenumfängen zu sein, da versucht wird, allgemeine Aussagen zu treffen, die sich nicht nur auf den gerade zugrunde liegenden Testdatensatz mit den zugehörigen Stichprobenumfängen, sondern auch auf die Gesamtpopulation beziehen.

Eine andere Möglichkeit, um mit den relativen Effekten (auch im unbalancierten Fall) arbeiten zu können, ist, die mittlere gewichtete Verteilungsfunktion durch eine mittlere *ungewichtete* Verteilungsfunktion

$$G(x) = \frac{1}{4} \sum_{i=1}^2 \sum_{s=1}^2 F_{is}(x)$$

zu ersetzen. Als Verteilungsfunktion wird in dieser Arbeit die normalisierte Version der Verteilungsfunktion  $F_{is}$  für  $i, s = 1, 2$  benutzt, die im Anhang auf Seite 115 formal definiert ist. Sonst wird oftmals mit der rechtsstetigen Verteilungsfunktion gearbeitet. In dieser Arbeit wird die Darstellung der normalisierten Version gewählt, damit die relativen Effekte der Daten mit stetiger und unstetiger Verteilungsfunktion einheitlich dargestellt werden können. Auch wenn im Folgenden lediglich von der Verteilungsfunktion die Rede ist, so sei immer die normalisierte Version derselben gemeint.

Damit definiert man den ungewichteten relativen Effekt.

**Definition 4.2.1 (Ungewichteter relativer Effekt)** Für beliebige unabhängige Zufallsvariablen  $X_{iks}$  mit  $X_{iks} \sim F_{is}$ ,  $i, s = 1, 2$  und  $k = 1, \dots, n_i$ , heißt

$$p_{is} = \int G dF_{is}$$

(mittlerer) ungewichteter relativer Effekt von  $F_{is}$  zu  $F_{11}, \dots, F_{2n_2}$ .

Wenn im Folgenden lediglich vom relativen Effekt die Rede sein wird, so ist in dieser Arbeit stets der soeben definierte ungewichtete relative Effekt gemeint. Der Zusatz *ungewichtet* bezieht sich demnach auf die Wahl der ungewichteten mittleren Verteilungsfunktion als mittlere Verteilungsfunktion. Für den vierdimensionalen Vektor der ungewichteten relativen Effekte gilt somit:

$$\mathbf{p} = (p_{11}, p_{12}, p_{21}, p_{22})^T = \int G d\mathbf{F}. \quad (4.10)$$

Aus der obigen Definition des relativen Effekts lässt sich ableiten, dass der relative Effekt die Abweichungen der einzelnen Faktorstufen vom Gesamtmittel der Verteilungsfunktionen beschreibt. Noch deutlicher wird dies bei Betrachtung der analogen Definition des relativen Effekts:

Seien  $X \sim F_{is}$  und  $\tilde{X} \sim G$  mit  $X$  und  $\tilde{X}$  unabhängig voneinander. Dann gilt

$$p_{is} = P(\tilde{X} < X) + \frac{1}{2} \cdot P(\tilde{X} = X). \quad (4.11)$$

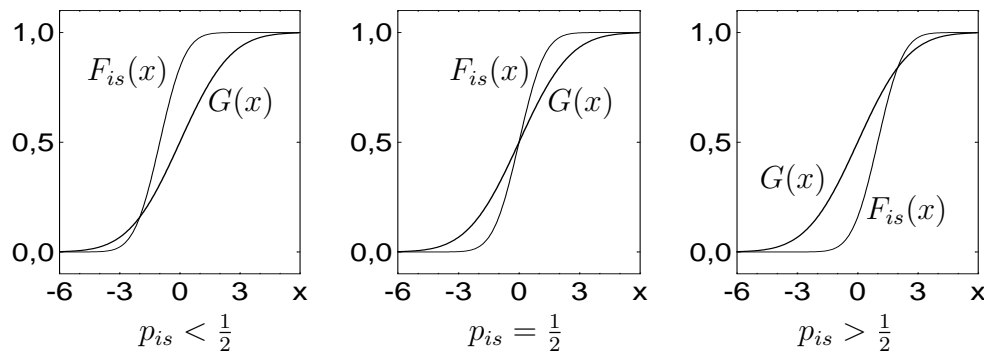


Abbildung 4.1: Relativer Effekt

Der Unterschied zu anderen Funktionalen, wie beispielsweise dem Erwartungswert, liegt darin, dass dieser für einzelne Verteilungen gebildet wird, während der relative Effekt Verteilungen zueinander ins Verhältnis setzt. Eine weitere Abweichung zu den bereits z.B. aus BRUNNER UND PURI (1996) bekannten *gewichteten* relativen Effekten liegt in der Definition des Gesamtmittels, denn dadurch ist der gewichtete relative Effekt im Gegensatz zum ungewichteten relativen Effekt abhängig von den Stichprobenumfängen. Mit der Darstellung des relativen Effekts aus Definition 4.2.1 ergibt sich auch eine anschauliche Erklärung desselben. Abbildung 4.1 aus BRUNNER UND LANGER (1999) soll zur Veranschaulichung dieser Größe dienen.

$p_{is}$  gibt die Tendenz von  $F_{is}$  zur mittleren ungewichteten Verteilungsfunktion  $G$  an. Der Einfachheit halber seien die Funktionen  $F_{is}$  und  $G$  wie in der Grafik als symmetrisch angenommen. Ist das Symmetriezentrum von  $F_{is}$  kleiner als das Symmetriezentrum von  $G$ , so ist  $F_{is}$  tendenziell kleiner als  $G$ . Das bedeutet  $p_{is} < \frac{1}{2}$ . Liegt entsprechend das Symmetriezentrum von  $F_{is}$  rechts des Symmetriezentrums von  $G$ , so ist  $F_{is}$  tendenziell größer als  $G$  und  $p_{is} > \frac{1}{2}$ . Falls keine Tendenz von  $F_{is}$  gegenüber  $G$  zu beobachten ist, d.h.  $F_{is}$  und  $G$  das gleiche Symmetriezentrum besitzen, gilt  $p_{is} = \frac{1}{2}$ . Die Begriffe 'tendenziell größer, kleiner oder gleich' sind aus der Arbeit von DOMHOF (2001) übernommen.

Durch den Verzicht auf eine Parametrisierung der Verteilung stehen in natürlicher Form keine Parameter mehr zur Verfügung, um Verteilungsunterschiede zu quantifizieren. In solchen Fällen kann man sich statistischer Funktionale bedienen, dem relativen Effekt. Es zeigt sich, dass die Interpretation desselben ähnlich der für die summarischen Messgrößen ist. Wie in Grafik 4.1 zu sehen ist, gibt der relative Effekt  $p_{is}$  nach (4.11) als Wahrscheinlichkeit die Tendenz der Verteilung  $F_{is}$  in Bezug auf die mittlere Verteilung  $G$  an. Der relative Effekt  $p_{is}$  ist im linearen Modell zu vergleichen mit dem entsprechenden Erwartungswert  $\mu_{is}$ , der sich aus den Behandlungs-, Residual- und Periodeneffekten additiv zusammensetzt. Im Gegensatz zum entsprechenden Erwartungswert  $\mu_{is}$  bewegt sich der relative Effekt  $p_{is}$  zwischen vorgegebenen Grenzen, da es sich um eine Wahrscheinlichkeit handelt.

## 4.2.2 Hypothesen

Die verschiedenen Effekte für den Cross-Over-Plan können über die in Definition 4.2.1 angegebenen Funktionale  $p_{is}$  angegeben werden. Da hier im Gegensatz zum linearen Modell die Effekte nicht in natürlicher Weise zur Hypothesenbildung vorliegen, stellt man die Hypothesen über die Verteilungsfunktionen oder entsprechende Funktionale. Je nach

zu klärender Fragestellung, wählt man die zu stellende Hypothese. Zunächst definiert man entsprechend den Effekten im linearen Modell die nichtparametrischen Effekte im Cross-Over-Plan. Wurden die Kontrastvektoren bisher auf den Vektor der Erwartungswerte  $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})^T$  angewandt, so bildet man nun die Effekte für das nichtparametrische Analogon mit Hilfe derselben Kontrastvektoren.

**Definition 4.2.2 (Nichtparametrische Effekte im COP)** *Bezeichne  $\mathbf{p}$  den Vektor der relativen Effekte aus (4.10). Die Kontrastvektoren seien wie in Abschnitt 4.1.2 definiert. Im nichtparametrischen COP heißt*

$$\begin{aligned} p^\lambda &:= \frac{1}{2} \mathbf{c}_\lambda \mathbf{p} = \frac{1}{2} (p_{11} + p_{12} - p_{21} - p_{22}) \text{ relativer Residual-Effekt,} \\ p^\pi &:= \frac{1}{2} \mathbf{c}_\pi \mathbf{p} = \frac{1}{2} (p_{11} - p_{12} + p_{21} - p_{22}) \text{ relativer Perioden-Effekt,} \\ p^\delta &:= \frac{1}{2} \mathbf{c}_\delta \mathbf{p} = \frac{1}{2} (p_{11} - p_{12} - p_{21} + p_{22}) \text{ relativer Cross-Over-Effekt.} \end{aligned}$$

Den reinen Behandlungseffekt kann man analog zur Darstellung im linearen Modell durch

$$p^\Phi := \frac{1}{2} (p_{11} - p_{21})$$

definieren, was dann zur später verwendeten Beziehung

$$p^\delta = 2p^\Phi - p^\lambda \quad (4.12)$$

führt. Diese Beziehung, d.h.

$$\text{Cross-Over-Effekt} = 2 \cdot (\text{reiner}) \text{ Behandlungseffekt} - \text{Überhangseffekt}$$

ist schon aus (4.9) aus dem linearen Modell auf Seite 27 bekannt.

Die soeben vorgestellten Effekte des nichtparametrischen Modells entsprechen sinngemäß den Effekten  $\delta, \pi, \lambda$  und  $\Phi$  aus dem linearen Modell. Um dies nachzuweisen, kann gezeigt werden, dass die nichtparametrischen Effekte monoton im jeweiligen Effekt des linearen Modells sind. Das bedeutet, dass bei wachsendem Residualeffekt  $\lambda$  im linearen Modell auch der relative Residual-Effekt  $p^\lambda$  ansteigt und umgekehrt. Man spricht dann von Monotonie von  $p^\lambda$  in  $\lambda$ . Der folgende Satz bestätigt diese Eigenschaft und zeigt die Äquivalenz der Hypothesen  $\lambda = 0$  und  $p^\lambda = 0$  im linearen Modell.

**Satz 4.2.3 (Monotonie von  $p^\lambda, p^\pi$  und  $p^\delta$ )** *Im linearen Modell gelten folgende Aussagen:*

1.  $p^\lambda(\lambda, \pi, \delta)$  ist monoton in  $\lambda$  und es gilt  $\lambda = 0 \iff p^\lambda = 0$ ,
2.  $p^\pi(\lambda, \pi, \delta)$  ist monoton in  $\pi$  und es gilt  $\pi = 0 \iff p^\pi = 0$ ,
3.  $p^\delta(\lambda, \pi, \delta)$  ist monoton in  $\delta$  und es gilt  $\delta = 0 \iff p^\delta = 0$ .

Der Beweis dieser Aussage befindet sich im Anhang auf Seite 129. Eine Folgerung dieses Satzes ist, dass die Bildung der Hypothesen über die relativen nichtparametrischen Effekte eine sinnvolle Verallgemeinerung der Hypothesen der ursprünglichen Effekte im linearen Modell darstellt.

Wie auch im linearen Modell können die Hypothesen sowohl direkt über die Effekte  $p^\lambda$ ,  $p^\pi$  und  $p^\delta$  als auch mit Hilfe der entsprechenden Kontrastvektoren  $\mathbf{c}_\delta$ ,  $\mathbf{c}_\lambda$  und  $\mathbf{c}_\pi$  formuliert werden, was offensichtlich äquivalent ist.

$$\begin{aligned} H_0^p(\lambda) : \mathbf{c}_\lambda \mathbf{p} = \mathbf{0} &\iff p^\lambda = 0 &\iff p_{11} + p_{12} = p_{21} + p_{22} \\ H_0^p(\pi) : \mathbf{c}_\pi \mathbf{p} = \mathbf{0} &\iff p^\pi = 0 &\iff p_{11} + p_{21} = p_{12} + p_{22} \\ H_0^p(\delta) : \mathbf{c}_\delta \mathbf{p} = \mathbf{0} &\iff p^\delta = 0 &\iff p_{11} + p_{22} = p_{12} + p_{21} \end{aligned}$$

Die Hypothesenstellung über diese relativen Effekte ist zwar allgemeiner als direkt über die Effekte  $\lambda$ ,  $\pi$  oder  $\delta$  des linearen Modells, aber dafür liegt die Interpretation der Hypothesen nicht unmittelbar auf der Hand. Welchen Sachverhalt stellen Hypothesen der Art  $\mathbf{c}\mathbf{p} = 0$  dar?

Dazu sei zunächst einmal betrachtet:

$$\begin{aligned} p_{is} &= \int G dF_{is} = P(Y < X_{is}) + \frac{1}{2}P(Y = X_{is}) \quad \text{mit} \\ Y &\sim G. \end{aligned}$$

Betrachtet man zunächst die Hypothese  $H_0^p(\lambda) : p_{11} + p_{12} = p_{21} + p_{22}$ . Dann ergibt sich für  $i = 1, 2$

$$\begin{aligned} p_{i1} + p_{i2} &= \int G d(F_{i1} + F_{i2}) \\ &= 2 \cdot \left( P(Y < X_i) + \frac{1}{2}P(Y = X_i) \right) \quad \text{mit} \\ X_i &\sim \frac{1}{2}(F_{i1} + F_{i2}). \end{aligned}$$

$X_i$  ist demnach eine Zufallsvariable, die verteilt ist nach der mittleren (ungewichteten) Verteilungsfunktion der Gruppe  $i$ . Mit  $i' \neq i$  folgt weiterhin für den gerade betrachteten Term:

$$\begin{aligned} p_{i1} + p_{i2} &= 2 \cdot \left( \int G d(F_{i1} + F_{i2}) / 2 \right) \\ &= 2 \cdot \left( \frac{1}{4}(F_{i1} + F_{i2} + F_{i'1} + F_{i'2}) d(F_{i1} + F_{i2}) / 2 \right) \\ &= 2 \cdot \left( \frac{1}{4} + \frac{1}{2} \left( P(X_{i'} < X_i) + \frac{1}{2}P(X_{i'} = X_i) \right) \right) \\ &= \frac{1}{2} + P(X_{i'} < X_i) + \frac{1}{2}P(X_{i'} = X_i). \end{aligned}$$

Damit ist die Nullhypothese  $\mathbf{c}_\lambda \mathbf{p} = 0$  gleichbedeutend mit

$$P(X_1 < X_2) + \frac{1}{2}P(X_1 = X_2) = \frac{1}{2},$$

d.h. die mittleren Verteilungsfunktionen der beiden Gruppen  $(F_{11} + F_{12})/2$  und  $(F_{21} + F_{22})/2$  sind nach Definition von DOMHOF (2001) tendenziell gleich. Eine Anschauung über diesen Sachverhalt wurde bereits bei Einführung der relativen Effekte gegeben.



Entsprechend ergibt sich für den Zeiteffekt, dass die Hypothese  $\mathbf{c}_\pi \mathbf{p} = 0$  gleichbedeutend ist mit der Aussage, dass die mittleren Verteilungsfunktionen der beiden Zeitpunkte  $(F_{11} + F_{21})/2$  und  $(F_{12} + F_{22})/2$  tendenziell gleich sind. Beim Cross-Over-Effekt gestaltet sich die Interpretation ein wenig schwerer. Allgemein lässt sich jedoch auch hier festhalten, dass die mittleren Verteilungsfunktionen  $(F_{11} + F_{22})/2$  und  $(F_{12} + F_{21})/2$  tendenziell gleich sind.

Wie auch sonst bei nichtparametrischen Verfahren üblich, können die Hypothesen ebenfalls über die Verteilungsfunktionen gestellt werden.

$$\begin{aligned} H_0^F(\lambda) : \mathbf{c}_\lambda \mathbf{F} = 0 &\iff F_{11} + F_{12} = F_{21} + F_{22} \\ H_0^F(\pi) : \mathbf{c}_\pi \mathbf{F} = 0 &\iff F_{11} + F_{21} = F_{12} + F_{22} \\ H_0^F(\delta) : \mathbf{c}_\delta \mathbf{F} = 0 &\iff F_{11} + F_{22} = F_{12} + F_{21} \end{aligned}$$

Zunächst einmal fällt folgender Zusammenhang zwischen dieser Hypothesenstellung und der zuvor besprochenen Hypothesenstellung über die relativen Effekte auf: Für alle Effekte ist die Hypothesenstellung über die Verteilungsfunktionen *ein Spezialfall* der Hypothesenstellung über die relativen Effekte, denn jeweils die Gleichheit der mittleren Verteilungsfunktionen ist eine Möglichkeit der tendenziellen Gleichheit zwischen diesen beiden Termen. Beispielsweise gilt unter der Voraussetzung von  $F_{11} + F_{12} = F_{21} + F_{22}$

$$p_{11} + p_{12} = P(X_2 < X_1) + \frac{1}{2}P(X_2 = X_1) = \frac{1}{2},$$

da die Verteilungsfunktionen für  $X_1$  und  $X_2$  identisch sind. Allgemein gilt folgender Zusammenhang zwischen den Hypothesen  $H_0^F$  und  $H_0^p$  im nichtparametrischen Modell. Die Hypothese über die Verteilungsfunktionen impliziert die Hypothese über die relativen Effekte, denn:

$$H_0^F : \mathbf{cF} = 0 \implies H_0^p : \mathbf{cp} = \mathbf{c} \int Gd\mathbf{F} = \int Gd(\mathbf{cF}) = 0$$

Umgekehrt gilt das jedoch nicht, da im nichtparametrischen Modell unter  $H_0^p$  die Verteilungen durchaus verschieden sein können, was beispielsweise mit den Skalenlokationsmodellen belegt werden kann. Diese Skalenlokationsmodelle sind symmetrische Verteilungen mit identischem Lagezentrum (Lokation), aber möglicherweise ungleichen Skalen. Ein mögliches Skalenlokationsmodell ist das klassische Behrens-Fisher-Problem. Die beiden unabhängigen Stichproben werden dabei als normalverteilt angenommen mit möglicherweise ungleichen Erwartungswerten und Varianzen. Unter der Nullhypothese  $p_1 = p_2$  (hier äquivalent zu  $\mu_1 = \mu_2$ ) sind zwar die Erwartungswerte gleich, die Varianzen der betrachteten Stichproben muss jedoch keineswegs gleich sein. Dieser Sachverhalt ist ausführlich in KULLE (1999) diskutiert.

Unter der Nullhypothese  $H_0^F : \mathbf{cF} = 0$  folgt weiterhin  $\mathbf{p} = 1/2 \cdot \mathbf{1}$ . Für das Zwei-Stichproben-Problem gibt es eine simple Anschauung. Die Hypothese über die relativen Effekte bedeutet in diesem Fall  $H_0^p : p_1 = p_2$ , also tendenzielle Gleichheit der Zufallsvariablen. In diesem Fall ist die Hypothese auch äquivalent zu  $H_0^\mu : \mu_1 = \mu_2$ .

### 4.2.3 Schätzer

Bezeichne  $\widehat{F}_{is}$  die empirische Verteilungsfunktion (siehe auch Definition A.3.6), die  $F_{is}$  schätzt, und analog dazu

$$\widehat{G}(x) = \frac{1}{4} \sum_{i=1}^2 \sum_{s=1}^2 \widehat{F}_{is}(x)$$

die mittlere empirische Verteilungsfunktion als Schätzer von  $G(x)$ . Beide Schätzer sind unter der Voraussetzung  $\min_i(n_i) \rightarrow \infty$  konsistent und erwartungstreu.

Die **Schätzer für die relativen Effekte** können definiert werden durch Einsetzen der empirischen Verteilungsfunktion und der mittleren gewichteten empirischen Verteilungsfunktion.

**Definition 4.2.4 (Schätzer für den relativen Effekt)** Sei  $\widehat{G}$  die mittlere empirische Verteilungsfunktion und  $\widehat{F}_{is}$  der Schätzer der Verteilungsfunktion  $F_{is}$ . Dann ist der Schätzer  $\widehat{p}_{is}$  des relativen Effekts  $p_{is}$  definiert als

$$\widehat{p}_{is} = \int \widehat{G} d\widehat{F}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \widehat{G}(X_{ik_s}) \quad \text{für alle } i, s = 1, 2.$$

Auch dieser Schätzer hat unter bestimmten Voraussetzungen die erwünschten Eigenschaften 1. Konsistenz und 2. Asymptotische Erwartungstreue, wie in Proposition A.3.7 auf Seite 116 im Anhang bewiesen wird.

Einzelne große Effekte werden bei diesem nichtparametrischen Konzept gedämpft, da diese Funktionale robust sind. Um dies nachzuweisen, benutzte DOMHOF (2001) die Definition des Grobe-Fehler Bruchpunkts. Dieser ist maximal, was folgendes bedeutet: Der Anteil der Beobachtungen, die durch Ausreißer ersetzt werden können, ohne dass der Schätzer alle in ihm enthaltene Information verliert, ist maximal. Ferner hätten selbst als robust geltende Schätzer, wie der empirische Median, nur einen Grobe-Fehler Bruchpunkt von 1/2.

Den Schätzer des relativen Effekts  $\widehat{p}_{is}$  kann man auch mit Hilfe von Rängen bzw. harmonischen Rängen darstellen. Ein Rang von einer Zufallsvariable beschreibt den Platz, den diese in der Reihenfolge der aufsteigend sortierten Zufallsvariablen einnimmt. In dieser Arbeit werden ausschließlich die sog. Mittel-Ränge verwendet. Bei Bindungen wird den entsprechenden Zufallsvariablen der mittlere Rang zugeordnet. Diesen Sachverhalt drückt die im folgenden veranschaulichte Mittel-Rang Definition aus. Als Beispiel betrachte man folgende geordnete Messreihe

$X_i$	1	2	2	2	4	7	7
$R_i$	1	3	3	3	5	6.5	6.5

wobei  $R_i$  den Rang der angeordneten Zufallsvariable  $X_i$  bezeichnet. Entsprechend ordnet man im Cross-Over-Design die Zufallsvariablen  $X_{is}$  für  $i, s = 1, 2$  der Größe nach aufsteigend an und vergibt dann die entsprechenden Ränge  $R_{is}$ .

Zusätzlich zu den soeben erklärten Mittel-Rängen gibt es noch Minimum- und Maximum-Ränge. Beim Minimum-Rang wird bei Bindungen jeweils der kleinste zu vergebene Rang

für alle betroffenen Werte vergeben. Entsprechend wird beim Maximum-Rang der maximal zu vergebene Rang vergeben. Die hohe Bedeutung der Mittel-Ränge kommt daher, dass die Summe und damit der Mittelwert der Ränge nicht von der Anzahl und dem Ausmaß der Bindungen abhängt.

Die Ränge dienen in dieser Arbeit in erster Linie als Übergang zu den harmonischen Rängen, die zur Darstellung der Schätzer des relativen ungewichteten Effekts benutzt werden. Durch Benutzung der Ränge für diesen Schätzer wird deutlich, warum der Schätzer invariant unter monotonen Transformationen ist und daher auch für ordinale Daten sehr gut geeignet ist. Ferner wird deutlich, dass der Schätzer relativ unempfindlich auf Ausreißer ist.

Neben der Unterscheidung der Ränge nach der Art des Umgangs mit Bindungen unterscheidet man (siehe Definition A.3.8) drei Arten von Rängen: Die Gesamt-Ränge, Intern-Ränge und Teil-Ränge.

- Der **Gesamt-Rang** bezeichnet im COP den (Mittel-)Rang  $R_{iks}$  der Beobachtung  $X_{iks}$  unter allen Beobachtungen aus beiden Gruppen zu beiden Zeitpunkten.
- Beim **Intern-Rang**  $R_{iks}^{(is)}$  betrachtet man ausschließlich die Beobachtungen des  $s$ -ten Zeitpunkts der  $i$ -ten Gruppe und bestimmt den Rang  $R_{iks}^{(is)}$  der Beobachtung  $X_{iks}$  innerhalb dieser Beobachtungen  $X_{i1s}, \dots, X_{in_1s}$ .
- Beim **Teil-Rang**  $R_{iks}^{(-i's')}$  bildet man den Rang von  $X_{iks}$  unter allen Beobachtungen mit Ausnahme der Beobachtungen der Gruppe  $i'$  zum Zeitpunkt  $s'$ .

Diese drei Arten der Rangdefinition werden in erster Linie dazu benötigt, die harmonischen Ränge darzustellen. Eine formale Definition befindet sich zusätzlich noch im Anhang auf Seite 117.

#### 4.2.4 Harmonische Ränge

Mit Hilfe der harmonischen Ränge lassen sich die Schätzer für den relativen ungewichteten Effekt oder die Varianzen gegenüber der eher formellen Darstellung über die Integrale vereinfacht aufschreiben und berechnen. Die harmonischen Ränge haben keine so einfache praktische Interpretation wie die Ränge; sie haben eine theoretische Bedeutung. Die harmonischen Ränge sind (noch) nicht in Computerprogrammen implementiert, man kann sie aber einfach über die Ränge berechnen. Der Zusammenhang dafür wird in Lemma 4.2.5 angegeben. Der Rang von  $X_{iks}$  wird mit im Folgenden mit  $R_{iks}$  bezeichnet, der harmonischen Rang von  $X_{iks}$  hingegen mit  $\Psi_{iks}$ .

Aufgrund der theoretischen Bedeutung der harmonischen Ränge wird an dieser Stelle auf die formale Definition verzichtet (siehe dafür Definition A.3.9 auf Seite A.3.9) und stattdessen der Zusammenhang angegeben, wie man die harmonischen Ränge aus den bereits bekannten Rängen gewinnen kann. Entsprechend den Rängen unterscheidet man auch bei den harmonischen Rängen die drei Arten Gesamt-, Intern- und Teil-Rang. Auch für die praktische Berechnung der harmonischen Ränge ist es wichtig, dass ein Zusammenhang zwischen Rängen und harmonischen Rängen angegeben werden kann.

Für eine übersichtliche Schreibweise wird an dieser Stelle noch das harmonische Mittel  $\tilde{n}$  der beiden Stichprobenumfänge eingeführt:

$$\frac{1}{\tilde{n}} := \frac{1}{4} \sum_{i=1}^2 \sum_{s=1}^2 \frac{1}{n_i} = \frac{1}{2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right).$$

**Lemma 4.2.5 (Zusammenhang zwischen Rängen und harmonischen Rängen)** Seien die Ränge und die harmonischen Ränge wie zuvor eingeführt. Dann bestehen für  $i, s = 1, 2$  und  $k = 1, \dots, n_i$  folgende Zusammenhänge:

1. für Ränge und harmonische Ränge

$$\Psi_{iks} = \sum_{(i',s') \neq (i,s)} \frac{\tilde{n}}{n_{i'}} \left( R_{iks} - R_{iks}^{(-i's')} \right) + \frac{\tilde{n}}{n_i} R_{iks}^{(is)}$$

2. für Internränge und harmonische Internränge

$$\Psi_{iks}^{(is)} = R_{iks}^{(is)}$$

3. für Teilränge und harmonische Teilränge mit  $(i, s) \neq (i', s')$

$$\Psi_{iks}^{(-i's')} = \sum_{\substack{(j,t) \neq (i',s') \\ (j,t) \neq (i,s)}} \frac{\tilde{n}}{n_j} \left( R_{iks} - R_{iks}^{(-jt)} \right) + \frac{\tilde{n}}{n_i} R_{iks}^{(is)}.$$

Für die praktische Berechnung werden die normierten Plazierungen noch eine große Hilfe sein. Durch diese normierten Plazierungen lassen sich empirische Verteilungsfunktionen sowie die ungewichtete mittlere Verteilungsfunktion mit Hilfe der harmonischen Ränge ausdrücken. Dadurch gelangt man anschließend unmittelbar zur Darstellung des Schätzers des relativen Effekts mit Hilfe der harmonischen Ränge.

**Lemma 4.2.6 (Normierten Plazierungen)** Bezeichne für  $i, s = 1, 2$  und  $k = 1, \dots, n_i$   $\hat{F}_{is}$  die empirische Verteilungsfunktion von  $X_{i1s}, \dots, X_{in_i s}$  und  $\hat{G}$  die mittlere ungewichtete empirische Verteilungsfunktion. Dann gilt:

$$\begin{aligned} \hat{G}(X_{iks}) &= \frac{1}{4\tilde{n}} \left( \Psi_{iks} - \frac{\tilde{n}}{2n_i} \right), \\ \hat{F}_{is}(X_{iks}) &= \frac{1}{n_i} \left( \Psi_{iks}^{(is)} - \frac{1}{2} \right) \text{ und} \\ \hat{F}_{i's'}(X_{iks}) &= \frac{1}{\tilde{n}} \left( \Psi_{iks} - \Psi_{iks}^{(-i's')} \right) \text{ für } (i', s') \neq (i, s). \end{aligned}$$

Jetzt kann eine einfache Darstellung des Schätzers für den relativen Effekt erfolgen. Dazu wird die Definition des Schätzers für den relativen Effekt aus 4.2.4 und Lemma 4.2.6 benötigt:

$$\hat{p}_{is} = \int \hat{G} d\hat{F}_{is} = \frac{1}{n_i} \sum_{k=1}^{n_i} \hat{G}(X_{iks}) = \frac{1}{4\tilde{n}} \left( \bar{\Psi}_{i.s} - \frac{\tilde{n}}{2n_i} \right).$$

### 4.2.5 Asymptotik und Grenzverteilungen

Bezeichne im weiteren Verlauf der Arbeit  $n = n_1 + n_2$  die Anzahl der Versuchseinheiten, die im Experiment vorkommen und  $N = 2n$  entsprechend die Anzahl der Beobachtungen insgesamt in beiden Perioden im Standard-COP.

Die wesentliche Grundlage der nachfolgenden Betrachtungen ist der asymptotische Äquivalenzsatz. Er besagt, dass  $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$  asymptotisch äquivalent ist zu einer Statistik, die sich in unabhängige Summanden zerlegen lässt. Varianzen und Kovarianzen dieser Vergleichsstatistik lassen sich dann ungleich einfacher bestimmen. Die Kovarianzmatrix (bzw. deren Schätzer) von  $\sqrt{N}\mathbf{c}\hat{\mathbf{p}}$  wird noch benötigt, damit die Teststatistiken zum Testen einzelner Effekte gebildet werden können.

Bezeichne im weiteren Verlauf der Arbeit

$$\mathbf{n}^{-1} := \begin{pmatrix} 1/n_1 \\ 1/n_2 \end{pmatrix} \otimes \mathbf{1}_2.$$

Die Einträge dieses Vektors sind demnach die reziproken Stichprobenumfänge der jeweiligen Zellen  $(i, s)$ .

Mit den bisher dargestellten Grundlagen soll die Verteilung für die Teststatistiken unter den Nullhypothesen  $H_0^F : \mathbf{c}\mathbf{F} = 0$  und  $H_0^p : \mathbf{c}\mathbf{p} = 0$  hergeleitet werden. Der Beweis der nun folgenden Ergebnisse befindet sich der Lesbarkeit halber wiederum im Anhang in Satz A.3.15 auf Seite 121.

Sei  $Y_{iks} = G(X_{iks})$ . Dann bezeichne im Weiteren

$$\hat{Y}_{iks} = \hat{G}(X_{iks}) = \frac{1}{4\tilde{n}} \left( \Psi_{iks} - \frac{\tilde{n}}{2n_i} \right) \quad (4.13)$$

die asymptotische Harmonische-Rang Transformation AHRT. Auch diese beiden Größen lassen sich kanonisch in einem Vektor  $\mathbf{Y}$  anordnen:

$$\begin{aligned} \mathbf{Y} &= (Y_{111}, \dots, Y_{2n_22})^T \quad \text{und} \\ \hat{\mathbf{Y}} &= (\hat{Y}_{111}, \dots, \hat{Y}_{2n_22})^T. \end{aligned}$$

Desweiteren sei

$$\begin{aligned} \bar{\mathbf{Y}}_{i.} &= (\bar{Y}_{i.1}, \bar{Y}_{i.2})^T \quad \text{mit} \quad \bar{Y}_{i.s} = \frac{1}{n_i} \sum_{i=1}^2 Y_{iks} \quad \text{und} \\ \bar{\mathbf{Y}}_{.} &= (\bar{\mathbf{Y}}_{1.}^T, \bar{\mathbf{Y}}_{2.}^T)^T. \end{aligned}$$

Unter der Voraussetzung, dass die Stichprobenumfänge gleichmäßig gegen unendlich streben und die Kovarianzmatrizen  $V_{i,n_i} = \text{Cov}(\bar{\mathbf{Y}}_{i.})$  konvergieren, kann die asymptotische Normalität der gewählten Teststatistik unter  $H_0^F : \mathbf{c}\mathbf{F} = 0$  und die Konsistenz des Varianzschätzers bewiesen werden.

**Satz 4.2.7 (Asymptotische Normalität unter  $H_0^F : \mathbf{c}\mathbf{F} = 0$ )**

$$\sqrt{N}\mathbf{c}\hat{\mathbf{p}} = \sqrt{N}\mathbf{c} \int \hat{G}d\hat{\mathbf{F}} = \sqrt{N}\mathbf{c}\hat{\bar{\mathbf{Y}}}_{.} = \frac{\sqrt{N}}{4\tilde{n}}\mathbf{c} \left( \bar{\Psi}_{.} - \frac{\tilde{n}}{2}\mathbf{n}^{-1} \right)$$

ist unter  $H_0^F : \mathbf{cF} = 0$  asymptotisch normalverteilt mit Erwartungswert 0 und Varianz  $\mathbf{cV}_N \mathbf{c}^T$ . Dabei lässt sich  $\mathbf{V}_N$  darstellen als

$$\mathbf{V}_N = \bigoplus_{i=1}^2 \frac{N}{n_i} \mathbf{V}_{i,n_i}$$

mit folgendem konsistenten Schätzer für  $\mathbf{V}_{i,n_i} = \text{Cov}(\bar{\mathbf{Y}}_i)$

$$\hat{\mathbf{V}}_{i,n_i} = \frac{1}{16\tilde{n}^2(n_i - 1)} \sum_{k=1}^{n_i} (\Psi_{ik} - \bar{\Psi}_i) (\Psi_{ik} - \bar{\Psi}_i)^T.$$

Dabei ist  $\Psi_{ik} = (\Psi_{ik1}, \Psi_{ik2})^T$  und entsprechend  $\bar{\Psi}_i = (\bar{\Psi}_{i.1}, \bar{\Psi}_{i.2})^T$ .

Genauer bedeutet dies, dass unter der Nullhypothese  $H_0^F : \mathbf{cF} = 0$  die Kovarianzmatrix  $\mathbf{V}_N$  nicht von  $\sqrt{N}\hat{\mathbf{p}}$  berechnet werden muss, sondern stattdessen auf die asymptotisch äquivalente Vergleichsstatistik  $\sqrt{N}\mathbf{Y}$  zurückgegriffen werden kann.  $\hat{\mathbf{V}}_{i,n_i}$  bezeichnet entsprechend den Schätzer der Kovarianzmatrix von  $\bar{\mathbf{Y}}_i = (\bar{Y}_{i.1}, \bar{Y}_{i.2})^T$ .

Mit Hilfe dieser Information ist es jetzt für die Nullhypothese  $H_0^F : \mathbf{cF} = 0$  möglich, die entsprechenden Teststatistiken anzugeben. Die Teststatistiken sind im folgenden Abschnitt 4.2.6 angegeben.

Zur Bildung der Teststatistiken unter der Hypothese  $H_0^p : \mathbf{cp} = 0$  muss man zunächst einen (komplizierteren) Varianzschätzer finden. Anschließend kann man die asymptotische Verteilung der Teststatistik herleiten. Dass der Varianzschätzer bei dieser Art der Hypothesenstellung komplexer wird, kommt daher, dass die Vergleichsstatistik unter  $H_0^F : \mathbf{cF} = 0$  eine wesentlich einfachere Struktur hat.

Werden die Hypothesen über die relativen Effekte gestellt, behilft man sich mit einer nicht ganz so einfachen Vergleichsstatistik zu  $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$

$$\mathbf{Z}_N := \sqrt{N} \left( \int G d\hat{\mathbf{F}} - \int \mathbf{F} d\hat{G} + \mathbf{1}_4 - 2\mathbf{p} \right).$$

Um Teststatistiken herleiten zu können, muss zunächst einmal die Kovarianzmatrix von  $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$  berechnet und geschätzt werden. Dieselbe erhält man wiederum einfacher über die eben genannte Vergleichsstatistik  $\mathbf{Z}_N$ . Der folgende Satz gilt unter den Voraussetzungen, dass die Stichprobenumfänge gleichmäßig gegen unendlich streben. Eine weitere Voraussetzung an die Kovarianzmatrix  $\mathbf{W}_N$  von  $\mathbf{Z}_N$  ist mehr formeller Natur und befindet sich mitsamt dem Beweis im Anhang auf Seite 125

**Satz 4.2.8 (Asymptotische Normalität unter  $H_0^p : \mathbf{cp} = 0$ )**

$$\sqrt{N}\mathbf{c}\hat{\mathbf{p}} = \sqrt{N}\mathbf{c} \int \hat{G} d\hat{\mathbf{F}} = \frac{\sqrt{N}}{4\tilde{n}} \mathbf{c} \left( \bar{\Psi} - \frac{\tilde{n}}{2} \mathbf{n}^{-1} \right)$$

ist unter  $H_0^p : \mathbf{cp} = 0$  asymptotisch normalverteilt mit Erwartungswert 0 und Varianz  $\mathbf{cW}_N \mathbf{c}^T$ . Dabei lässt sich  $\hat{\mathbf{W}}_N$  für  $j \neq i$  darstellen mit den Diagonalelementen

$$\hat{c}_{iiss} = \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \hat{Z}_{iks} - \hat{Z}_{i.s} \right)^2 + \frac{N}{16n_j(n_j - 1)} \sum_{k=1}^{n_j} \left( \hat{Z}_{j^{(-is)}} - \hat{Z}_{j.}^{(-is)} \right)^2.$$

Die Kovarianzschätzer innerhalb einer Gruppe  $i$  ergeben sich

$$\begin{aligned}\widehat{C}_{iiss'} &= \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right) \left( \widehat{Z}_{iks'} - \widehat{Z}_{i.s'} \right) \\ &\quad + \frac{N}{16n_j(n_j - 1)} \sum_{k=1}^{n_j} \left( \widehat{Z}_{jk}^{(-is)} - \widehat{Z}_j^{(-is)} \right) \left( \widehat{Z}_{jk}^{(-is')} - \widehat{Z}_j^{(-is')} \right) \quad j \neq i\end{aligned}$$

und zwischen den Gruppen für  $i \neq i'$

$$\begin{aligned}\widehat{C}_{ii'ss'} &= -\frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right) \left( \widehat{Z}_{ik}^{(-i's')} - \widehat{Z}_i^{(-i's')} \right) \\ &\quad - \frac{N}{16n_{i'}(n_{i'} - 1)} \sum_{k'=1}^{n_{i'}} \left( \widehat{Z}_{i'k's'} - \widehat{Z}_{i'.s'} \right) \left( \widehat{Z}_{i'k'}^{(-is)} - \widehat{Z}_{i'}^{(-is)} \right).\end{aligned}$$

$\widehat{W}_N$  ist ein konsistenter Schätzer für  $W_N$ .

Die schrittweise Herleitung dieses Kovarianzschätzers sowie die Konsistenz desselben befinden sich im Anhang ab Seite 121.

Wie aus dem vorangegangenen Satz abzulesen ist, lässt sich die Schätzung der Kovarianzmatrix im Wesentlichen auf die Schätzung der einzelnen Komponenten  $Z_{iks}$  und  $Z_{jk}^{(-is)}$  zurückführen, die sich mit Hilfe der harmonischen Ränge folgendermaßen ausdrücken lassen:

$$\begin{aligned}\widehat{Z}_{iks} &= 4 \cdot \widehat{G}(X_{iks}) - \sum_{t=1}^2 \widehat{F}_{is}(X_{ikt}) \\ &= \frac{1}{\widetilde{n}} \left( \Psi_{iks} - \frac{\widetilde{n}}{2n_i} \right) - \frac{1}{n_i} \left( \Psi_{iks}^{(is)} - \frac{1}{2} \right) - \frac{1}{\widetilde{n}} \left( \Psi_{ikt} - \Psi_{ikt}^{(-is)} \right) \quad \text{mit } t \neq s\end{aligned}$$

und

$$\begin{aligned}\widehat{Z}_{jk}^{(-is)} &= -\sum_{t=1}^2 \widehat{F}_{is}(X_{jkt}) \\ &= \frac{1}{\widetilde{n}} \sum_{t=1}^2 \left( \Psi_{jkt}^{(-is)} - \Psi_{jkt} \right) \quad \text{für } j \neq i.\end{aligned}$$

Damit hat man alle Grundlagen zusammen, um Teststatistiken unter der Nullhypothese  $H_0^p : \mathbf{c}p = 0$  herleiten zu können. Diese befinden sich in Abschnitt 4.2.7.

#### 4.2.6 Teststatistiken für $H_0^F$

Mit Hilfe der Ergebnisse des vorigen Abschnitts lassen sich die Teststatistiken allgemein über folgendes Resultat ausrechnen:

$$U_n^F = \frac{\sqrt{N} \mathbf{c} \widehat{p}}{\sqrt{\mathbf{c} \widehat{V}_N \mathbf{c}^T}} \quad \overset{\cdot}{\sim} \quad N(0, 1). \quad (4.14)$$

Dabei muss neben den Anforderungen für Satz 4.2.7 erfüllt sein:  $\mathbf{c}\widehat{\mathbf{V}}_N\mathbf{c}^T \geq c_0 > 0$ .

**Statistik für  $H_0^F(\lambda) : \bar{F}_{1.} = \bar{F}_{2.}$**

$$\begin{aligned} U_n^{\lambda,F} &= \frac{\sqrt{N}\mathbf{c}_\lambda\widehat{\mathbf{p}}}{\sqrt{\mathbf{c}_\lambda\widehat{\mathbf{V}}_N\mathbf{c}_\lambda^T}} \\ &= \frac{\bar{\Psi}_{1.1} + \bar{\Psi}_{1.2} - \bar{\Psi}_{2.1} - \bar{\Psi}_{2.2} - \left(\frac{\tilde{n}}{n_1} - \frac{\tilde{n}}{n_2}\right)}{\sqrt{\sum_{i=1}^2 \frac{\widehat{\sigma}_{i,F}^2}{n_i}}} \quad \text{mit} \\ \widehat{\sigma}_{i,F}^2 &= \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\Psi_{ik1} + \Psi_{ik2} - \bar{\Psi}_{i.1} - \bar{\Psi}_{i.2})^2 \end{aligned}$$

**Statistik für  $H_0^F(\pi) : \bar{F}_{.1} = \bar{F}_{.2}$**

$$\begin{aligned} U_n^{\pi,F} &= \frac{\sqrt{N}\mathbf{c}_\pi\widehat{\mathbf{p}}}{\sqrt{\mathbf{c}_\pi\widehat{\mathbf{V}}_N\mathbf{c}_\pi^T}} \\ &= \frac{\bar{\Psi}_{1.1} - \bar{\Psi}_{1.2} + \bar{\Psi}_{2.1} - \bar{\Psi}_{2.2}}{\sqrt{\sum_{i=1}^2 \frac{\widehat{\tau}_{i,F}^2}{n_i}}} \quad \text{mit} \\ \widehat{\tau}_{i,F}^2 &= \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\Psi_{ik1} - \Psi_{ik2} - \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2})^2 \end{aligned}$$

**Statistik für  $H_0^F(\delta) : F_{11} - F_{12} = F_{21} - F_{22}$**

$$\begin{aligned} U_n^{\delta,F} &= \frac{\sqrt{N}\mathbf{c}_\delta\widehat{\mathbf{p}}}{\sqrt{\mathbf{c}_\delta\widehat{\mathbf{V}}_N\mathbf{c}_\delta^T}} \\ &= \frac{\bar{\Psi}_{1.1} - \bar{\Psi}_{1.2} - \bar{\Psi}_{2.1} + \bar{\Psi}_{2.2}}{\sqrt{\sum_{i=1}^2 \frac{\widehat{\tau}_{i,F}^2}{n_i}}} \quad \text{mit} \\ \widehat{\tau}_{i,F}^2 &= \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\Psi_{ik1} - \Psi_{ik2} - \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2})^2 \end{aligned}$$

Unter der Nullhypothese  $H_0^F : \mathbf{c}\mathbf{F} = 0$  haben die Statistiken nach (4.14) asymptotisch eine Standardnormalverteilung. Für kleine Stichprobenumfänge approximiert man die Verteilung der Statistiken unter Nullhypothese mit Hilfe der  $t$ -Verteilung. Die Berechnungen befinden sich im Anhang ab Seite 127. Der Freiheitsgrad von  $U_n^{\lambda,F}$  ist

$$\frac{(\sum_{i=1}^2 \widehat{\sigma}_{i,F}^2/n_i)^2}{\sum_{i=1}^2 (\widehat{\sigma}_{i,F}^2/n_i)^2/(n_i - 1)}$$

und der Freiheitsgrad für die Statistiken  $U_n^{\pi,F}$  und  $U_n^{\delta,F}$  ergibt sich durch

$$\frac{(\sum_{i=1}^2 \widehat{\tau}_{i,F}^2/n_i)^2}{\sum_{i=1}^2 (\widehat{\tau}_{i,F}^2/n_i)^2/(n_i - 1)}.$$



### 4.2.7 Teststatistiken für $H_0^p$

Mit Hilfe der Ergebnisse des Abschnitts 4.2.5 und der Voraussetzung  $\mathbf{c}\widehat{\mathbf{W}}_N\mathbf{c}^T \geq c_0 > 0$  lassen sich die Teststatistiken allgemein über folgendes Resultat ausrechnen:

$$U_n^p = \frac{\sqrt{N}\mathbf{c}\widehat{\mathbf{p}}}{\sqrt{\mathbf{c}\widehat{\mathbf{W}}_N\mathbf{c}^T}} \underset{\sim}{\sim} N(0, 1). \quad (4.15)$$

**Statistik für  $H_0^p(\lambda) : \bar{p}_{.1} = \bar{p}_{.2}$ .**

$$\begin{aligned} U_n^{\lambda,p} &= \frac{\sqrt{N}\mathbf{c}_\lambda\widehat{\mathbf{p}}}{\sqrt{\mathbf{c}_\lambda\widehat{\mathbf{W}}_N\mathbf{c}_\lambda^T}} \\ &= \frac{\sqrt{N}/(4\tilde{n}) \cdot \left( \bar{\Psi}_{1.1} + \bar{\Psi}_{1.2} - \bar{\Psi}_{2.1} - \bar{\Psi}_{2.2} - \left( \frac{\tilde{n}}{n_1} - \frac{\tilde{n}}{n_2} \right) \right)}{\sqrt{\mathbf{c}_\lambda\widehat{\mathbf{W}}_N\mathbf{c}_\lambda^T}} \end{aligned}$$

**Statistik für  $H_0^p(\pi) : \bar{p}_{.1} = \bar{p}_{.2}$**

$$\begin{aligned} U_n^{\pi,p} &= \frac{\sqrt{N}\mathbf{c}_\pi\widehat{\mathbf{p}}}{\sqrt{\mathbf{c}_\pi\widehat{\mathbf{W}}_N\mathbf{c}_\pi^T}} \\ &= \frac{\sqrt{N}/(4\tilde{n}) \cdot (\Psi_{1.1} - \Psi_{1.2} + \Psi_{2.1} - \Psi_{2.2})}{\sqrt{\mathbf{c}_\pi\widehat{\mathbf{W}}_N\mathbf{c}_\pi^T}} \end{aligned}$$

**Statistik für  $H_0^p(\delta) : p_{11} - p_{12} = p_{21} - p_{22}$**

$$\begin{aligned} U_n^{\delta,p} &= \frac{\sqrt{N}\mathbf{c}_\delta\widehat{\mathbf{p}}}{\sqrt{\mathbf{c}_\delta\widehat{\mathbf{W}}_N\mathbf{c}_\delta^T}} \\ &= \frac{\sqrt{N}/(4\tilde{n}) \cdot (\bar{\Psi}_{1.1} - \bar{\Psi}_{1.2} - \bar{\Psi}_{2.1} + \bar{\Psi}_{2.2})}{\sqrt{\mathbf{c}_\delta\widehat{\mathbf{W}}_N\mathbf{c}_\delta^T}} \end{aligned}$$

Unter der Nullhypothese  $H_0^p : \mathbf{c}\mathbf{p} = 0$  haben die Statistiken asymptotisch eine Standardnormalverteilung. Für kleine Stichprobenumfänge approximiert man die Verteilung der Statistiken unter Nullhypothese mit Hilfe der  $t$ -Verteilung. Die Freiheitsgrade ergeben sich dabei nach der sogenannten Satterthwaite-Smith-Welch-Approximation (siehe ab Seite 127):

Der Freiheitsgrad für die Approximation mit der  $t$ -Verteilung von  $U_n^{\lambda,p}$ :

$$\frac{\left( \sum_{i=1}^2 \widehat{\sigma}_{i,p}^2 / n_i \right)^2}{\sum_{i=1}^2 (\widehat{\sigma}_{i,p}^2 / n_i)^2 / (n_i - 1)},$$

wobei

$$\begin{aligned} \widehat{\sigma}_{i,p}^2 &= \frac{1}{(n_i - 1)} \sum_{k=1}^{n_i} \left( -\Psi_{ik1} - \Psi_{ik2} + 2\Psi_{ik}^{(-j1)} + 2\Psi_{ik}^{(-j2)} \right. \\ &\quad \left. + \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2} - 2\bar{\Psi}_{i.}^{(-j1)} - 2\bar{\Psi}_{i.}^{(-j2)} \right)^2 \quad \text{mit } j \neq i. \end{aligned}$$

Der Freiheitsgrad für die Approximation mit der  $t$ -Verteilung von  $U_n^{\pi,p}$ :

$$\frac{\left(\sum_{i=1}^2 \hat{\tau}_{i,p}^2/n_i\right)^2}{\sum_{i=1}^2 (\hat{\tau}_{i,p}^2/n_i)^2/(n_i - 1)},$$

wobei

$$\begin{aligned} \hat{\tau}_{i,p}^2 &= \frac{1}{(n_i - 1)} \sum_{k=1}^{n_i} \left( \Psi_{ik1} - \Psi_{ik2} - 2\Psi_{ik}^{(-j1)} + 2\Psi_{ik}^{(-j2)} \right. \\ &\quad \left. - \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2} + 2\bar{\Psi}_i^{(-j1)} - 2\bar{\Psi}_i^{(-j2)} \right)^2 \quad \text{mit } j \neq i \end{aligned}$$

und der Freiheitsgrad für die Approximation mit der  $t$ -Verteilung von  $U_n^{\delta,p}$ :

$$\frac{\left(\sum_{i=1}^2 \hat{v}_{i,p}^2/n_i\right)^2}{\sum_{i=1}^2 (\hat{v}_{i,p}^2/n_i)^2/(n_i - 1)},$$

wobei

$$\begin{aligned} \hat{v}_{i,p}^2 &= \frac{1}{(n_i - 1)} \sum_{k=1}^{n_i} \left( \Psi_{ik1} - \Psi_{ik2} + 2\Psi_{ik}^{(-j1)} - 2\Psi_{ik}^{(-j2)} \right. \\ &\quad \left. - \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2} - 2\bar{\Psi}_i^{(-j1)} + 2\bar{\Psi}_i^{(-j2)} \right)^2 \quad \text{mit } j \neq i. \end{aligned}$$

#### 4.2.8 Eigenschaften der Teststatistiken

Mit den Verteilungsfunktionen hängen im linearen Modell auch die empirischen Verteilungsfunktionen sowie die Teststatistiken von den Parametern  $\lambda$ ,  $\Phi$  und  $\pi$  ab. Damit die auf  $p^\lambda$  basierende Statistik  $U_n^\lambda$  in einem linearen Modell einen sinnvollen Gruppeneffekt testen soll, muss die Teststatistik größer werden, wenn  $\lambda$  größer wird, d.h.  $U_n^\lambda$  sollte streng monoton in  $\lambda$  sein.

Falls für  $\lambda = 0$  der (asymptotische) Erwartungswert ebenfalls gleich 0 ist, deckt man mit einem auf  $U_n^\lambda$  basierenden Test die Alternativen der Form  $\lambda \neq 0$  auf, d.h. dieselben Alternativen, die man mit einem Verfahren für ein lineares Modell aufdecken würde. Die im nichtparametrischen Modell entwickelten Verfahren stellen dann eine sinnvolle Verallgemeinerung der bekannten Statistiken des linearen Modells dar.

Somit folgt aus Satz 4.2.3 auf Seite 31 unter Berücksichtigung der asymptotischen Erwartungstreue der relativen Effekte folgendes Lemma. Dabei steht  $U_n^\lambda$  stellvertretend für  $U_n^{\lambda,F}$  und  $U_n^{\lambda,p}$  sowie entsprechend für die anderen Statistiken.

**Lemma 4.2.9 (Asymptotische Erwartungstreue der Teststatistiken)** *Im linearen Marginalmodell gelten folgende Aussagen:*

1. Der asymptotische Erwartungswert von  $U_n^\lambda$  ist genau dann gleich 0, wenn  $\lambda = 0$  ist.
2. Der asymptotische Erwartungswert von  $U_n^\pi$  ist genau dann gleich 0, wenn  $\pi = 0$  ist.
3. Der asymptotische Erwartungswert von  $U_n^\delta$  ist genau dann gleich 0, wenn  $\delta = 0$  ist.

## 4.3 Auswertung der Beispiele

In diesem Abschnitt wird zunächst der Datensatz von PATEL (1983) mit zweierlei Methoden ausgewertet. Zunächst einmal wird das lineare Modell von LEHMACHER (1987) zugrunde gelegt und anschließend die neu entwickelten nichtparametrischen Verfahren auf dieses Beispiel angewendet. Die beiden Ansätze werden zuletzt miteinander verglichen. Im Abschnitt 4.3.2 wird dann eine mögliche Auswertung für den Datensatz von SENN (1993) vorgeschlagen.

### 4.3.1 Asthma-Studie II

Basierend auf dem Datensatz von PATEL (1983) können mit Hilfe der in diesem Kapitel gewonnenen Verfahren Teststatistiken berechnet werden.

Für dieses Beispiel soll die Fragestellung untersucht werden, ob ein Residual-, Perioden- oder Cross-Over-Effekt vorliegt. Die Kernfrage lautet selbstverständlich: Unterscheiden sich auf einem Niveau von 5% die beiden Behandlungen  $A$  und  $B$ ?

Für die Analyse des Beispiels soll zunächst einmal vorausgesetzt werden, dass das **lineare Modell** (3.1) gültig ist, d.h. die Additivität der einzelnen Effekte angenommen werden kann. Ferner sei die Normalverteilung der Daten vorausgesetzt. Um jetzt als erstes den Test auf einen Residualeffekt durchzuführen, bedient man sich den in Abschnitt 4.1.2.1 auf Seite 24 angegebenen Verfahren. Nach der Bildung der Individuumssummen  $s_{ik} = X_{ik1} + X_{ik2}$  lässt sich die Teststatistik wie folgt berechnen:

$$U_{LEH}^{\lambda} = \frac{\bar{s}_{1.} - \bar{s}_{2.}}{\hat{\sigma}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

mit  $\hat{\sigma}^2$  wie dort angegeben. Für den Varianzschätzer erhält man:

$$\hat{\sigma}^2 = 1.686,$$

was zur Teststatistik

$$\begin{aligned} U_{LEH}^{\lambda} &= \frac{3.263 - 4.287}{1.686} \cdot \sqrt{\frac{144}{24}} \\ &= -1.623 \end{aligned}$$

führt. Damit erhält man den  $p$ -Wert

$$p\text{-Wert} = 2 \cdot t^{-1}(U_{LEH}^{\lambda}, 22) = 0.125 > 0.05.$$

Auf einem Niveau von 5% kann demnach kein Residualeffekt nachgewiesen werden. Für die Statistiken für einen Periodeneffekt und einen Cross-Over-Effekt ergeben sich die Teststatistiken in analoger Art und Weise. Die entsprechenden  $p$ -Werte sind in Tabelle 4.2 niedergeschrieben.

Ein Periodeneffekt kann auf dem zugrunde gelegten Niveau von 5% ebenfalls nicht nachgewiesen werden, wohl aber (streng genommen) ein Cross-Over-Effekt bzw. verallgemeinerter Behandlungseffekt. Falls ein Nacheffekt a priori ausgeschlossen wurde, lässt sich (streng genommen) schlussfolgern, dass sich die beiden Behandlungen unterscheiden.

Tabelle 4.2:  $p$ -Werte der Asthma-Studie II - lineares Modell

	$p$ -Werte
$\lambda$	0.125
$\pi$	0.260
$\delta$	0.047

Lässt man die Voraussetzungen der Linearität und Verteilungsannahme fallen und wählt das **nichtparametrische Modell (GM)**, so geht man wie folgt vor. Dabei sollen die Hypothesen über die Verteilungsfunktionen formuliert sein:  $H_0^F : \mathbf{cF} = 0$ . Aufgrund des relativ geringen Stichprobenumfangs von  $n_1 = 8$  und  $n_2 = 9$  werden die Teststatistiken mit der  $t$ -Verteilung approximiert.

Zunächst soll die Frage beantwortet werden, ob ein Residualeffekt vorhanden ist. Es liegt im Interesse des Untersuchers, ob vielleicht die Wash-Out-Phase nicht adäquat war oder möglicherweise ein starker Nacheffekt der Behandlungen aus Periode 1 vorliegt. Die Teststatistik für diese Fragestellung ergibt sich aus Abschnitt 4.2.6 auf Seite 39. Die Rechnung wird hier einmal exemplarisch für die Teststatistik auf einen Residualeffekt durchgeführt. Die Nullhypothese  $H_0^F$  bedeutet in diesem Fall, dass die Summe der Verteilungsfunktionen in beiden Gruppen gleich sind. Es wurden alle Zwischenergebnisse, wie beispielsweise die harmonischen Rangmittel, mit dem Computer berechnet.

$$\begin{aligned}
 U_n^{\lambda,F} &= \frac{\bar{\Psi}_{1,1} + \bar{\Psi}_{1,2} - \bar{\Psi}_{2,1} - \bar{\Psi}_{2,2} - \left( \frac{\tilde{n}}{n_1} - \frac{\tilde{n}}{n_2} \right)}{\sqrt{\sum_{j=1}^2 \frac{\hat{\sigma}_j^2}{n_j}}} \\
 &= \frac{13.875 + 14.596 - 23.529 - 17.765 - \left( \frac{8.471}{8} - \frac{8.471}{9} \right)}{\sqrt{\frac{342.039}{8} + \frac{300.235}{9}}} = -1.483.
 \end{aligned}$$

Für das entsprechende Quantil der  $t$ -Verteilung wird noch der Freiheitsgrad benötigt:

$$\begin{aligned}
 fg &= \frac{\left( \sum_{i=1}^2 \hat{\sigma}_i^2 / n_i \right)^2}{\sum_{i=1}^2 (\hat{\sigma}_i^2 / n_i)^2 / (n_i - 1)} \\
 &= \frac{\left( \frac{342.039}{8} + \frac{300.235}{9} \right)^2}{\left( \frac{342.039}{8} \right)^2 / 7 + \left( \frac{300.235}{9} \right)^2 / 8} \\
 &= 14.475.
 \end{aligned}$$

Damit ergibt sich der  $p$ -Wert wie folgt:

$$p\text{-Wert} = 2 \cdot t^{-1} \left( U_n^{\lambda,F}; 14.475 \right) = 0.159.$$

Das bedeutet, dass auf einem Niveau von 5% gilt:  $0.159 > \alpha$ . Damit kann die Nullhypothese nicht verworfen werden, also kein Residualeffekt nachgewiesen werden. Das darf allerdings nicht dazu führen, diese Tatsache als Beweis für die Nullhypothese, also keinen Residualeffekt, zu interpretieren.

Die  $p$ -Werte für den Zeiteffekt und den Cross-Over-Effekt ergeben sich auf die gleiche Art und Weise. Alle relevanten  $p$ -Werte sind in Tabelle 4.3 zusammengefasst.

Tabelle 4.3:  $p$ -Werte der Asthma-Studie II - nichtparametrisches Modell

	$p$ -Werte
$\lambda$	0.159
$\pi$	0.143
$\delta$	0.066

Aus Tabelle 4.3 lässt sich folgendes ablesen: Neben dem bereits diskutierten Nacheffekt, bleiben die Ergebnisse für einen eventuellen Perioden- und Cross-Over-Effekt zu diskutieren. Mit einer Irrtumswahrscheinlichkeit von 5% kann man ebenso keinen Periodeneffekt nachweisen, was wiederum nicht dazu verleiten darf, diesen aufgrund dieses Tests auszuschließen. Der Test auf einen Cross-Over-Effekt ist ebenfalls nicht signifikant, d.h. dass (streng genommen) kein Unterschied in den Differenzen der Verteilungsfunktionen (erste minus zweite Periode) pro Gruppe angenommen werden kann. Selbst wenn a priori ein Residualeffekt ausgeschlossen werden kann, kann hier (ebenfalls streng genommen) nicht auf einen Unterschied der beiden Behandlungen geschlossen werden (Test auf einen reinen Behandlungsunterschied).

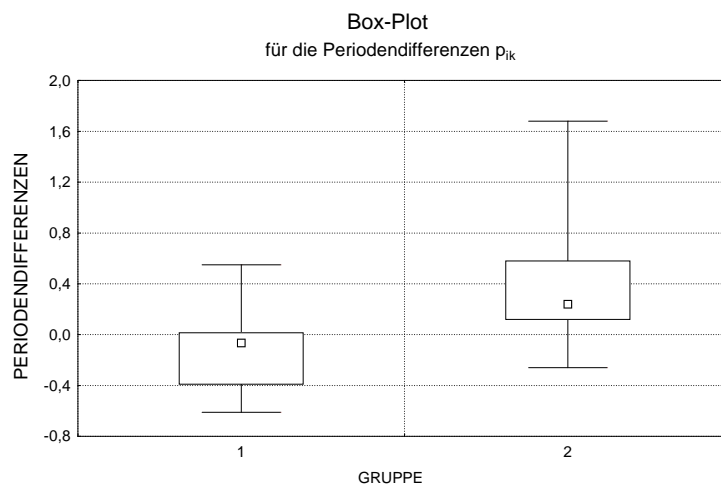
Stellt man abschließend für dieses Beispiel einen **Vergleich** zwischen den Ergebnissen der beiden Verfahren her, so stößt man zuerst auf unterschiedliche  $p$ -Werte für ein und dieselbe Fragestellung. Das hängt einfach damit zusammen, dass die Teststatistiken für diese unterschiedlichen Voraussetzungen sehr verschieden sind: die Teststatistik für das lineare Modell arbeitet mit den (Original-) Messwerten, während die Teststatistiken für das nichtparametrische Modell die Ränge als Berechnungsgrundlage wählen. In diesem Fall führt die Anwendung des linearen Modells in der Frage eines Cross-Over-Effekts zu einer Signifikanz, die bei Zugrundelegung des nichtparametrischen Modells nicht bestätigt werden kann.

Wenn jedoch die Voraussetzungen des unverbundenen  $t$ -Tests nicht erfüllt wären, könnte die Wahl des linearen Modells mit der Voraussetzung der Normalverteilung an dieser Stelle zu einer Fehlentscheidung hinsichtlich eines Behandlungseffekts führen. Dass die Voraussetzungen nicht erfüllt waren, kann man mit Hilfe von Abbildung 4.2 nachvollziehen. Weder ist eine ausgeprägte Symmetrie der Daten in beiden Gruppen zu erkennen, noch kann eine Gleichheit der Varianzen angenommen werden.

### 4.3.2 Asthma-Studie I

Zielvariable dieser Studie ist der Score, den der Untersuchungsleiter aufgrund der verschiedenen Lungenfunktionsparameter vergeben hat. Da dieses Beispiel demnach ordinale Daten enthält, kann das lineare Modell nach LEHMACHER (1987) nicht angewendet werden. Die Voraussetzung der Normalverteilung ist nicht erfüllt. Legt man das nichtparametrische Modell zugrunde, so hat man die Wahl zwischen Hypothesen, die über die Verteilungsfunktionen gestellt werden und Hypothesen, die über die relativen Effekte gestellt werden. In diesem Beispiel soll die Hypothese in der Form  $H_0^p : \mathbf{c}\mathbf{p} = 0$  über die relativen Effekte gewählt werden.

Auch hier wird aufgrund des geringen Stichprobenumfangs mit Hilfe der  $t$ -Verteilung gearbeitet. Die Teststatistiken lassen sich in diesem Falle aus Abschnitt 4.2.7 auf Seite 41 entnehmen. Die Berechnung verläuft analog, jedoch mit einer anderen Kovarianzma-

Abbildung 4.2: Box-Plot der Periodenunterschiede  $p_{ik}$  der Asthma-Studie IITabelle 4.4:  $p$ -Werte der Asthma-Studie I

	$p$ -Werte
$\lambda$	0.095
$\pi$	0.000
$\delta$	0.102

trix. Die resultierenden  $p$ -Werte sind in Tabelle 4.4 festgehalten. Der Test auf einen Residualeffekt ergibt, dass unter Berücksichtigung der Irrtumswahrscheinlichkeit nicht auf einen Unterschied der Nacheffekte in beiden Gruppen geschlossen werden kann. Starke Signifikanz hingegen weist der Test auf einen Periodeneffekt auf, d.h. dass sich die Summe der relativen Effekte zum ersten Zeitpunkt stark von denen des zweiten Zeitpunkts unterscheiden. Wenn man davon ausgeht, dass kein Residualeffekt vorhanden ist, dann ist ein reiner Periodeneffekt vorhanden, d.h. man hat hier einen Unterschied zu beiden Zeitpunkten unabhängig von der Behandlung. Ursachen können andere Ausgangsbedingungen, Gewöhnung der Probanden oder andere äußere Einflüsse sein, die sich über die Zeit ändern. Beispielsweise ist ja gerade Asthma nicht gänzlich unabhängig vom Wetter bzw. bestimmten daraus resultierenden Ereignissen wie beispielsweise Pollenflug o.ä. Ferner wird kein (verallgemeinerter) Behandlungsunterschied festgestellt, da der Test auf einen Cross-Over-Effekt nicht signifikant ist. Unter der Annahme, dass ein Residualeffekt ausgeschlossen werden kann, kann für die Behandlungen mit Formoterol und Salbutamol kein Unterschied nachgewiesen werden.

# Kapitel 5

## Der Nacheffekt

### 5.1 Allgemeines und Historisches

Ist ein Residualeffekt vorhanden, also ein Nacheffekt der Behandlung aus der ersten Periode, stellt sich das Problem, dass im Cross-Over-Plan unter Berücksichtigung der Informationen aus *beiden* Perioden kein erwartungstreuer Schätzer für den Behandlungseffekt vorliegt. Denn bei Annahme des linearen Modells nach LEHMACHER (1987) ist unter diesen Umständen der in Abschnitt 4.1.1.3 auf Seite 23 genannte Schätzer verzerrt um den Term  $-\hat{d}_\lambda/2$ . Ein erwartungstreuer Schätzer für den Behandlungsunterschied kann dann lediglich aus den Daten der ersten Periode gewonnen werden. Ebenso verhält es sich im nichtparametrischen Modell. Der Schätzer für den Behandlungseffekt  $p^\delta$  ist verzerrt um den Term  $-\hat{p}^\lambda/2$ , und daher steht ansonsten nur der Schätzer  $\hat{p}^\Phi = (\hat{p}_{11} - \hat{p}_{12})/2$  zur Verfügung. Dieser basiert lediglich auf den Daten der ersten Periode.

Das bedeutet, dass man nicht, wie z.B. beim klassischen Zwei-Stichproben-Test, unverzerrte Schätzer und damit auch Tests für den (reinen) Behandlungseffekt konstruieren kann. Diesen erhält man im COP lediglich dann, wenn man sich in einem solchen Fall auf die Daten der ersten Periode beschränkt oder die Annahme trifft, dass kein Überhangseffekt vorhanden ist. Im ersten Fall wird lediglich die Hälfte der erhobenen Daten für die Auswertung benutzt, und die Annahme, dass kein Überhangseffekt gegeben ist, ist in der Praxis häufig nicht zu begründen.

Viele Autoren, z.B. COX (1958), GRIZZLE (1965), HILLS UND ARMITAGE (1979) sowie BROWN (1980) diskutierten bereits das Cross-Over-Design und die Problematik mit dem Nacheffekt. Sie schlossen aus dem Vorhandensein des Nacheffekts, dass das Cross-Over-Design dann nicht anwendbar ist und stattdessen nur mit Hilfe von klassischen Zwei-Stichproben-Tests die Daten der ersten Periode ausgewertet werden könnten. WILLAN UND PATER (1986a) benutzten jenes Modell von Grizzle und Brown, um eine Alternative im Umgang mit dem Nacheffekt aufzuzeigen. Sie diskutierten das Verfahren von GRIZZLE (1965) und schlossen, dass im Falle eines Residualeffekts nicht gleich auf das Cross-Over-Verfahren verzichtet werden müsse. Basierend auf Untersuchung der Power der beiden gegenübergestellten Tests und des Standardfehlers für die zugrundegelegten Schätzer für einen Behandlungseffekt wurde abgeleitet, dass in der Praxis nur selten der Fall eintrete, in dem der Zwei-Stichproben-Test dem Cross-Over-Verfahren nach GRIZZLE (1965) vorzuziehen sei. Die möglicherweise entstehende Verzerrung sei in vielen Fällen durch die Reduktion der Varianz des Schätzers mehr als kompensiert.

LEHMACHER (1987) greift die Schwierigkeit des Vorhandenseins eines Residualeffekts auf: was tun, wenn ein Residualeffekt vorliegt? Dann sei nämlich die Beurteilung des Behandlungseffekts schwierig. Man müsse sich jedoch nicht gleich auf die Daten der ersten Periode beschränken, sondern könne auch die Interpretation modifizieren. Beschränkt man sich auf die Daten der ersten Periode, so erhalte man eine wesentlich höhere Varianz, da die intraindividuelle Variabilität nicht ausgeschaltet werde. Er konstatiert die Unterscheidung von positiven und negativen Residualeffekten aufgrund der Beziehung  $\delta = 2\Phi - \lambda$ . Aus einem positiven Residualeffekt folgere man die Verkleinerung des allgemeinen Behandlungsunterschieds in der zweiten Periode (Überhangseffekt) und entsprechend aus einem negativen Residualeffekt die Vergrößerung des allgemeinen Behandlungsunterschieds in der zweiten Periode (Entzugseffekt). Im ersten Fall unterschätze man den wahren Behandlungseffekt, im zweiten Fall überschätze man ihn. Diese Resultate wurden in Abschnitt 3.4 auch grafisch veranschaulicht. LEHMACHER (1987) unterscheidet weiterhin drei mögliche Vorgehensweisen: Test bei a priori fehlendem, vorhandenem und evtl. existierendem Residualunterschied.

1. Beim a priori *fehlenden Residualunterschied* werde der Zwei-Stichproben-*t*-Test auf die Behandlungsunterschiede angewandt oder entsprechend ohne Normalverteilungsannahme der Wilcoxon-Mann-Whitney-Test. Siehe dazu Abschnitt 4.1.2.3 auf Seite 26.
2. Bei a priori *vorhandenem Residualunterschied* müsse man auf die Daten der ersten Periode zurückgreifen oder einen verallgemeinerten Behandlungseffekt testen: Der Test auf einen Cross-Over-Effekt entsteht (Abschnitt 4.1.2.4). Die Hypothese  $H_0^\delta : \delta = 0$  bedeutet eine Äquivalenz der Behandlung in dem Sinne, dass die Erwartungswertverläufe pro Behandlung äquivalent sind. Die Alternative besagt dann, dass die Behandlungen nicht äquivalent sind, d.h. die Erwartungswertverläufe sind nicht parallel. Es sei jedoch noch einmal betont, dass in diesem Fall auf einen (verallgemeinerten) Behandlungseffekt getestet wird. Das bedeutet, dass der (reine) Behandlungseffekt und der Residualeffekt miteinander vermengt sind.
3. Bei *eventuell vorhandenem Residualeffekt* kann man neben dem Test auf einen (verallgemeinerten) Behandlungseffekt unter Umständen auch noch auf eine Tendenz der Behandlungen schließen. Die folgenden Überlegungen beziehen sich wiederum auf die Beziehung  $\delta = 2\Phi - \lambda$ . Ist ein negativer Residualeffekt und ein negativer Cross-Over-Effekt vorgegeben, so erhält man einen negativen Behandlungseffekt, d.h.  $\Phi_1 < \Phi_2$ . Entsprechend ergibt sich bei einem positiven Residualunterschied und positivem Cross-Over ein positiver Behandlungsunterschied  $\Phi_1 > \Phi_2$ . Einen solchen gerichteten Residualeffekt kann man a priori unter Umständen aus pharmakokinetischen Gründen annehmen, wenn beispielsweise bekannt ist, dass die Einnahme eines Medikaments die Wirkung eines anderen Medikaments hemmt oder verstärkt.

Zusätzlich zu den aus der Literatur betrachteten Vorgehensweisen sei hier noch eine Interpretationsmöglichkeit für das nichtparametrische Modell angegeben. Die Argumentation aus dem linearen Modell bei eventuell vorhandenem Residualeffekt aus der 3. Vorgehensweise lässt sich kanonisch übertragen. Aus der Darstellung  $p^\delta = 2p^\Phi - p^\lambda$  kann aufgrund



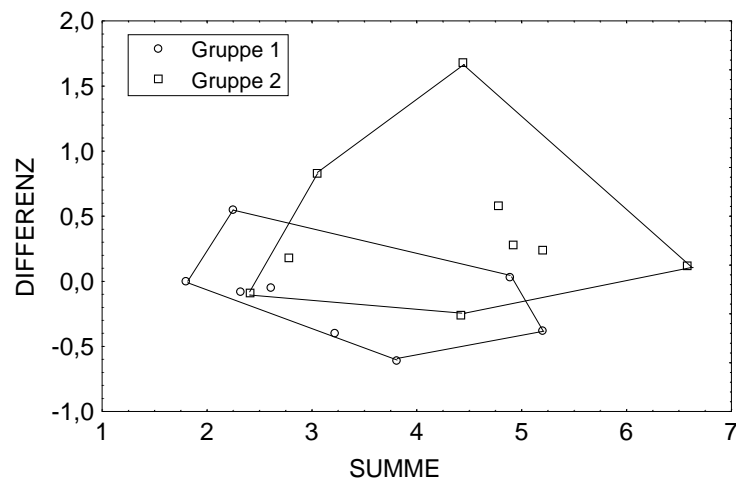


Abbildung 5.1: Differenzen - Summen Plot über die Messwerte für die Asthma-Studie II

der Monotonie der relativen Effekte folgender Schluss gezogen werden: Liegt ein positiver relativer Residualeffekt und ein positiver relativer Cross-Over-Effekt vor, so kann man auf einen positiven relativen Behandlungseffekt schließen. Das bedeutet per definitionem  $p_{11} > p_{21}$ , d.h. die Werte der ersten Behandlungsgruppe zum ersten Zeitpunkt sind tendenziell größer als die Werte der zweiten Behandlungsgruppe zum ersten Zeitpunkt. Man kann aber mit dieser Argumentation keine Signifikanzen nachweisen. Entsprechend verläuft die Schlussfolgerung bei negativem relativen Residualeffekt und negativem relativem Cross-Over-Effekt. Es wird unter diesen Umständen auf einen positiven relativen Behandlungseffekt geschlossen, was die folgende Aussage veranschaulicht: Die Werte der ersten Behandlungsgruppe zum ersten Zeitpunkt sind tendenziell kleiner als die Werte der zweiten Behandlungsgruppe zu demselben Zeitpunkt. Bei einer anderen Vorzeichenkombination von relativen Residual- und Cross-Over-Effekten kann nichts ausgesagt werden. Es stellt sich an dieser Stelle die Frage, ob man mit Hilfe von Tests oder anderen Verfahren bei möglicherweise vorhandenem Residualeffekt Aussagen über einen (reinen) Behandlungseffekt treffen kann.

Zunächst wird in Abschnitt 5.2 ein Vorschlag zur grafischen Bearbeitung des Problems gemacht, d.h. ohne statistische Tests wird eine Vorgehensweise erläutert mit der eventuelle Residual- bzw. Behandlungseffekte sichtbar gemacht werden. Anschließend wird die klassische Two-Stage-Analyse von GRIZZLE (1965) in Abschnitt 5.3 beschrieben und diskutiert. Nachdem zusätzlich ein neues, aber stark konservatives Verfahren für den Test auf einen Behandlungseffekt mit Hilfe von Konfidenzintervallen (siehe Abschnitt 5.4) vorgestellt wird, wird abschließend der Einsatz von Äquivalenztests (siehe Abschnitt 5.5) auch im Zusammenhang mit einer modifizierten Two-Stage-Analyse besprochen.

## 5.2 Grafische Bearbeitung des Problems

Für die folgende grafische Interpretation wird das lineare Modell (3.1) vorausgesetzt. Für jedes Individuum werden die Periodenunterschiede, d.h. die Differenzen pro Individuum von Zeitpunkt 1 zu Zeitpunkt 2 (y-Achse), gegen die Individuumssummen (x-Achse) ge-

geneinander aufgetragen. Je größer die Differenz beider Gruppen bezüglich der beiden Individuumssummen ist, desto eher kann auf einen Residualeffekt geschlossen werden. Wenn kein Residualeffekt vorhanden ist, so gilt: Je größer der Unterschied zwischen den Individuumsdifferenzen der beiden Gruppen, desto eher kann auf einen Behandlungseffekt geschlossen werden. Bei vorhandenem Residualeffekt wird entsprechend mit diesem Vorgang versucht, die Entscheidung über einen Cross-Over-Effekt zu treffen.

In Grafik 5.1 sind für die Asthma-Studie II die Summen und Differenzen der Messwerte pro Individuum aufgetragen. Wäre ein deutlicher Unterschied der beiden Gruppen auf der  $x$ -Achse zu beobachten, so deutet dies darauf hin, dass ein Residualeffekt vorhanden ist. In diesem Fall ist eine leichte Verschiebung auf der  $x$ -Achse zu beobachten, die jedoch in erster Linie auf *einen* großen Wert zurückzuführen ist. Daher würde man sich hier wohl nicht für einen Residualeffekt entscheiden. Ferner ist der Unterschied der Summen im Vergleich zum Unterschied in den Differenzen klein. Bezüglich der  $y$ -Achse ist ein deutlicher Unterschied sichtbar, weshalb man sich für einen Unterschied in den Behandlungen aussprechen könnte. Man beachte die Äquivalenz dieser Ergebnisse zu den Testergebnissen in Abschnitt 4.3.1 auf Seite 43.

Mit Hilfe dieser Vorgehensweise hätte jedoch die Asthma-Studie I nicht behandelt werden können. Das liegt daran, dass Summen und Differenzen dieser ordinalen Daten nicht sinnvoll gebildet werden können, und daher die Grundlage dieser Analyse nicht gegeben ist. Jedoch kann man diese Art der Analyse genauso auf die nichtparametrischen Effekte übertragen. Dazu erst einmal ein paar Sätze zur grundsätzlichen Vorgehensweise. Warum betrachtet man bei der Untersuchung auf einen Residualeffekt die Individuumssummen bzw. bei der Untersuchung auf einen Cross-Over-Effekt die Periodenunterschiede? Im Abschnitt 4.2.3 wird gezeigt, dass der Schätzer für den Residualunterschied aus der Differenz des Mittelwertes der Individuumssummen pro Gruppe zu berechnen ist. Entsprechendes gilt für den Cross-Over-Effekt. Da bei der Bildung der Teststatistik diese Schätzer im Zähler stehen, wird die Teststatistik (bei gleichbleibender Varianz) größer, wenn die jeweiligen Differenzen zwischen den beiden Gruppen größer werden. Und je größer die Teststatistik, desto eher kann auf einen entsprechenden Effekt geschlossen werden. Und genau diese Vorgehensweise wird beim zuvor erläuterten grafischen Verfahren angewendet. Um das grafische Verfahren nun auf das nichtparametrische Modell zu übertragen, werden die Pendanten zu den Individuumssummen und Periodenunterschieden benötigt. Die Effekte werden über den Kontrast  $c\hat{p}$  gebildet, d.h. für den Residualeffekt äquivalent mit  $\hat{p}_1 - \hat{p}_2$  und für den Cross-Over-Effekt mit  $\hat{p}_{11} + \hat{p}_{22} - \hat{p}_{12} - \hat{p}_{21}$ . Bezeichne des weiteren  $\hat{Y}_{iks}$  die AHRT als Schätzer von  $Y_{iks}$  aus Abschnitt 4.2.5 von  $X_{iks}$ , und sei  $\hat{Y}_{ik.} = \hat{Y}_{ik1} + \hat{Y}_{ik2}$ . Der Term  $\hat{Y}_{ik.}$  ist das Analogon zu den Individuumssummen  $s_{ik}$ , denn:

$$\begin{aligned} \hat{p}_i &= \hat{p}_{i1} + \hat{p}_{i2} \\ &= \frac{1}{n_i} \sum_{k=1}^{n_i} \left( \hat{G}(X_{ik1}) + \hat{G}(X_{ik2}) \right) \\ &= \frac{1}{n_i} \sum_{k=1}^{n_i} \left( \hat{Y}_{ik1} + \hat{Y}_{ik2} \right) \\ &= \frac{1}{n_i} \sum_{k=1}^{n_i} \hat{Y}_{ik.} \end{aligned}$$

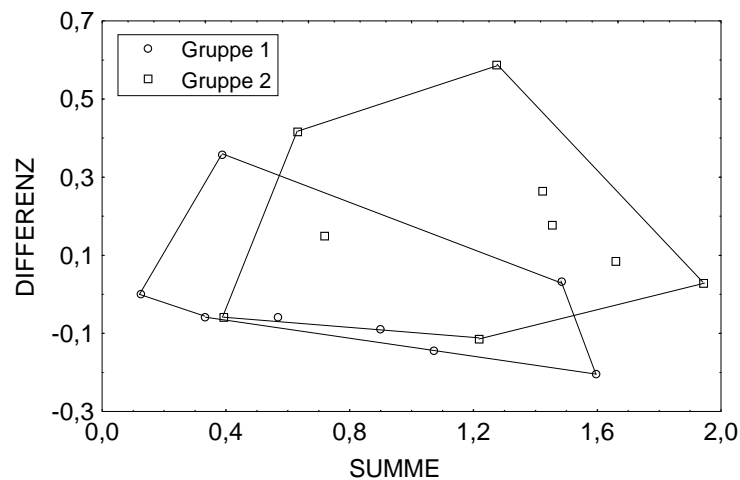


Abbildung 5.2: Differenzen - Summen Plot über die Rangtransformationen für die Asthma-Studie II

und mit der Differenz pro Gruppe der gemittelten  $\hat{Y}_{ik.}$  wird die Teststatistik gebildet. Mit den gleichen Überlegungen erhält man das Analogon zu den Periodenunterschieden  $p_{ik}$ , um den Cross-Over-Effekt zu testen:  $\hat{Y}_{ik.}^{DIFF} = \hat{Y}_{ik1} - \hat{Y}_{ik2}$ . Die weitere Analyse verläuft wie beim linearen Modell. Auf der  $y$ -Achse werden die  $\hat{Y}_{ik.}^{DIFF}$ 's aufgetragen und auf der  $x$ -Achse die Summe der AHRT's über beide Perioden  $\hat{Y}_{ik.}$ . Es ergibt sich die Grafik 5.2. Diese Darstellung ist für ordinale Daten geeignet, da die geplotteten Werte invariant unter monotonen Transformationen sind, also lediglich die Anordnung der Daten in dieser Grafik berücksichtigt wird.

Die Interpretation verläuft dann ebenfalls analog. Der Unterschied der beiden Gruppen bzgl. der Rangtransformationssummen ist zwar sichtbar, jedoch nicht gravierend und ähnelt stark den Relationen aus Grafik 5.1. Daher wird diese leichte Verschiebung eher nicht als unterschiedlicher Nacheffekt in beiden Gruppen interpretiert werden. Bei den Differenzen ist der Unterschied der beiden Gruppen schon sehr viel deutlicher. Jedoch ist dieser Effekt nicht so ausgeprägt wie in Grafik 5.1. Und genau das war ja auch in der Auswertung des Beispiels in Abschnitt 4.3.1 der Fall. Bei Annahme des linearen Modells mit Normalverteilung ergab sich ein Cross-Over-Effekt, während bei Anwendung des nichtparametrischen Modells nicht auf einen Nacheffekt geschlossen werden konnte.

Bei diesen Beobachtungen geht es nicht nur um die Illustration von Tatsachen, die man auch mit einem Test aufdecken kann, sondern in erster Linie um eine Übersicht über das Datenmaterial. Nach KENWARD UND JONES (1989) sollte diese Ergänzung zur rechnerischen Analyse der Daten routinemäßig bei der Auswertung von  $2 \times 2$ -Cross-Over-Designs verwendet werden. Ferner können diese Überlegungen unter Umständen den Test auf einen Residualeffekt ersetzen, ähnlich wie es in der Praxis häufig bei der Normalverteilung geschieht. Es wird in der Regel kein Test darauf gemacht, ob Daten normalverteilt sind oder nicht, sondern man betrachtet die Daten deskriptiv, z.B. in Form eines Histogramms oder Box-Plots und entscheidet sich dann für einen entsprechenden Test mit oder ohne Normalverteilungsannahme.

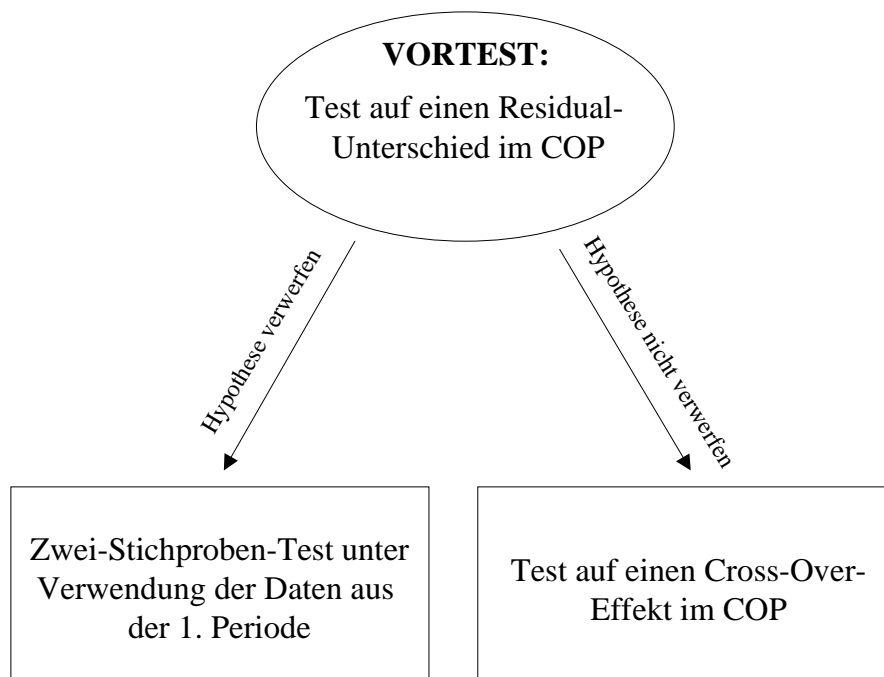


Abbildung 5.3: GRIZZLE'S Vortestmethode

### 5.3 Die Two-Stage-Analyse

Bei dieser zweistufigen Analyseform nach GRIZZLE (1965) wird zunächst ein Vortest auf einen Residualunterschied auf einem relativ hohen Niveau (10 – 20%) durchgeführt, und aufgrund dieses Ergebnisses wird weiter vorgegangen. Deshalb wird diese Methode auch *Vortest*-Methode genannt. Bei Ablehnung der Hypothese, also vorliegendem Residualeffekt, wird ein Test auf Behandlungsunterschied mit Daten aus der ersten Periode vorgeschlagen. Hingegen verweist GRIZZLE (1965) bei Nichtsignifikanz auf den Test auf Behandlungsunterschied. Diese Vorgehensweise ist in Grafik 5.3 veranschaulicht. Diese Analyseform wurde bewusst allgemein beschrieben, denn obwohl die Tests ursprünglich nur für das lineare Modell unter zusätzlicher Berücksichtigung der Normalverteilung der Daten gedacht waren, kann dieser Ansatz auch auf das nichtparametrische Modell übertragen werden. Der Vortest ist dann der Test auf einen Residualeffekt im nichtparametrischen Modell im COP. Bei Ablehnung der Hypothese wird in diesem Falle entweder nur auf einen Cross-Over-Effekt getestet oder wie im linearen Modell auf die Daten der zweiten Periode verzichtet und ein Wilcoxon-Mann-Whitney-Test auf die Daten der ersten Periode durchgeführt. Bei Nichtablehnung der Hypothese wird davon ausgegangen, dass kein Residualeffekt vorliegt und man wendet den Test auf einen Cross-Over-Effekt an, was in diesem Falle ein Test auf einen (reinen) Behandlungsunterschied ist.

Diese Analyseform birgt wesentliche Interpretationsfehler. Mit einer relativ hohen Wahrscheinlichkeit von 10 – 20% verwirft man die Nullhypothese des Vortests, obwohl sie richtig ist. Auf den Test auf einen Residualeffekt bezogen bedeutet das folgendes: Mit der gewählten Irrtumswahrscheinlichkeit entscheidet man sich für einen Residualeffekt,

obwohl keiner vorhanden ist, d.h. es wird dann fälschlicherweise von einem Residual-effekt ausgegangen. Dies hat zur Folge, dass nur noch die Hälfte der Daten, nämlich die Daten der ersten Periode, für die weitere Analyse benutzt werden können. Ferner geht man bei Nichtablehnung der Hypothese von deren Gültigkeit aus, was ebenfalls eine falsche Schlussfolgerung ist. Hierbei hat man sogar keine Kontrolle über die Fehlerwahrscheinlichkeit. Zur vertieften Problematik dieser Vorgehensweise siehe z.B. LEHMACHER (1987). Weitere Kritik am Verfahren übte FREEMAN (1989). Er geht aufgrund seiner Untersuchungen sogar soweit, zu sagen, dass die Two-Stage-Analyse irreführend und nicht zufriedenstellend und daher lediglich von historischem Interesse sei. Als erster machte er darauf aufmerksam, dass der Test auf Residualeffekt mit dem Test auf Behandlungseffekt mit Hilfe der Daten aus der ersten Periode hoch korreliert sind. Man überschreite bei dieser Vorgehensweise demnach das (Gesamt-)Niveau, wenn kein Residualeffekt gefunden wurde. Ebenso werde das Niveau bei vorhandenem Residualeffekt überschritten, insbesondere, wenn Baseline-Werte ignoriert werden. SENN (1993) fasste die Konsequenz von diesen Beobachtungen zusammen: Wenn der Überhangseffekt signifikant ist, dann ist der Parallelgruppenvergleich verzerrt. Wenn kein Überhangseffekt nachgewiesen werden kann, kann man nicht davon ausgehen, dass keiner vorhanden ist. Bei beiden Möglichkeiten erhalte man nicht zufriedenstellende Ergebnisse, daher sollte auf diese Two-Stage-Analyse verzichtet werden.

## 5.4 Konfidenzintervallverfahren

Nach BRUNNER UND NEUMANN (1987) kann mit Hilfe der sogenannten Konfidenzintervallmethode ein konservatives Verfahren zur Untersuchung auf einen Behandlungseffekt vorgestellt werden. Der Vorteil dieser Vorgehensweise liegt in erster Linie darin, dass man sich nicht im Vorfeld für oder gegen einen Residualeffekt entscheiden muss. Der Nachteil ist der, dass durch notwendige Abschätzungen dieses Verfahren von vornherein stark konservativ ist.

Aus der Darstellung  $p^\delta = 2p^\Phi - p^\lambda$  kann neben der bloßen Interpretation (siehe Abschnitt 5.1 auf Seite 47) aber auch ein konservatives Verfahren für den Test auf einen Behandlungseffekt hergeleitet werden. Es sei  $[p_U^\delta, p_O^\delta]$  das asymptotische  $1 - \alpha_\delta$ -Konfidenzintervall für den Cross-Over-Effekt mit

$$P(p_U^\delta \leq p^\delta \leq p_O^\delta) \geq 1 - \alpha_\delta$$

und entsprechend  $[p_U^\lambda, p_O^\lambda]$  das asymptotische  $1 - \alpha_\lambda$ -Konfidenzintervall für den Residualeffekt mit

$$P(p_U^\lambda \leq p^\lambda \leq p_O^\lambda) \geq 1 - \alpha_\lambda.$$

Bezeichne  $M_\delta$  die Menge der Versuchsergebnisse  $\omega$ , die zu einem Intervall führen, dass  $p^\delta$  enthält. Also

$$M_\delta := \{\omega | p_U^\delta(\omega) \leq p^\delta \leq p_O^\delta(\omega)\},$$

$$M_\lambda := \{\omega | p_U^\lambda(\omega) \leq p^\lambda \leq p_O^\lambda(\omega)\} \text{ und}$$

$$M_{2\Phi} := \{\omega | p_U^\lambda(\omega) + p_U^\delta(\omega) \leq p^\lambda + p^\delta \leq p_O^\lambda(\omega) + p_O^\delta(\omega)\}.$$

Mit diesen Bezeichnungen erhält man folgende Schlussfolgerungen:

$$M_\delta \cap M_\lambda \subseteq M_{2\Phi} \Rightarrow P(M_{2\Phi}) \geq P(M_\delta \cap M_\lambda).$$

Aus

$$P(M_\delta \cup M_\lambda) = P(M_\delta) + P(M_\lambda) - P(M_\delta \cap M_\lambda)$$

und  $P(M_\delta \cup M_\lambda) \leq 1$  folgt weiter

$$\begin{aligned} P(M_{2\Phi}) &\geq P(M_\delta) + P(M_\lambda) - P(M_\delta \cup M_\lambda) \\ &\geq P(M_\delta) + P(M_\lambda) - 1. \end{aligned}$$

Diese außerordentlich grobe Abschätzung, die an dieser Stelle benutzt wird, ist in erster Linie verantwortlich für die Konservativität des Verfahrens. Da weiterhin definitionsgemäß gilt

$$P(M_\delta) \geq 1 - \alpha_\delta \quad \text{und} \quad P(M_\lambda) \geq 1 - \alpha_\lambda,$$

erhält man die Abschätzung

$$P(M_{2\Phi}) \geq 1 - \alpha_\delta - \alpha_\lambda.$$

Da  $p^\delta + p^\lambda = 2p^\Phi$  ist, folgt

$$P(p_U^\Phi \leq p^\Phi \leq p_O^\Phi) \geq 1 - \alpha_\delta - \alpha_\lambda$$

mit  $p_U^\Phi := \frac{1}{2}(p_U^\delta + p_U^\lambda)$  und entsprechender Definition für  $p_O^\Phi$ .

Aus diesen Konklusionen erhält man einen (konservativen) Test für  $H_0^\Phi : p^\Phi = 0$ , indem man verwirft, wenn  $0 \notin [p_U^\Phi, p_O^\Phi]$ .

### Berechnung der Konfidenzintervallgrenzen

Die asymptotischen Konfidenzintervalle erhält man mit Hilfe von Satz 4.2.8 auf Seite 38 und dem für  $W_N$  konsistenten Schätzer  $\widehat{W}_N$  durch

$$u_{\alpha/2} \leq \frac{\sqrt{N}c(\widehat{\mathbf{p}} - \mathbf{p})}{\sqrt{c\widehat{W}_N c^T}} \leq u_{1-\alpha/2}$$

und damit allgemein die Ungleichung:

$$\sqrt{N}c\widehat{\mathbf{p}} - u_{1-\alpha/2}\sqrt{c\widehat{W}_N c^T} \leq \sqrt{N}c\mathbf{p} \leq \sqrt{N}c\widehat{\mathbf{p}} + u_{1-\alpha/2}\sqrt{c\widehat{W}_N c^T} \quad \text{bzw.}$$

$$\frac{1}{2}c\widehat{\mathbf{p}} - u_{1-\alpha/2}\sqrt{c\widehat{W}_N c^T}/(2 \cdot \sqrt{N}) \leq \frac{1}{2}c\mathbf{p} \leq \frac{1}{2}c\widehat{\mathbf{p}} + u_{1-\alpha/2}\sqrt{c\widehat{W}_N c^T}/(2 \cdot \sqrt{N}).$$

Da  $p^\lambda = \frac{1}{2}c_\lambda \mathbf{p}$  und entsprechend  $p^\delta = \frac{1}{2}c_\delta \mathbf{p}$ , folgt weiterhin

$$p_{O;U}^\lambda = \frac{1}{8\tilde{n}} \left( \overline{\Psi}_{1.1} + \overline{\Psi}_{1.2} - \overline{\Psi}_{2.1} - \overline{\Psi}_{2.2} - \left( \frac{\tilde{n}}{n_1} - \frac{\tilde{n}}{n_2} \right) \right) \pm \frac{1}{2\sqrt{N}} \sqrt{c_\lambda \widehat{W}_N c_\lambda^T} u_{1-\alpha_\lambda/2}$$

und

$$p_{O;U}^{\delta} = \frac{1}{8\tilde{n}}(\bar{\Psi}_{1.1} - \bar{\Psi}_{1.2} - \bar{\Psi}_{2.1} + \bar{\Psi}_{2.2}) \pm \frac{1}{2\sqrt{N}} \sqrt{\mathbf{c}_{\delta} \widehat{\mathbf{W}}_N \mathbf{c}_{\delta}^T} u_{1-\alpha_{\delta}/2}.$$

Entsprechend ergibt sich dann für das gesuchte Konfidenzintervall von  $p^{\Phi}$

$$\begin{aligned} p_{O;U}^{\Phi} &= \frac{1}{2}(p_{O;U}^{\lambda} + p_{O;U}^{\delta}) \\ &= \frac{1}{16\tilde{n}} \left( 2\bar{\Psi}_{1.1} - 2\bar{\Psi}_{2.1} - \left( \frac{\tilde{n}}{n_1} - \frac{\tilde{n}}{n_2} \right) \right) \pm \frac{1}{4\sqrt{N}} \left( \sqrt{\mathbf{c}_{\lambda} \widehat{\mathbf{W}}_N \mathbf{c}_{\lambda}^T} u_{1-\alpha_{\lambda}/2} \right) \\ &\quad \pm \frac{1}{4\sqrt{N}} \left( \sqrt{\mathbf{c}_{\delta} \widehat{\mathbf{W}}_N \mathbf{c}_{\delta}^T} u_{1-\alpha_{\delta}/2} \right). \end{aligned}$$

## 5.5 Äquivalenztests

Häufig soll im Sinne des Konsumenten von Medikamenten das Risiko, dass ein Medikament auf den Markt kommt, obwohl es eigentlich in seiner Wirkung nicht äquivalent zum Referenzmedikament war, kontrollierbar sein. Diese Kontrollgröße entspricht dem Fehler 2. Art bei den üblichen Zwei-Stichproben-Tests. Dieser Fehler wird, wenn die Nullhypothese auf Gleichheit formuliert ist, nicht kontrolliert. Um dem Abhilfe zu schaffen, ist das Entscheidungsproblem bei Äquivalenztests folgendermaßen formuliert: Die Nullhypothese ist *keine* Äquivalenz gegen die Alternative der Äquivalenz zweier Medikamente. Bei Verwerfen der Hypothese hat man dementsprechend die äquivalente Wirkung zweier Medikamente nachgewiesen. Eine wesentliche Annahme in humanpharmakologischen Bioäquivalenzstudien ist laut RÖHMEL (1998), dass der Äquivalenznachweis die Gleichwertigkeit bzgl. Wirksamkeit und Sicherheit impliziert. Daher erklärt sich auch die große Bedeutung dieser Studien bei der Entwicklung von Generika. Ferner birgt diese Art von Studie eine Zeit- und Kostenersparnis dadurch, dass bei Arzneimittelzulassungen unter Bezugnahme auf wirkstoffidentische Referenzprodukte ohne aufwendige klinische Prüfung eine Entscheidung getroffen werden kann.

HAUSCHKE et al. (1990) entwickelten eine nichtparametrische Prozedur für die Analyse von Bioäquivalenzstudien im Cross-Over-Design. Wie WESTLAKE (1988) vernachlässigten sie dabei Überhangseffekte. Unter Annahme des linearen Modells wurde bei HAUSCHKE et al. (1990) die Voraussetzung auf die Gleichheit der Fehlerterme in beiden Perioden und die (logarithmische) Normalverteilung der Zielvariablen fallengelassen.

In diesem Abschnitt soll ein Verfahren für das nichtparametrische Modell (GM) vorgestellt werden. D.h., es wird keine Linearität der Modellparameter vorausgesetzt sowie keine Voraussetzung an die zugrunde liegenden Verteilungsfunktionen gestellt.

In diesem Abschnitt sollen Tests vorgestellt werden für die folgenden Alternativen: *kein* relevanter Residualeffekt, *kein* relevanter Perioden-Effekt und *kein* relevanter verallgemeinerter Behandlungseffekt, d.h. Cross-Over-Effekt. Wird die Nullhypothese beispielsweise beim Test auf einen relevanten Nacheffekt verworfen, so kann im Hinblick auf ein (leicht modifiziertes) Two-Stage-Verfahren, ein relevanter Residualeffekt ausgeschlossen werden. Kann die Nullhypothese hingegen nicht verworfen werden, kann nicht auf Gleichheit der Residualeffekte geschlossen werden und somit könnte in einer anschließenden Analyse auf den Behandlungseffekt der Nacheffekt nicht ausgeschlossen werden.

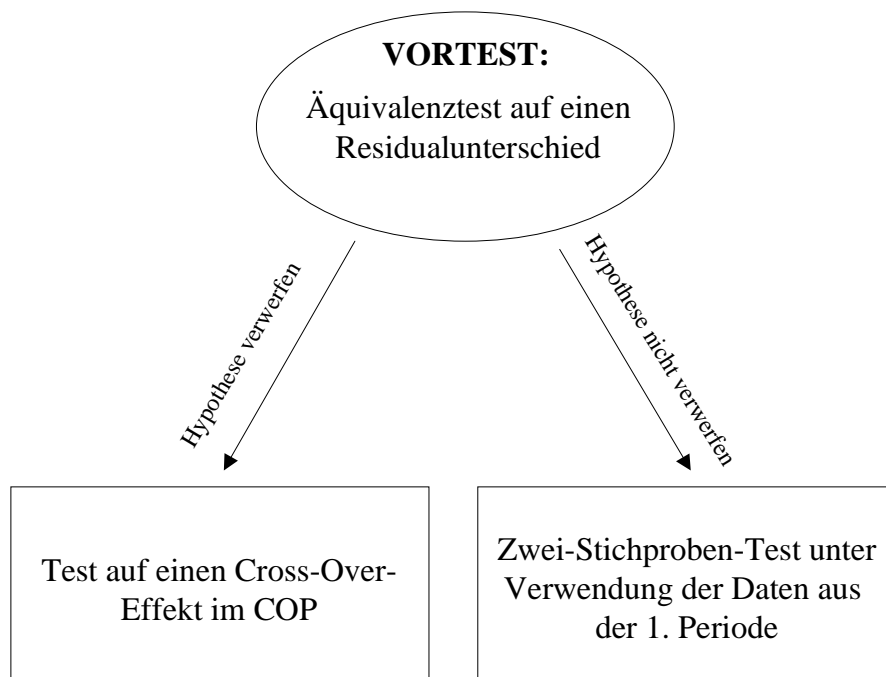


Abbildung 5.4: modifizierte Vortestmethode

Und genau dieses ermöglicht jetzt eine Alternative zur Two-Stage-Analyse von GRIZZLE (1965): die modifizierte Two-Stage-Analyse. Man kann dafür analog zu Grafik 5.3 in Abschnitt 5.3 Grafik 5.4 angeben.

Allgemein lässt sich für alle Testprobleme die Hypothesenstellung wie folgt formulieren. Dabei sei  $\theta_1 < 0$  und  $\theta_2 > 0$  fest und beliebig.

$$H_0^\theta : \mathbf{c}\mathbf{p} \leq \theta_1 \vee \mathbf{c}\mathbf{p} \geq \theta_2 \quad \text{vs.} \quad H_1^\theta : \theta_1 < \mathbf{c}\mathbf{p} < \theta_2$$

mit den jeweiligen Kontrastvektoren  $\mathbf{c}_\lambda$ ,  $\mathbf{c}_\pi$  oder  $\mathbf{c}_\delta$ .

Den Bereich  $\theta_1 < \mathbf{c}\mathbf{p} < \theta_2$  bezeichnet man auch als Sicherheitsbereich oder Bereich des nicht relevanten Unterschieds. Hier zeigt sich, dass es streng genommen nicht um *keinen* Residual-, Perioden- oder Cross-Over-Effekt geht, sondern einen so geringen Effekt, der nach Meinung des Testers nicht ins Gewicht fällt.

Der nächste Schritt ist die Berechnung eines  $(1 - 2\alpha) \cdot 100\%$ -Konfidenzintervalls für  $\mathbf{c}\mathbf{p}$ . Aus Satz 4.2.8 folgt mit  $S_N^2 \geq c_0 > 0$

$$\sqrt{N}\mathbf{c}(\hat{\mathbf{p}} - \mathbf{p})/S_N \quad \dot{\sim} \quad N(0; 1),$$

wobei  $S_N^2 = \mathbf{c}\mathbf{W}_N\mathbf{c}^T$  und

$$U_n^p = \frac{\sqrt{N}\mathbf{c}(\hat{\mathbf{p}} - \mathbf{p})}{S_N}.$$



Dann folgen aus

$$\begin{aligned} 1 - 2\alpha &= P(u_\alpha \leq U_n^p \leq u_{1-\alpha}) \\ &= P\left(S_N u_\alpha / \sqrt{N} \leq \mathbf{c}\hat{\mathbf{p}} - \mathbf{c}\mathbf{p} \leq S_N u_{1-\alpha} / \sqrt{N}\right) \\ &= P\left(\mathbf{c}\hat{\mathbf{p}} - u_{1-\alpha} S_N / \sqrt{N} \leq \mathbf{c}\mathbf{p} \leq \mathbf{c}\hat{\mathbf{p}} + u_{1-\alpha} S_N / \sqrt{N}\right) \end{aligned}$$

die Konfidenzintervallgrenzen

$$\begin{aligned} KI(u; 1 - 2\alpha) &= \mathbf{c}\hat{\mathbf{p}} - u_{1-\alpha} S_N / \sqrt{N} \quad \text{und} \\ KI(o; 1 - 2\alpha) &= \mathbf{c}\hat{\mathbf{p}} + u_{1-\alpha} S_N / \sqrt{N}. \end{aligned}$$

Die Nullhypothese  $H_0^\theta$  wird abgelehnt, wenn das Konfidenzintervall ganz in dem Intervall  $(\theta_1, \theta_2)$  enthalten ist, also gilt:

$$\begin{aligned} KI(u; 1 - 2\alpha) &> \theta_1 \quad \text{und} \\ KI(o; 1 - 2\alpha) &< \theta_2. \end{aligned}$$

Dieses Konfidenzintervall ließe sich auch auffassen als Schnitt zweier einseitiger  $(1 - \alpha)$ -Konfidenzintervalle für  $\mathbf{c}\mathbf{p}$ . Man kann dieses Testproblem demnach auch als zwei einseitige Testprobleme auffassen:

$$\begin{aligned} H_{01}^\theta : \mathbf{c}\mathbf{p} \leq \theta_1 \quad \text{vs.} \quad H_{11}^\theta : \mathbf{c}\mathbf{p} \geq \theta_1 \\ H_{02}^\theta : \mathbf{c}\mathbf{p} \geq \theta_2 \quad \text{vs.} \quad H_{12}^\theta : \mathbf{c}\mathbf{p} \leq \theta_2. \end{aligned}$$

In Anlehnung an BERGER (1982) gilt unter Zuhilfenahme des intersection-union-Prinzips: Werden die Nullhypothesen  $H_{01}^\theta$  und  $H_{02}^\theta$  auf dem Niveau  $\alpha$  verworfen, so ist das gleichbedeutend mit der Aussage, dass die Nullhypothese  $H_0^\theta = H_{01}^\theta \cup H_{02}^\theta$  auf dem Niveau  $\alpha$  verworfen wird, d.h.  $H_1^\theta = H_{11}^\theta \cap H_{12}^\theta$  angenommen wird.

Die Problematik bei dieser Vorgehensweise liegt insbesondere in der Wahl der Abweichungsparameter  $\theta_1$  und  $\theta_2$ . Dazu wurde Näheres in den Veröffentlichungen MUNZEL UND HAUSCHKE (2001) bzw. WELLEK UND HAMPEL (1999) gesagt.

Im Sinne von Bioäquivalenzstudien kann die zweite Stufe dieser modifizierten Two-Stage-Analyse entsprechend verändert werden, um einen *Bioäquivalenznachweis* für zwei Behandlungen zu erbringen. Dazu zunächst ein paar Erläuterungen. Damit ein Medikament wirksam ist, muss eine minimale Plasmakonzentration im Blut vorhanden sein. Bei zu hoher Plasmakonzentration wiederum können unerwünschte Nebenwirkungen die Folge sein. Für die Bioäquivalenz entscheidend sind die Gleichwertigkeit von Ausmaß und Geschwindigkeit der Resorption. Aus der Plasmaspiegelkurve lassen sich die Größen *AUC* (=area under the curve), die maximale Konzentration  $C_{MAX}$  oder der Zeitpunkt der maximalen Konzentration  $t_{MAX}(C_{MAX})$  gewinnen. Die Zielvariable in Bioäquivalenzstudien ist für gewöhnlich *AUC*. Damit werden die Konzentration-Zeit-Kurven verglichen als Indikator für die Wirksamkeit der Behandlung. Alternativ können auch  $C_{MAX}$  oder  $t_{MAX}(C_{MAX})$  gewählt werden. Wie andere Behandlungsvergleiche auch, können ebenso Bioäquivalenz-Studien als Cross-Over-Design angelegt werden mit Übernahme aller Eigenschaften desselben. Man muss lediglich die zweite Stufe der Analyse ebenfalls mit Äquivalenzttests durchführen.



# Kapitel 6

## Baseline-Werte

### 6.1 Einführung und Motivation

Im Cross-Over-Design bezeichnet man als Baseline-Werte Größen, die zusätzlich zu den verschiedenen Messwerten (unter Behandlung) erhoben werden (z.B. nach SENN 1993). Es werden an den Versuchseinheiten der Studie auch dann Werte erhoben, wenn keine aktuelle Behandlung einwirkt. Da es sich um Werte nicht unter Einfluss einer Behandlung handelt, bieten Baseline-Werte in erster Linie zusätzliche Hintergrundinformationen, aber es kann keine direkte Erkenntnis über die an den Patienten vorgenommene Behandlung gewonnen werden.

Man kann drei verschiedene Formen von Baseline-Werten im *AB/BA*-Design unterscheiden:

- Werte, die vor Beginn der ersten Behandlung erhoben werden,
- Werte, die vor der zweiten Behandlungen erhoben werden und
- Werte, die nach Abschluss der zweiten Behandlung festgestellt wurden.

Die ersten beiden Baseline-Werte werden auch als *Vorwerte* bezeichnet, da sie *vor* den Behandlungen erhoben werden. Sei im Folgenden eine Erläuterung der soeben genannten Baseline-Werte gegeben.

Nach KENWARD UND JONES (1989) sei lediglich der zuerst genannte Baseline-Wert, der sogenannte Run-In-Wert, ein echter Baseline-Wert, denn nur hier bestehe mit Sicherheit noch kein Einfluss einer Behandlung. Vorteile dieses Baseline-Wertes sind, dass man die zusätzliche Information gut interpretieren kann. Beispielsweise kann man sich durch einen Vergleich der jeweiligen Behandlungswerte mit den entsprechenden Vorwerten ein Bild von der Größenordnung der Behandlungswirkung machen. Es ist also eine Quantifizierung der individuellen Behandlungswirkung möglich; der Unterschied zu keiner Behandlung wird bei jedem Individuum sichtbar. Eine angenehme Begleiterscheinung bei Erhebung der Baseline-Werte ist die Gewöhnung von Proband und Untersucher an die Studie. Dadurch wird unter Umständen die Datenqualität der beiden Behandlungsperioden verbessert.

Der zusätzliche Baseline-Wert vor der Behandlung in der zweiten Periode am Ende der Wash-Out-Phase, kann zum Beispiel dazu dienen, nachzusehen, ob die Länge der Wash-Out-Phase angemessen ist. Dies geschieht, indem überprüft wird, ob die Verteilung der

beiden Baseline-Werte zu beiden Zeitpunkten (vor Run-In und vor Wash-Out) Unterschiede aufweisen. Oft wird daran erinnert, dass der zweite Baseline-Wert kein wahrer ist, da er bereits durch die Behandlung in der ersten Periode beeinflusst sein kann. Wenn das Verhältnis der Länge der Wash-Out-Phase sehr groß ist im Vergleich zum Verhältnis der Länge der Zeit zwischen der Erhebung des zweiten Baseline-Werts und des zweiten Messwerts, so könnte man davon ausgehen, dass auf diese beiden Werte der gleiche Nacheffekt der Behandlung aus der ersten Periode wirkt. Würde man vom zweiten Messwert den zweiten Baselinewert abziehen, so ließe sich der Residualeffekt herauskürzen. Ist die Wash-Out-Phase jedoch von vornherein aus z.B. ethischen Gründen sehr kurz im Vergleich zur Behandlungsphase, so kann auf den zweiten Baseline-Wert verzichtet werden. Die dritte und letzte Möglichkeit ist ein Baseline-Wert nach der zweiten Behandlungsperiode, auf den an dieser Stelle jedoch nicht weiter eingegangen wird. Diskussionen dieser dritten Art befinden sich unter anderem in RATKOWSKI et al. (1993) und SENN (1993). Um eine Entscheidung zu treffen, ob Baseline-Werte im Einzelfall nutzbringend sind oder nicht, muss in erster Linie die Verhältnismäßigkeit von Kosten bzw. Aufwand der Erhebung der zusätzlichen Werte und Nutzbarkeit der Baseline-Werte überprüft werden.

An die Baseline-Werte sind allgemeine Voraussetzungen geknüpft. Es ist zunächst einmal für das im Folgenden auf Baseline-Werte erweiterte lineare Modell unerlässlich, dass bei Erhebung der Baseline-Werte die *gleiche Messgröße* unter Verwendung der *gleichen Skala* wie zu den Messzeitpunkten verwendet wird. Das klingt wie eine Selbstverständlichkeit, kann jedoch unter Umständen schwer zu verwirklichen sein. Ein Beispiel für Schwierigkeiten mit der gleichen Skala ist folgendes: Wenn die Messungen von Baseline-Werten zu Hause und die von Messzeitpunkten in der Klinik stattfinden, muss unter Umständen auf unterschiedliche Geräte zurückgegriffen werden. Das liegt darin begründet, dass manche Geräte nicht für den privaten Gebrauch geeignet sind. Das Messen der Lungenfunktion mit Spirometer und Flowmeter ist ein Beispiel dafür, denn das Flowmeter misst die Milliliter pro Zeiteinheit und das Spirometer das Lungenvolumen. Die Messzeitpunkte dürfen zudem nicht zu weit auseinander liegen. Die Variation der Baseline-Werte darf sich ferner nicht wesentlich von der der Messwerte unterscheiden.

Im Folgenden werden zwei Ansätze zur Berücksichtigung von Baseline-Werten vorgestellt. Sowohl für das lineare Modell als auch das nichtparametrische Modell werden in erster Linie Verfahren zur Berücksichtigung von zwei Baseline-Werten vorgestellt. Es handelt sich dabei um Vorwerte, die jeweils vor den Behandlungen der ersten und zweiten Periode erhoben werden. Im folgenden Abschnitt 6.2 wird für diese Problemstellung ein Verfahren vorgestellt, dass die Messwerte unter Behandlung mit den entsprechenden Baseline-Werten adjustiert. Für das in dieser Arbeit betrachtete nichtparametrische Modell wird anschließend in Abschnitt 6.3 die Verwendung der nichtparametrischen Kovarianzanalyse vorgeschlagen, wie sie von SIEMER (1999) eingeführt wurde. Dabei werden die Baseline-Werte als spezielle Kovariablen angesehen.

## 6.2 Lineares Modell

In diesem Abschnitt werden ausgewählte Verfahren vorgestellt, bei denen es möglich ist, zusätzliche Baseline-Werte in die Analyse mit einfließen zu lassen. Zum Aufdecken eines Effekts kann dabei letztendlich auf die bereits in Abschnitt 4.1.2 auf Seite 24 vorgestell-

ten Tests zurückgegriffen werden. Dabei wird zunächst einmal insbesondere auf den Fall zweier Baseline-Werte jeweils vor den Behandlungen eingegangen und anschließend der Spezialfall eines einzelnen Baseline-Werts vor Beginn der ersten Behandlung betrachtet. Werden in einem Versuch zwei Baseline-Werte erhoben, einer vor der ersten Behandlung und einer vor der zweiten Behandlung, so sind das zwei Vorwerte für dieselbe Person, die sich über die Zeit ändern. Das bedeutet, dass für jeden Messzeitpunkt unter Behandlung ein *eigener* Baseline-Wert zur Verfügung steht. Ein Beispiel, wo dieser Fall nicht eintritt, wäre beispielsweise die Erhebung eines einzelnen Baseline-Wertes vor Beginn der ersten Behandlung, möglicherweise zur Gewöhnung des Patienten an die Studienbedingung. SENN (1993) stellt folgendes Verfahren zur Berücksichtigung von zwei Baseline-Werten vor.

Wie bereits erwähnt, besteht die Möglichkeit, die bereits hergeleiteten Tests aus Abschnitt 4.1.2 zu verwenden. Jedoch werden als Ausgangsmesswerte nicht die Messwerte der beiden Behandlungen in den Perioden benutzt, sondern jeweils die um den Baseline-Wert adjustierten Werte. Dafür wird an dieser Stelle die Erläuterung gebraucht. Das bedeutet für das hier zugrunde liegende Versuchsdesign aus Tabelle 6.1 folgendes:

Tabelle 6.1: Versuchsaufbau mit zwei Baseline-Werten vor Behandlung

Gruppe		Zeitpunkt			
		Run-In	Periode 1	Wash-Out	Periode 2
$i = 1$	$k = 1$	$X_{111}^{(BL)}$	$X_{111}$	$X_{112}^{(BL)}$	$X_{112}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$k = n_1$	$X_{1n_1}^{(BL)}$	$X_{1n_11}$	$X_{1n_12}^{(BL)}$	$X_{1n_12}$
$i = 2$	$k = 1$	$X_{211}^{(BL)}$	$X_{211}$	$X_{212}^{(BL)}$	$X_{212}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$k = n_2$	$X_{2n_21}^{(BL)}$	$X_{2n_21}$	$X_{2n_22}^{(BL)}$	$X_{2n_22}$

Dieses Schema liegt beispielsweise der Asthma-Studie II zugrunde, die in Abschnitt 2.2 beschrieben wurde. Für die ursprünglichen Messwerte wird das bereits bekannte Modell (3.1) mit multivariater Normalverteilung nach LEHMACHER (1987) (siehe Seite 17) zugrunde gelegt:

$$X_{iks} = \mu + \Phi_{2-\delta_{is}} + \pi_s + \delta_{2s}\lambda_i + \epsilon_{iks}$$

wobei  $i = 1, 2$ ,  $k = 1, \dots, n_i$  und  $s = 1, 2$  und den Effekten wie an genannter Stelle angegeben. Dieses Modell kann für die Baseline-Werte zum erweiterten linearen Modell ergänzt werden. Für die zusätzlich erhobenen Baseline-Werte nimmt man allgemein die folgende Struktur an:

$$X_{iks}^{(BL)} = \nu + \theta_s + \delta_{2s}\tau_i + \zeta_{iks} \quad (6.1)$$

wobei  $i = 1, 2$ ,  $k = 1, \dots, n_i$  und  $s = 1, 2$  und  $(\zeta_{ik1}, \zeta_{ik2})$  multivariat normalverteilt mit Erwartungswertvektor  $\mathbf{0}$  und beliebiger (nichtsingulärer) Kovarianzmatrix. Dabei seien die Effekte

- $\nu$  : mittlerer Erwartungswert der Baseline-Werte
- $\theta_s$  : Effekt des  $s$ -ten Baseline-Wertes vor Periode  $s$
- $\tau_i$  : Residualeffekt der Behandlung in Gruppe  $i$  in der zweiten Periode am Ende der Wash-Out-Phase
- $\zeta_{iks}$  : unabhängiger Versuchsfehler innerhalb der Baseline-Werte.

Zwischen den Versuchsfehlern  $\epsilon_{iks}$  und  $\zeta_{iks}$  sind beliebige Abhängigkeiten zugelassen. Während sich alle bekannten Effekte in der Gleichung (6.1) wiederfinden, taucht an dieser Stelle der Behandlungseffekt  $\Phi$  aus dem linearen Modell ohne Baseline-Werte nicht auf. Das hat seine Begründung in der Tatsache, dass diese Baseline-Werte nicht unter Behandlung erhoben werden. Dabei ist zu beachten, dass der zweite Vorwert durchaus bei einer zu kurzen Wash-Out-Phase noch die Behandlung aus der ersten Periode reflektieren könnte. Dies wird durch die Hinzunahme eines möglichen Residualeffekts  $\tau_i$  berücksichtigt.

Bei dem hier beschriebenen Ansatz geht man davon aus, dass die interindividuelle Variabilität geringer wird, wenn man die Messwerte mit den Baseline-Werten adjustiert, da bei hoch korrelierten Werten für die Schätzer der Effekte eine geringere Varianz erwartet wird, wenn die folgenden Differenzen als Beobachtungen benutzt werden. Bezeichne

$$X_{iks}^{adj} := X_{iks} - X_{iks}^{(BL)} \quad \text{mit } i, s = 1, 2, k = 1, \dots, n_i$$

den um den entsprechenden Baseline-Wert adjustierten Messwert. Man erhält mit Hilfe dieser Definition sowie dem um Baseline-Werte erweiterten linearen Modell das sogenannte *adjustierte Modell*:

$$X_{iks}^{adj} = \tilde{\mu} + \Phi_{2-\delta_{is}} + \tilde{\pi}_s + \delta_{2s}\tilde{\lambda}_i + \tilde{\epsilon}_{iks}, \quad (6.2)$$

wobei  $(\tilde{\epsilon}_{ik1}, \tilde{\epsilon}_{ik2})$  multivariat normalverteilt ist mit  $E(\tilde{\epsilon}_{iks}) = 0$  für  $i, s = 1, 2$  und  $k = 1, \dots, n_i$  und

$$\begin{aligned} \tilde{\mu} &= \mu - \nu, \\ \tilde{\pi}_s &= \pi_s - \theta_s, \\ \tilde{\lambda}_i &= \lambda_i - \tau_i \quad \text{und} \\ \tilde{\epsilon}_{iks} &= \epsilon_{iks} - \zeta_{iks}. \end{aligned}$$

Dieses Modell hat genau die gleiche Struktur wie das von Lehmaner angegebene lineare Modell: Durch die Adjustierung hat man wieder zwei Versuchsgruppen mit jeweils zwei Zeitpunkten. Wie man soeben gesehen hat, bleiben die Additivität einzelner Effekte sowie auch die Normalverteilungsannahme von der Adjustierung unangetastet. Da die Voraussetzungen für die in Abschnitt 4.1 angegebenen Testverfahren erfüllt sind, können diese auch eingesetzt werden. Die Ergebnisse sind beispielsweise für den Behandlungseffekt (wenn kein Residualeffekt vorhanden) wie bisher zu interpretieren, da der Behandlungseffekt in das adjustierte Modell (6.2) unverändert einfließt. Jedoch sind die anderen Effekte adjustiert, und daher fragt man beim Residual-, Perioden- und Cross-Over-Effekt streng genommen nach dem Unterschied in den adjustierten Effekten, z.B.  $\tilde{\pi}_1 = \tilde{\pi}_2$  für den Periodeneffekt. Oftmals werden an dieser Stelle weitere Annahmen getroffen. Beispielsweise geht man häufig von der Gleichheit der Periodeneffekte bei zusammengehörigen Werten  $X_{iks}$  und  $X_{iks}^{(BL)}$  aus, so dass  $\tilde{\pi} = 0$ . Dies macht laut WILLAN UND PATER (1986b)

Sinn, wenn der zeitliche Unterschied zwischen Baseline- und passendem Messwert sehr gering im Vergleich zur Länge der Wash-Out-Phase ist. Entsprechend kann man für den Residualeffekt argumentieren.

Das adjustierte Modell ließe sich unter zusätzlichen Annahmen noch vereinfachen. Wenn beispielsweise  $\tilde{\pi}_s = 0$  und  $\tilde{\lambda}_i = 0$  für  $i, s = 1, 2$  wären, dann ließen sich Tests für den reinen Behandlungsunterschied herleiten, da sich dann das Modell zu

$$X_{iks}^{adj} = \tilde{\mu} + \Phi_{2-\delta_{is}} + \tilde{\epsilon}_{iks}$$

verkürzt. Diese Annahmen sind auch so realitätsfern nicht, denn  $\tilde{\pi}_1 = \tilde{\pi}_2 = 0$  bedeutet, dass der Zeiteffekt für den Messwert zum Zeitpunkt  $s$  und den entsprechenden Baseline-Wert gleich ist. Diese Voraussetzung kann gerechtfertigt sein, wenn der zeitliche Unterschied zwischen diesen beiden Messungen marginal ist. Analog verhält es sich mit der Annahme  $\tilde{\lambda}_1 = \tilde{\lambda}_2 = 0$ .

Je nach Verteilungs-, Modellannahme bzw. Korrelation zwischen Vorwert und Messwert fallen Effizienzvergleiche für Verfahren mit und ohne Baseline-Werte unterschiedlich aus. Der generelle Vorteil dieser Vorgehensweise gegenüber der Handhabung der Tests ohne Baseline-Werte ist, dass bei entsprechend hoher Korrelation zwischen den zusammengehörigen Baseline- und Messwerten eine Verringerung des Versuchsfehlers der entsprechenden Schätzer eintritt. Siehe dazu beispielsweise LEHMACHER (1987). Bei einer entsprechend geringen Zeitspanne zwischen diesen beiden Werten im Vergleich zur Länge der Wash-Out-Phase wird diese Tatsache trotzdem gemeinhin vorausgesetzt, selten jedoch überprüft. Demzufolge kann keine allgemeine Aussage über die sinnvolle Hinzunahme von Baseline-Werten getroffen werden, sondern es muss im Einzelfall erwogen werden, ob die zusätzliche Erhebung von Vorwerten einen Gewinn für das Verfahren bringt.

Selbst wenn mit Hilfe der Messwerte keine Effizienzsteigerung erlangt wird: Informationen über eine adäquate Wash-Out-Phase erhält man in jedem Fall. Selbst wenn jedoch der Einsatz von Baseline-Werten vorteilhaft ist, löst das nicht die Probleme, die sich durch einen eventuellen Nacheffekt ergeben.

Eine weitere und der Ansicht vieler Statistiker, wie beispielsweise SENN (1993), nach bessere Möglichkeit der Berücksichtigung von Vorwerten bietet die (parametrische) Kovarianzanalyse ANCOVA. Auf die Beschreibung derselben wird an dieser Stelle verzichtet und es sei auf die Literatur KIRK (1982) oder HUITEMA (1980) verwiesen.

Sei nun noch der Spezialfall betrachtet, dass nur ein Baseline-Wert vor der ersten Periode zur Verfügung steht. Dann ändert sich dieser Baseline-Wert nicht über die Zeit, was bedeutet, dass er Vorwert für den Messwert der ersten und der zweiten Periode ist. Wenn lediglich ein Baseline-Wert vor der ersten Behandlung erhoben wurde, erzielt man mit der Adjustierung beider Messwerte mit demselben Baseline-Wert gemäß

$$X_{iks}^{adj} = X_{iks} - X_{ik}^{(BL)} \quad \text{mit} \quad X_{ik}^{(BL)} = X_{ik1}^{(BL)} = X_{ik2}^{(BL)}$$

in zwei von drei Fällen überhaupt keinen Effekt, da sich die Baseline-Werte bei der Differenzenbildung  $c_1. - c_2.$  und  $p_1. - p_2.$  wieder herauskürzen. Lediglich der Test auf einen Residualeffekt ändert sich, da dort mit den Individuumssummen  $s_{ik}$  gearbeitet wird. Die Varianz des Fehlerterms kann sich laut LEHMACHER (1987) bei entsprechend niedriger Korrelation sogar erhöhen. Ebenfalls erfährt man an dieser Stelle, dass der Behandlungs- und Residualunterschied auch durch einen (zweidimensionalen) Vergleich der adjustierten Messwerte  $X_{ik1}^{adj} = X_{ik1} - X_{ik1}^{(BL)}$  mit den Messwerten der zweiten Periode  $X_{ik2}$

erreicht werden könnte, was bei entsprechend hoher Korrelation der Daten wiederum zu einem Effizienzgewinn führe. Durch die Schwierigkeit der Interpretation der entsprechenden Schätzer und somit der Testergebnisse aufgrund der ungleichen Ausgangswerte  $X_{ik1}^{adj}$  und  $X_{ik2}$ , soll an dieser Stelle von dieser Vorgehensweise abgeraten werden.

Wie bei zwei erhobenen Baseline-Werten muss die Qualität eines solchen Baseline-Werts in erster Linie auch in der bereits erläuterten zusätzlichen Hintergrundinformation gesehen werden. Dazu gehören die Überprüfung der Randomisation genauso wie vor allem auch die Abschätzung der Behandlungswirkung und als vermeintlich größten Pluspunkt: die Akklimatisierung der Probanden an die Studie.

Ferner gilt bei zwei wie auch bei einem Baseline-Wert: Bei ordinalen Daten, wie beispielsweise in der Asthma-Studie I, können keine sinnvollen Differenzen der Beobachtungen gebildet werden, da diese Differenz nicht sinnvoll zu interpretieren ist. Die adjustierten Messwerte  $X_{iks}^{adj}$  können dann demnach nicht gebildet werden. Einen Ausweg bietet die im nächsten Abschnitt beschriebene nichtparametrische Kovarianzanalyse. Die beiden Vorwerte werden bei dieser Analyseform allgemein als Kovariable aufgefasst und fließen als solche in die Analyse mit ein.

## 6.3 Nichtparametrisches Modell

Die Berücksichtigung von Baseline-Werten im nichtparametrischen Modell soll in dieser Arbeit mit Hilfe der **nichtparametrischen Kovarianzanalyse** geschehen. Dabei werden die Baseline-Werte als Kovariable aufgefasst. Das bringt neben den Vorteilen durch das sogleich in Abschnitt 6.3.1 eingeführte allgemeinere Modell im Vergleich zum linearen Modell auf Basis der adjustierten Messwerte auch den Vorteil, dass die Baseline-Werte durch Verwendung derselben als Kovariable nicht mehr auf der gleichen Skala gemessen werden müssen. Eigentlich, was für diese Arbeit aber nicht relevant ist, braucht es sich nicht einmal um die gleiche Messgröße handeln.

Zunächst einmal ein paar einführende Worte zur Kovarianzanalyse:

Mit einem Experiment soll überprüft werden, ob bestimmte Faktoren einen Einfluss auf die Zielvariable haben. Alle anderen Größen, die einen Einfluss haben könnten, sollten konstant gehalten werden. In der Praxis können selbstverständlich nicht alle Faktoren konstant gehalten werden, z.B. das Alter der Versuchspersonen. Es gibt Einflussfaktoren, die nicht einmal bekannt sind. Die Größen, die nicht kontrolliert, aber mitbeobachtet werden können, heißen Kovariablen. Ein Teil der Varianz des Fehlerterms kann durch die Berücksichtigung der Kovariablen im Modell erklärt werden. Daher ist ein Ziel der Kovarianzanalyse, die Streuung des Versuchsfehlers zu vermindern, um dadurch die Macht der Tests zu erhöhen.

### 6.3.1 Notation und Modell

Die theoretischen Ausführungen der folgenden Kapitel gehen im Wesentlichen auf die Arbeit von SIEMER (1999) zurück. Die Ergebnisse werden auf das nichtparametrische Modell im COP übertragen. Auf die Beweise wird daher in diesem Abschnitt verzichtet. Auch sie befinden sich sämtlich in SIEMER (1999).

Die Bezeichnung der Zielvariablen  $X_{iks}$  für  $i, s = 1, 2$  und  $k = 1, \dots, n_i$  ist dieselbe wie



zuvor. Die Baseline-Werte werden mit dem Hochindex ( $BL$ ) kenntlich gemacht.  $X_{iks}^{(BL)}$  bezeichnet somit den Baseline-Wert, der zu der entsprechenden Messung  $X_{iks}$  gehört und vorher erhoben wurde. Für  $s = 1$  ist  $X_{ik1}^{(BL)}$  der Run-In-Wert der  $k$ -ten Versuchseinheit aus der  $i$ -ten Gruppe, und für  $s = 2$  ist  $X_{ik2}^{(BL)}$  der Wert, der für die gleiche Versuchseinheit nach der Wash-Out-Phase, also vor der Behandlung in der zweiten Periode, erhoben wurde.

Das in SIEMER (1999) angegebene gemischte Modell mit Kovariablen vereinfacht sich für die Cross-Over-Versuchsanlage wie folgt.

**Definition 6.3.1 (Das gemischte Modell mit Kovariablen (GMK))** Seien für  $i = 1, 2$  und  $k = 1, \dots, n_i$  unabhängige Zufallsvektoren

$$\begin{aligned}\mathbf{X}_{ik} &= (X_{ik1}, X_{ik2})^T \sim \mathbf{F}_i \quad \text{und} \\ \mathbf{X}_{ik}^{(BL)} &= (X_{ik1}^{(BL)}, X_{ik2}^{(BL)})^T \sim \mathbf{F}_i^{(BL)}\end{aligned}$$

gegeben. Für die Einträge gelte

$$\begin{aligned}X_{iks} &\sim F_{is} \quad \text{und für die Baseline-Werte} \\ X_{iks}^{(BL)} &\sim F_{is}^{(BL)}\end{aligned}$$

mit  $i, s = 1, 2$  und  $k = 1, \dots, n_i$ ,

d.h. die Verteilungsfunktionen hängen nicht von den Versuchswiederholungen ab, sondern von der Faktorstufenkombination und der Kovariablen. Zwischen den Komponenten der Vektoren

$$(X_{iks}, X_{iks}^{(BL)})^T \quad i, s = 1, 2, \quad k = 1, \dots, n_i,$$

sind beliebige Abhängigkeiten zugelassen und die Vektoren selbst sind voneinander unabhängig.

Die Zufallsvariablen und Verteilungsfunktionen werden in Vektoren angeordnet:

$$\begin{aligned}\widetilde{\mathbf{X}} &= (\mathbf{X}^T, \mathbf{X}^{(BL)T})^T \\ &= (X_{111}, \dots, X_{2n_22}, X_{111}^{(BL)}, \dots, X_{2n_22}^{(BL)})^T \quad \text{und} \\ \widetilde{\mathbf{F}} &= (\mathbf{F}^T, \mathbf{F}^{(BL)T})^T \\ &= (F_{11}, \dots, F_{22}, F_{11}^{(BL)}, \dots, F_{22}^{(BL)})^T.\end{aligned}$$

Die Indizierung innerhalb der Vektoren erfolgt wie im Folgenden auch zuerst nach der Unterscheidung Messwert oder Baseline-Wert ( $BL$ ), dann nach der Gruppe ( $i = 1, 2$ ), anschließend nach der Zeit ( $s = 1, 2$ ) und zuletzt nach der Versuchseinheit ( $k = 1, \dots, n_i$ ). Vektoren wie beispielsweise  $\widetilde{\mathbf{X}}$  und  $\widetilde{\mathbf{F}}$ , die zunächst den Vektor der Messgrößen und anschließend den Vektor der Baseline-Werte enthalten, werden mit dem Zeichen  $\sim$  über dem fettgedruckten Vektor-Buchstaben kenntlich gemacht.

Sämtliche Größen wie der relative Behandlungseffekt, Ränge, harmonische Ränge und Schätzer für die Verteilungsfunktion, mittlere Verteilungsfunktion sowie der relative Effekt lassen sich wie im nichtparametrischen Modell (GM) für die Stichproben  $\mathbf{X} = (X_{iks})_{i,s,k}$  auch separat für die *Baseline-Stichproben*  $\mathbf{X}^{(BL)} = (X_{iks}^{(BL)})_{i,s,k}$  definieren. Man betrachtet diese beiden Stichproben, d.h. die Messwerte unter Behandlung und die Baseline-Werte zunächst einmal als getrennte Datensätze und bildet die interessierenden Größen jeweils wie im Abschnitt 4.2 angegeben. Um den Unterschied bei den ermittelten Größen zu verdeutlichen, wird immer, wenn es um die Baseline-Werte geht, mit dem Hochindex  $(BL)$  gearbeitet. Die Interpretation aller Größen ist die Gleiche wie im nichtparametrischen Modell ohne Kovariablen. Durch die getrennte Betrachtung dieser beiden Stichproben erklärt sich, dass bei der Kovarianzanalyse die Skala der Baseline-Werte nicht mit der Skala der Messwerte übereinstimmen muss. Diese beiden Werte werden nicht direkt miteinander in Bezug gesetzt wie dies bei der Bildung der adjustierten Werte der Fall ist.

$G^{(BL)}$  bezeichne die mittlere gewichtete Verteilungsfunktion aus der Baseline-Stichprobe, und damit bezeichnet

$$p_{is}^{(BL)} = \int G^{(BL)} dF_{is}^{(BL)}$$

den relativen ungewichteten Effekt für  $i, s = 1, 2$  und  $k = 1, \dots, n_i$  etc.

Die relativen Effekte beider Stichproben fasst man wieder in einem Vektor zusammen

$$\begin{aligned} \tilde{\mathbf{p}} &= \left( \mathbf{p}^T, \mathbf{p}^{(BL)T} \right)^T \\ &= \left( p_{111}, \dots, p_{2n_22}, p_{111}^{(BL)}, \dots, p_{2n_22}^{(BL)} \right)^T. \end{aligned}$$

Zur besseren Unterscheidung von  $\mathbf{p} = (p_{11}, p_{12}, p_{21}, p_{22})^T$  wird der *gemeinsame* Vektor von  $\mathbf{p}$  und  $\mathbf{p}^{(BL)}$  mit  $\tilde{\mathbf{p}}$  bezeichnet. Der gemeinsame Vektor der Schätzer des relativen Effekts wird entsprechend gebildet und mit  $\hat{\tilde{\mathbf{p}}}$  bezeichnet.

Sei  $Y_{iks}^{(BL)} = G^{(BL)}(X_{iks}^{(BL)})$  und bezeichne im Weiteren analog zum nichtparametrischen Modell (GM) ohne Kovariablen

$$\hat{Y}_{iks}^{(BL)} = \hat{G}^{(BL)}(X_{iks}^{(BL)})$$

die asymptotische Harmonische-Rang Transformation AHRT für die Baseline-Messwerte. Auch diese Rangtransformationen lassen sich kanonisch in einem Vektor anordnen:

$$\begin{aligned} \tilde{\mathbf{Y}} &= \left( Y_{111}, \dots, Y_{2n_22}, Y_{111}^{(BL)}, \dots, Y_{2n_22}^{(BL)} \right)^T \quad \text{und} \\ \hat{\tilde{\mathbf{Y}}} &= \left( \hat{Y}_{111}, \dots, \hat{Y}_{2n_22}, \hat{Y}_{111}^{(BL)}, \dots, \hat{Y}_{2n_22}^{(BL)} \right)^T. \end{aligned}$$

Nachdem zunächst die Stichprobe mit den Messwerten und die Baseline-Stichprobe getrennt betrachtet wurden, stellt sich an dieser Stelle die Frage, wie diese beiden Stichproben jetzt miteinander verbunden bzw. zueinander in Beziehung gesetzt werden können.

### 6.3.2 Regressionsmodell

Um die Streuung der Zielvariablen, die unter anderem durch die Baseline-Werte verursacht wird, zu reduzieren und damit die Macht der Statistiken zu verbessern, hat LANGER (1998) die Existenz einer Zerlegung der Zielvariablen in der asymptotischen Rangtransformation der folgenden Form eingeführt (übertragen auf die ungewichtete mittlere Verteilungsfunktion):

$$Y_{iks} = \gamma Y_{iks}^{(BL)} + Y_{iks}^{reg} \quad \text{für alle } i, s = 1, 2, k = 1, \dots, n_i,$$

wobei  $Y_{iks}^{reg}$  eine nicht beobachtbare Zufallsvariable ist. Anschaulich bedeutet dies unter der zusätzlichen Voraussetzung  $E(Y_{iks}^{reg}) = 0$ : Wendet man eine lineare Regression durch den Ursprung auf die unabhängige Größe *Baseline-Wert* und die abhängige Größe *Messwert* an, so bezeichnet  $Y_{iks}^{reg}$  die individuelle Abweichung der Beobachtung  $Y_{iks}$  von der Regressionsgeraden. Durch den Regressionskoeffizienten  $\gamma$  wird der Einfluss der transformierten Kovariablen auf die transformierte Zielvariable beschrieben. In der zuvor beschriebenen Veranschaulichung bedeutet dies die Steigung der Geraden. Die Voraussetzungen an den Regressionskoeffizienten sowie die Veranschaulichung dieses Ansatzes werden in LANGER (1998) diskutiert.

Durch Umformung und mit  $\beta = -\gamma$  ergibt sich

$$Y_{iks}^{reg} = Y_{iks} + \beta Y_{iks}^{(BL)} \quad \text{für alle } i, s = 1, 2, k = 1, \dots, n_i. \quad (6.3)$$

Basierend auf diesem Regressionsmodell kann der *bereinigte Schätzer*  $\hat{p}_{is}^{reg}$  der ungewichteten relativen Behandlungseffekte gebildet werden. Durch die Beziehung  $\hat{p}_{is} = \sum_{k=1}^{n_i} \hat{Y}_{iks} = \hat{\bar{Y}}_{i,s}$  ergibt sich mit Hilfe der Gleichung (6.3)

$$\hat{p}_{is}^{reg} = \hat{p}_{is} + \beta \hat{p}_{is}^{(BL)}$$

ein Schätzer der durch den Einfluss der Kovariablen bereinigten ungewichteten relativen Effekte

$$p_{is}^{reg} = p_{is} + \beta p_{is}^{(BL)}$$

für alle  $i, s = 1, 2$ .

Die empirische Größe  $\hat{p}_{is}^{reg}$  hat die angenehmen Eigenschaften der Konsistenz und asymptotischen Erwartungstreue im Bezug auf  $p_{is}^{reg}$ . Ferner ist  $\hat{p}_{is}^{reg}$  die gewichtete Summe von Integralen und damit liegt dieser Schätzer im Allgemeinen nicht im Intervall  $(0, 1)$  wie die relativen Effekte, die sich sonst hinter dieser Notation verbergen. Somit kann die Interpretation der relativen Effekte aus Abschnitt 4.2.1 ab Seite 28 nicht einfach übernommen werden. Zur weiteren Erläuterung dieser bereinigten relativen Effekte sei auf SIEMER (1999) verwiesen.

Sei

$$\beta = (1, \beta).$$

Bezeichnen

$$\begin{aligned} \mathbf{p}^{reg} &= (\beta \otimes \mathbf{I}_4) \tilde{\mathbf{p}} = \mathbf{p} + \beta \mathbf{p}^{(BL)} \quad \text{und} \\ \hat{\mathbf{p}}^{reg} &= (\beta \otimes \mathbf{I}_4) \hat{\tilde{\mathbf{p}}} = \hat{\mathbf{p}} + \beta \hat{\mathbf{p}}^{(BL)} \end{aligned}$$

den Vektor der bereinigten relativen Effekte bzw. des Schätzers.  $\mathbf{p}^{reg}$  setzt sich offenbar aus dem bereits bekannten Vektor der relativen Effekte und dem Produkt aus dem Koeffizienten  $\beta$  und dem relativen Effekt der Baseline-Stichprobe zusammen. Mit letzterem soll der Teil der Verzerrung des relativen Effekts herausgerechnet werden, der durch die Kovariablen verursacht wurde. Entsprechendes gilt für den Schätzer  $\hat{\mathbf{p}}^{reg}$

Um Teststatistiken mit Hilfe der Statistik  $\sqrt{N}\mathbf{c}\hat{\mathbf{p}}^{reg}$  bilden zu können, benötigt man die Verteilung der Statistik sowie die Kovarianzmatrix unter den beiden betrachteten Hypothesen  $H_0^{p^{reg}} : \mathbf{c}\mathbf{p}^{reg} = 0$  und  $H_0^F : \mathbf{c}\mathbf{F} = 0$ . Diese Resultate erhält man wie in Abschnitt 4.2.1 über entsprechende Vergleichsstatistiken.

Für folgende Berechnungen wird häufig der Vektor

$$\tilde{\mathbf{Y}} := \left( \bar{Y}_{1.1}, \dots, \bar{Y}_{2.2}, \bar{Y}_{1.1}^{(BL)}, \dots, \bar{Y}_{2.2}^{(BL)} \right)^T \quad (6.4)$$

mit  $\bar{Y}_{i.s} = 1/n_i \cdot \sum_{k=1}^{n_i} Y_{iks}$  bzw.  $\bar{Y}_{i.s}^{(BL)} = 1/n_i \cdot \sum_{k=1}^{n_i} Y_{iks}^{(BL)}$  gebraucht.  $\bar{\mathbf{Y}}_i$  sei der Vektor  $\bar{\mathbf{Y}}_i = (\bar{Y}_{i.1}, \bar{Y}_{i.2})$  und entsprechend  $\bar{\mathbf{Y}}_i^{(BL)}$ .

### 6.3.3 Asymptotik und Grenzverteilungen

Mit Hilfe des Asymptotischen Äquivalenzsatzes für das (GMK) erhält man wie bei der Herleitung der Statistiken für das nichtparametrische Modell eine Teststatistik, die zu  $\sqrt{N}(\hat{\mathbf{p}}^{reg} - \mathbf{p}^{reg})$  äquivalent ist. Die Vergleichsstatistik wird wiederum benutzt, um die Verteilungsaussagen für diese Statistik zu treffen. Genauer bedeutet dies, dass unter der Nullhypothese  $H_0^F : \mathbf{c}\mathbf{F} = 0$  die Kovarianzmatrix  $\mathbf{V}_N^{reg}$  nicht von  $\sqrt{N}\hat{\mathbf{p}}^{reg}$  berechnet werden muss, sondern stattdessen der Einfachheit halber auf die Vergleichsstatistik  $\sqrt{N}\tilde{\mathbf{Y}}$  von  $\sqrt{N}\hat{\mathbf{p}}$  zurückgegriffen werden kann.

Bezeichne  $\tilde{\mathbf{V}}_N = \text{Cov}(\sqrt{N}\tilde{\mathbf{Y}})$  die Kovarianzmatrix der Vergleichsstatistik unter  $H_0^F : \mathbf{c}\mathbf{F} = 0$ , wobei  $\tilde{\mathbf{V}}_N$  die folgende Blockstruktur hat:

$$\tilde{\mathbf{V}}_N = \begin{pmatrix} \mathbf{V}_N & \mathbf{V}_N^{(COV)} \\ \mathbf{V}_N^{(COV)^T} & \mathbf{V}_N^{(BL)} \end{pmatrix}.$$

Aufgrund der durch das Modell vorgegebenen Abhängigkeitsstruktur der Zufallsvariablen können die einzelnen Blöcke als Kroneckersumme dargestellt werden:

$$\mathbf{V}_N = \bigoplus_{i=1}^2 \frac{N}{n_i} \mathbf{V}_{i,n_i} \quad \mathbf{V}_N^{(BL)} = \bigoplus_{i=1}^2 \frac{N}{n_i} \mathbf{V}_{i,n_i}^{(BL)} \quad \mathbf{V}_N^{(COV)} = \bigoplus_{i=1}^2 \frac{N}{n_i} \mathbf{V}_{i,n_i}^{(COV)}$$

mit  $\mathbf{V}_{i,n_i}^{(COV)} = \text{Cov}(\sqrt{N}\bar{\mathbf{Y}}_i, \sqrt{N}\bar{\mathbf{Y}}_i^{(BL)})$  und  $\mathbf{V}_{i,n_i}$  (und entsprechend  $\mathbf{V}_{i,n_i}^{(BL)}$ ) wie in Abschnitt 4.2.5 definiert.

Für den Beweis von Verteilungsaussagen unter der Hypothese über die Verteilungsfunktionen  $H_0^F : \mathbf{c}\mathbf{F} = 0$  wird die Modellvoraussetzung von LANGER (1998) benötigt:

**Voraussetzung 6.3.2 (MV)**  $F_{is}^{(BL)} \equiv F^{(BL)}$  für alle  $i, s = 1, 2$ .

Diese Voraussetzung impliziert, dass die Verteilung der Kovariablen gleich ist. Die Verteilungen der Baseline-Werte unterscheiden sich dann weder in Bezug auf die Gruppe, noch in Bezug auf den Zeitpunkt.

Unter den weiteren Voraussetzungen, dass die Stichprobenumfänge gleichmäßig gegen unendlich streben und die Eigenwerte der Kovarianzmatrix konvergieren, kann folgendes Resultat bewiesen werden.

**Satz 6.3.3 (Asymptotische Normalität unter  $H_0^F : \mathbf{cF} = 0$ )**  $\sqrt{N}\widehat{\mathbf{cP}}^{reg}$  ist unter  $H_0^F : \mathbf{cF} = 0$  asymptotisch normalverteilt mit Erwartungswert 0 und Varianz  $\mathbf{cV}_N^{reg}\mathbf{c}^T$ . Dabei lässt sich  $\mathbf{V}_N^{reg}$  darstellen als

$$\begin{aligned}\mathbf{V}_N^{reg} &= \mathbf{V}_N + \beta\mathbf{V}_N^{(COV)T} + \beta\mathbf{V}_N^{(COV)} + \beta^2\mathbf{V}_N^{(BL)} \\ &= \bigoplus_{i=1}^2 \frac{N}{n_i} \left( \mathbf{V}_{i,n_i} + \beta\mathbf{V}_{i,n_i}^{(COV)T} + \beta\mathbf{V}_{i,n_i}^{(COV)} + \beta^2\mathbf{V}_{i,n_i}^{(BL)} \right).\end{aligned}$$

mit

$$\begin{aligned}\widehat{\mathbf{V}}_{i,n_i} &= \frac{1}{16\tilde{n}^2(n_i - 1)} \sum_{k=1}^{n_i} (\Psi_{ik} - \bar{\Psi}_i) (\Psi_{ik} - \bar{\Psi}_i)^T \\ \widehat{\mathbf{V}}_{i,n_i}^{(BL)} &= \frac{1}{16\tilde{n}^2(n_i - 1)} \sum_{k=1}^{n_i} \left( \Psi_{ik}^{(BL)} - \bar{\Psi}_i^{(BL)} \right) \left( \Psi_{ik}^{(BL)} - \bar{\Psi}_i^{(BL)} \right)^T \quad \text{und} \\ \widehat{\mathbf{V}}_{i,n_i}^{(COV)} &= \frac{1}{16\tilde{n}^2(n_i - 1)} \sum_{k=1}^{n_i} (\Psi_{ik} - \bar{\Psi}_i) \left( \Psi_{ik}^{(BL)} - \bar{\Psi}_i^{(BL)} \right)^T\end{aligned}$$

konsistente Schätzer und  $\Psi_{ik} = (\Psi_{ik1}, \Psi_{ik2})^T$ ,  $\bar{\Psi}_i = (\bar{\Psi}_{i.1}, \bar{\Psi}_{i.2})^T$  und entsprechend für die Baseline-Werte.

Mit Hilfe dieser Information ist es jetzt für die Nullhypothese  $H_0^F : \mathbf{cF} = 0$  möglich, die entsprechenden Teststatistiken anzugeben. Die Teststatistiken befinden sich im folgenden Abschnitt 6.3.4.

In den Arbeiten von LANGER (1998) und SIEMER (1999) sind verschiedene Hypothesen zum Testen von Effekten, wenn Kovariablen vorhanden sind, diskutiert. Für gewöhnlich werden die Hypothesen über die Verteilungsfunktion gestellt  $H_0^F : \mathbf{cF} = 0$ , wozu im (GMK) die Modellvoraussetzung (MV) benötigt wird. Diese Hypothesenstellung bereitet Probleme, wenn  $\mathbf{cF} \neq 0$ , aber  $\mathbf{cP} = 0$  gilt. Die verbleibende sinnvolle Hypothese ist  $H_0^{p^{reg}} : \mathbf{cP}^{reg} = 0$ , wenn man die Modellvoraussetzung (MV) umgehen möchte. Unter dieser Voraussetzung wären die Hypothesen  $H_0^p : \mathbf{cP} = 0$  und  $H_0^{p^{reg}} : \mathbf{cP}^{reg} = 0$  identisch, da dann  $p_{is}^{(BL)} \equiv 1/2$  für alle  $i, s = 1, 2$ . Würde man die Modellvoraussetzung fallen lassen, so kann im Modell ohne Kovariablen analog zum nichtparametrischen Modell (GM) die Hypothese in der Form  $H_0^p : \mathbf{cP} = 0$  gestellt werden. Der eventuell gefundene Effekt kann bei Mitaufnahme von Kovariablen in das Modell auch durch diese mitbeobachteten Werte verursacht sein, weshalb diese Art der Hypothesenstellung durch die Hypothesenstellung über  $\mathbf{P}^{reg}$  ersetzt wird. Damit wird versucht die entstandene Verzerrung in die richtige Richtung zu korrigieren. Da weiterhin die Hypothese  $H_0^{p^{reg}} : \mathbf{cP}^{reg} = 0$  mit

Hilfe der ungewichteten relativen Effekte gebildet wird, hängt diese Hypothese zudem nicht von den Stichprobenumfängen ab.

Bezeichne

$$\mathbf{Z}_N^{reg} := \sqrt{N} (\boldsymbol{\beta}^T \otimes \mathbf{I}_4) \tilde{\mathbf{Z}}_N$$

$$\text{mit } \tilde{\mathbf{Z}}_N := \begin{pmatrix} \int G d\hat{\mathbf{F}} - \int \mathbf{F} d\hat{G} \\ \int G^{(BL)} d\hat{\mathbf{F}}^{(BL)} - \int \mathbf{F}^{(BL)} d\hat{G}^{(BL)} \end{pmatrix} + \mathbf{1}_8 - 2\mathbf{p}.$$

Im Modell (GMK) gilt unter der Voraussetzung, dass die Stichprobenumfänge gleichmäßig gegen unendlich streben:

$$\sqrt{N} (\hat{\mathbf{p}}^{reg} - \mathbf{p}^{reg}) \doteq \mathbf{Z}_N^{reg}. \quad (6.5)$$

Insbesondere gilt unter der Hypothese  $H_0^{p^{reg}} : \mathbf{c}\mathbf{p}^{reg} = 0$

$$\sqrt{N} \mathbf{c}\hat{\mathbf{p}}^{reg} \doteq \sqrt{N} \mathbf{c} (\boldsymbol{\beta}^T \otimes \mathbf{I}_4) \begin{pmatrix} \int G d\hat{\mathbf{F}} - \int \mathbf{F} d\hat{G} \\ \int G^{(BL)} d\hat{\mathbf{F}}^{(BL)} - \int \mathbf{F}^{(BL)} d\hat{G}^{(BL)} \end{pmatrix}.$$

Um Teststatistiken für die verschiedenen Effekte herleiten zu können, benötigt man die Kovarianzmatrix und die Verteilung der Vergleichsstatistik  $\mathbf{Z}_N^{reg}$ . Aufgrund der asymptotischen Äquivalenz aus (6.5) gelten Verteilungsaussagen sowie auch die Kovarianzmatrix  $\mathbf{W}_N^{reg}$  asymptotisch für beide Statistiken.

$\tilde{\mathbf{W}}_N = \text{Cov}(\tilde{\mathbf{Z}}_N)$  hat dieselbe Blockstruktur wie die Matrix  $\tilde{\mathbf{V}}_N$ , die unter der Hypothese  $H_0^F : \mathbf{c}\mathbf{F} = 0$  hergeleitet wurde:

$$\tilde{\mathbf{W}}_N = \begin{pmatrix} \mathbf{W}_N & \mathbf{W}_N^{(COV)} \\ \mathbf{W}_N^{(COV)^T} & \mathbf{W}_N^{(BL)} \end{pmatrix}.$$

Im folgenden Satz wird die Verteilungsaussage für die Statistik  $\sqrt{N}(\hat{\mathbf{p}}^{reg} - \mathbf{p}^{reg})$  formuliert. Für die Verteilungsaussage müssen, wie in den vorangegangenen Abschnitten, Forderungen an die Kovarianzmatrix gestellt werden. Unter den Voraussetzungen, dass die Stichprobenumfänge gleichmäßig gegen unendlich streben und die Eigenwerte der Kovarianzmatrix konvergieren, kann folgendes Resultat bewiesen werden. Für die exakte Formulierung der Voraussetzungen sei auf SIEMER (1999) verwiesen.

**Satz 6.3.4 (Asymptotische Normalität unter  $H_0^{p^{reg}} : \mathbf{c}\mathbf{p}^{reg} = 0$ )** *Mit den genannten Voraussetzungen ist  $\sqrt{N}(\hat{\mathbf{p}}^{reg} - \mathbf{p}^{reg})$  asymptotisch multivariat normalverteilt mit Erwartungswertvektor  $\mathbf{0}$  und Kovarianzmatrix*

$$\mathbf{W}_N^{reg} = (\boldsymbol{\beta}^T \otimes \mathbf{I}_4) \tilde{\mathbf{W}}_N (\boldsymbol{\beta}^T \otimes \mathbf{I}_4)^T.$$

*Unter der Hypothese  $H_0^{p^{reg}} : \mathbf{c}\mathbf{p}^{reg} = 0$  ist  $\sqrt{N}\mathbf{c}\hat{\mathbf{p}}^{reg}$  asymptotisch normalverteilt mit Erwartungswert  $0$  und Varianz*

$$\mathbf{c}\mathbf{W}_N^{reg}\mathbf{c}^T = \mathbf{c} (\boldsymbol{\beta}^T \otimes \mathbf{I}_4) \tilde{\mathbf{W}}_N (\boldsymbol{\beta}^T \otimes \mathbf{I}_4)^T \mathbf{c}^T.$$

Dabei lässt sich  $\widehat{\mathbf{W}}_N$  darstellen mit den Diagonalelementen

$$\widehat{c}_{iiss} = \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right)^2 + \frac{N}{16n_j(n_j - 1)} \sum_{k=1}^{n_j} \left( \widehat{Z}_{jk}^{(-is)} - \widehat{Z}_j^{(-is)} \right)^2$$

für  $j \neq i$  bzw.  $\widehat{c}_{iiss}^{(BL)}$  analog. Ferner gilt:

$$\begin{aligned} \widehat{c}_{iiss}^{(COV)} &= \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right) \left( \widehat{Z}_{iks}^{(BL)} - \widehat{Z}_{i.s}^{(BL)} \right) \\ &+ \frac{N}{16n_j(n_j - 1)} \sum_{k=1}^{n_j} \left( \widehat{Z}_{jk}^{(-is)} - \widehat{Z}_j^{(-is)} \right) \left( \widehat{Z}_{jk}^{(BL)(-is)} - \widehat{Z}_j^{(BL)(-is)} \right) \quad j \neq i. \end{aligned}$$

Die Kovarianzschätzer innerhalb einer Gruppe  $i$  ergeben sich zu

$$\begin{aligned} \widehat{c}_{iiss'} &= \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right) \left( \widehat{Z}_{iks'} - \widehat{Z}_{i.s'} \right) \\ &+ \frac{N}{16n_j(n_j - 1)} \sum_{k=1}^{n_j} \left( \widehat{Z}_{jk}^{(-is)} - \widehat{Z}_j^{(-is)} \right) \left( \widehat{Z}_{jk}^{(-is')} - \widehat{Z}_j^{(-is')} \right) \quad j \neq i \end{aligned}$$

und für  $i \neq i'$  zu

$$\begin{aligned} \widehat{c}_{ii'ss'} &= \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right) \left( \widehat{Z}_{ik}^{(-i's')} - \widehat{Z}_i^{(-i's')} \right) \\ &+ \frac{N}{16n_{i'}(n_{i'} - 1)} \sum_{k'=1}^{n_{i'}} \left( \widehat{Z}_{i'k's'} - \widehat{Z}_{i'.s'} \right) \left( \widehat{Z}_{i'k'}^{(-is)} - \widehat{Z}_{i'}^{(-is)} \right). \end{aligned}$$

Entsprechend ergeben sich die Kovarianzschätzer  $\widehat{c}_{iiss'}^{(BL)}$  und  $\widehat{c}_{ii'ss'}^{(BL)}$ . Ferner ergeben sich die analogen Kovarianzschätzer für  $\widehat{c}_{iiss'}^{(COV)}$  und  $\widehat{c}_{ii'ss'}^{(COV)}$  zu

$$\begin{aligned} \widehat{c}_{iiss'}^{(COV)} &= \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right) \left( \widehat{Z}_{iks'}^{(BL)} - \widehat{Z}_{i.s'}^{(BL)} \right) \\ &+ \frac{N}{16n_j(n_j - 1)} \sum_{k=1}^{n_j} \left( \widehat{Z}_{jk}^{(-is)} - \widehat{Z}_j^{(-is)} \right) \left( \widehat{Z}_{jk}^{(BL)(-is')} - \widehat{Z}_j^{(BL)(-is')} \right) \quad j \neq i \end{aligned}$$

und für  $i \neq i'$  zu

$$\begin{aligned} \widehat{c}_{ii'ss'}^{(COV)} &= \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right) \left( \widehat{Z}_{ik}^{(BL)(-i's')} - \widehat{Z}_i^{(BL)(-i's')} \right) \\ &+ \frac{N}{16n_{i'}(n_{i'} - 1)} \sum_{k'=1}^{n_{i'}} \left( \widehat{Z}_{i'k's'} - \widehat{Z}_{i'.s'} \right) \left( \widehat{Z}_{i'k'}^{(BL)(-is)} - \widehat{Z}_{i'}^{(BL)(-is)} \right). \end{aligned}$$

$\widehat{\mathbf{W}}_N$  ist ein konsistenter Schätzer für  $\widetilde{\mathbf{W}}_N$ .

Die genaue Herleitung des Kovarianzschätzers sowie der Beweis dieses Satzes befinden sich für den allgemeineren Fall von beliebig vielen Kovariablen in SIEMER (1999).

Die Schätzung der für die Teststatistiken unter  $H_0^{p^{reg}} : \mathbf{c}\mathbf{p}^{reg} = 0$  benötigten Kovarianzmatrix  $\widehat{\mathbf{W}}_N^{reg}$  lässt sich im Wesentlichen auf die Schätzung der Komponenten  $Z_{iks}$  ( $Z_{iks}^{(BL)}$ ) und  $Z_{jk}^{(-is)}$  ( $Z_{jk}^{(BL)(-is)}$ ) zurückführen. Diese empirischen Größen lassen sich aus den Stichproben berechnen und mit Hilfe der harmonischen Ränge ausdrücken.  $Z_{iks}$  und  $Z_{jk}^{(-is)}$  sind bereits aus Abschnitt 4.2.5 ab Seite 37 bekannt. Entsprechend ergeben sich die Größen für die Baseline-Stichprobe:

$$\begin{aligned}\widehat{Z}_{iks}^{(BL)} &= 4 \cdot \widehat{G}^{(BL)}(X_{iks}^{(BL)}) - \sum_{t=1}^2 \widehat{F}_{is}^{(BL)}(X_{ikt}^{(BL)}) \\ &= \frac{1}{\widetilde{n}} \left( \Psi_{iks}^{(BL)} - \frac{\widetilde{n}}{2n_i} \right) - \frac{1}{n_i} \left( \Psi_{iks}^{(BL)(is)} - \frac{1}{2} \right) - \frac{1}{\widetilde{n}} \left( \Psi_{ikt}^{(BL)} - \Psi_{ikt}^{(BL)(-is)} \right)\end{aligned}$$

mit  $t \neq s$  und

$$\begin{aligned}\widehat{Z}_{jk}^{(BL)(-is)} &= - \sum_{t=1}^2 \widehat{F}_{is}^{(BL)}(X_{jkt}^{(BL)}) \\ &= \frac{1}{\widetilde{n}} \sum_{t=1}^2 \left( \Psi_{jkt}^{(BL)(-is)} - \Psi_{jkt}^{(BL)} \right) \text{ für } j \neq i.\end{aligned}$$

Mit diesen Hilfsmitteln können analog zu den Abschnitten 4.2.6 und 4.2.7 ohne die Berücksichtigung von Baseline-Werten, Teststatistiken für das nichtparametrische Modell mit Kovariablen (GMK) angegeben werden.

### 6.3.4 Teststatistiken für $H_0^F$

Mit Hilfe der Ergebnisse des vorigen Abschnitts lassen sich unter der Voraussetzung  $\mathbf{c}\widehat{\mathbf{V}}_N^{reg} \mathbf{c}^T \geq c_0 > 0$  die Teststatistiken für die Effekte des COP mit Baseline-Werten allgemein über folgendes Resultat ausrechnen:

$$U_n^{F,reg} = \frac{\sqrt{N} \mathbf{c}\widehat{\mathbf{p}}^{reg}}{\sqrt{\mathbf{c}\widehat{\mathbf{V}}_N^{reg} \mathbf{c}^T}} \quad \rightsquigarrow \quad N(0, 1). \quad (6.6)$$

Der Zusatz *reg* im Exponenten der Teststatistik im Unterschied zu den Statistiken im Modell ohne Baseline-Werte aus (4.14) suggeriert die Berücksichtigung der Vorwerte in diesen Berechnungen.

Generell war eine Voraussetzung für die Herleitung dieser Statistiken unter der Hypothese  $H_0^F : \mathbf{c}\mathbf{F} = 0$  die Gleichheit der Verteilungen der Baseline-Werte (MV):

$$X_{iks}^{(BL)} \sim F_{is} \equiv F \quad \text{für alle } i, s = 1, 2, k = 1, \dots, n_i.$$

Dass die Verteilungen über die Gruppe gleich sein sollen, kann ohne weiteres angenommen werden, da die Individuen den Gruppen randomisiert zugeteilt werden. Die Voraussetzung, dass die Verteilung über die Zeit gleich sein soll, impliziert jedoch folgendes: Es



darf weder ein Einfluss einer Behandlung, der Zeit oder ein Nacheffekt auf die Baseline-Messwerte einwirken. Diese Annahme kann unter Umständen realitätsfern sein, weshalb in einem solchen Falle noch die Teststatistiken aus dem Abschnitt 6.3.5 für die Hypothesen  $H_0^{P^{reg}} : \mathbf{c}p^{reg} = 0$  zur Verfügung stehen.

**Statistik für  $H_0^{F,reg}(\lambda) : \bar{F}_{1.} = \bar{F}_{2.}$**

$$\begin{aligned} U_n^{\lambda,F,reg} &= \frac{\sqrt{N}\mathbf{c}_\lambda\hat{\mathbf{p}}^{reg}}{\sqrt{\mathbf{c}_\lambda\hat{\mathbf{V}}_N^{reg}\mathbf{c}_\lambda^T}} \\ &= \frac{\sqrt{N}\mathbf{c}_\lambda\hat{\mathbf{p}} + \beta\sqrt{N}\mathbf{c}_\lambda\hat{\mathbf{p}}^{(BL)}}{\sqrt{\frac{N}{16\bar{n}^2}\sum_{i=1}^2\frac{\hat{\sigma}_{i,F}^{reg^2}}{n_i}}} \text{ mit} \\ \hat{\sigma}_{i,F}^{reg^2} &= \frac{\hat{\sigma}_{i,F}^2}{n_i} + \beta^2\frac{\hat{\sigma}_{i,F}^{(BL)^2}}{n_i} + 2\beta\frac{\hat{\sigma}_{i,F}^{(COV)^2}}{n_i}, \quad \text{wobei} \\ \hat{\sigma}_{i,F}^2 &= \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\Psi_{ik1} + \Psi_{ik2} - \bar{\Psi}_{i.1} - \bar{\Psi}_{i.2})^2 \end{aligned}$$

und  $\hat{\sigma}_{i,F}^{(BL)^2}$  entsprechend. Ferner ist:

$$\hat{\sigma}_{i,F}^{(COV)^2} = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\Psi_{ik1} + \Psi_{ik2} - \bar{\Psi}_{i.1} - \bar{\Psi}_{i.2}) \left( \Psi_{ik1}^{(BL)} + \Psi_{ik2}^{(BL)} - \bar{\Psi}_{i.1}^{(BL)} - \bar{\Psi}_{i.2}^{(BL)} \right).$$

**Statistik für  $H_0^{F,reg}(\pi) : \bar{F}_{1.} = \bar{F}_{2.}$**

$$\begin{aligned} U_n^{\pi,F,reg} &= \frac{\sqrt{N}\mathbf{c}_\pi\hat{\mathbf{p}}^{reg}}{\sqrt{\mathbf{c}_\pi\hat{\mathbf{V}}_N^{reg}\mathbf{c}_\pi^T}} \\ &= \frac{\sqrt{N}\mathbf{c}_\pi\hat{\mathbf{p}} + \beta\sqrt{N}\mathbf{c}_\pi\hat{\mathbf{p}}^{(BL)}}{\sqrt{\frac{N}{16\bar{n}^2}\sum_{i=1}^2\frac{\hat{\tau}_{i,F}^{reg^2}}{n_i}}} \text{ mit} \\ \hat{\tau}_{i,F}^{reg^2} &= \frac{\hat{\tau}_{i,F}^2}{n_i} + \beta^2\frac{\hat{\tau}_{i,F}^{(BL)^2}}{n_i} + 2\beta\frac{\hat{\tau}_{i,F}^{(COV)^2}}{n_i} \quad \text{wobei} \\ \hat{\tau}_{i,F}^2 &= \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\Psi_{ik1} - \Psi_{ik2} - \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2})^2 \end{aligned}$$

und  $\hat{\tau}_{i,F}^{(BL)^2}$  entsprechend. Ferner ist:

$$\hat{\tau}_{i,F}^{(COV)^2} = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\Psi_{ik1} - \Psi_{ik2} - \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2}) \left( \Psi_{ik1}^{(BL)} - \Psi_{ik2}^{(BL)} - \bar{\Psi}_{i.1}^{(BL)} + \bar{\Psi}_{i.2}^{(BL)} \right).$$

**Statistik für  $H_0^{F,reg}(\delta) : F_{11} - F_{12} = F_{21} - F_{22}$**

$$\begin{aligned} U_n^{\delta,F,reg} &= \frac{\sqrt{N} \mathbf{c}_\delta \widehat{\mathbf{p}}^{reg}}{\sqrt{\mathbf{c}_\delta \widehat{\mathbf{V}}_N^{reg} \mathbf{c}_\delta^T}} \\ &= \frac{\sqrt{N} \mathbf{c}_\delta \widehat{\mathbf{p}} + \beta \sqrt{N} \mathbf{c}_\delta \widehat{\mathbf{p}}^{(BL)}}{\sqrt{\frac{N}{16\bar{n}^2} \sum_{i=1}^2 \frac{\widehat{\tau}_{i,F}^{reg^2}}{n_i}}} \end{aligned}$$

mit  $\widehat{\tau}_{i,F}^{reg^2}$  wie zuvor.

Unter der Nullhypothese  $H_0^F : \mathbf{c}\mathbf{F} = 0$  haben die Statistiken nach (6.6) asymptotisch eine Standardnormalverteilung. Für kleine Stichprobenumfänge approximiert man die Verteilung der Statistiken unter Nullhypothese mit Hilfe der  $t$ -Verteilung. Der Freiheitsgrad von  $U_n^{\lambda,F,reg}$  ist

$$\frac{\left( \sum_{i=1}^2 \widehat{\sigma}_{i,F}^{reg^2} / n_i \right)^2}{\sum_{i=1}^2 (\widehat{\sigma}_{i,F}^{reg^2} / n_i)^2 / (n_i - 1)}$$

und der Freiheitsgrad für die Statistiken  $U_n^{\pi,F,reg}$  und  $U_n^{\delta,F,reg}$  ergibt sich durch

$$\frac{\left( \sum_{i=1}^2 \widehat{\tau}_{i,F}^{reg^2} / n_i \right)^2}{\sum_{i=1}^2 (\widehat{\tau}_{i,F}^{reg^2} / n_i)^2 / (n_i - 1)}.$$

### 6.3.5 Teststatistiken für $H_0^{p,reg}$

Unter der Voraussetzung  $\mathbf{c}\widehat{\mathbf{W}}_N^{reg} \mathbf{c}^T \geq c_0 > 0$  lassen sich mit Hilfe der Ergebnisse des Abschnitts 6.3.3 die Teststatistiken allgemein über folgendes Resultat ausrechnen:

$$U_n^{p,reg} = \frac{\sqrt{N} \mathbf{c}\widehat{\mathbf{p}}^{reg}}{\sqrt{\mathbf{c}\widehat{\mathbf{W}}_N^{reg} \mathbf{c}^T}} \overset{\sim}{\sim} N(0, 1). \quad (6.7)$$

**Statistik für  $H_0^{p,reg}(\lambda) : \bar{p}_{1.}^{reg} = \bar{p}_{2.}^{reg}$**

$$\begin{aligned} U_n^{\lambda,p,reg} &= \frac{\sqrt{N} \mathbf{c}_\lambda \widehat{\mathbf{p}}^{reg}}{\sqrt{\mathbf{c}_\lambda \widehat{\mathbf{W}}_N^{reg} \mathbf{c}_\lambda^T}} \\ &= \frac{\sqrt{N} \mathbf{c}_\lambda \widehat{\mathbf{p}} + \beta \sqrt{N} \mathbf{c}_\lambda \widehat{\mathbf{p}}^{(BL)}}{\sqrt{\mathbf{c}_\lambda \widehat{\mathbf{W}}_N^{reg} \mathbf{c}_\lambda^T}} \end{aligned}$$

**Statistik für  $H_0^{p,reg}(\pi) : \bar{p}_{1.}^{reg} = \bar{p}_{2.}^{reg}$**

$$\begin{aligned} U_n^{\pi,p,reg} &= \frac{\sqrt{N} \mathbf{c}_\pi \widehat{\mathbf{p}}^{reg}}{\sqrt{\mathbf{c}_\pi \widehat{\mathbf{W}}_N^{reg} \mathbf{c}_\pi^T}} \\ &= \frac{\sqrt{N} \mathbf{c}_\pi \widehat{\mathbf{p}} + \beta \sqrt{N} \mathbf{c}_\pi \widehat{\mathbf{p}}^{(BL)}}{\sqrt{\mathbf{c}_\pi \widehat{\mathbf{W}}_N^{reg} \mathbf{c}_\pi^T}} \end{aligned}$$

**Statistik für**  $H_0^{p,reg}(\delta) : p_{11}^{reg} - p_{12}^{reg} = p_{21}^{reg} - p_{22}^{reg}$

$$\begin{aligned} U_n^{\delta,p,reg} &= \frac{\sqrt{N} \mathbf{c}_\delta \widehat{\mathbf{p}}^{reg}}{\sqrt{\mathbf{c}_\delta \widehat{\mathbf{W}}_N^{reg} \mathbf{c}_\delta^T}} \\ &= \frac{\sqrt{N} \mathbf{c}_\delta \widehat{\mathbf{p}} + \beta \sqrt{N} \mathbf{c}_\delta \widehat{\mathbf{p}}^{(BL)}}{\sqrt{\mathbf{c}_\delta \widehat{\mathbf{W}}_N^{reg} \mathbf{c}_\delta^T}} \end{aligned}$$

Unter der Nullhypothese  $H_0^{p,reg} : \mathbf{c}\mathbf{p}^{reg} = 0$  haben die Statistiken asymptotisch eine Standardnormalverteilung. Für kleine Stichprobenumfänge approximiert man die Verteilung der Statistiken unter Nullhypothese mit Hilfe der  $t$ -Verteilung.

Der Freiheitsgrad für die Approximation mit der  $t$ -Verteilung von  $U_n^{\lambda,p,reg}$ :

$$\frac{\left( \sum_{i=1}^2 \widehat{\sigma}_{i,p}^{reg^2} / n_i \right)^2}{\sum_{i=1}^2 (\widehat{\sigma}_{i,p}^{reg^2} / n_i)^2 / (n_i - 1)},$$

wobei

$$\begin{aligned} \widehat{\sigma}_{i,p}^{reg^2} &= \frac{\widehat{\sigma}_{i,p}^2}{n_i} + \beta^2 \frac{\widehat{\sigma}_{i,p}^{(BL)^2}}{n_i} + 2\beta \frac{\widehat{\sigma}_{i,p}^{(COV)^2}}{n_i}, \quad \text{wobei} \\ \widehat{\sigma}_{i,p}^2 &= \frac{1}{(n_i - 1)} \sum_{k=1}^{n_i} \left( -\Psi_{ik1} - \Psi_{ik2} + 2\Psi_{ik}^{(-j1)} + 2\Psi_{ik}^{(-j2)} \right. \\ &\quad \left. + \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2} - 2\bar{\Psi}_i^{(-j1)} - 2\bar{\Psi}_i^{(-j2)} \right)^2 \quad \text{mit } j \neq i \end{aligned}$$

und  $\widehat{\sigma}_{i,p}^{(BL)^2}$  und  $\widehat{\sigma}_{i,p}^{(COV)^2}$  entsprechend.

Der Freiheitsgrad für die Approximation mit der  $t$ -Verteilung von  $U_n^{\pi,p,reg}$ :

$$\frac{\left( \sum_{i=1}^2 \widehat{\tau}_{i,p}^{reg^2} / n_i \right)^2}{\sum_{i=1}^2 (\widehat{\tau}_{i,p}^{reg^2} / n_i)^2 / (n_i - 1)},$$

mit

$$\begin{aligned} \widehat{\tau}_{i,p}^{reg^2} &= \frac{\widehat{\tau}_{i,p}^2}{n_i} + \beta^2 \frac{\widehat{\tau}_{i,p}^{(BL)^2}}{n_i} + 2\beta \frac{\widehat{\tau}_{i,p}^{(COV)^2}}{n_i}, \quad \text{wobei} \\ \widehat{\tau}_{i,p}^2 &= \frac{1}{(n_i - 1)} \sum_{k=1}^{n_i} \left( \Psi_{ik1} - \Psi_{ik2} - 2\Psi_{ik}^{(-j1)} + 2\Psi_{ik}^{(-j2)} \right. \\ &\quad \left. - \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2} + 2\bar{\Psi}_i^{(-j1)} - 2\bar{\Psi}_i^{(-j2)} \right)^2 \quad \text{mit } j \neq i. \end{aligned}$$

Entsprechend werden  $\widehat{\tau}_{i,p}^{(BL)^2}$  und  $\widehat{\tau}_{i,p}^{(COV)^2}$  gebildet.

Der Freiheitsgrad für die Approximation mit der  $t$ -Verteilung von  $U_n^{\delta,p,reg}$ :

$$\frac{\left( \sum_{i=1}^2 \widehat{v}_{i,p}^{reg^2} / n_i \right)^2}{\sum_{i=1}^2 (\widehat{v}_{i,p}^{reg^2} / n_i)^2 / (n_i - 1)},$$

mit

$$\begin{aligned}\widehat{v}_{i,p}^{reg^2} &= \frac{\widehat{v}_{i,p}^2}{n_i} + \beta^2 \frac{\widehat{v}_{i,p}^{(BL)^2}}{n_i} + 2\beta \frac{\widehat{v}_{i,p}^{(COV)^2}}{n_i}, \quad \text{wobei} \\ \widehat{v}_{i,p}^2 &= \frac{1}{(n_i - 1)} \sum_{k=1}^{n_i} \left( \Psi_{ik1} - \Psi_{ik2} + 2\Psi_{ik}^{(-j1)} - 2\Psi_{ik}^{(-j2)} \right. \\ &\quad \left. - \overline{\Psi}_{i.1} + \overline{\Psi}_{i.2} - 2\overline{\Psi}_{i.}^{(-j1)} + 2\overline{\Psi}_{i.}^{(-j2)} \right)^2 \quad \text{mit } j \neq i.\end{aligned}$$

Entsprechend werden  $\widehat{v}_{i,p}^{(BL)^2}$  und  $\widehat{v}_{i,p}^{(COV)^2}$  gebildet.

## 6.4 Auswertung der Asthma-Studie II

Es sind lediglich Baseline-Werte für die Asthma-Studie II vorhanden. In diesem Beispiel steht jeweils zusätzlich zu den  $FEV_1$ -Werten unter den beiden Behandlungen ein (*Run – In*)-Wert, der vor der Behandlung der ersten Periode erhoben wurde, und ein Wert am Ende der *Wash – Out*-Phase, vor der Behandlung der zweiten Periode, zur Verfügung.

### 6.4.1 Lineares Modell

Die Additivität der Effekte sowie ein Lokationsmodell sollen an dieser Stelle für die Daten angenommen werden können. Die Normalverteilungsannahme scheint jedoch nicht gerechtfertigt zu sein. Unter diesen Voraussetzungen entscheidet man sich für die semi-parametrische Vorgehensweise von KOCH (1972).

Die Vorgehensweise für diese Problemstellung und die Nullhypothese  $H_0^F : \mathbf{cF} = 0$  ist in Abschnitt 4.1 beschrieben worden. Anstatt der Ausgangsbeobachtungen  $X_{iks}$ ,  $i, s = 1, 2$  und  $k = 1, \dots, n_i$ , werden für diesen Test die adjustierten Werte  $X_{iks}^{adj} = X_{iks} - X_{iks}^{(BL)}$  benutzt. Siehe dazu Abschnitt 6.2. In Tabelle 6.2 sind die Ergebnisse der entsprechenden Statistiken für das lineare Modell zu entnehmen. Es ergibt sich folgendes: Auf einem Niveau von 5% kann kein Residualeffekt nachgewiesen werden. Aber man erhält bei einer Irrtumswahrscheinlichkeit von 5% einen Periodeneffekt und auch einen Cross-Over-Effekt bzw. verallgemeinerten Behandlungseffekt. Falls ein Nacheffekt a priori ausgeschlossen wurde, lässt sich schlussfolgern, dass sich die beiden Behandlungen unterscheiden. An dieser Stelle ist jedoch zu beachten, dass sich die Interpretation des Perioden- und Residualeffekts geändert hat. Unter Annahme von  $\tilde{\lambda} = 0$  (kein Residualeffekt) bleibt die Interpretation vom Behandlungseffekt erhalten. Dies wurde bereits in Abschnitt 6.2 diskutiert. Deshalb ist die Vergleichbarkeit zu entsprechenden Tests ohne Berücksichtigung von Baseline-Werten nicht uneingeschränkt sinnvoll.

### 6.4.2 Nichtparametrisches Modell

Bei Zugrundelegung des nichtparametrischen Modells soll *nicht* die Annahme getroffen werden können, dass die Baseline-Werte alle die gleiche Verteilung haben. Die Modellvoraussetzung (MV) sei demnach nicht erfüllt und die zu testenden Hypothesen müssen über die Effekte  $p_{is}^{reg}$  formuliert werden.

Tabelle 6.2:  $p$ -Werte der Asthma-Studie II - lineares Modell

	$p$ -Werte
$\lambda$	0.501
$\pi$	0.027
$\delta$	0.021

Tabelle 6.3:  $p$ -Werte der Asthma-Studie II - nichtparametrisches Modell

	$p$ -Werte
$\lambda$	0.546
$\pi$	0.028
$\delta$	0.044

Die Teststatistiken lassen sich in diesem Falle aus Abschnitt 6.3.5 entnehmen. Die Berechnung erfolgte mit Hilfe eines Computers. Die resultierenden  $p$ -Werte sind in Tabelle 6.3 festgehalten.

Danach erhält man ein ähnliches Bild wie unter Verwendung des Ansatzes zur Berücksichtigung von Baseline-Werten im linearen Modell. Auf einem Niveau von 5% kann kein Residualeffekt nachgewiesen werden. Aber man erhält bei einer Irrtumswahrscheinlichkeit von 5% einen Periodeneffekt und auch einen Cross-Over-Effekt bzw. verallgemeinerten Behandlungseffekt. Falls ein Nacheffekt a priori ausgeschlossen wurde, lässt sich schlussfolgern, dass sich die beiden Behandlungen unterscheiden. Auch hier muss angemerkt werden, dass die Interpretation der Effekte und Hypothesen nicht analog zur Interpretation im nichtparametrischen Modell (GM) ohne Hinzunahme von Baseline-Werten ist.



# Kapitel 7

## Simulation

In diesem Kapitel werden die asymptotischen Eigenschaften der in dieser Arbeit vorgestellten Teststatistiken durch Computersimulationen untersucht. Dazu wird in den Abschnitten jeweils die Vorgehensweise bei der Simulation erläutert und anschließend die einzelnen Simulationsergebnisse in Tabellenform angegeben. Bei der Simulation für die Power werden die Ergebnisse zusätzlich grafisch dargestellt.

### 7.1 Niveausimulation

Die Niveausimulation soll aufdecken, wie gut das Niveau bzw. die Irrtumswahrscheinlichkeit  $\alpha$  eines Tests von dem jeweiligen Test eingehalten wird. Ein Test zum Niveau  $\alpha$  wird als *konservativ* bezeichnet, wenn unter Hypothese das Niveau unterschritten wird, demnach weniger oft verworfen wird als notwendig. Bei Überschreitung des Niveaus hingegen spricht man von einem *anti-konservativen* (oder *liberalen*) Testverhalten.

Wenn nichts anderes im Text erwähnt ist, werden standardnormalverteilte Zufallsvariablen erzeugt. Die speziellen Abhängigkeiten des Cross-Over-Designs werden ebenfalls nachempfunden. Als Indikator für den Grad der Abhängigkeit dient der Parameter  $\rho$ . Je höher der betragsmäßige Wert, desto abhängiger die Variablen. Bei der Simulation werden zwei zunächst unabhängige Variablen  $X_1 \sim N(0, 1)$  und  $Y \sim N(0, 1)$  in Abhängigkeit gesetzt durch folgende Vorschrift:

$$X_2 = \frac{\rho \cdot X_1 + Y}{\sqrt{1 + \rho^2}}. \quad (7.1)$$

Damit hat man zwei abhängige Variablen, z.B. für zwei Zeitpunkte, erzeugt und  $X_2 \sim N(0, 1)$ . Dieses Verfahren wird für beide Gruppen unabhängig voneinander benutzt.

Die Niveausimulationen gliedern sich in fünf Teile. Als Erstes werden die neuen Statistiken *ohne Baseline-Werte* unter verschiedenen Bedingungen zu bestimmten Niveaus getestet. Danach wird die Niveaueinhaltung der neu entwickelten Teststatistiken mit der bei klassischen Verfahren für das lineare Modell verglichen. Anschließend werden Baseline-Werte in die Simulationen mit einbezogen. Zuletzt werden zusätzlich der Äquivalenztest und das Konfidenzintervallverfahren auf die Einhaltung des Niveaus überprüft. Für die Powersimulation wird diese Gliederung beibehalten, jedoch wird auf die Überprüfung des Konfidenzintervallverfahrens verzichtet.

Tabelle 7.1: Niveausimulation des Abhängigkeitsparameters bei einem Stichprobenumfang von 20/30,  $\alpha = 5\%$ , kein Effekt

		$H_0^F : \mathbf{cF} = 0$					
		Normalverteilung			$t$ -Verteilung		
$\rho$		$U_n^{\lambda,F}$	$U_n^{\pi,F}$	$U_n^{\delta,F}$	$U_n^{\lambda,F}$	$U_n^{\pi,F}$	$U_n^{\delta,F}$
-4		0.942	0.941	0.942	0.950	0.947	0.949
-3		0.945	0.941	0.943	0.952	0.949	0.949
-2		0.948	0.937	0.939	0.953	0.945	0.946
-1		0.941	0.941	0.942	0.948	0.947	0.949
0		0.946	0.945	0.942	0.953	0.952	0.949
1		0.942	0.941	0.940	0.949	0.947	0.947
2		0.947	0.950	0.940	0.954	0.956	0.946
3		0.947	0.944	0.945	0.953	0.953	0.952
4		0.942	0.945	0.945	0.950	0.950	0.952
		$H_0^p : \mathbf{cp} = 0$					
		Normalverteilung			$t$ -Verteilung		
$\rho$		$U_n^{\lambda,p}$	$U_n^{\pi,p}$	$U_n^{\delta,p}$	$U_n^{\lambda,p}$	$U_n^{\pi,p}$	$U_n^{\delta,p}$
-4		0.960	0.937	0.938	0.965	0.943	0.944
-3		0.959	0.937	0.937	0.964	0.944	0.944
-2		0.956	0.933	0.935	0.963	0.939	0.942
-1		0.946	0.937	0.939	0.952	0.944	0.945
0		0.945	0.943	0.940	0.952	0.951	0.947
1		0.939	0.948	0.949	0.946	0.956	0.955
2		0.945	0.965	0.959	0.950	0.970	0.965
3		0.945	0.970	0.967	0.950	0.975	0.971
4		0.939	0.971	0.973	0.947	0.975	0.977

### 7.1.1 Simulation ohne Baseline-Werte

Bei der Simulation ohne Berücksichtigung der Baseline-Werte wird zunächst einmal für ein konstantes Niveau (5%) und einen festen Stichprobenumfang von 20 bzw. 30 pro Gruppe die Robustheit der Teststatistiken bzgl. des Abhängigkeitsparameters  $\rho$  untersucht. Dazu wurde  $\rho$  (von -4 bis 4) variiert unter den o.g. Bedingungen mit 10000 Wiederholungen simuliert und die 95%-Werte in Tabelle 7.1 abgetragen. Es wurde kein Effekt zugrunde gelegt.

Unter der Hypothese  $H_0^F : \mathbf{cF} = 0$  bleiben die Teststatistiken unbeeindruckt vom Grad der Abhängigkeit, d.h. die Einhaltung des Niveaus ändert sich nicht mit schwächerer oder stärkerer Abhängigkeitsstruktur. Die für die Hypothese  $H_0^p : \mathbf{cp} = 0$  konzipierten Tests zeigen hingegen ein anderes Verhalten auf. Hier ist stets eine leichte Tendenz zu beobachten. Entweder ein Übergang von konservativ (bei negativem Abhängigkeitsparameter) zu antikonservativ (bei positivem Abhängigkeitsparameter) wie bei  $U_n^{\lambda,p}$  oder andersrum wie bei  $U_n^{\pi,p}$  oder  $U_n^{\delta,p}$ .

Sei der Abhängigkeitsparameter  $\rho$  auf 2 festgelegt. Es wurden die neuen Teststatistiken für das nichtparametrische Modell (GM) aus den Abschnitten 4.2.6 und 4.2.7 auf den Sei-



ten 39 bzw. 41 und folgende für den Perioden-, den Residual- und den Cross-Over-Effekt auf die Einhaltung des Niveaus überprüft. Dabei wurde stets mit der Normalverteilung und der  $t$ -Verteilung approximiert. Alle Simulationen sind mit den vorgegebenen Irrtumswahrscheinlichkeiten 1%, 5% und 10% durchgeführt. Dabei sind die Stichprobenumfänge 6/8, 9/11, 13/17, 20/30 und 45/50 untersucht worden. Die Hypothesen wurden simuliert für den Fall, dass kein Effekt, ein Residual-, ein Perioden- oder ein Behandlungseffekt vorliegt. Es wurden jeweils 10000 Simulationen durchgeführt.

Bei jeder Simulation wurden die geeigneten Teststatistiken berechnet und die resultierenden  $p$ -Werte mit dem zugehörigen Niveau verglichen und gemäß der Entscheidungsregel der Test verworfen oder nicht. Der Quotient aus der Summe der nicht verworfenen Tests und der Anzahl der Simulationen (hier immer 10000) ergibt jeweils die angegebene Zahl in den Tabellen. Um die Einhaltung des Niveau beurteilen zu können, muss diese mit dem vorgegebenen Wert  $1 - \alpha$  verglichen werden.

Aus Tabelle 7.2 ersieht man, zunächst einmal für die Nullhypothese  $H_0^F : \mathbf{cF} = 0$ , dass die Approximation der Normalverteilung bei den kleinen Stichprobenumfängen noch antikonservativ ist, für größere Stichprobenumfänge ab ca. durchschnittlich 15-20 Versuchseinheiten pro Gruppe jedoch sehr gut. Die Approximation mit der  $t$ -Verteilung hingegen stellt sich bereits bei einem durchschnittlichen Stichprobenumfang von 7 Versuchseinheiten pro Gruppe als sehr gut heraus. Bei Untersuchung der Teststatistiken unter der Hypothese  $H_0^p : \mathbf{cp} = 0$  mit Hilfe der Approximation mit der Normalverteilung zeigt sich folgendes: Die Teststatistik für den Residualeffekt verhält sich im Gegensatz zu den Statistiken der anderen Effekte antikonservativ und hält erst ab durchschnittlich ca. 25 Versuchseinheiten pro Gruppe das Niveau annehmbar gut ein. Die Teststatistiken für den Zeit- und Cross-Over-Effekt sind bei kleinen Stichprobenumfängen konservativ und erreichen auch erst bei sehr hohen Stichprobenumfängen ca. 45 Versuchseinheiten pro Gruppe ein gutes Niveau. Die Approximation mit der  $t$ -Verteilung ist durchweg schlechter. Lediglich beim Residualeffekt hat man einen leichten Vorteil. Dort erhält man bereits ab einem Stichprobenumfang von ca. 10 Versuchseinheiten pro Gruppe sinnvolle Ergebnisse.

Es ist weiterhin von Interesse, wie sich die jeweils anderen Teststatistiken bei speziell variiertes Effektstärke ändern. Diese Frage wird mit Hilfe der Tabellen 7.3, 7.5 und 7.7 beantwortet. Wenn dann ein Effekt vorhanden ist, so wird nachgesehen, wie sich bei Variation der Stichprobenumfänge das Niveau der entsprechend anderen Teststatistiken verhält. Dabei werden jeweils die Teststatistiken untersucht, die nichts direkt mit dem zugrunde gelegten Effekt zu tun haben. Bei Annahme eines Periodeneffekts werden demnach nur die Statistiken zur Aufdeckung des Residualeffekts und des Behandlungseffekts (Cross-Over) überprüft.

Zunächst wird in Tabelle 7.3 deutlich, dass unter variierendem Zeiteffekt die Statistiken stabil sind. Während jedoch unter  $H_0^F$  keine besondere Tendenz zu beobachten ist, zeigt sich für den Test auf den Cross-Over-Effekt unter  $H_0^p$ , dass  $U_n^{\delta,p}$  zunächst konservativ und dann mit größer werdender Effektstärke von  $\pi$  zunehmend antikonservativ wird.

Legt man den vorhandenen Zeiteffekt auf  $\pi = 0.4$  fest und variiert die Stichprobenumfänge, so wird in Tabelle 7.4 folgendes deutlich: Man kann ablesen, dass, zunächst einmal für die Nullhypothese  $H_0^F : \mathbf{cF} = 0$ , die Approximation der Normalverteilung bei kleinen Stichprobenumfängen noch antikonservativ ist und ab ca. 15-20 Versuchseinheiten pro Gruppe gute Ergebnisse liefert. Dies gilt sowohl für die Teststatistiken für einen Residualeffekt als auch für die Teststatistiken für einen Behandlungseffekt. Die Ap-

Tabelle 7.2: Niveausimulation bei verschiedenen Stichprobenumfängen, wenn kein Effekt vorhanden ist

		$H_0^F : \mathbf{cF} = 0$					
		Normalverteilung			$t$ -Verteilung		
$n_1/n_2$	$1 - \alpha$	$U_n^{\lambda,F}$	$U_n^{\pi,F}$	$U_n^{\delta,F}$	$U_n^{\lambda,F}$	$U_n^{\pi,F}$	$U_n^{\delta,F}$
6/8	0.90	0.873	0.872	0.865	0.900	0.903	0.898
	0.95	0.922	0.922	0.921	0.950	0.951	0.951
	0.99	0.971	0.974	0.973	0.988	0.992	0.991
9/11	0.90	0.883	0.878	0.877	0.901	0.901	0.894
	0.95	0.933	0.933	0.930	0.950	0.951	0.947
	0.99	0.979	0.982	0.978	0.987	0.991	0.989
13/17	0.90	0.887	0.891	0.888	0.901	0.902	0.902
	0.95	0.937	0.941	0.938	0.948	0.951	0.950
	0.99	0.982	0.984	0.984	0.989	0.991	0.990
20/30	0.90	0.893	0.892	0.895	0.900	0.899	0.904
	0.95	0.947	0.944	0.944	0.952	0.953	0.952
	0.99	0.986	0.986	0.988	0.990	0.990	0.992
45/50	0.90	0.899	0.895	0.892	0.902	0.898	0.995
	0.95	0.952	0.949	0.943	0.954	0.951	0.946
	0.99	0.989	0.991	0.986	0.990	0.992	0.987
		$H_0^p : \mathbf{cp} = 0$					
		Normalverteilung			$t$ -Verteilung		
$n_1/n_2$	$1 - \alpha$	$U_n^{\lambda,p}$	$U_n^{\pi,p}$	$U_n^{\delta,p}$	$U_n^{\lambda,p}$	$U_n^{\pi,p}$	$U_n^{\delta,p}$
6/8	0.90	0.863	0.931	0.926	0.888	0.953	0.951
	0.95	0.909	0.968	0.964	0.934	0.982	0.982
	0.99	0.957	0.991	0.992	0.975	0.998	0.999
9/11	0.90	0.877	0.929	0.923	0.895	0.945	0.940
	0.95	0.927	0.966	0.964	0.941	0.979	0.975
	0.99	0.971	0.995	0.993	0.981	0.998	0.997
13/17	0.90	0.885	0.922	0.922	0.896	0.934	0.934
	0.95	0.933	0.966	0.963	0.942	0.975	0.972
	0.99	0.977	0.995	0.994	0.984	0.997	0.997
20/30	0.90	0.890	0.914	0.920	0.898	0.920	0.927
	0.95	0.944	0.961	0.961	0.949	0.967	0.967
	0.99	0.983	0.993	0.994	0.988	0.995	0.996
45/50	0.90	0.898	0.909	0.905	0.902	0.912	0.907
	0.95	0.950	0.955	0.951	0.953	0.958	0.953
	0.99	0.988	0.993	0.990	0.990	0.994	0.991

Tabelle 7.3: Niveausimulation bei verschiedenen starken Perioden-Effekten

$\pi$	$H_0^F : cF = 0$				$H_0^P : cp = 0$			
	Normalverteilung		$t$ -Verteilung		Normalverteilung		$t$ -Verteilung	
	$U_n^{\lambda,F}$	$U_n^{\delta,F}$	$U_n^{\lambda,F}$	$U_n^{\delta,F}$	$U_n^{\lambda,P}$	$U_n^{\delta,P}$	$U_n^{\lambda,P}$	$U_n^{\delta,P}$
0	0.942	0.947	0.950	0.953	0.939	0.963	0.946	0.968
0.1	0.945	0.945	0.951	0.953	0.942	0.960	0.948	0.966
0.2	0.940	0.944	0.947	0.950	0.937	0.958	0.944	0.962
0.3	0.942	0.944	0.949	0.950	0.940	0.954	0.945	0.959
0.4	0.944	0.945	0.952	0.953	0.940	0.952	0.947	0.958
0.5	0.942	0.948	0.949	0.954	0.939	0.950	0.946	0.957
0.6	0.941	0.944	0.948	0.950	0.938	0.945	0.945	0.951
0.7	0.944	0.940	0.952	0.946	0.943	0.939	0.949	0.945
0.8	0.943	0.944	0.950	0.952	0.940	0.942	0.947	0.951
0.9	0.941	0.942	0.948	0.948	0.939	0.940	0.946	0.946
1	0.944	0.942	0.950	0.950	0.943	0.942	0.948	0.949

Tabelle 7.4: Niveausimulation bei verschiedenen Stichprobenumfängen, wenn ein Perioden-Effekt vorhanden ist

$n_1/n_2$	$1 - \alpha$	$H_0^F : cF = 0$				$H_0^P : cp = 0$			
		Normalvertlg		$t$ -Verteilung		Normalvertlg		$t$ -Verteilung	
		$U_n^{\lambda,F}$	$U_n^{\delta,F}$	$U_n^{\lambda,F}$	$U_n^{\delta,F}$	$U_n^{\lambda,P}$	$U_n^{\delta,P}$	$U_n^{\lambda,P}$	$U_n^{\delta,P}$
6/8	0.90	0.873	0.873	0.900	0.904	0.867	0.889	0.895	0.917
	0.95	0.920	0.925	0.948	0.953	0.916	0.936	0.941	0.959
	0.99	0.972	0.975	0.988	0.991	0.963	0.977	0.980	0.991
9/11	0.90	0.876	0.878	0.895	0.897	0.876	0.895	0.893	0.913
	0.95	0.938	0.932	0.945	0.951	0.926	0.943	0.943	0.958
	0.99	0.975	0.981	0.987	0.990	0.971	0.984	0.982	0.992
13/17	0.90	0.887	0.884	0.889	0.898	0.891	0.897	0.901	0.910
	0.95	0.938	0.936	0.947	0.947	0.938	0.947	0.951	0.956
	0.99	0.984	0.983	0.989	0.989	0.983	0.986	0.988	0.991
20/30	0.90	0.892	0.891	0.901	0.898	0.895	0.904	0.902	0.910
	0.95	0.944	0.945	0.950	0.952	0.941	0.949	0.947	0.956
	0.99	0.987	0.987	0.990	0.991	0.983	0.990	0.987	0.993
45/50	0.90	0.896	0.898	0.900	0.902	0.899	0.904	0.902	0.907
	0.95	0.944	0.947	0.948	0.951	0.946	0.948	0.948	0.951
	0.99	0.987	0.989	0.989	0.990	0.987	0.990	0.988	0.991

Tabelle 7.5: Niveausimulation bei verschiedenen starken Behandlungs-Effekten

$\Phi$	$H_0^F : \mathbf{cF} = 0$				$H_0^p : \mathbf{cp} = 0$			
	Normalverteilung		$t$ -Verteilung		Normalverteilung		$t$ -Verteilung	
	$U_n^{\lambda,F}$	$U_n^{\pi,F}$	$U_n^{\lambda,F}$	$U_n^{\pi,F}$	$U_n^{\lambda,p}$	$U_n^{\pi,p}$	$U_n^{\lambda,p}$	$U_n^{\pi,p}$
0	0.943	0.942	0.951	0.949	0.940	0.958	0.947	0.963
0.1	0.946	0.946	0.952	0.953	0.942	0.961	0.948	0.967
0.2	0.944	0.945	0.951	0.953	0.942	0.960	0.947	0.966
0.3	0.942	0.946	0.948	0.953	0.939	0.956	0.944	0.961
0.4	0.943	0.944	0.948	0.951	0.939	0.951	0.946	0.956
0.5	0.942	0.941	0.949	0.948	0.939	0.944	0.946	0.950
0.6	0.942	0.940	0.951	0.948	0.940	0.942	0.947	0.948
0.7	0.945	0.945	0.952	0.951	0.943	0.945	0.950	0.952
0.8	0.943	0.944	0.950	0.951	0.941	0.943	0.948	0.948
0.9	0.943	0.944	0.951	0.952	0.941	0.943	0.948	0.950
1	0.944	0.943	0.952	0.950	0.943	0.943	0.950	0.949

proximation mit der  $t$ -Verteilung stellt sich in beiden Fällen bereits auch bei einem durchschnittlichen Stichprobenumfang von 7 Versuchseinheiten pro Gruppe als sehr gut heraus. Bei Untersuchung der Teststatistiken unter der Hypothese  $H_0^p : \mathbf{cp} = 0$  verhalten sich die Teststatistiken bei Approximation mit der Normalverteilung für beide getesteten Effekte leicht antikonservativ. Die Teststatistik auf einen Behandlungseffekt kann ab einem durchschnittlichen Stichprobenumfang von ca. 10 in der Praxis angewendet werden, während sich für die Statistik zur Aufdeckung eines Residualeffekt wenigstens 20 Versuchseinheiten pro Gruppe empfehlen. Zur Approximation mit der  $t$ -Verteilung: Die Teststatistik für den Cross-Over-Effekt ist bei kleinen Stichprobenumfängen konservativ und erreicht bei einem Stichprobenumfang von ca. 15 Versuchseinheiten pro Gruppe ein annehmbares Niveau. Das Niveau beim Test auf einen Residual-Effekt hingegen wird von Anfang an, d.h. ab ca. 7 Versuchseinheiten pro Gruppe, gut eingehalten.

In Tabelle 7.5 wird deutlich, dass auch unter variierendem Behandlungseffekt die Statistiken stabil sind. Es ergibt sich das gleiche Bild wie unter variierendem Zeiteffekt. Die leichte Tendenz aus Tabelle 7.3 wiederholt sich, d.h.  $U_n^{\pi,p}$  ist zunächst konservativ und mit zunehmender Effektstärke immer antikonservativer.

Ist ein Behandlungseffekt vorhanden ( $\Phi = 0.4$ ), so sind in Tabelle 7.6 die Simulationsergebnisse bei variierendem Stichprobenumfang festgehalten. Es lässt sich im Vergleich zu Tabelle 7.2, wo kein Effekt zugrunde gelegt wurde, kein nennenswerter Unterschied der Statistiken unter  $H_0^F : \mathbf{cF} = 0$  für einen Zeit- und Residualeffekt feststellen. Die Approximation mit der  $t$ -Verteilung ist bereits bei kleinstmöglichen Stichprobenumfängen sehr gut und die Normalverteilung greift spätestens ab einem durchschnittlichen Stichprobenumfang von ca. 15-20. Unter  $H_0^p$  sind die Resultate ein wenig abweichend. Die Approximation mit der Normalverteilung greift beim Test auf einen Residualeffekt eher langsam, so dass man erst ab einem Stichprobenumfang von wenigstens 40 pro Gruppe vernünftige Ergebnisse erzielt. Der Test auf einen Zeiteffekt hingegen hat bereits ab ca. 10 Versuchseinheiten pro Gruppe die gewünschten Eigenschaften bzgl. der Niveaueinhaltung. Da macht es auch nichts, dass bei dieser Statistik die Approximation mit der

Tabelle 7.6: Niveausimulation bei verschiedenen Stichprobenumfängen, wenn ein Behandlungs-Effekt vorhanden ist

		$H_0^F : \mathbf{cF} = 0$				$H_0^P : \mathbf{cp} = 0$			
		Normalvertlg		$t$ -Verteilung		Normalvertlg		$t$ -Verteilung	
$n_1/n_2$	$1 - \alpha$	$U_n^{\lambda,F}$	$U_n^{\pi,F}$	$U_n^{\lambda,F}$	$U_n^{\pi,F}$	$U_n^{\lambda,p}$	$U_n^{\pi,p}$	$U_n^{\lambda,p}$	$U_n^{\pi,p}$
6/8	0.90	0.875	0.869	0.904	0.901	0.867	0.889	0.891	0.918
	0.95	0.926	0.923	0.950	0.952	0.913	0.938	0.937	0.961
	0.99	0.971	0.976	0.988	0.992	0.958	0.979	0.977	0.993
9/11	0.90	0.887	0.878	0.906	0.900	0.881	0.894	0.898	0.912
	0.95	0.934	0.929	0.951	0.948	0.928	0.943	0.943	0.959
	0.99	0.981	0.980	0.990	0.991	0.973	0.985	0.984	0.993
13/17	0.90	0.887	0.878	0.906	0.900	0.881	0.894	0.896	0.911
	0.95	0.940	0.938	0.950	0.950	0.933	0.947	0.946	0.957
	0.99	0.984	0.982	0.989	0.989	0.979	0.987	0.985	0.992
20/30	0.90	0.883	0.891	0.896	0.898	0.887	0.899	0.893	0.908
	0.95	0.940	0.940	0.947	0.947	0.936	0.947	0.944	0.954
	0.99	0.985	0.986	0.988	0.988	0.982	0.988	0.986	0.891
45/50	0.90	0.901	0.895	0.904	0.898	0.901	0.899	0.904	0.903
	0.95	0.946	0.947	0.948	0.949	0.945	0.950	0.947	0.951
	0.99	0.988	0.990	0.989	0.991	0.987	0.991	0.989	0.993

$t$ -Verteilung eher langsam greift und erst ab einem Stichprobenumfang von ca. 25 pro Gruppe als sinnvoll einsetzbar erscheint. Hingegen ist  $U_n^{\lambda,p}$  bereits ab durchschnittlich 10 Versuchseinheiten pro Gruppe gut nutzbar.

In Tabelle 7.7 wird deutlich, dass auch unter variierendem Residualeffekt die Statistiken stabil sind. Es ergibt sich das gleiche Bild wie unter variierendem Zeiteffekt und Behandlungseffekt, d.h. mit wachsender Effektstärke ein leichter Trend von konservativ zu antikonservativ bei den Teststatistiken  $U_n^{\pi,p}$ .

Simulationen, deren Ergebnisse in Tabelle 7.8 dargestellt sind, wurde ein fester Residualeffekt von  $\lambda = 0.4$  zugrunde gelegt. Um die Teststatistiken für die Nullhypothese  $H_0^F : \mathbf{cF} = 0$  anwenden zu können, braucht man unter Verwendung der Approximation mit der  $t$ -Verteilung lediglich 7 Versuchseinheiten pro Gruppe. Die Approximation mit der Normalverteilung ist erst ab einem Stichprobenumfang von ca. 20 Versuchseinheiten pro Gruppe akzeptabel. Teststatistiken für die Nullhypothese  $H_0^P : \mathbf{cp} = 0$  können erst ab einem Stichprobenumfang von wenigstens 20 pro Gruppe eingesetzt werden, wobei hier die Approximation mit der Normalverteilung bessere Ergebnisse bringt als die Approximation mit der  $t$ -Verteilung.

### 7.1.2 Simulationsvergleiche mit den Verfahren des linearen Modells

In diesem Abschnitt geht es um den Vergleich der klassischen Verfahren von Lehmaner (parametrisch) und Koch (semiparametrisch) unter Annahme des linearen Modells mit den in dieser Arbeit neu entwickelten Verfahren für das nichtparametrische Modell (GM). Dabei wird der Schwerpunkt derart gelegt, was passiert, wenn die Voraussetzungen der

Tabelle 7.7: Niveausimulation bei verschieden starken Residual-Effekten

	$H_0^F : \mathbf{cF} = 0$		$H_0^p : \mathbf{cp} = 0$	
	Normalverteilung	$t$ -Verteilung	Normalverteilung	$t$ -Verteilung
$\lambda$	$U_n^{\pi,F}$	$U_n^{\pi,F}$	$U_n^{\pi,p}$	$U_n^{\pi,p}$
0	0.947	0.953	0.963	0.968
0.1	0.945	0.953	0.962	0.968
0.2	0.942	0.950	0.960	0.965
0.3	0.945	0.952	0.959	0.966
0.4	0.939	0.947	0.952	0.959
0.5	0.945	0.951	0.956	0.963
0.6	0.944	0.951	0.954	0.961
0.7	0.942	0.949	0.950	0.957
0.8	0.944	0.952	0.953	0.958
0.9	0.941	0.949	0.945	0.953
1	0.944	0.950	0.947	0.953

Tabelle 7.8: Niveausimulation bei verschiedenen Stichprobenumfängen, wenn ein Residual-Effekt vorhanden ist

		$H_0^F : \mathbf{cF} = 0$		$H_0^p : \mathbf{cp} = 0$	
		Normalverteilung	$t$ -Verteilung	Normalverteilung	$t$ -Verteilung
$n_1/n_2$	$1 - \alpha$	$U_n^{\pi,F}$	$U_n^{\pi,F}$	$U_n^{\pi,p}$	$U_n^{\pi,p}$
6/8	0.90	0.876	0.905	0.918	0.941
	0.95	0.922	0.952	0.959	0.977
	0.99	0.974	0.991	0.989	0.996
9/11	0.90	0.878	0.900	0.917	0.933
	0.95	0.932	0.949	0.960	0.972
	0.99	0.980	0.990	0.991	0.997
13/17	0.90	0.886	0.889	0.913	0.927
	0.95	0.941	0.952	0.960	0.970
	0.99	0.985	0.990	0.992	0.996
20/30	0.90	0.890	0.897	0.907	0.917
	0.95	0.941	0.948	0.955	0.962
	0.99	0.986	0.990	0.992	0.994
45/50	0.90	0.895	0.898	0.904	0.908
	0.95	0.948	0.951	0.955	0.958
	0.99	0.989	0.990	0.991	0.993

Tabelle 7.9: Niveausimulation der verschiedenen Statistiken bei Normalverteilung, kein Effekt

		Residualeffekt			
$n_1/n_2$	$U_n^{\lambda,F}$	$U_n^{\lambda,p}$	$U_{LEH}^{\lambda}$	$U_{KOCH}^{\lambda}$	
6/8	0.950	0.938	0.945	0.954	
9/11	0.950	0.941	0.948	0.954	
13/17	0.948	0.944	0.949	0.949	
20/30	0.948	0.946	0.952	0.952	
45/50	0.952	0.950	0.950	0.948	
		Periodeneffekt			
$n_1/n_2$	$U_n^{\pi,F}$	$U_n^{\pi,p}$	$U_{LEH}^{\pi}$	$U_{KOCH}^{\pi}$	
6/8	0.951	0.967	0.949	0.958	
9/11	0.951	0.966	0.949	0.952	
13/17	0.951	0.966	0.951	0.953	
20/30	0.953	0.961	0.947	0.948	
45/50	0.951	0.955	0.951	0.952	
		Cross-Over-Effekt			
$n_1/n_2$	$U_n^{\delta,F}$	$U_n^{\delta,p}$	$U_{LEH}^{\delta}$	$U_{KOCH}^{\delta}$	
6/8	0.951	0.968	0.946	0.956	
9/11	0.947	0.966	0.950	0.953	
13/17	0.950	0.963	0.948	0.951	
20/30	0.952	0.958	0.950	0.948	
45/50	0.946	0.951	0.948	0.948	

Verteilungsannahmen verletzt sind. Dabei wird bei den neu entwickelten Statistiken jeweils automatisch die beste Approximation (mit  $t$ - oder Normalverteilung) gewählt. Um die Übersichtlichkeit zu wahren und den Vergleich in den Vordergrund zu stellen, sei das Niveau einheitlich auf  $\alpha = 5\%$  festgelegt.

Tabelle 7.9 zeigt den Vergleich der Teststatistiken, wenn sowohl die Normalverteilung der Daten als auch das additive Modell der Effekte zugrunde gelegt werden kann. Ferner sei kein Effekt vorhanden.

Da für alle Teststatistiken sämtliche Voraussetzungen erfüllt sind, ist es nicht überraschend, dass sich die Teststatistiken  $U_n^F$ ,  $U_{LEH}$  und  $U_{KOCH}$  bezüglich der Niveaueinhaltung kaum unterscheiden. Generell schlechter verhalten sich die Teststatistiken unter  $H_0^p$ .

Im Folgenden werden die Voraussetzung insofern abgewandelt, dass (siehe Tabelle 7.10) die Daten nicht als normalverteilt angenommen werden, sondern eine Gamma-Verteilung zugrunde gelegt wird. Zweitens folgen die Daten einer linksschiefen diskreten Verteilung mit einer Ordinalskala der Ordnung 5 (siehe Tabelle 7.11). Bei Verletzung der Normalverteilungsannahme erzielt man keine veränderte Reaktion insbesondere der Lehmaner-Statistik  $U_{LEH}^{\lambda}$ . Alle Statistiken zeigen sich auch bei diskreter Verteilung der Zufallsvariablen stabil. Es gibt keine nennenswerten Unterschiede zwischen  $U_n^{\lambda,F}$  und den Statistiken von Lehmaner und Koch.

Tabelle 7.10: Niveausimulation der verschiedenen Statistiken bei Gammaverteilung, kein Effekt

		Residualeffekt			
$n_1/n_2$	$U_n^{\lambda,F}$	$U_n^{\lambda,p}$	$U_{LEH}^\lambda$	$U_{KOCH}^\lambda$	
6/8	0.946	0.936	0.945	0.954	
9/11	0.951	0.944	0.948	0.954	
13/17	0.947	0.944	0.949	0.949	
20/30	0.946	0.952	0.952	0.952	
45/50	0.950	0.948	0.948	0.950	
		Periodeneffekt			
$n_1/n_2$	$U_n^{\pi,F}$	$U_n^{\pi,p}$	$U_{LEH}^\pi$	$U_{KOCH}^\pi$	
6/8	0.948	0.960	0.949	0.958	
9/11	0.950	0.954	0.949	0.952	
13/17	0.949	0.953	0.951	0.953	
20/30	0.946	0.951	0.947	0.948	
45/50	0.949	0.950	0.951	0.952	
		Cross-Over-Effekt			
$n_1/n_2$	$U_n^{\delta,F}$	$U_n^{\delta,p}$	$U_{LEH}^\delta$	$U_{KOCH}^\delta$	
6/8	0.951	0.955	0.946	0.956	
9/11	0.948	0.953	0.950	0.953	
13/17	0.952	0.952	0.948	0.951	
20/30	0.951	0.954	0.950	0.948	
45/50	0.947	0.951	0.948	0.948	

### 7.1.3 Simulation mit Baseline-Werten

In diesem Abschnitt werden die Ergebnisse für Verfahren unter Berücksichtigung von Baseline-Werten im linearen Modell und nichtparametrischen Modell beschrieben. Es werden die in Abschnitt 6.2 beschriebenen Statistiken für zwei Baseline-Werte im linearen Modell unter verschiedenen Bedingungen simuliert. Ebenso geht es um die Niveaueinhaltung der Statistiken der Kovarianzanalyse (nichtparametrischer Ansatz) aus Abschnitt 6.3. Der Abhängigkeitsparameter wurde bei diesen Simulationen wiederum auf  $\rho = 2$  festgelegt. Es wurden die üblichen Stichprobenumfangskombinationen zu den Irrtumswahrscheinlichkeiten 1%, 5% und 10% untersucht. Die Baseline-Werte werden wie die fiktiven Messwerte zufällig erzeugt und durch Erweiterung der eingangs erläuterten Vorgehensweise (7.1) zueinander in Beziehung gesetzt.

Im linearen Modell wurde wie in Abschnitt 6.2 beschrieben verfahren: Als Ausgangswerte werden die adjustierten Werte  $X_{iks}^{adj}$  benutzt, d.h. von jedem Messwert unter Behandlung wurde der dazugehörige vorher erhobene Baseline-Wert abgezogen:

$$X_{iks}^{adj} = X_{iks} - X_{iks}^{(BL)}.$$

Es wird an dieser Stelle die Niveaueinhaltung überprüft, wenn auf diese Werte die Verfahren basierend auf dem linearen Modell nach LEHMACHER (1987) bzw. semiparametrisch KOCH (1972), angewendet werden. Damit die Vergleichbarkeit zu den Simulationen ohne Baseline-Werte gewährt bleibt, wurden die Parameter analog gewählt. Das



Tabelle 7.11: Niveausimulation der verschiedenen Statistiken bei diskreter Verteilung, kein Effekt

		Residualeffekt			
$n_1/n_2$	$U_n^{\lambda,F}$	$U_n^{\lambda,p}$	$U_{LEH}^{\lambda}$	$U_{KOCH}^{\lambda}$	
6/8	0.947	0.938	0.953	0.953	
9/11	0.948	0.944	0.951	0.953	
13/17	0.950	0.945	0.950	0.951	
20/30	0.949	0.947	0.947	0.948	
45/50	0.950	0.948	0.949	0.948	
		Periodeneffekt			
$n_1/n_2$	$U_n^{\pi,F}$	$U_n^{\pi,p}$	$U_{LEH}^{\pi}$	$U_{KOCH}^{\pi}$	
6/8	0.955	0.954	0.955	0.962	
9/11	0.953	0.956	0.956	0.957	
13/17	0.952	0.954	0.952	0.953	
20/30	0.948	0.951	0.946	0.950	
45/50	0.949	0.950	0.947	0.949	
		Cross-Over-Effekt			
$n_1/n_2$	$U_n^{\delta,F}$	$U_n^{\delta,p}$	$U_{LEH}^{\delta}$	$U_{KOCH}^{\delta}$	
6/8	0.957	0.958	0.952	0.959	
9/11	0.954	0.956	0.954	0.956	
13/17	0.949	0.953	0.951	0.952	
20/30	0.946	0.950	0.948	0.949	
45/50	0.949	0.951	0.949	0.947	

bedeutet der Abhängigkeitsparameter ist  $\rho = 2$  und es wurden dieselben Stichprobenumfangskombinationen wie im Abschnitt zuvor untersucht. Es werden die verfügbaren Teststatistiken für Residual-, Perioden- und Cross-Over Effekt betrachtet. Da kein Nacheffekt angenommen wird, sind die Tests auf den Cross-Over-Effekt im Folgenden der Test auf einen reinen Behandlungseffekt. Im linearen Modell ist dieser im Gegensatz zu den anderen betrachteten Effekten in seiner ursprünglichen Interpretation erhalten geblieben. Aus Tabelle 7.12 folgt, dass bei der zuvor erklärten Methode zur Berücksichtigung von zwei Baseline-Werten im linearen Modell ab bereits 7 Versuchseinheiten pro Gruppe mit vernünftigen Ergebnissen gerechnet werden kann. An dieser Stelle muss natürlich erwähnt sein, dass diese Simulation so angelegt ist, dass alle Voraussetzungen der Verfahren von Lehmaner und Koch erfüllt sind. Das bedeutet, dass unter Additivität der Parameter und Normalverteilung simuliert wurde.

Berücksichtigt man die beiden Baseline-Werte mit Hilfe der Kovarianzanalyse im nicht-parametrischen Modell, so erhält man die in Tabelle 7.13 niedergeschriebenen Resultate. Es wurden hierbei der Übersichtlichkeit zuliebe nicht die Teststatistiken mit den verschiedenen Approximationen mit Normalverteilung oder  $t$ -Verteilung getrennt aufgeführt, sondern jeweils die abgetragen, die das Niveau am besten einhält.

Bei zusätzlicher Berücksichtigung von Baseline-Werten im nichtparametrischen Modell mit Kovariablen (GMK) sind die Statistiken grundsätzlich ein wenig langsamer in ihrer Konvergenz gegen das vorgegebene Niveau als ohne die Berücksichtigung von Baseline-

Tabelle 7.12: Niveausimulation für Verfahren zur Berücksichtigung von Baseline-Werten im linearen Modell

$n_1/n_2$	$1 - \alpha$	$U_{LEH}^\lambda$	$U_{KOCH}^\lambda$	$U_{LEH}^\pi$	$U_{KOCH}^\pi$	$U_{LEH}^\delta$	$U_{KOCH}^\delta$
6/8	0.90	0.898	0.890	0.901	0.888	0.904	0.895
	0.95	0.951	0.956	0.950	0.956	0.951	0.958
	0.99	0.990	0.992	0.991	0.992	0.990	0.991
9/11	0.90	0.900	0.903	0.898	0.903	0.896	0.879
	0.95	0.950	0.954	0.946	0.952	0.950	0.952
	0.99	0.990	0.993	0.990	0.993	0.991	0.994
13/17	0.90	0.901	0.898	0.898	0.893	0.904	0.897
	0.95	0.950	0.954	0.946	0.950	0.951	0.954
	0.99	0.991	0.993	0.988	0.990	0.991	0.993
20/30	0.90	0.904	0.905	0.903	0.906	0.899	0.898
	0.95	0.951	0.953	0.955	0.954	0.949	0.947
	0.99	0.990	0.992	0.990	0.991	0.988	0.989
45/50	0.90	0.899	0.896	0.899	0.898	0.899	0.897
	0.95	0.946	0.949	0.950	0.950	0.950	0.949
	0.99	0.989	0.991	0.991	0.992	0.990	0.990

Tabelle 7.13: Niveausimulation für Verfahren zur Berücksichtigung von Baseline-Werten im nichtparametrischen Modell

$n_1/n_2$	$1 - \alpha$	$U_n^{F,\lambda,reg}$	$U_n^{p,\lambda,reg}$	$U_n^{F,\pi,reg}$	$U_n^{p,\pi,reg}$	$U_n^{F,\delta,reg}$	$U_n^{p,\delta,reg}$
6/8	0.90	0.873	0.881	0.869	0.913	0.906	0.913
	0.95	0.929	0.933	0.928	0.956	0.954	0.954
	0.99	0.979	0.974	0.983	0.992	0.993	0.983
9/11	0.90	0.883	0.895	0.880	0.911	0.903	0.909
	0.95	0.937	0.945	0.936	0.954	0.951	0.953
	0.99	0.986	0.988	0.988	0.987	0.991	0.985
13/17	0.90	0.889	0.900	0.891	0.908	0.901	0.908
	0.95	0.937	0.945	0.947	0.952	0.953	0.952
	0.99	0.988	0.990	0.990	0.988	0.992	0.992
20/30	0.90	0.894	0.903	0.894	0.905	0.898	0.906
	0.95	0.944	0.952	0.945	0.950	0.951	0.951
	0.99	0.990	0.991	0.989	0.990	0.990	0.992
45/50	0.90	0.896	0.902	0.897	0.903	0.898	0.903
	0.95	0.947	0.952	0.948	0.952	0.949	0.951
	0.99	0.990	0.991	0.990	0.991	0.990	0.991

Werten. Die Statistiken unter  $H_0^F : \mathbf{cF} = 0$  für den Residualeffekt sind ab einem Stichprobenumfang von wenigstens 15 Versuchseinheiten pro Gruppe einsatzbereit, während unter der Nullhypothese über den bereinigten relativen Effekt  $H_0^{reg} : \mathbf{cp}^{reg} = 0$  sogar durchschnittlich 10 Versuchseinheiten pro Gruppe ausreichend sind. Ähnlich verhält es sich bei den Tests auf die beiden anderen Effekte. Unabhängig von der Hypothesenstellung benötigt man für eine gute Niveaueinhaltung ca. 15-20 Versuchseinheiten pro Gruppe, wenn man auf einen Periodeneffekt testet. Test auf Cross-Over-Effekte, bzw. in diesem Fall reine Behandlungseffekte, können bei guter Niveaueinhaltung ab wenigstens durchschnittlich 10 Versuchseinheiten pro Gruppe verwendet werden. Ohne Berücksichtigung weiterer Kriterien würde man aufgrund dieser Untersuchung unter Umständen von einer zusätzlichen Erhebung von Baseline-Werten im Cross-Over-Verfahren absehen. Denn wozu sollten mehr Messungen für weniger Niveaueinhaltung gezahlt werden. Jedoch sollte unter allen Umständen auch die Macht des Tests in die Beurteilung eines Vergleichs mit einfließen. Dazu muss Abschnitt 7.2.3 beachtet werden.

#### 7.1.4 Simulation des Äquivalenzverfahrens

Der in Abschnitt 5.5 auf Seite 55 vorgestellte Äquivalenztest soll hier auf die Einhaltung des Niveaus geprüft werden. Hier werden auch höhere Stichprobenumfänge untersucht, da die Konvergenz etwas langsamer als bei den zuvor betrachteten Tests ist. Die Simulationen werden jeweils für die Effektstärken 0.4 durchgeführt. Als kritische Werte  $\theta_1$  bzw.  $\theta_2$  wird der Rand der Hypothese gewählt, d.h.  $\mathbf{cp}$  für die entsprechenden Parameter berechnet und  $\theta_1$  oder  $\theta_2$ , je nach Vorzeichen, gesetzt. Lag das jeweils aus den Daten geschätzte Konfidenzintervall innerhalb der vorgegebenen Grenzen  $[\theta_1, \theta_2]$ , so wurde die Nullhypothese verworfen. In der Tabelle ist zu den entsprechenden Niveaus der Anteil der nicht verworfenen Hypothesen im Verhältnis zu allen Schleifendurchläufen (10000) angegeben, d.h. die tatsächlichen Werte für  $1 - \alpha$ . Es ergeben sich die in Tabelle 7.14 festgehaltenen Werte. Die Spalte  $U_n^\lambda$  enthält den entsprechenden Quotienten, wenn  $\lambda = 0.4$ , die Spalte  $U_n^\pi$ , wenn  $\pi = 0.4$  etc. Die Niveaueinhaltung ist beim Residualeffekt am schlechtesten. Das Verfahren ist in diesem Fall unter Hypothese stark konservativ und konvergiert sehr langsam, so dass man diese Verfahren nicht unter einem Stichprobenumfang von wenigstens 50 pro Gruppe beginnen sollte. Gute Ergebnisse erzielt man aber erst ab ungefähr 70 pro Gruppe. Die anderen Statistiken für den Perioden- und Cross-Over-Effekt sind bereits ab ungefähr 30 (Periodeneffekt) bzw. 50 (Cross-Over-Effekt) in der Praxis einsetzbar.

#### 7.1.5 Simulation des Konfidenzintervallverfahrens

Um Faktoren wie ungleiche und niedrige Stichprobenumfänge auszuschließen wurde ein relativ hoher Stichprobenumfang von 50 pro Gruppe festgelegt. Die Simulationen wurden für verschiedene Kombinationen des Niveaus des Residualeffekts  $\alpha_\lambda$  und des Cross-Over-Effekts  $\alpha_\delta$  durchgeführt. Die Ergebnisse befinden sich in Tabelle 7.15. In Klammern hinter dem simulierten Niveau steht jeweils das vorgegebene Niveau. Unter Zuhilfenahme der Abbildung 7.1 zeigt sich deutlich die Konservativität des Verfahrens. Überall wird das Niveau unterschritten. Je höher die einzelnen Niveaus, also das Niveau für den Residualeffekt bzw. den Cross-Over-Effekt, vorausgesetzt werden, desto gravierender ist die Abweichung vom vorgegebenen Niveau. Dies verdeutlicht die Lage der Ebenen zueinan-

Tabelle 7.14: Simulation des Niveaus beim Äquivalenzverfahren unter Berücksichtigung verschiedener Stichprobenumfänge

$n_1/n_2$	$1 - \alpha$	$U_n^{\lambda,p}$	$U_n^{\pi,p}$	$U_n^{\delta,p}$
6/8	0.90	1.000	0.931	0.930
	0.95	1.000	0.960	0.959
	0.99	1.000	0.988	0.986
9/11	0.90	1.000	0.929	0.925
	0.95	1.000	0.960	0.959
	0.99	1.000	0.988	0.987
13/17	0.90	1.000	0.922	0.924
	0.95	1.000	0.958	0.959
	0.99	1.000	0.989	0.988
20/30	0.90	0.980	0.911	0.921
	0.95	1.000	0.952	0.958
	0.99	1.000	0.987	0.989
30/40	0.90	0.932	0.907	0.909
	0.95	0.999	0.953	0.953
	0.99	1.000	0.987	0.987
45/50	0.90	0.914	0.908	0.913
	0.95	0.969	0.957	0.955
	0.99	1.000	0.990	0.989
60/70	0.90	0.911	0.908	0.911
	0.95	0.955	0.957	0.953
	0.99	1.000	0.989	0.987
90/110	0.90	0.904	0.908	0.906
	0.95	0.952	0.957	0.952
	0.99	0.993	0.989	0.990

der. Interessant ist, dass dies nicht gleichmäßig für die beiden vorgegebenen Niveaus geschieht, sondern der Einfluss des Niveaus des Cross-Over-Effekts geringer ist. Demnach trägt dieser Faktor den größeren Anteil an der konservativen Charakteristik des Verfahrens.

## 7.2 Powersimulationen

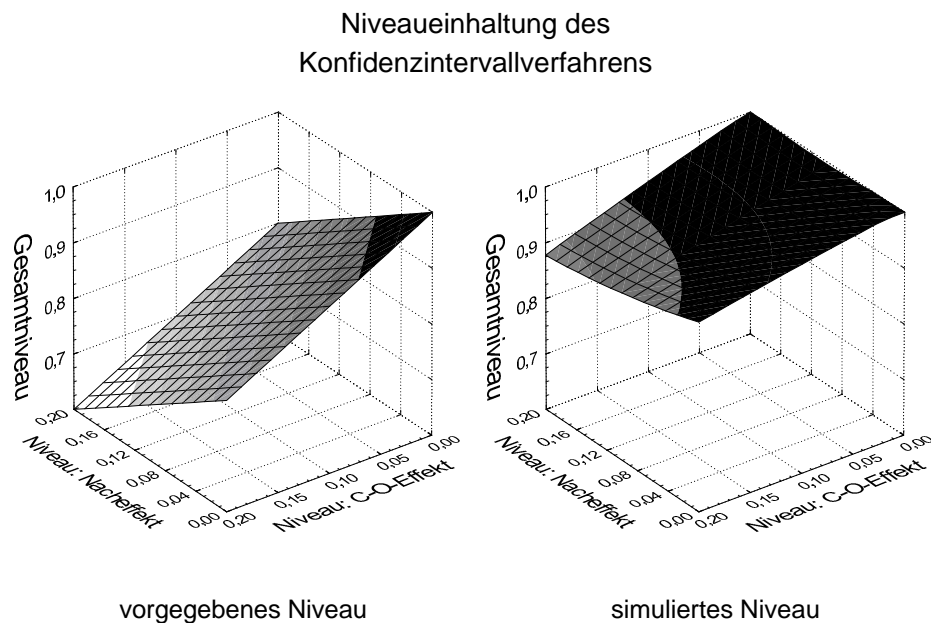
Die Power soll aufdecken, mit welcher Wahrscheinlichkeit die Alternativhypothese aufgedeckt wird, wenn sie gilt. Das bedeutet, dass es hier um die Wahrscheinlichkeit geht, mit der eine falsche Hypothese auch als solche erkannt wird. Die Power berechnet sich aus dem Fehler 2. Art  $\beta$  derart, dass gilt:

$$power = 1 - \beta.$$

Die zunehmende Verletzung der Hypothese wird simuliert, indem der entsprechende Effekt immer stärker wird, d.h. von 0 (unter Hypothese) bis zu dem Wert bei dem die Power

Tabelle 7.15: Niveausimulation für das Konfidenzintervallverfahren mit  $n_i \equiv 50$ 

		$\alpha_\delta$				
		0.01	0.05	0.1	0.15	0.2
$\alpha_\lambda$	0.01	0.997 (0.980)	0.995 (0.940)	0.996 (0.890)	0.996 (0.840)	0.995 (0.790)
	0.05	0.986 (0.940)	0.980 (0.900)	0.975 (0.850)	0.974 (0.800)	0.970 (0.750)
	0.10	0.974 (0.890)	0.960 (0.850)	0.953 (0.800)	0.947 (0.750)	0.943 (0.700)
	0.15	0.957 (0.840)	0.939 (0.800)	0.927 (0.750)	0.920 (0.700)	0.914 (0.650)
	0.20	0.938 (0.790)	0.919 (0.750)	0.900 (0.700)	0.893 (0.650)	0.875 (0.600)

Abbildung 7.1: Niveausimulation für das Konfidenzintervallverfahren mit  $n_i \equiv 50$ 

gleich 1 ist. Mit zunehmender Abweichung erwartet man, dass die Aufdeckung der Alternative immer wahrscheinlicher wird, demnach also die Power gegen 1 konvergiert.

Zuerst werden standardnormalverteilte Zufallsvariablen erzeugt. Die speziellen Abhängigkeiten des Cross-Over-Designs werden wie bei der Niveausimulation ebenfalls linear nachempfunden. Als Indikator für den Grad der Abhängigkeit dient der Parameter  $\rho$ . Je höher der betragsmäßige Wert, desto abhängiger die Variablen. Hier gilt wie bei der Niveausimulation durchgehend  $\rho = 2$ .

Die Powersimulationen gliedern sich in vier Teile. Zunächst wird in Abschnitt 7.2.1 die Power der neu entwickelten Teststatistiken *ohne Baseline-Werte* unter verschiedenen Bedingungen getestet und anschließend in 7.2.2 der Vergleich zu den Verfahren des linearen Modells gezogen. Danach werden in Abschnitt 7.2.3 auch die Baseline-Werte in die Simulationen mit einbezogen. Letztendlich muss in 7.2.4 das Äquivalenzverfahren auf den Prüfstand. Auf die Powersimulation des Konfidenzintervall-Verfahrens wird an dieser Stelle verzichtet, da unter anderem auch das Niveau nicht gut eingehalten wurde.

Tabelle 7.16: Powersimulation für die Statistiken  $U_n^\pi$ 

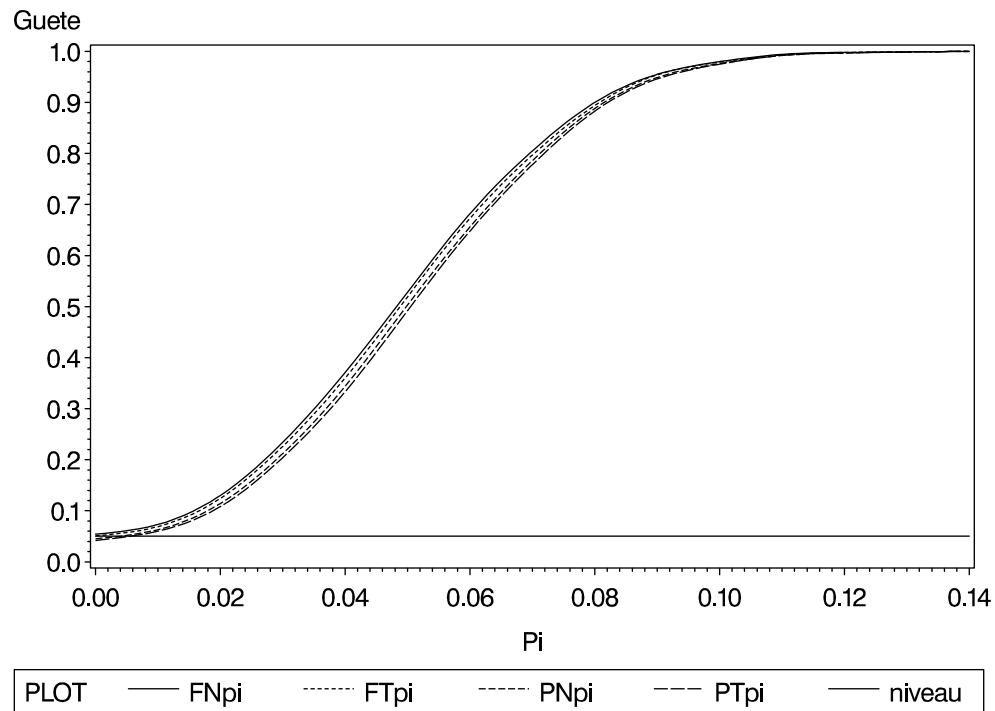
$\pi$	$H_0^F : \mathbf{cF} = 0$		$H_0^p : \mathbf{cp} = 0$	
	$U_n^{\pi,F}(N)$	$U_n^{\pi,F}(t)$	$U_n^{\pi,p}(N)$	$U_n^{\pi,p}(t)$
0	0.054	0.051	0.045	0.042
0.01	0.073	0.069	0.063	0.060
0.02	0.130	0.124	0.114	0.108
0.03	0.234	0.226	0.212	0.204
0.04	0.371	0.360	0.344	0.333
0.05	0.528	0.519	0.503	0.492
0.06	0.682	0.673	0.656	0.648
0.07	0.805	0.797	0.786	0.778
0.08	0.900	0.894	0.888	0.883
0.09	0.955	0.954	0.949	0.946
0.10	0.980	0.979	0.976	0.975
0.11	0.994	0.993	0.992	0.992
0.12	0.998	0.998	0.997	0.997
0.13	0.999	0.999	0.999	0.999
0.14	1	1	1	1

### 7.2.1 Simulation ohne Baseline-Werte

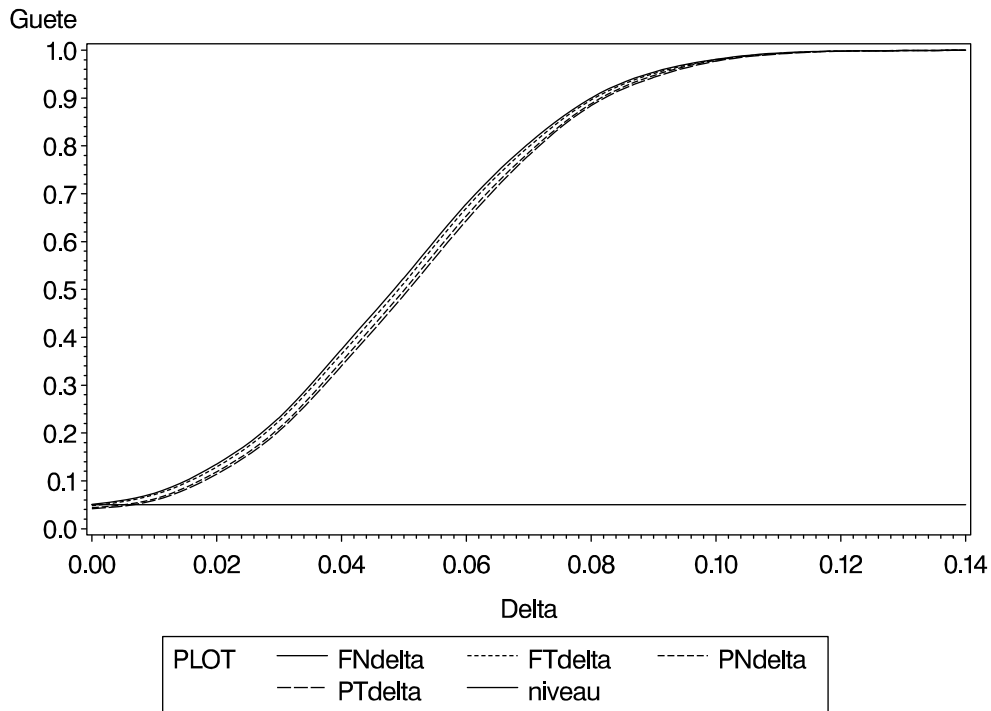
Zunächst wird der Fall betrachtet, dass der Zeiteffekt  $\pi$  von 0 bis 0.14 ansteigt. Es liegt eine Abhängigkeit ( $\rho = 2$ ) zwischen den beiden Werten einer Versuchseinheit zugrunde und ein Stichprobenumfang von jeweils 50 pro Gruppe. Für den Verschiebungsparameter  $\pi \in (0, 0.01, \dots, 0.14)$  wird die Power ausgerechnet. Tabelle 7.16 zeigt die Ergebnisse. Generell unterscheiden sich die Teststatistiken zur Überprüfung des Zeiteffekts nur marginal. Unter  $H_0^F : \mathbf{cF} = 0$  hat man dabei noch eine unwesentlich günstigere Position als bei der Hypothesenstellung über die relativen Effekte, was aber auch aus der günstigeren Niveaueinhaltung resultiert. Die Unterschiede der Approximationen mit Normal- und  $t$ -Verteilung sind praktisch nicht mehr vorhanden. Grafik 7.2 veranschaulicht diesen Sachverhalt.

In Tabelle 7.17 lassen sich die Ergebnisse für die Powersimulation bei immer stärker werdendem Behandlungseffekt  $\Phi$  ablesen. Die anderen Effekte werden auf 0 gesetzt. Die Bedingungen sind wie bei der vorigen Simulation für den Zeiteffekt bezüglich des Stichprobenumfangs und des Grades der Abhängigkeit. Auch die Simulationsergebnisse unterscheiden sich nur minimal, d.h. die oben genannten Beschreibungen übertragen sich auf diesen Fall. Grafik 7.3 veranschaulicht diesen Sachverhalt.

Auch bei der Simulation für die Statistiken zum Aufdecken eines Residualeffekts wurden die selben Ausgangsbedingungen wie zuvor gewählt; ein Abhängigkeitsparameter  $\rho = 2$  und ein Stichprobenumfang von jeweils 50 pro Gruppe. Im Gegensatz zu den Statistiken zum Aufdecken des Zeit- und Cross-Over-Effekts konvergieren diese Teststatistiken deutlich langsamer gegen 1, d.h. sie haben eine durchweg schlechtere Power. Hier kann weder von einem Unterschied der Statistiken bzgl. der Verteilung noch bzgl. der Hypothese die Rede sein. Diese Ergebnisse befinden sich in Tabelle 7.18 und sind in Grafik 7.4 dargestellt.

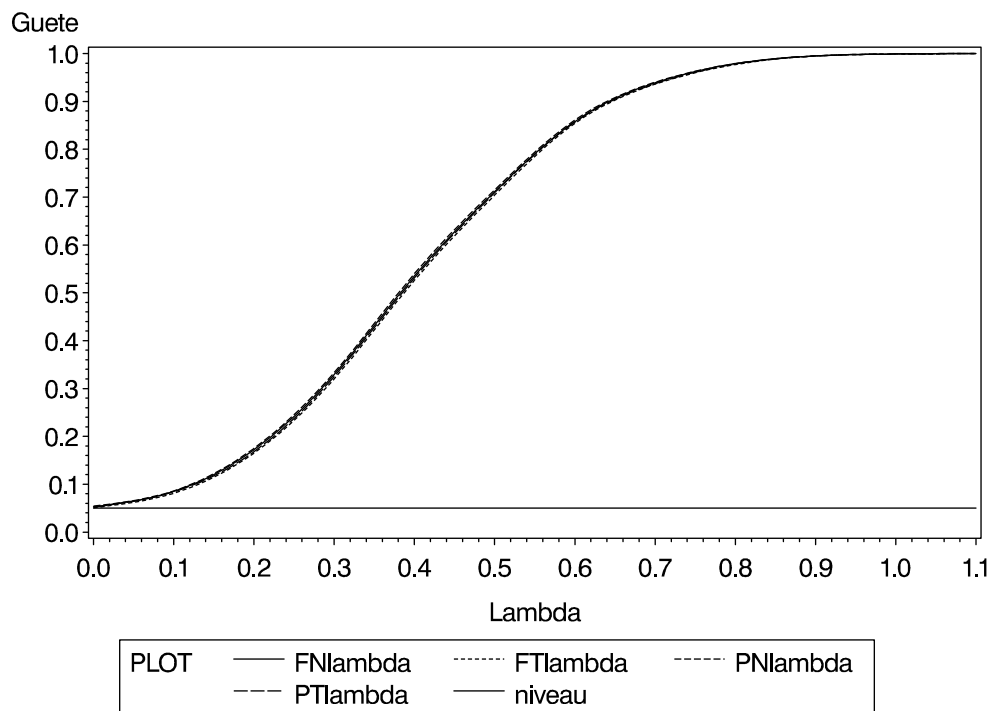
Abbildung 7.2: Powersimulation für die Statistiken  $U_n^\pi$ Tabelle 7.17: Powersimulation für die Statistiken  $U_n^\delta$ 

$\Phi$	$H_0^F : c\mathbf{F} = 0$		$H_0^p : c\mathbf{p} = 0$	
	$U_n^\delta(N)$	$U_n^\delta(t)$	$U_n^\delta(N)$	$U_n^\delta(t)$
0	0.051	0.048	0.044	0.042
0.01	0.074	0.071	0.062	0.059
0.02	0.135	0.129	0.119	0.114
0.03	0.232	0.225	0.210	0.204
0.04	0.375	0.365	0.349	0.340
0.05	0.524	0.513	0.498	0.488
0.06	0.679	0.670	0.654	0.644
0.07	0.805	0.798	0.786	0.780
0.08	0.900	0.896	0.888	0.884
0.09	0.954	0.951	0.947	0.943
0.1	0.981	0.980	0.979	0.977
0.11	0.994	0.993	0.993	0.992
0.12	0.998	0.998	0.998	0.998
0.13	0.999	0.999	0.999	0.999
0.14	1	1	1	1

Abbildung 7.3: Powersimulation für die Statistiken  $U_n^\delta$ Tabelle 7.18: Powersimulation für die Statistiken  $U_n^\lambda$ 

$\lambda$	$H_0^F : cF = 0$		$H_0^P : cp = 0$	
	$U_n^\lambda(N)$	$U_n^\lambda(t)$	$U_n^\lambda(N)$	$U_n^\lambda(t)$
0	0.053	0.051	0.054	0.052
0.1	0.085	0.081	0.086	0.084
0.2	0.172	0.165	0.175	0.168
0.3	0.329	0.320	0.333	0.325
0.4	0.535	0.526	0.540	0.531
0.5	0.711	0.704	0.715	0.709
0.6	0.858	0.854	0.861	0.857
0.7	0.939	0.936	0.940	0.937
0.8	0.979	0.977	0.979	0.978
0.9	0.995	0.995	0.995	0.995
1.0	0.999	0.999	0.999	0.999
1.1	1	1	1	1



Abbildung 7.4: Powersimulation für die Statistiken  $U_n^\lambda$ 

### 7.2.2 Simulationsvergleiche mit den Verfahren des linearen Modells

Während im entsprechenden Abschnitt 7.1.2 beim Niveauvergleich der Verfahren im nicht-parametrischen Modell (GM) mit den Verfahren des linearen Modells kaum Unterschiede bezüglich der Einhaltung des Niveaus erkennbar wurden, zeigen die Powersimulationen die Brüchigkeit der alten Verfahren, wenn entsprechende Voraussetzungen an die Verteilung nicht erfüllt sind. Es stellt sich zunächst einmal heraus (siehe Tabellen 7.19 und 7.20), dass sich alte und neue Statistiken bei Zugrundelegung der Normalverteilung kaum in ihrer Power unterscheiden. Das bedeutet, dass die Statistiken  $U_n^{\lambda,F}$ ,  $U_n^{\lambda,p}$ ,  $U_{LEH}^\lambda$  und  $U_{KOCH}^\lambda$  nur unwesentlich voneinander abweichen. Ebenso verhält es sich innerhalb der anderen Effekte. Die Effekte untereinander sind insofern unterschiedlich, dass die Statistiken zur Aufdeckung eines Residualeffekts generell eine deutlich schlechtere Power besitzen als Statistiken zur Aufdeckung eines Zeit- oder Cross-Over-Effekts. Diese Eigenschaft deckt sich offensichtlich mit dem Verhalten bei der Powersimulation ohne Berücksichtigung von Baseline-Werten im vorangegangenen Abschnitt.

Die Tabellen 7.21 und 7.22 zeigen auf, dass die Statistiken von Grizzle im Modell von Lehmacher  $U_{LEH}$  am meisten Macht verlieren. Das ist darauf zurückzuführen, dass durch die Zugrundelegung der Gamma-Verteilung die Voraussetzung der Normalverteilung für diese Statistiken nicht mehr erfüllt ist. Kochs Statistiken konvergieren zwar auch langsamer als die neuen Statistiken auf Basis des nichtparametrischen Modells, haben aber keinen deutlichen Abstand von ihnen. In Abbildung 7.5 ist dargestellt, wie sich die Teststatistiken für das nichtparametrische Modell unter  $H_0^F : \mathbf{cF} = 0$ , also  $U_n^{\lambda,F}$ ,  $U_n^{\pi,F}$  und  $U_n^{\delta,F}$ , von den Statistiken von Lehmacher für die entsprechenden Effekte unterscheiden. Dabei wird in erster Linie der Unterschied zwischen Statistiken für den Residualeffekt



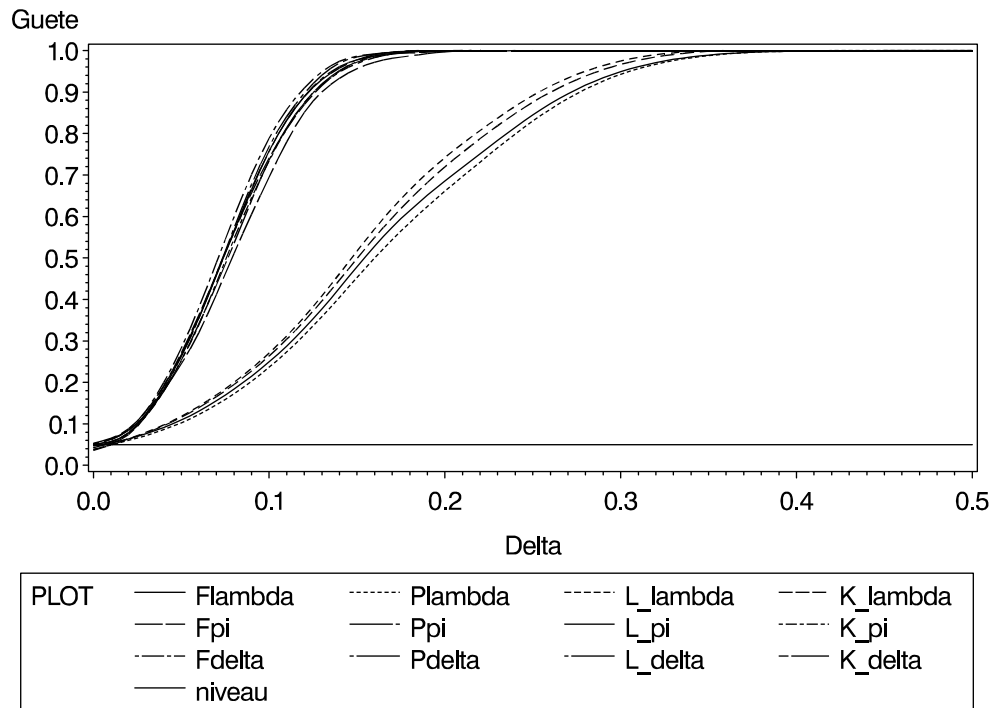


Abbildung 7.5: Powersimulation im Vergleich für die Statistiken des linearen und nicht-parametrischen Modells

und den Statistiken für die beiden anderen Effekte deutlich. Ferner fällt auf, dass die Statistiken des linearen Modells bei jedem Effekt eine geringere Power als die Statistiken des nichtparametrischen Modells haben.

Die Tabellen 7.23 und 7.24 zeigen die entsprechenden Simulationsergebnisse bei schiefer diskreter Verteilung. Die Statistiken, die für das nichtparametrische Modell (GM) entwickelt wurden, werden nicht schlechter als unter Gamma-verteilten Zufallsvariablen. Insgesamt wird verglichen damit in erster Linie die Konvergenz der Verfahren auf einen Perioden- bzw. Cross-Over-Effekt schlechter. Insgesamt halten sich die Statistiken von Lehmacher und Koch erstaunlich gut, wenn man bedenkt, dass nicht einmal mehr die Stetigkeit der Verteilungen gegeben ist.

### 7.2.3 Simulation mit Baseline-Werten

Die Simulation unter Berücksichtigung von zwei Vorwerten funktioniert prinzipiell wie die Simulation ohne die Berücksichtigung von Baseline-Werten. Alles Wesentliche bzgl. der zusätzlichen Erzeugung der Baseline-Werte und die Modellierung des Abhängigkeitsprozesses wurde bereits im Abschnitt der Niveausimulationen beschrieben. Die zunehmende Entfernung von den Nullhypothesen wurde simuliert, indem der entsprechende Effekt schrittweise angehoben wurde bis die Power gleich 1 betrug. In den Tabellen ist jeweils in der linken Spalte die zugrundeliegende Effektgröße und daneben die dann für die verschiedenen Teststatistiken resultierende Power abgetragen.

In Tabelle 7.25 wurde die Power für die Statistiken auf einen Residualeffekt abgetragen. Tabelle 7.26 widmet sich der Untersuchung der Power der Statistiken für Perioden- bzw.





Tabelle 7.25: Powersimulation mit Baseline-Werten von Statistiken für einen Residualeffekt

$\lambda$	$U_n^{F,\lambda,reg}$	$U_n^{p,\lambda,reg}$	$U_{LEH}^\lambda$	$U_{KOCH}^\lambda$
0	0.052	0.047	0.053	0.052
0.05	0.101	0.092	0.097	0.094
0.10	0.241	0.223	0.263	0.251
0.15	0.453	0.426	0.508	0.485
0.20	0.685	0.665	0.753	0.728
0.25	0.861	0.846	0.910	0.893
0.30	0.952	0.948	0.978	0.971
0.35	0.987	0.985	0.996	0.995
0.40	0.997	0.997	0.999	0.999
0.45	0.999	0.999	1	1
0.50	1	1	1	1

Cross-Over-Effekte. Dabei werden jeweils parallel die Verfahren für das nichtparametrische Modell (GMK) und das (erweiterte) lineare Modell dargestellt, so dass an dieser Stelle gleich ein Vergleich zwischen diesen gezogen werden kann. Die Statistiken auf Basis des nichtparametrischen Modells (GMK) basieren auf der Kovarianzanalyse, d.h. die Baseline-Werte werden als Kovariable aufgefasst. Die Power für das in Abschnitt 6.2 ab Seite 60 angegebene Verfahren für das lineare Modell basiert auf der Adjustierung der ursprünglichen Messwerte durch die Baseline-Werte. Es sei jedoch noch einmal festgehalten, dass dann lediglich die Interpretation des Behandlungseffekts im Vergleich zu den Verfahren des Standard-COPs gleich bleibt. Aufgrund der Adjustierung ändern sich die anderen Effekte bzgl. ihrer Interpretation. Ebenso ändert sich durch die Testgröße  $\hat{p}^{reg}$  die Interpretation der Effekte im nichtparametrischen Modell (GMK) im Vergleich zur Analyse des Standard-COPs auf Basis des nichtparametrischen Modells (GM). Siehe dazu Abschnitt 6.3 ab Seite 64. In den folgenden Simulationen wurden die Parameter wie folgt gewählt. Der Vergleichbarkeit wegen wurden die Parameter  $\rho = 2$ ,  $n_i \equiv 50$  und  $\alpha = 5\%$  wie bei der Simulation ohne Baseline-Werte gewählt. Für die Power der Teststatistiken zur Aufdeckung des Zeit- und Cross-Over-Effekts gilt das im vorigen Abschnitt gesagte. In Tabelle 7.25 sind die Ergebnisse der Powersimulation für die Statistiken auf einen Residualeffekt eingetragen. Der Unterschied zwischen den einzelnen Teststatistiken ist nicht sehr groß. Eine marginal bessere Macht haben die Tests, die für das lineare Modell entwickelt wurden. Betrachtet man diese Ergebnisse zusammen mit den Resultaten aus Abschnitt 7.2.1 ab Seite 94, so zeigt sich, dass die Power durch die zusätzliche Berücksichtigung von Baseline-Werten doppelt so schnell gegen 1 konvergiert als ohne Berücksichtigung der beiden Vorwerte pro Versuchseinheit.

Tabelle 7.26 zeigt die entsprechenden Ergebnisse für die Statistiken des Zeit- und Cross-Over-Effekts. Beim Zeiteffekt sind die Unterschiede zwischen den Statistiken  $U_{LEH}^\pi$ ,  $U_{KOCH}^\pi$  und  $U_n^{F,\pi,reg}$  sehr gering. Ein wenig schlechter ist die Macht bei den Teststatistiken, die unter  $H_0^{p^{reg}} : \mathbf{c}p^{reg} = 0$  entwickelt wurden. Die Teststatistiken auf einen Cross-Over-Effekt testen aufgrund der Abwesenheit eines Nacheffekts den (reinen) Behandlungseffekt. Hier sind die Statistiken des nichtparametrischen Modells denen des linearen Modells marginal überlegen. Für alle Statistiken auf einen Perioden- oder Cross-

Tabelle 7.26: Powersimulation mit Baseline-Werten von Statistiken für Perioden- und Cross-Over-Effekt

$\pi \Phi$	$U_n^{F,\pi,reg}$	$U_n^{p,\pi,reg}$	$U_{LEH}^\pi$	$U_{KOCH}^\pi$	$U_n^{F,\delta,reg}$	$U_n^{p,\delta,reg}$	$U_{LEH}^\delta$	$U_{KOCH}^\delta$
0	0.053	0.047	0.051	0.047	0.052	0.048	0.051	0.047
0.02	0.088	0.074	0.078	0.077	0.086	0.078	0.085	0.084
0.04	0.192	0.180	0.190	0.179	0.199	0.183	0.184	0.178
0.06	0.362	0.320	0.355	0.339	0.380	0.355	0.359	0.341
0.08	0.563	0.511	0.566	0.546	0.599	0.572	0.562	0.535
0.10	0.741	0.695	0.758	0.734	0.789	0.769	0.757	0.736
0.12	0.870	0.851	0.889	0.874	0.910	0.897	0.890	0.874
0.14	0.950	0.933	0.958	0.948	0.974	0.970	0.961	0.952
0.16	0.987	0.972	0.990	0.983	0.992	0.991	0.987	0.984
0.18	0.994	0.987	0.998	0.997	0.999	0.999	0.997	0.996
0.20	0.999	0.997	0.999	0.999	1	1	0.999	0.999
0.22	1	0.999	1	1	1	1	1	1
0.22	1	1	1	1	1	1	1	1

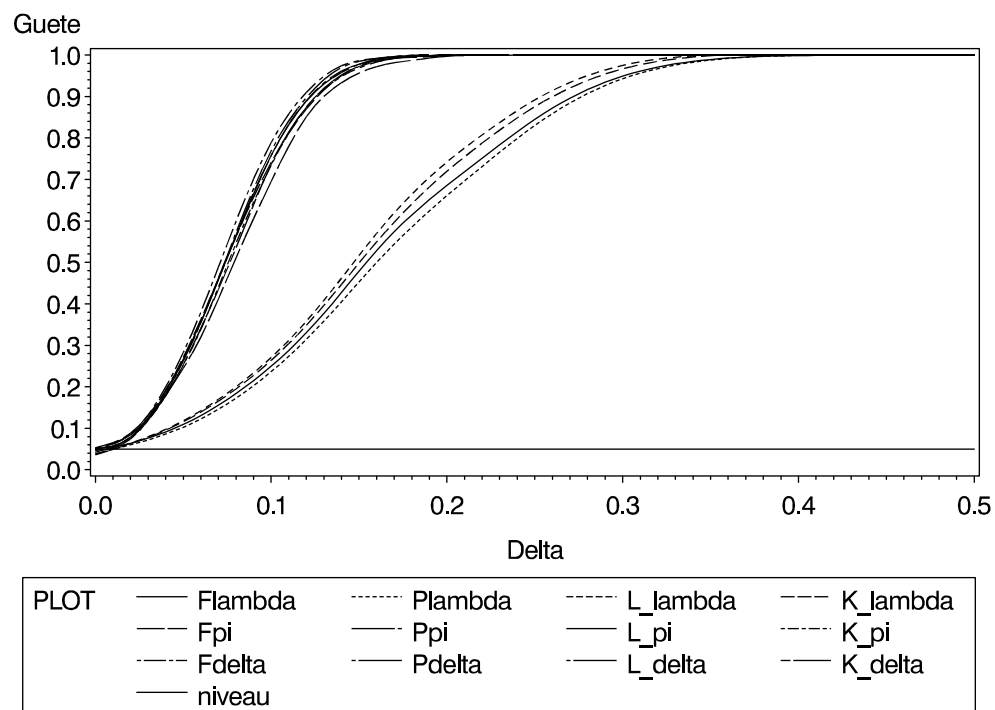


Abbildung 7.6: Powersimulation für alle Statistiken unter Berücksichtigung von Baseline-Werten

Tabelle 7.27: Powersimulation für das Äquivalenzverfahren bei  $n_i \equiv 50$ 

	$\lambda$	$\pi$	$\delta$
0.4	0.034	0.050	0.048
0.35	0.057	0.364	0.374
0.30	0.080	0.852	0.847
0.25	0.109	0.992	0.991
0.20	0.146	1	1
0.15	0.169	1	1
0.10	0.188	1	1
0.05	0.219	1	1
0.00	0.223	1	1

Over-Effekt gilt: die Macht dieser Tests unter Berücksichtigung von Baseline-Werten konvergiert stets ungefähr doppelt so schnell wie die bei den Statistiken ohne die Berücksichtigung von Baseline-Werten.

Abbildung 7.6 fasst die Charakteristiken der soeben geschilderten Ergebnisse noch einmal zusammen. Insbesondere wird deutlich, dass sich im Wesentlichen zwei Bündel gebildet haben. Die Verfahren, die bezüglich der Macht deutlich hinter den anderen zurückliegen, sind die zum Aufdecken eines Residualeffekts (rechts im Bild). Die Unterschiede aller anderen Statistiken wirken dagegen eher unwesentlich.

## 7.2.4 Simulation des Äquivalenzverfahrens

Die Power wurde zu Beginn des Abschnitts über die Powersimulation definiert. Da das Äquivalenzverfahren eine andere Art der Hypothesenstellung als herkömmliche Verfahren verwendet, beschreibt diese Größe hier folgendes: Die Wahrscheinlichkeit, dass eine tatsächlich vorliegende (zuvor als solche definierte) Gleichheit der Effekte erkannt wird. Der Abhängigkeitsparameter  $\rho$  wird wieder auf 2 festgelegt und es wird ein balancierter Stichprobenumfang von 50 angenommen. Die zunehmende Verletzung der Hypothese wird derart simuliert, dass die Zufallsvariablen mit einem immer geringeren Effekt (0.4, 0.35, ..., 0) versehen werden. Mit zunehmender Abweichung von der Hypothese, also dem Unterschied, sollte die Alternative erkannt werden, demnach die Power gegen 1 streben. In Tabelle 7.27 wird ein Abweichen von der Hypothese beim Zeit- und beim Behandlungseffekt sehr schnell erkannt, während die Power selbst bei Nichtvorhandensein des Residualeffekts unter 25% liegt. Um die Geschwindigkeit der Konvergenz bei abnehmenden Residualeffekt besser einschätzen zu können, werden für diesen Effekt zusätzlich noch die Stichprobenumfänge  $n_i \equiv 100$  und  $n_i \equiv 150$  simuliert. Die Ergebnisse sind in Tabelle 7.28 festgehalten. Erwartungsgemäß wird die Konvergenz schneller, jedoch ist die Power erst bei einem Stichprobenumfang von 150 akzeptabel. In Grafik 7.7 werden diese Ergebnisse noch einmal veranschaulicht.



Tabelle 7.28: Powersimulation für das Äquivalenzverfahren für  $\lambda$  bei  $n_i \equiv 100$  und  $n_i \equiv 150$

	$n_i \equiv 100$	$n_i \equiv 150$
0.4	0.047	0.048
0.35	0.090	0.106
0.30	0.160	0.197
0.25	0.250	0.338
0.20	0.372	0.512
0.15	0.498	0.682
0.10	0.615	0.800
0.05	0.706	0.894
0.00	0.731	0.913

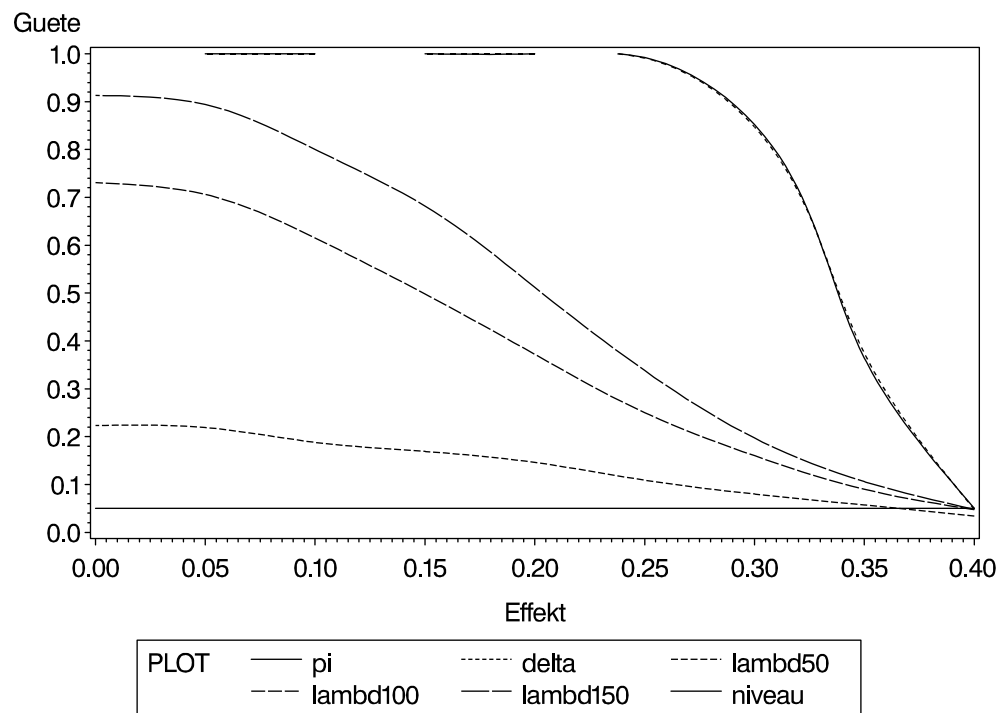


Abbildung 7.7: Powersimulation für das Äquivalenzverfahren



# Kapitel 8

## Zusammenfassung und Ausblick

### Zusammenfassung

Zu Beginn dieser Arbeit wurden zwei Beispiele von Asthma-Studien vorgestellt anhand derer der Aufbau eines Cross-Over-Plans mit und ohne Baseline-Werte erläutert wurde. Anschließend wurden in Kapitel 3 die in dem Versuchsdesign des Standard-COPs auftretenden Effekte, Residual-, Perioden- und Cross-Over-Effekt, beschrieben. Ferner wurden zwei in dieser Arbeit betrachtete Modelle vorgestellt. Zum Einen das lineare Modell nach GRIZZLE (1965), welches die Additivität der einzelnen auftretenden Effekte und nach LEHMACHER (1987) die multivariate Normalverteilung für den Fehlerterm voraussetzt. Zum Anderen wurde das wenig restriktive nichtparametrische Modell (GM) vorgestellt für das in dieser Arbeit Statistiken zum Testen der einzelnen Effekte entwickelt wurden. Die Einfachheit dieses Modells besteht darin, dass jegliche Voraussetzungen an die Verteilungsfunktionen der Stichproben fallengelassen werden und lediglich die Ein-Punkt-Verteilung ausgeschlossen werden muss. Daher können mit dem hier zugrunde gelegten Modell auch beispielsweise Daten mit unstetigen Verteilungsfunktionen berücksichtigt werden, was unter anderem im Hinblick auf ordinale Daten sehr wichtig ist.

In Kapitel 4 werden nach der Vorstellung der Vorgehensweise im linearen Modell Teststatistiken für das nichtparametrische Modell (GM) im Cross-Over-Plan hergeleitet, so dass vorhandene Zeit-, Residual- und Cross-Over-Effekte mit deren Hilfe aufgedeckt werden können. Dies ist sowohl für die Nullhypothesen, die über die Verteilungsfunktionen formuliert werden, möglich, als auch dann durchführbar, wenn man die Hypothesen weiter fasst und zwar über die relativen Effekte. Genauer wird in dieser Arbeit mit dem im Vergleich zum relativen *gewichteten* Effekt etwas neueren relativen *ungewichteten* Effekt gearbeitet. Die Unterscheidung liegt in der Verwendung des Mittels der Verteilungsfunktionen. Im ersten Fall wird das mit den Stichprobenumfängen gewichtete Mittel und im zweiten das arithmetische Mittel der Verteilungsfunktionen benutzt.

Das Vorhandensein eines Nacheffekts führt dazu, dass ein reiner Behandlungsunterschied mit diesen Verfahren nicht mehr aufgedeckt werden kann. Aber schon die Entscheidung für oder gegen einen Residualeffekt kann mit Fehlern behaftet sein. Als Problemlösung wurden neben bereits bestehenden Ideen das Konfidenzintervallverfahren (Abschnitt 5.4) und die modifizierte Two-Stage-Analyse mittels Äquivalenzverfahren (Abschnitt 5.5) untersucht. Ersteres ist durch die Konstruktion jedoch schon stark konservativ.

Wird der Standard-COP durch Baseline-Werte ergänzt, kann unter Umständen ein Verbesserung Effizienz der Verfahren eintreten. Das Kapitel 6 stellt ein Verfahren im li-

nearen Modell vor, welches mit um die Vorwerte adjustierten Messwerten arbeitet und anschließend eine Möglichkeit, solche Zusatzinformationen im nichtparametrischen Modell zu berücksichtigen. Dies geschieht durch die Einführung der Kovarianzanalyse nach SIEMER (1999) auf Basis des nichtparametrischen Modells mit Kovariablen (GMK). Wie beim nichtparametrischen Modell (GM) ist es hier möglich, zweierlei Hypothesen zu betrachten. Unter der Modellvoraussetzung (MV), dass alle Vorwerte der gleichen Verteilung folgen, können Verteilungsaussagen über Teststatistiken unter der Hypothese  $H_0^F : cF = 0$  angegebene werden. Ebenso lassen sich Teststatistiken auf Basis des bereinigten relativen Effekts unter der Hypothese  $H_0^{reg} : cp^{reg} = 0$  angeben.

Die theoretischen Überlegungen dieser Arbeit wurden in Kapitel 7 mit Hilfe von Computersimulationen auf die Eigenschaften hinsichtlich der Niveaueinhaltung und der Konvergenz der Power mit wachsendem Abweichen von der Hypothese untersucht. Dabei ergab sich folgendes: Die in dieser Arbeit entwickelten neuen Verfahren können, obwohl weniger Voraussetzungen benötigt werden, sowohl im Niveau als auch bei der Power mit den Verfahren im linearen Modell von GRIZZLE (1965) bzw. LEHMACHER (1987) und KOCH (1972) mithalten. Bei Abweichung von der Normalverteilung haben die in dieser Arbeit entwickelten Verfahren unter Umständen Vorteile gegenüber den Verfahren des linearen Modells in der Power, vor allem, wenn diese noch die Normalverteilung voraussetzen. Bei der Überprüfung der neuen Verfahren ergab sich weiterhin: Die Tests auf einen Residual-, Zeit- oder Cross-Over-Effekt im Cross-Over-Plan lassen sich bei einer Hypothesenstellung über die Verteilungsfunktionen bereits bei einem kleinstmöglichen Stichprobenumfang von ca. 7-10 Versuchseinheiten pro Gruppe durchführen. Je nach vorliegenden Effekten greift auch spätestens ab einem Stichprobenumfang von ca. 20 pro Gruppe die Approximation mit der Normalverteilung. Bei der Hypothesenstellung über die relativen Effekte braucht man für die Niveaueinhaltung schon wenigstens 40 bis 50 Versuchseinheiten, Tests auf einen Residualeffekt nur etwa 10. Bezüglich der Power gibt es keine nennenswerten Unterschiede zwischen den Hypothesen und Approximationen. Die Simulationen mit Baseline-Werten unter Voraussetzung des linearen Modells zeigen im Vergleich der Tests auf einen Behandlungseffekt, dass die Hinzunahme von Baseline-Werten für die Niveaueinhaltung keinen Zugewinn bringen, bei jedoch zusätzlich doppeltem Gewinn in der Power. Ferner können die Baseline-Werte auch zur Überprüfung der Randomisierung, einer adäquaten Länge der Wash-Out-Phase oder zur Gewöhnung der Probanden an die Testsituation eingesetzt werden. Von einem Einsatz des Konfidenzintervallverfahrens ist aufgrund seiner starken Konservativität abzuraten. Wenn überhaupt keine Aussage über einen eventuell vorhandenen Residualeffekt im Vorfeld getroffen werden kann, könnte dieses Verfahren notfalls bei Irrtumswahrscheinlichkeiten kleiner als 1% eingesetzt werden, um einen Behandlungseffekt aufzudecken. Sollte aus Gründen der Fragestellung zum Nachweis für *keinen* Effekt, z.B. bei der Einführung eines neuen Medikaments oder als 1. Stufe des modifizierten Two-Stage-Verfahrens, das Äquivalenzverfahren zum Einsatz kommen, so sollte man einen entsprechend höheren Stichprobenumfang einplanen. Ca. 30-40 Individuen pro Gruppe sind für den Einsatz eines Äquivalenztests auf keinen Zeit- oder Cross-Over-Effekt einzuplanen und noch einmal nahezu doppelt so viele, wenn man einen Nachweis für keinen Residualeffekt erhalten möchte. Um jedoch auch eine adäquate Power erwarten zu können, muss der Stichprobenumfang zur Aufdeckung eines Residualeffekts noch einmal deutlich auf wenigstens 150 Individuen pro Gruppe angehoben werden. Die Two-Stage-Analyse bietet dennoch die bessere Alterna-

tive einer Two-Stage-Analyse als Grizzles Vortestmethode, da letzteres Verfahren von der Konstruktion her das Niveau schon gar nicht einhalten kann, da bei Nichtablehnung der Hypothese auf deren Gültigkeit geschlossen wird und dieses für den anschließenden Test vorausgesetzt wird. Jedoch sollen auch die Punkte nicht unerwähnt bleiben, die beim Umgang der modifizierten Two-Stage-Analyse zu beachten sind. Dies ist zum einen die Wahl des Intervalls und folgende Tatsache: Beim Äquivalenztest möchte man strenggenommen nicht das Nichtvorhandensein eines Effekts nachweisen, sondern dass der Effekt so klein ist, dass er innerhalb eines vorgegebenen Bereichs liegt. Verwirft man die Hypothese, hat man also einen beliebig kleinen Effekt. Der daran anschließende Test der modifizierten Two-Stage-Analyse setzt jedoch keinen Residualeffekt voraus, was das Niveau dieses Tests beeinflusst. Eine weitere (nichtvalidierte) Möglichkeit zur Überprüfung bietet unter Umständen der Summen-Differenzen Plot. Ohne Test könnte man sich hier schon im Vorfeld davon überzeugen, ob ein Residualeffekt vorhanden ist oder nicht. Ferner könnte man anschließend noch ein Urteil über einen möglichen Cross-Over- bzw. Behandlungseffekt fällen.

Zu den Vorteilen eines Cross-Over-Plans gehört die Reduktion des Stichprobenumfangs durch die Elimination der inter-individuellen Variabilität. Jeder Patient dient als seine eigene Kontrolle. Wie bereits vielfach in der Literatur festgehalten, wirkt sich in erster Linie ein Residual-Effekt nachteilig aus. Dann ist eine weitere Analyse des Versuchsplans erschwert (hohe Stichprobenumfänge, höhere Unsicherheit) oder gar nicht erst möglich. Das Cross-Over-Verfahren ist immer dann sinnvoll einsetzbar, wenn die Ausgangssituation in der zweiten Periode (trotz Behandlung oder Medikation) die gleiche wie bei der ersten Periode ist. Dazu wird in erster Linie ein konstanter Krankheitsverlauf gefordert sowie auch Kenntnis über die Pharmakodynamik der eingesetzten Behandlungen, damit die Wash-out-Phase adäquat gewählt werden kann. Klassische Einsatzbereiche sind Studien für chronische Krankheiten. Ansonsten kann man mit einer entsprechend hohen Stichprobenzahl auf die modifizierte Two-Stage-Analyse zurückgreifen oder grafisch arbeiten.

#### **Ausblick**

Viele Fragestellungen können auf Basis dieser Arbeit weiterverfolgt werden. Beispielsweise die Erweiterung der Analysemethoden auf einen erweiterten Cross-Over-Plan, sog. ABCD-Designs. Das sind Verfahren, wo eben nicht zwei, sondern mehrere Verfahren miteinander verglichen werden sollen. Das Problem von missing-values ist noch zu bearbeiten, und die Übertragbarkeit (gruppen-)sequenzieller und adaptiver Verfahren auf die hergeleiteten Verfahren könnte vielleicht den Vorteil der Verringerung der Probandenzahl noch erhöhen. Denkbar wäre auch die Entwicklung von Verfahren basierend auf dem Bayes-Ansatz.



# Anhang A

## Theorie

### A.1 Notation

In diesem Abschnitt soll die Notation eingeführt und die Begriffe definiert werden, die im Verlauf der Arbeit vorausgesetzt werden. Dazu zählen die Schreibweise für Vektoren, Matrizen und deren Rechenoperationen genauso wie verwendete Konvergenzbegriffe.

#### A.1.1 Vektoren und Matrizen

**Matrizen** und **Vektoren** werden im Folgenden mit fett gedruckten Buchstaben und Symbolen bezeichnet. So ist z.B.  $\mathbf{I}_d$  die Einheitsmatrix der Dimension  $d \times d$ . Ferner stellt  $\mathbf{1}_d$  einen  $d \times 1$ -dimensionalen Einservektor dar. Die  $d \times d$ -Einsermatrix wird mit  $\mathbf{J}_d = \mathbf{1}_d \mathbf{1}'_d$  bezeichnet. Die Nullmatrix bzw. der Nullvektor wird entsprechend mit  $\mathbf{0}$  bezeichnet. Allgemein werden Matrizen mit großen fettgedruckten Buchstaben und Vektoren mit kleinen fettgedruckten Buchstaben beschrieben.

Für zwei beliebige Matrizen

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pq} \end{pmatrix} \in \mathbb{R}^{p \times q} \quad \text{und} \quad \mathbf{B} = \begin{pmatrix} b_{11} & \cdots & b_{1s} \\ \vdots & & \vdots \\ b_{r1} & \cdots & b_{rs} \end{pmatrix} \in \mathbb{R}^{r \times s}$$

heißt

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B} \\ \vdots & & \vdots \\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B} \end{pmatrix} \in \mathbb{R}^{(p \cdot r) \times (q \cdot s)}$$

das Kronecker-Produkt und

$$\mathbf{A} \oplus \mathbf{B} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \in \mathbb{R}^{(p+r) \times (q+s)}$$

die Kronecker-Summe.

Mit  $\mathbf{A}^{-1}$  wird die Inverse von  $\mathbf{A}$  bezeichnet. Ferner bezeichne  $\mathbf{A}^T$  die Transponierte der Matrix  $\mathbf{A}$ .

## A.1.2 Konvergenz

Folgende Formen der **Konvergenz** werden in dieser Arbeit verwendet, z.B. wenn es um den Nachweis der Konsistenz geht. Die folgenden Begriffsdefinitionen basieren auf den Ausführungen von BAUER (1991). Sei  $X_n(\omega) \in \mathbb{R}$  und  $X : \mathbb{R} \rightarrow \mathbb{R}$  eine reellwertige Zufallsvariable.

- **Stochastische Konvergenz/ Konvergenz in Wahrscheinlichkeit:** Eine Folge von Zufallsvariablen  $X_n$  heißt stochastisch konvergent gegen  $X$  (in Zeichen  $X_n \xrightarrow{P} X$ ), wenn gilt:

$$\lim_{n \rightarrow \infty} P \{ |X_n - X| \geq \epsilon \} = 0 \quad \forall \epsilon > 0.$$

- **Fast sichere Konvergenz:** Eine Folge  $X_n$  heißt fast sicher konvergent gegen  $X$  (in Zeichen  $X_n \xrightarrow{f.s.} X$ ), wenn gilt:

$$P \left( \limsup_{n \rightarrow \infty} \{ |X_n - X| > \epsilon \} \right) = 0 \quad \forall \epsilon > 0.$$

- **Konvergenz im quadratischen Mittel:** Eine Folge  $X_n$  heißt konvergent in  $\mathcal{L}_2$ -Norm gegen  $X$  (in Zeichen  $X_n - X \xrightarrow{\mathcal{L}_2} 0$ ), wenn gilt:

$$\lim_{n \rightarrow \infty} E ( |X_n - X|^2 ) = 0.$$

- **Konvergenz in Verteilung:** Sei  $g$  stetig und beschränkt und  $P_n, P$  W-Maße auf  $\mathbb{R}$ . Konvergiert eine Folge von Verteilungen  $P_n$  derart gegen  $P$ , dass  $\lim_{n \rightarrow \infty} \int g dP_n = \int g dP$  gilt, so spricht man von Konvergenz in Verteilung (in Zeichen  $P_n \xrightarrow{\mathcal{L}} P$ ).

**Anmerkung:** Sowohl aus fast sicherer als auch aus  $\mathcal{L}_2$ -Konvergenz folgt stochastische Konvergenz. Aus dieser wiederum folgt die Konvergenz in Verteilung.

In dieser Arbeit wird der Begriff der **asymptotischen Äquivalenz** und der **asymptotischen Verteilung** benutzt. Eine genauere Definition dieser grundlegenden Begriffe findet sich in DOMHOF (1999).

**Definition A.1.1 (Asymptotisch Äquivalent)** Die Zufallsvariablen  $X_n$  und  $Y_n$  heißen *asymptotisch äquivalent* (in Zeichen  $\doteq$ ), wenn ihre Differenz für  $n \rightarrow \infty$  stochastisch gegen Null konvergiert.

Aus der asymptotischen Äquivalenz zweier Folgen von Zufallsvariablen folgt, dass diese für große Stichprobenumfänge näherungsweise die gleiche Verteilung haben. Hat nun die Folge von Zufallsvariablen  $(X_n)_{n \geq 0}$  die Verteilung  $F$  und gilt  $Y_n \doteq X_n$ , so sagt man:  $Y_n$  ist asymptotisch  $F$ -verteilt (in Zeichen  $\dot{\sim}$ ), wobei  $F$  an dieser Stelle eine beliebige Verteilungsfunktion ist.

Die in dieser Arbeit auftauchenden Integrale sind **Lebesgue-Stieltjes-Integrale**. Zur Definition des Integrals siehe z.B. HOFFMANN-JORGENSEN (1994).



## A.2 Grundlegende Lemmata

Im nichtparametrischen Modell (GM) gelten die folgenden Aussagen:

**Lemma A.2.1** Für alle  $(i, s) \neq (i', s')$  gilt:

$$E(c(X_{iks} - X_{i'k's'})) = \int F_{i's'} dF_{is} \quad \text{für alle } i, s = 1, 2, k = 1, \dots, n_i.$$

**Beweis:**

Siehe BRUNNER (1997).

**Lemma A.2.2 (Jensens Ungleichung für Erwartungswerte)** Sei  $X$  eine Zufallsvariable mit endlichem Erwartungswert  $E(X)$  und  $g$  eine konvexe Funktion. Dann gilt:

$$g(E(X)) \leq E(g(X)).$$

**Beweis:**

Siehe LOEVE (1977). □

**Lemma A.2.3 (Jensens Ungleichung für die quadratische Funktion)** Seien  $X_1, \dots, X_n$  beliebige reellwertige Zufallsvariablen. Dann gilt:

$$\left( \sum_{i=1}^n X_i \right)^2 \leq n \sum_{i=1}^n X_i^2.$$

**Beweis:**

Siehe LOEVE (1977). □

### A.2.1 Grenzwertsätze

**Satz A.2.4**  $S$  sei die  $\sigma$ -Algebra der Borel-Mengen und  $(P_n)_{n \in \mathbb{N}}$  eine Folge von Wahrscheinlichkeitsverteilungen auf  $S$ . Dann sind äquivalent:

- Es existiert ein  $P$ , so dass  $(P_n)_{n \in \mathbb{N}}$  schwach gegen  $P$  konvergiert.
- Für jede Linearform  $\lambda$  auf  $\mathbb{R}^k$  konvergiert die Folge der eindimensionalen Verteilungen  $(P_n \lambda^{-1})_{n \in \mathbb{N}}$  schwach gegen einen Limes.

**Beweis:**

Siehe VARADARAJAN (1958). □

**Korollar A.2.5** Mit der Notation aus Satz A.2.4 sind äquivalent:

- $(P_n)_{n \in \mathbb{N}}$  konvergiert schwach gegen eine Normalverteilung auf  $\mathbb{R}^k$ .
- Für jede Linearform  $\lambda$  auf  $\mathbb{R}^k$  konvergiert die Folge der eindimensionalen Verteilungen  $(P_n \lambda^{-1})_{n \in \mathbb{N}}$  schwach gegen eine Normalverteilung.

**Beweis:**

Siehe MUNZEL (1996). □

**Satz A.2.6 (Lindeberg-Feller)** Für jedes  $n$  seien  $X_{1,n}, \dots, X_{n,n}$  unabhängige Zufallsvariablen mit

$$X_{i,n} \sim F_{i,n}, E(X_{i,n}) = \mu_{i,n} \quad \text{und} \quad 0 < \text{Var}(X_{i,n}) = \sigma_{i,n}^2 < \infty$$

für alle  $i = 1, \dots, n$ . Sei ferner

$$\sigma_n^2 = \sum_{i=1}^n \sigma_{i,n}^2.$$

Dann hat

$$\frac{1}{\sigma_n} \sum_{i=1}^n (X_{i,n} - \mu_{i,n})$$

asymptotisch eine Standardnormalverteilung, wenn für alle  $\epsilon > 0$  gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_n^2} \sum_{i=1}^n \int_{|x - \mu_{i,n}| > \epsilon \sigma_n} (x - \mu_{i,n})^2 dF_{i,n} = 0.$$

**Beweis:**

Siehe GNEDENKO (1962). □

### A.3 Definitionen, Sätze und Beweise

Dieses Kapitel enthält im Wesentlichen die Theorie, die benötigt wird, um die Verteilungen der neuen Teststatistiken für das nichtparametrische Modell herzuleiten. Es kann zum Nachschlagen für Details der Berechnungen aus Kapitel 4 genutzt werden oder auch als theorieorientiertes Kapitel im Ganzen betrachtet werden. Teils bekannte und auf das hiesige Modell angewandte Beweise werden dennoch meistens ausformuliert, damit dem Leser das nahtlose Nachvollziehen der Sachverhalte möglich gemacht wird. Die Basis dieser Ausführungen ist das nichtparametrische Modell.

**Definition A.3.1 (Nichtparametrisches Modell (GM))** Gegeben seien für  $i = 1, 2$  und  $k = 1, \dots, n_i$  unabhängige Zufallsvektoren

$$\mathbf{X}_{ik} = (X_{ik1}, X_{ik2})^T \sim \mathbf{F}_i.$$

Für die Einträge gelte:

$$X_{iks} \sim F_{is} \quad (\text{A.1})$$

mit  $i, s = 1, 2$  und  $k = 1, \dots, n_i$ .

Zwischen den Einträgen des Vektors  $\mathbf{X}_{ik}$  seien beliebige Abhängigkeiten zugelassen.

**Definition A.3.2 (Verteilungsfunktion)** Sei das Modell wie in (GM) definiert. Dann heißt für die Zufallsvariable  $X_{iks}$  mit  $i, s = 1, 2$  und  $k = 1, \dots, n_i$

$$\begin{aligned} F_{is}^-(x) &= P(X_{iks} < x) \quad \text{links-stetige,} \\ F_{is}^+(x) &= P(X_{iks} \leq x) \quad \text{rechts-stetige und} \\ F_{is}(x) &= \frac{1}{2} [F_{is}^+(x) + F_{is}^-(x)] \quad \text{normalisierte} \end{aligned}$$

Version der Verteilungsfunktion.

$\mathbf{F} = (F_{11}, F_{12}, F_{21}, F_{22})^T$  bezeichnet den Vektor der Verteilungsfunktionen. In dieser Arbeit wird die Darstellung der normalisierten Version gewählt.

Um die relativen Effekte bilden zu können, wird noch die mittlere ungewichtete Verteilungsfunktion benötigt. Der relative Effekt vergleicht, grob gesagt, die einzelnen Verteilungsfunktionen mit dieser mittleren Verteilungsfunktion.

**Definition A.3.3 (mittlere ungewichtete Verteilungsfunktion)** Sei  $F_{is}$  für  $i, s = 1, 2$  die zuvor definierte normalisierte Version der Verteilungsfunktion. Dann heißt

$$G(x) = \frac{1}{4} \sum_{i=1}^2 \sum_{s=1}^2 F_{is}(x)$$

mittlere ungewichtete Verteilungsfunktion.

**Definition A.3.4 (Ungewichteter relativer Effekt)**  $X_{iks}$  seien unabhängige Zufallsvariablen mit  $X_{iks} \sim F_{is}$ ,  $i, s = 1, 2$  und  $k = 1, \dots, n_i$ , heißt

$$p_{is} = \int G dF_{is}$$

(mittlerer) ungewichteter relativer Effekt von  $F_{is}$  zu  $F_{11}, \dots, F_{22}$ .

Es gilt:

$$\frac{1}{\bar{n}} := \frac{1}{4} \sum_{i=1}^2 \sum_{s=1}^2 \frac{1}{n_i} = \frac{1}{2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right).$$

Jetzt wird die Definition der Zählfunktion als Basis für Rangdefinitionen und Bildung für Schätzer der empirischen Verteilungsfunktion angegeben.

**Definition A.3.5 (Zählfunktion)** Die Funktion

$$\begin{aligned} c^-(x) &= \begin{cases} 0 & , x \leq 0 \\ 1 & , x > 0 \end{cases} \quad \text{heißt links-stetige,} \\ c^+(x) &= \begin{cases} 0 & , x < 0 \\ 1 & , x \geq 0 \end{cases} \quad \text{rechts-stetige und} \\ c(x) &= \frac{1}{2} [c^+(x) + c^-(x)] \quad \text{normalisierte} \end{aligned}$$

Version der Zählfunktion.

Wichtig für den Einsatz der Zählfunktion bei ordinalen Daten ist, dass die Differenzen bei Ausdrücken wie  $c(X_{iks} - X_{j\ell t})$  immer qualitativ zu verstehen sind. Daher kann man die Zählfunktion auch als Vergleichsfunktion betiteln. Im Falle ordinaler Daten interessiert lediglich die Relation, ob die Differenz negativ oder positiv ist.

Konsistente und erwartungstreue Schätzer für die verschiedenen Versionen der Verteilungsfunktionen erhält man durch Einsetzen der drei Versionen der Zählfunktion.

**Definition A.3.6 (Empirische Verteilungsfunktion)** Es liege das Modell (GM) zugrunde. Für eine Stichprobe  $X_{i1s}, \dots, X_{in_i s}$  mit  $i = 1, 2$  und  $s = 1, 2$  von Beobachtungen heißt die Funktion

$$\begin{aligned} \widehat{F}_{is}^-(x) &= \frac{1}{n_i} \sum_{k=1}^{n_i} c^-(x - X_{iks}) \quad \text{links-stetige,} \\ \widehat{F}_{is}^+(x) &= \frac{1}{n_i} \sum_{k=1}^{n_i} c^+(x - X_{iks}) \quad \text{rechts-stetige und} \\ \widehat{F}_{is}(x) &= \frac{1}{n_i} \sum_{k=1}^{n_i} c(x - X_{iks}) = \frac{1}{2} [\widehat{F}_{is}^+(x) + \widehat{F}_{is}^-(x)] \quad \text{normalisierte} \end{aligned}$$

Version der empirischen Verteilungsfunktion von  $X_{i1s}, \dots, X_{in_i s}$ .

Entsprechend wird die mittlere ungewichtete Verteilungsfunktion  $G$  geschätzt durch die **mittlere ungewichtete empirische Verteilungsfunktion**

$$\widehat{G}(x) = \frac{1}{4} \sum_{i=1}^2 \sum_{s=1}^2 \widehat{F}_{is}(x) \quad (\text{A.2})$$

der empirischen Verteilungsfunktionen  $\widehat{F}_{11}, \dots, \widehat{F}_{22}$ . Die Konsistenz und Erwartungstreue der empirischen Verteilungsfunktion übertragen sich auf  $\widehat{G}(x)$ . Damit kann auch die asymptotische Erwartungstreue und Konsistenz der Schätzer des relativen Effekts nachgewiesen werden.

**Proposition A.3.7 (asymptotische Erwartungstreue und Konsistenz von  $\widehat{p}_{is}$ )** Es liege das Modell (GM) zugrunde. Bezeichne  $\widehat{F}_{is}$  die empirische Verteilungsfunktion der Stichprobe  $X_{i1s}, \dots, X_{in_i s}$  und  $\widehat{G}$  die mittlere ungewichtete empirische Verteilungsfunktion. Wenn  $\min_i(n_i) \rightarrow \infty$ , dann gilt für alle  $i, s = 1, 2$ :

$$\begin{aligned} \widehat{p}_{is} - p_{is} &\xrightarrow{\mathcal{L}_2} 0 \quad \text{und} \\ E(\widehat{p}_{is}) &\rightarrow p_{is}, \end{aligned}$$

**Beweis:**

Die  $\mathcal{L}_2$ -Konvergenz folgt in erster Linie mit Hilfe der Jensen-Ungleichung A.2.3 und unter Berücksichtigung der Unabhängigkeitsstruktur.

$$\begin{aligned}
& E(\widehat{p}_{is} - p_{is})^2 \\
&= E\left(\int [\widehat{G} - G] d\widehat{F}_{is} + \int G d[\widehat{F}_{is} - F_{is}]\right)^2 \\
&\leq 2E\left(\int [\widehat{G} - G] d\widehat{F}_{is}\right)^2 + 2E\left(\int G d[\widehat{F}_{is} - F_{is}]\right)^2 \\
&\leq \frac{2}{n_i} \sum_{k=1}^{n_i} E\left(\widehat{G}(X_{iks}) - G(X_{iks})\right)^2 \\
&\quad + \frac{2}{n_i^2} \sum_{k=1}^{n_i} \sum_{\ell=1}^{n_i} E\left[\left(G(X_{iks}) - \int G dF_{is}\right) \left(G(X_{i\ell s}) - \int G dF_{is}\right)\right] \\
&\leq \frac{2}{n_i} \sum_{k=1}^{n_i} E\left(\widehat{G}(X_{iks}) - G(X_{iks})\right)^2 + \frac{2}{n_i^2} \sum_{k=1}^{n_i} E\left(G(X_{iks}) - \int G dF_{is}\right)^2 \\
&\leq \frac{2}{n_i} + \frac{2}{n_i} = \frac{4}{n_i} \longrightarrow 0.
\end{aligned}$$

Wegen A.2.2 folgt aus

$$[E(\widehat{p}_{is} - p_{is})]^2 \leq E(\widehat{p}_{is} - p_{is})^2$$

auch die asymptotische Erwartungstreue von  $\widehat{p}_{is}$ :

$$E(\widehat{p}_{is}) - p_{is} \longrightarrow 0.$$

□

Da sich die Schätzer der relativen Effekte auch mit Hilfe von Rängen dargestellt werden können, werden dieselben zunächst einmal formal definiert.

**Definition A.3.8 (Ränge)** *Im Modell (GM) heißt für  $i, s = 1, 2$  und  $k = 1, \dots, n_i$*

$$R_{iks} = \frac{1}{2} + \sum_{j=1}^2 \sum_{\ell=1}^{n_j} \sum_{t=1}^2 c(X_{iks} - X_{j\ell t})$$

der (Gesamt-)Rang von  $X_{iks}$  unter allen  $N = 2(n_1 + n_2)$  Zufallsvariablen  $X_{111}, \dots, X_{2n_22}$ ,

$$R_{iks}^{(is)} = \frac{1}{2} + \sum_{\ell=1}^{n_i} c(X_{iks} - X_{i\ell s})$$

der Internrang von  $X_{iks}$  innerhalb aller Zufallsvariablen mit der Faktorstufenkombination  $(i, s)$ :  $X_{i1s}, \dots, X_{in_i s}$  und

$$R_{iks}^{(-i's')} = \frac{1}{2} + \sum_{(j,t) \neq (i',s')} \sum_{\ell=1}^{n_j} c(X_{iks} - X_{j\ell t})$$

der Teilrang von  $X_{iks}$  unter allen  $N - n_{i'}$  Zufallsvariablen ohne die Faktorstufenkombination  $(i', s')$ .

Für die Darstellung der Schätzer der ungewichteten relativen Effekte können die harmonischen Ränge verwendet werden. Zunächst einmal zur Definition der harmonischen Ränge:

**Definition A.3.9 (Harmonische Ränge)** Im Modell (GM) heißt für  $i, s = 1, 2$  und  $k = 1, \dots, n_i$

$$\Psi_{iks} = \frac{\tilde{n}}{2n_i} + \sum_{j=1}^2 \sum_{\ell=1}^{n_j} \sum_{t=1}^2 \frac{\tilde{n}}{n_j} c(X_{iks} - X_{j\ell t})$$

der harmonische (Gesamt-)Rang von  $X_{iks}$  unter allen  $N = 2(n_1 + n_2)$  Zufallsvariablen  $X_{111}, \dots, X_{2n_22}$ ,

$$\Psi_{iks}^{(is)} = R_{iks}^{(is)} = \frac{1}{2} + \sum_{\ell=1}^{n_i} c(X_{iks} - X_{i\ell s})$$

der harmonische Internrang von  $X_{iks}$  innerhalb aller Zufallsvariablen mit der Faktorstufenkombination  $(i, s)$  und

$$\Psi_{iks}^{(-i's')} = \frac{\tilde{n}}{2n_i} + \sum_{(j,t) \neq (i',s')} \sum_{\ell=1}^{n_j} \frac{\tilde{n}}{n_j} c(X_{iks} - X_{j\ell t})$$

der harmonische Teilrang von  $X_{iks}$  unter allen  $N - n_{i'}$  Zufallsvariablen ohne die Faktorstufenkombination  $(i', s')$ .

Im balancierten Fall ( $n_1 = n_2$ ) stimmen die harmonischen Ränge mit den 'normalen' Rängen überein.

Praktisch gewinnt man die harmonischen Ränge jedoch nicht aus der Definition, sondern mit Hilfe der leicht zu bekommenden (Standard-)Ränge. Daher sei an dieser Stelle der Zusammenhang zwischen Rängen und harmonischen Rängen angegeben.

**Lemma A.3.10 (Zusammenhang zwischen Rängen und harmonischen Rängen)** Seien die Ränge wie in A.3.8 definiert und die harmonischen Ränge wie zuvor eingeführt. Dann bestehen für  $i, s = 1, 2$  und  $k = 1, \dots, n_i$  folgende Zusammenhänge:

1. für Ränge und harmonische Ränge

$$\Psi_{iks} = \sum_{(i',s') \neq (i,s)} \frac{\tilde{n}}{n_{i'}} \left( R_{iks} - R_{iks}^{(-i's')} \right) + \frac{\tilde{n}}{n_i} R_{iks}^{(is)}$$

2. für Internränge und harmonische Internränge

$$\Psi_{iks}^{(is)} = R_{iks}^{(is)}$$

3. für Teilräume und harmonische Teilräume mit  $(i, s) \neq (i', s')$

$$\Psi_{iks}^{(-i's')} = \sum_{\substack{(j,t) \neq (i',s') \\ (j,t) \neq (i,s)}} \frac{\tilde{n}}{n_j} \left( R_{iks} - R_{iks}^{(-jt)} \right) + \frac{\tilde{n}}{n_i} R_{iks}^{(is)}.$$

**Beweis:**

zu 1.

$$\begin{aligned} \Psi_{iks} &= \frac{\tilde{n}}{2n_i} + \sum_{j=1}^2 \sum_{\ell=1}^{n_j} \sum_{t=1}^2 \frac{\tilde{n}}{n_j} c(X_{iks} - X_{j\ell t}) \\ &= \frac{\tilde{n}}{2n_i} + \sum_{(j,t) \neq (i,s)} \sum_{\ell=1}^{n_j} \frac{\tilde{n}}{n_j} c(X_{iks} - X_{j\ell t}) + \frac{\tilde{n}}{n_i} \sum_{\ell=1}^{n_i} c(X_{iks} - X_{i\ell s}) \\ &= \sum_{(j,t) \neq (i,s)} \frac{\tilde{n}}{n_j} \left( R_{iks} - R_{iks}^{(-jt)} \right) + \frac{\tilde{n}}{n_i} R_{iks}^{(is)} \end{aligned}$$

zu 3.

$$\begin{aligned} \Psi_{iks}^{(-i's')} &= \frac{\tilde{n}}{2n_i} + \sum_{\substack{(j,t) \neq (i',s') \\ (j,t) \neq (i,s)}} \sum_{\ell=1}^{n_j} \frac{\tilde{n}}{n_j} c(X_{iks} - X_{j\ell t}) \\ &= \frac{\tilde{n}}{2n_i} + \sum_{\substack{(j,t) \neq (i',s') \\ (j,t) \neq (i,s)}} \sum_{\ell=1}^{n_j} \frac{\tilde{n}}{n_j} c(X_{iks} - X_{j\ell t}) + \sum_{\ell=1}^{n_i} \frac{\tilde{n}}{n_i} c(X_{iks} - X_{i\ell s}) \end{aligned}$$

□

Die normierten Plazierungen werden in erster Linie für die vereinfachte Darstellung der Schätzer benötigt. Daraus folgt dann auch unmittelbar die Darstellung des Schätzers des relativen Effekts mit Hilfe der harmonischen Ränge.

**Lemma A.3.11 (Normierten Plazierungen)** *Bezeichne für  $i, s = 1, 2$  und  $k = 1, \dots, n_i$   $\hat{F}_{is}$  die empirische Verteilungsfunktion von  $X_{i1s}, \dots, X_{in_i s}$  und  $\hat{G}$  die mittlere ungewichtete empirische Verteilungsfunktion. Dann gilt:*

$$\begin{aligned} \hat{G}(X_{iks}) &= \frac{1}{4\tilde{n}} \left( \Psi_{iks} - \frac{\tilde{n}}{2n_i} \right), \\ \hat{F}_{is}(X_{iks}) &= \frac{1}{n_i} \left( \Psi_{iks}^{(is)} - \frac{1}{2} \right) \text{ und} \\ \hat{F}_{i's'}(X_{iks}) &= \frac{1}{\tilde{n}} \left( \Psi_{iks} - \Psi_{iks}^{(-i's')} \right) \text{ für } (i', s') \neq (i, s). \end{aligned}$$

**Beweis:**

$$\begin{aligned} \hat{G}(X_{iks}) &= \frac{1}{4} \sum_{j=1}^2 \sum_{t=1}^2 \hat{F}_{jt}(X_{iks}) = \frac{1}{4} \sum_{j=1}^2 \sum_{t=1}^2 \frac{1}{n_j} \sum_{\ell=1}^{n_j} c(X_{iks} - X_{j\ell t}) = \\ &= \frac{1}{4\tilde{n}} \sum_{j=1}^2 \sum_{t=1}^2 \frac{\tilde{n}}{n_j} \sum_{\ell=1}^{n_j} c(X_{iks} - X_{j\ell t}) = \frac{1}{4\tilde{n}} \left( \Psi_{iks} - \frac{\tilde{n}}{2n_i} \right) \end{aligned}$$

$$\widehat{F}_{is}(X_{iks}) = \frac{1}{n_i} \sum_{\ell=1}^{n_i} c(X_{iks} - X_{i\ell s}) + \frac{1}{2n_i} - \frac{1}{2n_i} = \frac{1}{n_i} \left( \Psi_{iks}^{(is)} - \frac{1}{2} \right)$$

$$\begin{aligned} \widehat{F}_{i's'}(X_{iks}) &= 4\widehat{G}(X_{iks}) - \sum_{(j,t) \neq (i',s')} \widehat{F}_{jt}(X_{iks}) = \frac{1}{\widetilde{n}} \sum_{j=1}^2 \sum_{t=1}^2 \frac{\widetilde{n}}{n_j} \sum_{\ell=1}^{n_j} c(X_{iks} - X_{j\ell t}) \\ &= \frac{1}{\widetilde{n}} \sum_{(j,t) \neq (i',s')} \sum_{\ell=1}^{n_j} \frac{\widetilde{n}}{n_\ell} c(X_{iks} - X_{j\ell t}) = \frac{1}{\widetilde{n}} \left( \Psi_{iks} - \Psi_{iks}^{(-i's')} \right) \end{aligned}$$

□

Grundlegend für die Berechnungen der Kovarianzmatrix und somit auch für die Bildung von Teststatistiken für entsprechende Effekte unter Hypothese sowie unter Alternative ist der asymptotische Äquivalenzsatz. Unter der folgenden Voraussetzung erlaubt er die Berechnungen von Varianzen und Kovarianzen über die Vergleichsstatistik.

**Voraussetzung A.3.12 (V1)**  $N \rightarrow \infty$ , so dass ein  $N_0 \in \mathbb{N}$  existiert mit  $\frac{N}{n_i} \leq N_0 < \infty$  für alle  $i = 1, 2$ .

**Satz A.3.13 (Asymptotischer Äquivalenzsatz)** Es liegt das Modell (GM) zugrunde. Unter Voraussetzung (V1) folgt

$$\sqrt{N} \int \widehat{G}d(\widehat{\mathbf{F}} - \mathbf{F}) - \sqrt{N} \int Gd(\widehat{\mathbf{F}} - \mathbf{F}) \xrightarrow{\mathcal{L}_2} 0,$$

und somit sind  $\sqrt{N} \int \widehat{G}d(\widehat{\mathbf{F}} - \mathbf{F})$  und  $\sqrt{N} \int Gd(\widehat{\mathbf{F}} - \mathbf{F})$  bzw.  $\sqrt{N}(\widehat{\mathbf{p}} - \int Gd\widehat{\mathbf{F}})$  und  $\sqrt{N}(\int \widehat{G}d\widehat{\mathbf{F}} - \mathbf{p})$  asymptotisch äquivalent.

**Beweis:**

Dies ist ein Spezialfall von SIEMER (1999).

□

Mit Hilfe des asymptotischen Äquivalenzsatzes gilt unter der Nullhypothese  $H_0^F : \mathbf{cF} = 0$  dann:

$$\sqrt{N}\mathbf{c}\widehat{\mathbf{p}} - \sqrt{N}\mathbf{c} \left( \int Gd\widehat{\mathbf{F}} \right) \xrightarrow{\mathcal{L}_2} 0.$$

Auf Basis des folgenden Satzes können die Teststatistiken unter  $H_0^F : \mathbf{cF} = 0$  gebildet werden. Dafür wird jedoch noch eine weitere Voraussetzung benötigt:

**Voraussetzung A.3.14 (V2)**  $\text{Cov}(\overline{\mathbf{Y}}_i) = \mathbf{V}_{i,n_i} \xrightarrow{n_i \rightarrow \infty} \mathbf{V}_i$ ,

d.h. die Kovarianzmatrix  $\mathbf{V}_{i,n_i}$  konvergiert mit wachsendem Stichprobenumfang.



**Satz A.3.15 (Asymptotische Normalität unter  $H_0^F : \mathbf{c}\mathbf{F} = 0$ )** Falls (V1) und (V2) gilt, dann

$$\sqrt{N}\mathbf{c}\hat{\mathbf{p}} = \sqrt{N}\mathbf{c} \int \hat{G}d\hat{\mathbf{F}} = \sqrt{N}\mathbf{c}\widehat{\mathbf{Y}} = \frac{\sqrt{N}}{4\tilde{n}}\mathbf{c} \left( \overline{\Psi} - \frac{\tilde{n}}{2}\mathbf{n}^{-1} \right) =$$

ist unter  $H_0^F : \mathbf{c}\mathbf{F} = 0$  asymptotisch normalverteilt mit Erwartungswert 0 und Varianz  $\mathbf{c}\mathbf{V}_N\mathbf{c}^T$ . Dabei lässt sich  $\mathbf{V}_N$  darstellen als

$$\mathbf{V}_N = \bigoplus_{i=1}^2 \frac{N}{n_i} \mathbf{V}_{i,n_i}$$

mit folgendem konsistenten Schätzer für  $\mathbf{V}_{i,n_i} = \text{Cov}(\overline{\mathbf{Y}}_i)$

$$\widehat{\mathbf{V}}_{i,n_i} = \frac{1}{16\tilde{n}^2(n_i - 1)} \sum_{k=1}^{n_i} (\Psi_{ik} - \overline{\Psi}_i) (\Psi_{ik} - \overline{\Psi}_i)^T.$$

Dabei ist  $\Psi_{ik} = (\Psi_{ik1}, \Psi_{ik2})^T$  und entsprechend  $\overline{\Psi}_i = (\overline{\Psi}_{i,1}, \overline{\Psi}_{i,2})^T$ .

**Beweis:**

Siehe SIEMER (1999). □

Aus dem asymptotischen Äquivalenzsatz A.3.13 folgt unter  $H_0^P$

$$\sqrt{N}\mathbf{c}\hat{\mathbf{p}} \doteq \sqrt{N}\mathbf{c} \left( \int Gd\hat{\mathbf{F}} - \int \mathbf{F}d\hat{G} \right). \quad (\text{A.3})$$

Bezeichne im Weiteren

$$\mathbf{Z}_N = \sqrt{N} \left( \int Gd\hat{\mathbf{F}} - \int \mathbf{F}d\hat{G} + \mathbf{1} - 2\mathbf{p} \right)$$

die Vergleichsstatistik von  $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$  und  $\mathbf{Z}_N = (Z_{11}, Z_{12}, Z_{21}, Z_{22})^T$ .

Um Teststatistiken herleiten zu können, muss zunächst einmal die Kovarianzmatrix berechnet und v.a. geschätzt werden. Die Kovarianzmatrix der Vergleichsstatistik erhält man einfacher. Dazu wird im Folgenden eine Komponente der Vergleichsstatistik betrachtet. Sei  $j \neq i$  und  $s' \neq s$ :

$$\begin{aligned}
Z_{N, is} &= \sqrt{N} \left( \int G d\widehat{F}_{is} - \int F_{is} d\widehat{G} + 1 - 2p_{is} \right) \\
&= \sqrt{N} \left( \frac{1}{n_i} \sum_{k=1}^{n_i} G(X_{iks}) - \frac{1}{4} \sum_{j=1}^2 \sum_{t=1}^2 \frac{1}{n_{i'}} \sum_{\ell=1}^{n_{i'}} F_{is}(X_{i'\ell t}) \right) + \sqrt{N} (1 - 2p_{is}) \\
&= \sqrt{N} \left( \frac{1}{n_i} \sum_{k=1}^{n_i} \left( G(X_{iks}) - \frac{1}{4} \sum_{t=1}^2 F_{is}(X_{ikt}) \right) - \frac{1}{4} \sum_{t=1}^2 \frac{1}{n_j} \sum_{\ell=1}^{n_j} F_{is}(X_{j\ell t}) \right) \\
&\quad + \sqrt{N} (1 - 2p_{is}) \quad j \neq i \\
&= \frac{\sqrt{N}}{4} \left( \frac{1}{n_i} \sum_{k=1}^{n_i} \left( \sum_{(j,t) \neq (i,s)} F_{jt}(X_{iks}) - F_{is}(X_{iks'}) \right) \right. \\
&\quad \left. + \frac{1}{n_j} \sum_{\ell=1}^{n_j} \left( - \sum_{t=1}^2 F_{is}(X_{j\ell t}) \right) \right) \sqrt{N} (1 - 2p_{is}) \quad j \neq i \\
&= \frac{\sqrt{N}}{4} \left( \frac{1}{n_i} \sum_{k=1}^{n_i} Z_{iks} + \frac{1}{n_j} \sum_{k=1}^{n_j} Z_{jk}^{(-is)} \right) + \sqrt{N} (1 - 2p_{is}) \quad j \neq i
\end{aligned}$$

mit

$$\begin{aligned}
Z_{iks} &= 4 \cdot G(X_{iks}) - \sum_{t=1}^2 F_{is}(X_{ikt}) \\
&= \sum_{(j,t) \neq (i,s)} F_{jt}(X_{iks}) - F_{is}(X_{iks'}) \quad \text{mit } s' \neq s, \\
Z_{jk}^{(-is)} &= - \sum_{t=1}^2 F_{is}(X_{jkt}) \quad \text{mit } j \neq i.
\end{aligned} \tag{A.4}$$

Man hat demnach eine Komponente der Vergleichsstatistik  $Z_{N, is}$  für  $\sqrt{N}(\widehat{p}_{is} - p_{is})$  gefunden, die sich in unabhängige Summanden  $Z_{iks}$  und  $Z_{jk}^{(-is)}$  mit  $j \neq i$  zerlegen lässt. Mit diesen Vorkenntnissen ist die asymptotische Normalität von  $\sqrt{N}(\widehat{p}_{is} - p_{is})$  aus dem zentralen Grenzwertsatz offensichtlich.

Zunächst werden die Varianzen betrachtet.

$$\mathbf{W}_N = (c_{ii'ss'})_{\substack{i, i' = 1, 2 \\ s, s' = 1, 2}}$$

Die Varianzen  $c_{iiss} = \text{Var}(Z_{N, is})$  werden zerlegt:

$$\begin{aligned}
c_{iiss} &= \text{Var} \left( \sqrt{N} \left( \int G d\widehat{F}_{is} - \int F_{is} d\widehat{G} \right) \right) \\
&= \frac{N}{16} \text{Var} \left( \frac{1}{n_i} \sum_{k=1}^{n_i} Z_{iks} + \frac{1}{n_j} \sum_{k=1}^{n_j} Z_{jk}^{(-is)} \right) \quad j \neq i \\
&= \frac{N}{16n_i} \alpha_{iiss} + \frac{N}{16n_j} \alpha_{jj-is-is} \quad j \neq i,
\end{aligned}$$

wobei

$$\begin{aligned}\alpha_{iiss} &= \text{Var}(Z_{i1s}) \text{ und} \\ \alpha_{jj-is-is} &= \text{Var}(Z_{j1}^{(-is)}) \quad j \neq i.\end{aligned}$$

Die Kovarianzen zerfallen in der gleichen Weise ( $j \neq i$  und  $j' \neq i'$ ):

$$\begin{aligned}c_{ii'ss'} &= \text{Cov} \left( \frac{\sqrt{N}}{4} \left( \frac{1}{n_i} \sum_{k=1}^{n_i} Z_{iks} + \frac{1}{n_j} \sum_{k=1}^{n_j} Z_{jk}^{(-is)} \right), \right. \\ &\quad \left. \frac{\sqrt{N}}{4} \left( \frac{1}{n_{i'}} \sum_{k'=1}^{n_{i'}} Z_{i'k's'} + \frac{1}{n_{j'}} \sum_{k'=1}^{n_{j'}} Z_{j'k'}^{(-i's')} \right) \right) \\ &= \frac{N}{16n_i n_{i'}} \sum_{k=1}^{n_i} \sum_{k'=1}^{n_{i'}} \text{Cov}(Z_{iks}, Z_{i'k's'}) + \frac{N}{16n_i} \sum_{k=1}^{n_i} \frac{1}{n_{j'}} \sum_{k'=1}^{n_{j'}} \text{Cov}(Z_{iks}, Z_{j'k'}^{(-i's')}) \\ &\quad + \frac{N}{16n_{i'}} \sum_{k'=1}^{n_{i'}} \frac{1}{n_j} \sum_{k=1}^{n_j} \text{Cov}(Z_{i'k's'}, Z_{jk}^{(-is)}) \\ &\quad + \frac{N}{16n_j n_{j'}} \sum_{k=1}^{n_j} \sum_{k'=1}^{n_{j'}} \text{Cov}(Z_{jk}^{(-is)}, Z_{j'k'}^{(-i's')}).\end{aligned}$$

Zunächst einmal kann man unterscheiden nach Kovarianzen innerhalb einer Gruppe und außerhalb, also die beiden Fälle  $i = i'$  und  $i \neq i'$ .

$$\begin{aligned}\underline{i = i'} \\ c_{iiss'} &= \frac{N}{16n_i} \text{Cov}(Z_{i1s}, Z_{i1s'}) + \frac{N}{16n_j} \text{Cov}(Z_{j1}^{(-is)}, Z_{j1}^{(-is')}) \quad j \neq i \\ &= \frac{N}{16n_i} \alpha_{iiss'} + \frac{N}{16n_j} \alpha_{jj-is-is'} \quad j \neq i\end{aligned}$$

$$\begin{aligned}\underline{i \neq i'} \\ c_{ii'ss'} &= \frac{N}{16n_i} \text{Cov}(Z_{i1s}, Z_{i1}^{(-i's')}) + \frac{N}{16n_{i'}} \text{Cov}(Z_{i'1s'}, Z_{i'1}^{(-is)}) \\ &= \frac{N}{16n_i} \alpha_{iis-i's'} + \frac{N}{16n_{i'}} \alpha_{i'i's'-is}.\end{aligned}$$

Es treten dabei die folgenden Kovarianzen auf:

$$\begin{aligned}\alpha_{iiss'} &= \text{Cov}(Z_{i1s}, Z_{i1s'}) \\ \alpha_{jj-is-is} &= \text{Cov}(Z_{j1}^{(-is)}, Z_{j1}^{(-is')}) \quad \text{mit } j \neq i \\ \alpha_{iis-i's'} &= \text{Cov}(Z_{i1s}, Z_{i1}^{(-i's')}) \quad \text{mit } i \neq i'\end{aligned}$$

Die Kovarianzmatrix  $\mathbf{W}_N$  wird durch

$$\widehat{\mathbf{W}}_N = (\widehat{c}_{ii'ss'})_{\substack{i, i' = 1, 2 \\ s, s' = 1, 2}}$$

geschätzt. Dieser Schätzer zerfällt wie die Kovarianzmatrix in einzelne Komponenten, d.h.

$$\widehat{c}_{iiss} = \frac{N}{16n_i} \widehat{\alpha}_{iiss} + \frac{N}{16n_j} \widehat{\alpha}_{jj-is-is} \quad j \neq i.$$

Analog ergibt sich für die Kovarianzen

$$\widehat{c}_{iiss'} = \frac{N}{16n_i} \widehat{\alpha}_{iiss'} + \frac{N}{16n_j} \widehat{\alpha}_{jj-is-is'} \quad j \neq i$$

$$\widehat{c}_{i'is'} = \frac{N}{16n_i} \widehat{\alpha}_{iis-i's'} + \frac{N}{16n_{i'}} \widehat{\alpha}_{i'i's'-is}.$$

Dabei werden die einzelnen Komponenten wie folgt geschätzt (mit  $j \neq i$ ):

$$\begin{aligned} \widehat{\alpha}_{iiss} &= \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right)^2, \\ \widehat{\alpha}_{jj-is-is} &= \frac{1}{n_j - 1} \sum_{k=1}^{n_j} \left( \widehat{Z}_{jk}^{(-is)} - \widehat{Z}_j^{(-is)} \right)^2, \\ \widehat{\alpha}_{iiss'} &= \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right) \left( \widehat{Z}_{iks'} - \widehat{Z}_{i.s'} \right), \\ \widehat{\alpha}_{iis-i's'} &= \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left( \widehat{Z}_{iks} - \widehat{Z}_{i.s} \right) \left( \widehat{Z}_{ik}^{(-i's')} - \widehat{Z}_i^{(-i's')} \right) \quad \text{und} \\ \widehat{\alpha}_{jj-is-is'} &= \frac{1}{n_j - 1} \sum_{k=1}^{n_j} \left( \widehat{Z}_{jk}^{(-is)} - \widehat{Z}_j^{(-is)} \right) \left( \widehat{Z}_{jk}^{(-is')} - \widehat{Z}_j^{(-is')} \right). \end{aligned}$$

Die beobachtbaren Zufallsvariablen  $\widehat{Z}_{iks}$  und  $\widehat{Z}_{jk}^{(-is)}$  können mit Hilfe der harmonischen Ränge berechnet werden:

$$\begin{aligned} \widehat{Z}_{iks} &= 4 \cdot \widehat{G}(X_{iks}) - \sum_{t=1}^2 \widehat{F}_{is}(X_{ikt}) \\ &= \frac{1}{\widetilde{n}} \left( \Psi_{iks} - \frac{\widetilde{n}}{2n_i} \right) - \frac{1}{n_i} \left( \Psi_{iks}^{(is)} - \frac{1}{2} \right) - \frac{1}{\widetilde{n}} \left( \Psi_{ikt} - \Psi_{ikt}^{(-is)} \right) \quad \text{mit } t \neq s \end{aligned}$$

und

$$\begin{aligned} \widehat{Z}_{jk}^{(-is)} &= - \sum_{t=1}^2 \widehat{F}_{is}(X_{jkt}) \\ &= - \frac{1}{\widetilde{n}} \sum_{t=1}^2 \left( \Psi_{jkt} - \Psi_{jkt}^{(-is)} \right) \\ &= \frac{1}{\widetilde{n}} \sum_{t=1}^2 \left( \Psi_{jkt}^{(-is)} - \Psi_{jkt} \right) \quad \text{für } j \neq i. \end{aligned}$$

Mit Hilfe der folgenden Annahme lässt sich dann die multivariate Normalität zeigen.

**Voraussetzung A.3.16 (V3)** Es existiert eine Konstante  $c_0 > 0$ , so dass  $\mathbf{c}\mathbf{W}_N\mathbf{c}^T \geq c_0 > 0$ .

Für die Bildung der Teststatistiken unter der Hypothese  $H_0^p : \mathbf{c}\mathbf{p} = 0$  ist dieser nachfolgender Satz existenziell.

**Satz A.3.17** Unter den Annahmen (VI) und (V3) gilt folgendes:

$$\sqrt{N}\mathbf{c}\hat{\mathbf{p}} = \sqrt{N}\mathbf{c} \int \hat{G}d\hat{\mathbf{F}} = \frac{\sqrt{N}}{4\tilde{n}}\mathbf{c} \left( \bar{\Psi} - \frac{\tilde{n}}{2}\mathbf{n}^{-1} \right)$$

unter  $H_0^p : \mathbf{c}\mathbf{p} = 0$  asymptotisch normalverteilt mit Erwartungswert 0 und Varianz  $\mathbf{c}\mathbf{W}_N\mathbf{c}^T$ . Dabei lässt sich  $\widehat{\mathbf{W}}_N$  für  $j \neq i$  darstellen mit den Diagonalelementen

$$\hat{c}_{iiss} = \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \hat{Z}_{iks} - \hat{Z}_{i.s} \right)^2 + \frac{N}{16n_j(n_j - 1)} \sum_{k=1}^{n_j} \left( \hat{Z}_{jk}^{(-is)} - \hat{Z}_j^{(-is)} \right)^2.$$

Die Kovarianzschätzer innerhalb einer Gruppe  $i$  ergeben sich

$$\begin{aligned} \hat{c}_{iiss'} &= \frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \hat{Z}_{iks} - \hat{Z}_{i.s} \right) \left( \hat{Z}_{iks'} - \hat{Z}_{i.s'} \right) \\ &+ \frac{N}{16n_j(n_j - 1)} \sum_{k=1}^{n_j} \left( \hat{Z}_{jk}^{(-is)} - \hat{Z}_j^{(-is)} \right) \left( \hat{Z}_{jk}^{(-is')} - \hat{Z}_j^{(-is')} \right) \quad j \neq i \end{aligned}$$

und zwischen den Gruppen für  $i \neq i'$

$$\begin{aligned} \hat{c}_{i'i'ss'} &= -\frac{N}{16n_i(n_i - 1)} \sum_{k=1}^{n_i} \left( \hat{Z}_{iks} - \hat{Z}_{i.s} \right) \left( \hat{Z}_{ik}^{(-i's')} - \hat{Z}_i^{(-i's')} \right) \\ &- \frac{N}{16n_{i'}(n_{i'} - 1)} \sum_{k'=1}^{n_{i'}} \left( \hat{Z}_{i'k's'} - \hat{Z}_{i'.s'} \right) \left( \hat{Z}_{i'k'}^{(-is)} - \hat{Z}_{i'}^{(-is)} \right). \end{aligned}$$

$\widehat{\mathbf{W}}_N$  ist ein konsistenter Schätzer für  $\mathbf{W}_N$ .

**Beweis:**

Aus dem asymptotischen Äquivalenzsatz A.3.13 auf Seite 120 folgt, dass  $\sqrt{N}\mathbf{c}(\hat{\mathbf{p}} - \mathbf{p})$  und  $\mathbf{c}\mathbf{Z}_N$  asymptotisch äquivalent sind und demnach Verteilungsaussagen asymptotisch für beide Terme gelten. Daher gilt asymptotisch

$$E(\mathbf{Z}_N) \longrightarrow \mathbf{0}$$

und folglich auch

$$E \left( \int Gd\hat{\mathbf{F}} - \int \hat{\mathbf{F}}d\hat{G} \right) \longrightarrow 2\mathbf{p} - \mathbf{1}.$$

Betrachtet man demnach den Term mit  $j \neq i$

$$\begin{aligned}
\mathbf{cZ}_N &= \sqrt{N}\mathbf{c} \left( \int Gd\widehat{\mathbf{F}} - \int \mathbf{F}d\widehat{G} - (2\mathbf{p} - \mathbf{1}) \right) \\
&= \sqrt{N} \sum_{i=1}^2 \sum_{s=1}^2 c_{is} \left( \frac{1}{4n_i} \sum_{k=1}^{n_i} Z_{iks} + \frac{1}{4n_j} \sum_{\ell=1}^{n_j} Z_{j\ell}^{(-is)} - (2p_{is} - 1) \right) \\
&= \sqrt{N} \sum_{i=1}^2 \sum_{s=1}^2 \left( \frac{1}{4n_i} \sum_{k=1}^{n_i} c_{is} Z_{iks} + \frac{1}{4n_i} \sum_{k=1}^{n_i} c_{js} Z_{ik}^{(-js)} - 2c_{is}p_{is} \right) \\
&= \sum_{i=1}^2 \sum_{k=1}^{n_i} \sum_{s=1}^2 \left( \frac{\sqrt{N}}{4n_i} c_{is} Z_{iks} + \frac{\sqrt{N}}{4n_i} c_{js} Z_{ik}^{(-js)} \right) \\
&\quad - \left( \sum_{i=1}^2 \sum_{k=1}^{n_i} \sum_{s=1}^2 \frac{2}{n_i} \sqrt{N} c_{is} p_{is} \right),
\end{aligned}$$

wobei  $c_{is}$  für  $i, s = 1, 2$  die Einträge des Kontrastvektors  $\mathbf{c}$  sind und für jeden Kontrastvektor definitionsgemäß gilt  $\mathbf{c}\mathbf{1} = 0$ .

Mit den Vereinbarungen

$$\begin{aligned}
\zeta_{ik} &= \sum_{s=1}^2 \left( \frac{\sqrt{N}}{4n_i} c_{is} Z_{iks} + \frac{\sqrt{N}}{4n_i} c_{js} Z_{ik}^{(-js)} \right) \quad \text{und} \quad (\text{A.5}) \\
\mu_{is} &= \frac{\sqrt{4N}}{n_i} \sum_{s=1}^2 c_{is} p_{is}
\end{aligned}$$

folgt

$$\mathbf{cZ}_N = \sum_{i=1}^2 \sum_{k=1}^{n_i} (\zeta_{ik} - \mu_{is}).$$

Da die Variablen  $Z_{iks}$  und  $Z_{j\ell}^{(-is)}$  definitionsgemäß beschränkt und unabhängig sind und weiterhin nach Voraussetzung (V1) gilt  $N/n_i \leq N_0 < \infty$ , ist auch die Zufallsvariable  $\zeta_{ik}$  beschränkt und unabhängig. Um nun die Voraussetzungen für den Satz von Lindeberg-Feller A.2.6 zu erfüllen, muss lediglich noch gezeigt werden, dass die Varianz von 0 weg beschränkt ist, was aber nach Voraussetzung (V3) gewährleistet ist, und  $\zeta_{ik}$  mit wachsenden Stichprobenumfängen gegen Null streben. Dies ist mit Hilfe der Definition derselben aus (A.5) gültig.

Damit erhält man das gewünschte Ergebnis, dass  $\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p})$  bzw. die Vergleichsstatistik  $\mathbf{Z}_N$  asymptotisch normalverteilt ist mit Erwartungswert 0. Die Kovarianzmatrix  $\mathbf{W}_N$  wurde bereits im Vorfeld berechnet, und die Konsistenz von  $\widehat{\mathbf{W}}_N$  für  $\mathbf{W}_N$  beweist man mit den gleichen Grundlagen und Techniken wie unter  $H_0^F$ .

□

### Bestimmung der Quantile der $t$ -Verteilung für die Teststatistiken

Früher als die Approximation der Statistiken mit der Normalverteilung greift bei kleineren Stichprobenumfängen meist die Approximation mit der  $t$ -Verteilung. Nach den Techniken von SMITH (1936), WELCH (1938) und SATTERTHWAITTE (1946) werden für

die entsprechenden Quantile der  $t$ -Verteilung die Varianzen bzw. empirischen Varianzen von  $\sqrt{N}\mathbf{c} \int Gd\hat{\mathbf{F}}$  für  $H_0^F$  und von  $\sqrt{N}\mathbf{c}(\int Gd\hat{\mathbf{F}} - \int \mathbf{F}dG)$  für  $H_0^p$  benötigt. In DOM-HOF (2001) ist dieser Weg auch aufgezeigt, und es werden verschiedene Problematiken besprochen.

Sei  $Y_{iks} = G(X_{iks})$ . Aus

$$\begin{aligned} \text{Var}(\sqrt{N}\mathbf{c}_\lambda \int Gd\hat{\mathbf{F}}) &= \frac{N}{n_1} \text{Var}(Y_{111} + Y_{112}) + \frac{N}{n_2} \text{Var}(Y_{211} + Y_{212}) \\ &= \sum_{i=1}^2 \frac{N}{n_i} \sigma_i^2 \quad \text{mit} \\ \sigma_i^2 &= \text{Var}(Y_{i11} + Y_{i12}), \quad i = 1, 2 \end{aligned}$$

folgt unter Zuhilfenahme des zugehörigen Schätzers

$$\hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\Psi_{ik1} + \Psi_{ik2} - \bar{\Psi}_{i.1} - \bar{\Psi}_{i.2})^2 \quad (\text{A.6})$$

der Freiheitsgrad für  $U_n^{\lambda,F}$ :

$$\frac{(\sum_{i=1}^2 \hat{\sigma}_i^2/n_i)^2}{\sum_{i=1}^2 (\hat{\sigma}_i^2/n_i)^2/(n_i - 1)}.$$

Entsprechend folgt der Freiheitsgrad für  $U_n^{\pi,F}$  mit

$$\begin{aligned} \text{Var}(\sqrt{N}\mathbf{c}_\pi \int Gd\hat{\mathbf{F}}) &= \frac{N}{n_1} \text{Var}(Y_{111} - Y_{112}) + \frac{N}{n_2} \text{Var}(Y_{211} - Y_{212}) \\ &= \sum_{i=1}^2 \frac{N}{n_i} \tau_i^2 \quad \text{mit} \\ \tau_i^2 &= \text{Var}(Y_{i11} - Y_{i12}), \quad i = 1, 2 \end{aligned}$$

und des zugehörigen Schätzers

$$\hat{\tau}_i^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\Psi_{ik1} - \Psi_{ik2} + \bar{\Psi}_{i.1} - \bar{\Psi}_{i.2})^2$$

zu:

$$\frac{(\sum_{i=1}^2 \hat{\tau}_i^2/n_i)^2}{\sum_{i=1}^2 (\hat{\tau}_i^2/n_i)^2/(n_i - 1)}.$$

Der Freiheitsgrad der  $t$ -Verteilung der Statistik  $U_n^{\delta,F}$  ist genauso, da die Varianz von  $\text{Var}(\sqrt{N}\mathbf{c}_\delta \int Gd\hat{\mathbf{F}})$  ebenfalls gleich  $\sum_{i=1}^2 \frac{N}{n_i} \tau_i^2$  ist.

Für  $H_0^p$  erhält man Varianzschätzer, die sich aufgrund des zusätzlichen Terms in der Vergleichsstatistik etwas komplizierter gestalten. Aus

$$\begin{aligned} & \text{Var} \left( \sqrt{N} \mathbf{c}_\lambda \left( \int G d\hat{\mathbf{F}} + \int \mathbf{F} d\hat{G} \right) \right) \\ &= \frac{N}{16n_1} \text{Var} (Z_{111} + Z_{112} - Z_{11}^{(-21)} - Z_{11}^{(-22)}) \\ & \quad + \frac{N}{16n_2} \text{Var} (Z_{211} + Z_{212} - Z_{21}^{(-11)} - Z_{21}^{(-12)}) \\ &= \frac{N}{16} \sum_{i=1}^2 \frac{\sigma_{i,p}^2}{n_i} \quad \text{mit} \\ \sigma_{i,p}^2 &= \text{Var} (Z_{i11} + Z_{i12} - Z_{j1}^{(-i1)} - Z_{j1}^{(-i2)}), \quad i = 1, 2 \quad \text{und} \quad j \neq i \end{aligned}$$

folgt unter Zuhilfenahme des zugehörigen Schätzers

$$\begin{aligned} \hat{\sigma}_{i,p}^2 &= \frac{1}{(n_i - 1)\tilde{n}^2} \sum_{k=1}^{n_i} \left( -\Psi_{ik1} - \Psi_{ik2} + 2\Psi_{ik}^{(-j1)} + 2\Psi_{ik}^{(-j2)} \right. \\ & \quad \left. + \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2} - 2\bar{\Psi}_i^{(-j1)} - 2\bar{\Psi}_i^{(-j2)} \right)^2 \quad \text{mit} \quad j \neq i \end{aligned}$$

der Freiheitsgrad für die Approximation mit der  $t$ -Verteilung von  $U_n^{\lambda,p}$ :

$$\frac{(\sum_{i=1}^2 \hat{\sigma}_{i,p}^2/n_i)^2}{\sum_{i=1}^2 (\hat{\sigma}_{i,p}^2/n_i)^2/(n_i - 1)}$$

Entsprechend ergibt sich der Freiheitsgrad für  $U_n^{\pi,p}$  zu

$$\frac{(\sum_{i=1}^2 \hat{\tau}_{i,p}^2/n_i)^2}{\sum_{i=1}^2 (\hat{\tau}_{i,p}^2/n_i)^2/(n_i - 1)}$$

mit

$$\begin{aligned} \hat{\tau}_{i,p}^2 &= \frac{1}{(n_i - 1)\tilde{n}^2} \sum_{k=1}^{n_i} \left( \Psi_{ik1} - \Psi_{ik2} - 2\Psi_{ik}^{(-j1)} + 2\Psi_{ik}^{(-j2)} \right. \\ & \quad \left. - \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2} + 2\bar{\Psi}_i^{(-j1)} - 2\bar{\Psi}_i^{(-j2)} \right)^2 \quad \text{mit} \quad j \neq i \end{aligned}$$

und der Freiheitsgrad für  $U_n^{\delta,p}$  zu

$$\frac{(\sum_{i=1}^2 \hat{v}_{i,p}^2/n_i)^2}{\sum_{i=1}^2 (\hat{v}_{i,p}^2/n_i)^2/(n_i - 1)}$$

mit

$$\begin{aligned} \hat{v}_{i,p}^2 &= \frac{1}{(n_i - 1)\tilde{n}^2} \sum_{k=1}^{n_i} \left( \Psi_{ik1} - \Psi_{ik2} + 2\Psi_{ik}^{(-j1)} - 2\Psi_{ik}^{(-j2)} \right. \\ & \quad \left. - \bar{\Psi}_{i.1} + \bar{\Psi}_{i.2} - 2\bar{\Psi}_i^{(-j1)} + 2\bar{\Psi}_i^{(-j2)} \right)^2 \quad \text{mit} \quad j \neq i. \end{aligned}$$



Folgender Satz ist wichtig, um zu zeigen, dass die hergeleiteten Statistiken eine sinnvolle Erweiterung der bisher bestehenden Theorie für das lineare Modell sind. In dieser Arbeit wurde sich im Wesentlichen auf das Modell von LEHMACHER (1987) bezogen. In diesem Satz wird jedoch nicht die Normalverteilung, sondern lediglich eine streng monotone Verteilungsfunktion vorausgesetzt.

**Satz A.3.18 (Monotonie von  $p^\lambda, p^\pi$  und  $p^\delta$ )** *Im linearen Marginalmodell mit streng monotoner Verteilungsfunktion gelten folgende Aussagen:*

1.  $p^\lambda(\lambda, \pi, \delta)$  ist streng monoton in  $\lambda$ , und  $p^\lambda$  ist genau dann gleich 0, wenn  $\lambda = 0$  ist.
2.  $p^\pi(\lambda, \pi, \delta)$  ist streng monoton in  $\pi$ , und  $p^\pi$  ist genau dann gleich 0, wenn  $\pi = 0$  ist.
3.  $p^\delta(\lambda, \pi, \delta)$  ist streng monoton in  $\delta$ , und  $p^\delta$  ist genau dann gleich 0, wenn  $\delta = 0$  ist.

**Beweis:**

Sei  $\epsilon_{iks} \sim F(x)$  für  $i = 1, 2, s = 1, 2$  und  $k = 1, \dots, n_i$ , wobei  $F$  eine streng monotone Verteilungsfunktion ist. Im linearen Modell ergeben sich die einzelnen Verteilungsfunktionen aufgrund des Lokationsmodells und der Additivität der Parameter wie folgt:

$$\begin{aligned} F_{11}(x) &= F(x - \mu - \Phi - \pi), \\ F_{12}(x) &= F(x - \mu + \Phi + \pi - \lambda), \\ F_{21}(x) &= F(x - \mu + \Phi - \pi), \\ F_{22}(x) &= F(x - \mu - \Phi + \pi + \lambda). \end{aligned}$$

Mit den folgenden Definitionen können die relativen Residual-, Zeit- und Cross-Over-Effekte ausgedrückt werden:

$$\begin{aligned} p_s^\lambda &= \int F_{1s} dF_{2s}, \\ p_i^\pi &= \int F_{i1} dF_{i2}, \\ d_1 &= \int F_{12} dF_{21} \text{ und} \\ d_2 &= \int F_{11} dF_{22}. \end{aligned}$$

Exemplarisch erhält man den relativen Residualeffekt in der Form

$$\begin{aligned} p^\lambda &= \frac{1}{2} (p_{11} + p_{12} - p_{21} - p_{22}) \\ &= \frac{1}{2} \left( \int G dF_{11} + \int G dF_{12} - \int G dF_{21} - \int G dF_{22} \right) \\ &= \frac{1}{2} \left( -\frac{1}{2} \int F_{11} dF_{21} - \frac{1}{2} \int F_{11} dF_{22} - \frac{1}{2} \int F_{12} dF_{21} - \frac{1}{2} \int F_{12} dF_{22} + 4 \right) \\ &= \frac{1}{4} (2 - p_1^\lambda - p_2^\lambda - d_1 - d_2). \end{aligned}$$

Bezeichne  $K(x) := \int F(x+y)dF(y)$  mit der Eigenschaft  $K(x) + K(-x) = 1$ . Diese Eigenschaft ergibt sich mit Hilfe der Transformationsformel und den Regeln der partiellen Integration:

$$\begin{aligned} K(x) + K(-x) &= \int F(x+y)dF(y) + \int F(-x+y)dF(y) \\ &= \int F(x+y)dF(y) + \int F(y)dF(x+y) \\ &= \int F(x+y)dF(y) + 1 - \int F(x+y)dF(y) = 1. \end{aligned}$$

$K(x)$  ist streng monoton steigend, denn: Sei  $x_1 < x_2$ . Dann folgt

$$K(x_2) - K(x_1) = \int [F(x_2+y) - F(x_1+y)] dF(y) > 0$$

aufgrund der Eigenschaften von  $F$  und BILLINGSLEY (1995), Theorem 15.2.(ii).  
Dann folgt

$$\begin{aligned} p_1^\lambda &= \int F_{11}dF_{21} \\ &= \int F(x - \mu - \Phi - \pi)dF(x - \mu + \Phi - \pi) \\ &= \int F(-2\Phi + x - \mu + \Phi - \pi)dF(x - \mu + \Phi - \pi) \\ &= K(-2\Phi). \end{aligned}$$

Ebenso ergeben sich  $p_2^\lambda = K(2\Phi - 2\lambda)$ ,  $d_1 = K(2\pi - \lambda)$  und  $d_2 = K(-2\pi - \lambda)$ . Mit diesen Vorüberlegungen liegt der Nachweis der strengen Monotonie von  $p^\lambda$  in  $\lambda$  nahe:

$$\begin{aligned} p^\lambda &= \frac{1}{4} (-p_1^\lambda - p_2^\lambda - d_1 - d_2) \\ &= \frac{1}{4} (-K(-2\Phi) - K(2\Phi - 2\lambda) - K(2\Phi - \lambda) - K(-2\Phi - \lambda) + 2). \end{aligned}$$

Aus dieser Darstellung folgt mit Hilfe der strengen Monotonie von  $K(x)$  die Behauptung der strengen Monotonie von  $p^\lambda$  in  $\lambda$ .

Um die Äquivalenz der Hypothesen  $p^\lambda = 0$  und  $\lambda = 0$  nachzuweisen, genügt es zu zeigen:

$\lambda = 0 \implies p^\lambda = 0$ : Dazu gilt zunächst einmal

$$\begin{aligned} p_1^\lambda + p_2^\lambda &= K(-2\Phi) + K(2\Phi - 2\lambda) \\ &= K(-2\Phi) - K(2\lambda - 2\Phi) + 1. \end{aligned} \tag{A.7}$$

Aus der Eigenschaft

$$\begin{aligned} \lambda = 0 &\implies -2\Phi = 2\lambda - 2\Phi \\ &\implies K(-2\Phi) - K(2\lambda - 2\Phi) = 0 \end{aligned}$$

folgt mit Hilfe von (A.7)

$$p_1^\lambda + p_2^\lambda = 1.$$

Für  $p^\pi$  und  $p^\delta$  verlaufen die Beweise analog.

□

# Literaturverzeichnis

- [1] M. G. AKRITAS UND S. F. ARNOLD (1994). Fully nonparametric hypotheses for factorial designs I: Multivariate repeated measures designs *Journal of the American Statistical Association* 89, S. 336-343.
- [2] H. BAUER (1991). Wahrscheinlichkeitstheorie. *W. de Gruyter, Berlin*.
- [3] R. L. BERGER (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* 24, S. 295-300.
- [4] P. BILLINGSLEY (1995). Probability and Measure. *Wiley, New York*.
- [5] B. W. BROWN (1980). The crossover experiment for clinical trials. *Biometrics* 36, S. 69-79.
- [6] E. BRUNNER (1997). Theory of Rank Tests in Factorial Designs. *Vorlesungsskript, Universität Göttingen*.
- [7] E. BRUNNER UND F. LANGER (1999). Nichtparametrische Analyse longitudinaler Daten. *Oldenbourg Verlag, München*.
- [8] E. BRUNNER, U. MUNZEL UND L. PURI (1999). Rank-Score Tests in Factorial Designs with Repeated Measures. *Journal of Multivariate Analysis* 70, S. 286-317.
- [9] E. BRUNNER UND N. NEUMANN (1987). Non-Parametric Methods for the 2-Period-Cross-Over-Design under weak Model Assumptions. *Biometrical Journal* 29, S. 907-920.
- [10] E. BRUNNER UND L. PURI (1996). Nonparametric Methods in Design and Analysis of Experiments. *Handbook of Statistics* 13, North-Holland, S. 631-703.
- [11] W.G. COCHRAN (1939). Long-term agricultural experiments (with discussion). *Journal of the Royal Statistical Society, B*, 6, 104-48.
- [12] R.G. CORNELL (1980). Evaluation of bioavailability data using nonparametric statistics. *Drug Absorption and Disposition: Statistical Considerations*, S. 51-57.
- [13] D. R. COX (1958). Planning of Experiments. *Wiley, New York*.
- [14] A. CUSHNY UND A. PEEBLES (1905). The action of optical isomers . *Journal of Physiology* 32, S. 510ff..

- [15] S. DOMHOF (1999). Rangverfahren mit unbeschränkten Score-Funktionen in faktoriellen Versuchsplänen. *Diplomarbeit, Universität Göttingen*
- [16] S. DOMHOF (2001). Nichtparametrische Relative Effekte. *Dissertation, Universität Göttingen*
- [17] R. A. FISHER (1925). Statistical Methods for Research Workers. *Statistical Methods, Experimental Design and Scientific Inference, Oxford University Press, Oxford.*
- [18] P. R. FREEMAN (1989). The performance of the two-stage analysis of two-treatment, two-period cross-over trials. *Statistics in Medicine, 8, 1421-32.*
- [19] B. V. GNEDENKO (1962). The Theory of Probability. *Chelsea Publishing, New York.*
- [20] A. GRIEVE UND S. SENN (1998). Estimating Treatment Effects in Clinical Cross-Over Trials. *Journal of Biopharmaceutical Statistics 8, 2, S. 191-234.*
- [21] J. E. GRIZZLE (1965). The two-period change-over design and its use in clinical trials. *Biometrics 21, S. 469-480.*
- [22] D. HAUSCHKE (1999). Biometrische Methoden zur Planung und Auswertung von Sicherheitsstudien. *Habilitationsschrift Biometrie, Universität Dortmund.*
- [23] D. HAUSCHKE, V. E. STEINIANS UND E. DILETTI (1990). A Distribution-free procedure for the statistical analysis of bioequivalence studies. *International Journal of Clinical Pharmacology, Therapy and Toxicology 28, 2, S. 72-78.*
- [24] M. HILLS UND P. ARMITAGE (1979). The two-period crossover clinical trial. *British Journal of Clinical Pharmacology 8, S. 7-20.*
- [25] J. HOFFMANN-JORGENSEN (1994). Probability with a view toward statistics. *Chapman & Hall, New York.*
- [26] B. E. HUITEMA (1980). The Analysis of Covariances and Alternatives. *Wiley, New York.*
- [27] M. G. KENWARD UND B. JONES (1989). Design and Analysis of Cross-Over-Trials. *Chapman and Hall, London.*
- [28] R. E. KIRK (1982). Experimental Designs. *Brooks/Cole, 2nd edition.*
- [29] G. G. KOCH (1972). The use of nonparametric methods in the statistical analysis of the two-period change-over design. *Biometrics 28, S. 577-94.*
- [30] B. KULLE (1999). Das nichtparametrische Behrens-Fisher-Problem. *Diplomarbeit, Universität Göttingen*
- [31] F. LANGER (1998). Die Berücksichtigung von Kovariablen im nichtparametrischen gemischten Modell. *Dissertation, Universität Göttingen.*
- [32] W. LEHMACHER (1987). Verlaufskurven und Cross-Over. *Springer-Verlag, Berlin.*

- [33] J. LIEBIG (1847). Chemistry and its Application to Agriculture and Physiology. *Taylor and Walton, London.*
- [34] M. LOÈVE (1977). Probability Theory I. *Springer, Berlin, 4th edition.*
- [35] U. MUNZEL (1996). Multivariate nichtparametrische Verfahren für feste Faktoren in mehrfaktoriellen Versuchsanlagen. *Dissertation, Universität Göttingen.*
- [36] U. MUNZEL UND D. HAUSCHKE (2001). A nonparametric test for proving noninferiority in clinical trials with ordered categorical data. *preprint.*
- [37] H.I. PATEL (1983). Use of Baseline Measurements in the Two-Period Cross-Over Design. *Communications in Statistics - Theory and Methods 12, S. 2693-712.*
- [38] D.A. RATKOWSKI, M.A. EVANS UND ALLDREDGE (1993). Cross-Over Experiments, Design, Analysis and Application. *Marcel Dekker, New York, 1993.*
- [39] J. RÖHMEL (1998). Therapeutic equivalence investigations: statistical considerations. *Statistics in medicine, 17, S. 1703-1714.*
- [40] D. SALSBURG (1999). Why Analysis of Variance is Inappropriate for Multiclinic Trials. *Controlled Clinical Trials, 20, S. 453-468.*
- [41] F. E. SATTERTHWAITE (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2, S. 110-114.*
- [42] B. SCHNEIDER (1983). Crossover Designs and Repeated Measurements. *Neuropsychobiology 10, S. 49-55.*
- [43] S. SENN (1993). Cross-Over Trials in Clinical Research. *Wiley, New York.*
- [44] A. SIEMER (1999). Die Berücksichtigung von heterogen verteilten Kovariablen in einem nichtparametrischen Modell. *Diplomarbeit, Universität Göttingen.*
- [45] H. F. SMITH (1936). The problem of comparing the results of two experiments with unequal errors. *Journal of the Council for Scientific and Industrial Research (Australia), 9, S. 211-212.*
- [46] STUDENT (1908). The probable Error of a Mean. *Biometrika 6, S. 1-25.*
- [47] H. J. TRAMPISCH (1995). Editorial. *Biometrics in the Pharmaceutical Industry, 7, S. 1f.*
- [48] V. S. VARADARAJAN (1958). A useful Convergence Theorem. *Sankhya, 20, S. 221f.*
- [49] B. L. WELCH (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29, S. 350-362.*
- [50] S. WELLEK UND B. HAMPEL (1999). A distribution-free two-sample equivalence test allowing for tied observations. *Biometrical Journal, 41, S. 171-186.*

- [51] W. J. WESTLAKE (1988). Bioavailability and bioequivalence of pharmaceutical formulations. *Biopharmaceutical Statistics for Drug Development of pharmaceutical formulations*, Marcel Dekker, New York, S. 329-352.
- [52] A. R. WILLAN UND J. L. PATER (1986a). Carryover and the Two-Period Crossover Clinical Trial. *Biometrics* 42, S. 593-599.
- [53] A. R. WILLAN UND J. L. PATER (1986b). Using baseline measurements in the two-period crossover clinical trial. *Controlled Clinical Trials* 7, S. 282-289.
- [54] H. ZIMMERMANN UND V. W. RAHLFS (1980). Model Building and Testing for the Change-Over Design. *Biometrical Journal* 22, S. 197-210.







## Lebenslauf

Am 21. Januar 1975 wurde ich als erstes Kind von Gudrun Kulle, geb. Bartsch, und Gerhard Kulle in Stuttgart geboren.

In den Jahren 1981 bis 1985 besuchte ich nacheinander die Grundschulen in Stuttgart, Höxter und Witzenhausen. Nach zwei Jahren Förderstufe 1985-87 in Witzenhausen wechselte ich dann auf den gymnasialen Zweig der Rhenanus-Schule in Bad Sooden-Allendorf, wo ich nach dem Besuch der Mittelstufe von 1987-91 und der gymnasialen Oberstufe von 1991-94 meine Schullaufbahn mit der allgemeinen Hochschulreife beendete.

Zum Wintersemester 1994/95 nahm ich mein Studium der Fachrichtung Wirtschaftsmathematik an der Georg-August-Universität Göttingen auf. Die Diplom-Vorprüfung legte ich im April 1996 ab. Vom Wintersemester 1996/97 bis einschließlich Sommersemester 1998 war ich als studentische Hilfskraft am Institut für Wirtschaftsinformatik II beschäftigt. Im Juli 1997 nahm ich meine Tätigkeit als studentische Hilfskraft in der Abteilung Medizinische Statistik auf, die ich bis zum Abschluss meines Studiums im Oktober 1999 fortführte. Während dieser Zeit fertigte ich unter Anleitung von Prof. Dr. Brunner die Diplomarbeit mit dem Titel *Nichtparametrisches Behrens-Fisher-Problem* an.

Seit November 1999 bin ich als wissenschaftliche Mitarbeiterin an der Abteilung Medizinische Statistik angestellt. Des Weiteren arbeitete ich seither an der vorliegenden Arbeit, die der Leiter der Abteilung, Prof. Dr. Brunner, betreute.