ANWENDUNG VON MULTIPLEN TESTPROZEDUREN IM LINEAREN REGRESSIONSMODELL

Raimund ALT*

Forschungsbericht/
Research Memorandum No. 237

Dezember 1986

* Scholar am Institut für Höhere Studien, Abteilung Mathematische Methoden und Computerverfahren von 1984 - 1986.

VORWORT

In der 1984 im '*Handbook of Econometrics*' erschienenen Arbeit von Savin "Multiple Hypotheses Testing" wurde der Versuch unternommen, die zahlenmäßig sicher nicht sehr umfangreiche ökonometrische Literatur über das simultane Testen linearer Hypothesen zusammenzufassen. Im Mittelpunkt der Darstellung stand dabei die Anwendung des Bonferroni-Tests und des Scheffè-Tests im klassischen linearen Regressionsmodell.

Erstaunlicherweise fehlt in dieser Arbeit jeglicher Hinweis auf die sog. sequentiell verwerfenden Testprozeduren, deren Entwicklung in den 70er Jahren vor allem auf die Arbeiten von Holm(1979) und Marcus/Peritz/Gabriel (1976) zurückzuführen ist und die zu einem beträchtlichen Aufschwung der Theorie multipler Testprozeduren, die wohl zu den Domänen der Biometrie und Medizinischen Statistik gehört, führte.
Entsprechende Hinweise auf derartige Testprozeduren fehlen insbesondere auch in den Arbeiten von Phillips(1984), Kiviet/ Phillips(1985) und Phillips/McCabe(1985), in denen die Anwendung multipler Spezifikationstests behandelt wird.
Diese neuartigen Testverfahren scheinen offensichtlich noch keinen Eingang in die ökonometrische Literatur gefunden zu haben, obwohl sich hier durchaus zahlreiche Anwendungsmöglichkeiten bieten würden.

In der vorliegenden Arbeit habe ich nun versucht, diese Testprozeduren auf einige spezielle ökonometrische Testprobleme anzuwenden.
Dabei ergaben sich insgesamt folgende Schwerpunkte:

(1) Einführung in die Grundlagen multipler Testprobleme.

(2) Darstellung des Abschlußprinzips.

(3) Anwendung von sequentiell verwerfenden Testprozeduren auf elementare lineare Hypothesen.

(4) Präsentation der Ergebnisse einer Reihe von Monte Carlo Simulationen, bei denen unter anderem reale ökonomische Zeitreihen verwendet wurden.

(5) Anwendung von sequentiell verwerfenden Testprozeduren auf spezielle hierarchische Testprobleme (nested hypotheses).

An dieser Stelle möchte ich mich bei all denen bedanken, die durch zahlreiche Hinweise und Vorschläge wesentlich zur Fertigstellung dieser Arbeit beigetragen haben.
Mein besonderer Dank gilt dabei vor allem meinem Betreuer Herrn Dr.R.Kunst, Herrn Prof.P.Hackl(Wien) und Herrn Prof.P.Bauer(Köln).
Herrn Dr.A.Jäger danke ich sehr herzlich für seine Unterstützung bei der Zusammenstellung der Simulationsdesigns.
Nicht zuletzt möchte ich mich auch bei Herrn Prof. W.Krämer (Hannover) bedanken, der mich auf mögliche Anwendungen multipler Testprozeduren in der Ökonometrie aufmerksam machte.

ZUSAMMENFASSUNG

Die vorliegende Arbeit behandelt die Anwendung multipler Testprozeduren im linearen Regressionsmodell.
Nach einer Einführung in die Grundlagen multipler Testprobleme und einer ausführlichen Darstellung des Abschlußprinzips wird gezeigt, auf welche Weise Abschlußtests zur Überprüfung elementarer linearer Hypothesen verwendet werden können.
Für empirische Vergleichszwecke wurden eine Reihe von Simulationsexperimenten durchgeführt, deren Ergebnisse eine deutliche Überlegenheit der sequentiell verwerfenden Testprozeduren gegenüber herkömmlichen multiplen Tests wie etwa dem Bonferroni-Test oder dem Scheffè-Test aufweisen.
Schließlich wird gezeigt, wie das Abschlußprinzip bei Vorliegen sog. hierarchischer Teststrukturen durchgeführt werden kann. Dabei wird auch eine verallgemeinerte Teststruktur vorgestellt, die einige bereits bekannte Teststrukturen als Spezialfälle enthält.

SUMMARY

This paper deals with the application of multiple test procedures in the linear regression model.
The first part of the paper comprises the foundations of multiple test problems and an exact presentation of the so called closure principle. Then it will be shown how to apply closed test procedures to elementary linear hypotheses.
A number of simulation experiments were performed to compare different test procedures empirically. The results show the superiority of the sequentially rejective test procedures over conventional multiple tests like the Bonferroni test or the Scheffè test.
Finally the application of the closure principle to hierarchical test structures is demonstrated. In this context a generalized test structure is introduced, which contains some known test structures as special cases.

INHALT

I. EINLEITUNG.

Bei der Anwendung statistischer Tests treten häufig Situationen auf, in denen man nicht so sehr an der Überprüfung einer einzelnen Hypothese, als vielmehr an der "gleichzeitigen" (simultanen) Überprüfung mehrerer Hypothesen interessiert ist.
Betrachtet man z.B. bei einem linearen Regressionsmodell von der Form

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 + u_t \qquad t=1,\dots,T$$

das Testproblem

$$\begin{aligned} H_0&: \beta_1=\beta_2=0 \\ H_1&: \beta_1 \neq 0 \text{ oder } \beta_2 \neq 0 \quad , \end{aligned} \tag{1}$$

so ist im Falle einer Ablehnung der Hypothese H_0 durch den entsprechenden F-Test nicht klar, welche der Hypothesen

$$H_{01}: \beta_1=0 \quad \text{bzw.} \quad H_{02}: \beta_2=0$$

ebenfalls abzulehnen ist.
Ist man nun vor allem an der Überprüfung der Hypothesen H_{01} und H_{02} interessiert, so erscheint es sinnvoller, an Stelle des (einfachen) Testproblems (1) das entsprechende multiple Testproblem (2) heranzuziehen:

$$\begin{aligned} H_{01}&: \beta_1=0 \qquad H_{11}: \beta_1 \neq 0 \\ H_{02}&: \beta_2=0 \qquad H_{12}: \beta_2 \neq 0 \quad . \end{aligned} \tag{2}$$

Die Überprüfung der beiden Hypothesen H_{01} und H_{02} erfolgt dann jeweils durch einen einfachen t-Test.

Diese Vorgangsweise des mehrfachen oder multiplen Testens (zwei oder mehr Hypothesen werden auf der Grundlage eines einzigen Modells getestet) ist allerdings vom statistischen Standpunkt aus nicht ganz unproblematisch.
Überprüft man nämlich die beiden Hypothesen H_{01} und H_{02} jeweils mit einem Niveau-α-Test, so ist zwar die Wahrscheinlichkeit, bei einer gegebenen Hypothese einen Fehler 1.Art zu begehen (höchstens) gleich α, aber die Wahrscheinlichkeit, bei wenigstens

einer der beiden Hypothesen einen derartigen Fehler zu begehen, ist i.a. größer als α.
Dieser Umstand aber stellt eine eindeutige Verletzung des klassischen Testprinzips dar, nach dem für das gesamte Testproblem ein vorgegebenes Signifikanzniveau α einzuhalten ist.
Eine korrekte Behandlung des multiplen Testproblems (2) würde also zunächst ein "globales" Signifikanzniveau α erfordern. Erst dann sollten die Signifikanzniveaus für die einfachen Testprobleme $H_{01} : H_{11}$ bzw. $H_{02} : H_{12}$ festgelegt werden (in Abhängigkeit von α).
(Hier ist eine gewisse Analogie zur Theorie der simultanen Gleichungssysteme zu erkennen: Während dort mehrere lineare Gleichungen simultan zu lösen sind, sind bei multiplen Testproblemen mehrere Hypothesen simultan zu testen.)

Da die gemeinsame Verteilung der Teststatistiken i.a. nicht bekannt sein dürfte und somit (zur Berechnung der Signifikanzniveaus der einfachen Testprobleme) häufig auf entsprechende Abschätzungen, wie z.B. die Bonferroni-Ungleichung, zurückgegriffen werden muß, wird das vorgegebene Signifikanzniveau α allerdings nur selten exakt erreicht werden. Insbesondere wird die Verwendung unabhängiger Tests nur in Ausnahmefällen möglich sein.

Eine interessante Variante des Testproblems (2), bei der grundsätzlich jede Hypothese zum Niveau α getestet werden kann, ergibt sich durch Kombination der beiden obigen Testprobleme. Das dadurch entstehende sequentielle Testverfahren kann wie folgt beschrieben werden:

Man überprüfe zunächst die Hypothese $H_0: \beta_1=\beta_2=0$.
Wird H_0 nicht abgelehnt, so gilt auch keine der Hypothesen H_{01} und H_{02} als abgelehnt und das Testverfahren ist beendet. Andernfalls sind diese beiden Hypothesen explizit mit Hilfe von t-Tests zu überprüfen.

Dieses multiple Testverfahren besitzt folgende Eigenschaft:
Die Wahrscheinlichkeit dafür, wenigstens eine der wahren Nullhypothesen abzulehnen, ist stets kleiner oder gleich α.
(Unabhängig davon, wieviele Nullhypothesen tatsächlich wahr sind!)

Dies ist übrigens ein einfaches Beispiel eines sogenannten

Abschlußtests. Dieser Test wird in Kapitel III. noch ausführlich behandelt werden.

Derartige multiple Testprobleme treten bekanntlich in den verschiedensten statistischen "Disziplinen" auf, so vor allem in der Biometrie und in der Medizinischen Statistik (etwa: multiple Vergleiche bei der Überprüfung der Wirksamkeit verschiedener Medikamente). Und es sind dort auch seit einigen Jahrzehnten eine Reihe von Verfahren bekannt, die sich auf multiple Testprobleme anwenden lassen. (Einen Überblick über diese Verfahren findet man z.B. in Miller(1966,1981))

Durch die Entwicklung der sequentiell verwerfenden Testprozeduren (Abschlußtests) in den 70er Jahren gelten jedoch viele dieser "klassischen" Verfahren als überholt bzw. können durch sequentiell verwerfende Varianten verbessert werden.
Als grundlegend für die Entwicklung dieser Testprozeduren haben sich dabei besonders die Arbeiten von Gabriel(1969), Einot/Gabriel(1975), Naik(1975), Holm(1979) und Marcus/Peritz/Gabriel (1976) erwiesen.
Eine Zusammenfassung und übersichtliche Darstellung der wichtigsten Resultate dieser Arbeiten findet man in Sonnemann(1981 1982) oder Hommel(1984,1985).

Abgesehen von den erwähnten "linearen" Testproblemen gibt es noch eine ganze Reihe weiterer (multipler) Testprobleme, die insbesondere für ökonometrische Anwendungen von Interesse sind. So weisen z.B. Dhrymes et al.(1972) u.a. auf die Bedeutung sogenannter 'hierarchischer Testprobleme' hin. Dabei geht es um die sequentielle Überprüfung einer Folge hierarchisch angeordneter Hypothesen ("nested hypotheses").
Typische Testprobleme dieser Art treten etwa auf bei der Bestimmung des Grades einer polynomialen Regressionsfunktion oder bei der sequentiellen Durchführung von Spezifikationstests, wie sie z.B. in Mizon(1977), Phillips(1984), Kiviet/Phillips (1985) oder Phillips/McCabe(1985) beschrieben wird.
Das zugrundeliegende Testverfahren, das im wesentlichen auf dem Ansatz von Anderson(1962,1971) beruht, wird z.B. von Phillips(1984), auf der Grundlage eines einfachen "switching regression" -Modells

$$y_1 = X_1 \beta_1 + u_1 \qquad 1 \leq t \leq T_1$$

$$y_2 = X_2\beta_2 + u_2 \qquad\qquad T_1+1 \le t \le T$$

mit $u_t = \rho u_{t-1} + \varepsilon_t$ für $t=1,...,T$, $Var(u_t)=\sigma_1^2$ für $t=1,...,T_1$ und $Var(u_t)=\sigma_2^2$ für $t=T_1+1,..,T$.

auf das folgende multiple Testproblem angewandt:

$$H_{01}: \rho=0$$
$$H_{02}: \rho=0,\ \sigma_1^2=\sigma_2^2$$
$$H_{03}: \rho=0,\ \sigma_1^2=\sigma_2^2,\ \beta_1=\beta_2 \quad .$$

(Eine ausführliche Darstellung dieses Testverfahrens sowie der sequentiell verwerfenden Variante von Bauer/Hackl(1986) findet man in Kapitel V.)

Ein weiteres hierarchisches Testproblem erhält man, wenn man z.B. im Modell

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 + u_t \qquad\qquad t=1,...,T$$
$$\text{mit } u_t = \rho u_{t-1} + \varepsilon_t$$

die Nullhypothesen

$$H_{01}: \rho=0$$
$$H_{02}: \rho=0, \beta_1=0 \qquad\qquad H_{03}: \rho=0, \beta_2=0$$

betrachtet.
Wie schon beim Testproblem (2), so stellt sich auch bei den hierarchischen Testproblemen die Frage nach der "Aufteilung" der Signifikanzniveaus auf die einzelnen Hypothesentests.
(Siehe dazu Kapitel V.)

Die bereits erwähnten sequentiell verwerfenden Testprozeduren scheinen in der Ökonometrie noch weitgehend unbekannt zu sein. Insbesondere fehlt in den Arbeiten von Mizon(1977), Phillips(1984) usw. sowie bei dem Übersichtsartikel von Savin(1984) über das multiple Testen linearer Hypothesen jeglicher Hinweis auf derartige Verfahren .
In den folgenden Kapiteln sollen daher vor allem die theoretischen Grundlagen dieser Testprozeduren sowie einige typische ökonometrische Anwendungen behandelt werden.

II. GRUNDBEGRIFFE DER TESTTHEORIE.

Im Mittelpunkt dieses Kapitels stehen die Grundbegriffe multipler Testprozeduren (multiples Testproblem, multipler Test, Test zum globalen Niveau α, Test zum multiplen Niveau α, Kohärenz, Konsonanz). Obwohl bereits im vorigen Kapitel von einigen dieser Begriffe Gebrauch gemacht wurde, ist es für die weiteren Ausführungen erforderlich, die einzelnen Begriffe durch möglichst exakte Definitionen einzuführen.

Da sich multiple Testprozeduren aus einfachen Tests zusammensetzen, sollen zunächst die wichtigsten Grundbegriffe einfacher Testprobleme wiederholt werden.

1. Einfache Testprobleme.

Ausgangspunkt eines jeden statistischen Tests ist die Durchführung eines sogenannten "statistischen Experiments".
Die Grundlage eines solchen Experiments bildet der Vektor $X = (X_1, \ldots, X_n)$ der zu beobachtenden Zufallsvariablen. Die einzelnen Realisationen $x = (x_1, \ldots, x_n)$ des Zufallsvektors X werden Stichproben genannt. Bezeichnet man den Stichprobenraum (d.h. die Menge aller möglichen Stichproben) mit S und die zu einem Parameter θ des Parameterraumes Γ zugehörige Verteilung mit P_θ, dann läßt sich ein statistisches Experiment kurz in der Form $(S, \{P_\theta: \theta \in \Gamma\})$ schreiben.
Unter einem statistischen Raum versteht man dann das Tripel $(S, \mathcal{B}, \{P_\theta: \theta \in \Gamma\})$, wobei $\mathcal{B}$ die zugrundeliegende σ-Algebra der Ereignisse bedeutet.

Eine "Testsituation" zeichnet sich nun dadurch aus, daß eine oder mehrere Annahmen (Hypothesen) bezüglich der (zumindest teilweise) unbekannten Verteilung von X zu überprüfen ist/sind. Insbesondere spricht man von einem einfachen Testproblem, falls nur eine einzige Hypothese H_0 zu überprüfen ist. Diese Hypothese, die üblicherweise als Nullhypothese bezeichnet wird, ist stets eine nichtleere Teilmenge des Parameterraumes Γ, mit $H_0 \neq \Gamma$.
Ein einfaches Testproblem wird üblicherweise mit

$$H_0 : H_1$$

abgekürzt, wobei $H_1 := \Gamma - H_0$ die entsprechende Alternativhypothese darstellt. Definitionsgemäß bilden dann H_0 und H_1 eine Partition (Zerlegung) des Parameterraumes Γ, d.h.

$$\begin{aligned} H_0, H_1 &\subset \Gamma \\ H_0, H_1 &\neq \emptyset \\ H_0 \cap H_1 &= \emptyset \\ H_0 \cup H_1 &= \Gamma \end{aligned}$$

Der "wahre", aber unbekannte Parameter θ der Zufallsvariablen X liegt somit in genau einer der beiden Mengen H_0 bzw. H_1 .
Die eigentliche Aufgabe besteht jetzt darin, auf Grund einer Stichprobe $x \in S$ eine Entscheidung bezüglich H_0 zu treffen. Diese Entscheidung beruht (zusammen mit einer vorgegebenen Irrtumswahrscheinlichkeit α) auf einem geeigneten statistischen Test.

Def.: Unter einem (einfachen) statistischen Test versteht man eine (meßbare) Abbildung $\delta : S \to \{0,1\}$.

Dabei bedeutet "meßbar" insbesondere, daß die Mengen $\{x: \delta(x)=0\}$ und $\{x: \delta(x)=1\}$ als Ereignisse vorausgesetzt werden.

Das Ergebnis eines derartigen Tests ist wie folgt zu interpretieren:
Gilt $\delta(x)=1$, so ist die Hypothese H_0 (d.h. die Annahme "$\theta \in H_0$") zu verwerfen, andernfalls, also bei $\delta(x)=0$, wird H_0 nicht verworfen.
Üblicherweise wird ein statistischer Test über eine entsprechende Teststatistik (eine auf dem Stichprobenraum S definierte reellwertige Zufallsvariable T) definiert, so z.B. in der Form

$$\delta(x) = \begin{cases} 1 & T(x) \geq c \\ 0 & T(x) < c \end{cases} \qquad (3)$$

Die Hypothese H_0 wird also genau dann abgelehnt, wenn T(x) einen bestimmten Wert (hier: c) "überschreitet".
Aus naheliegenden Gründen wird die Menge $\{\delta=0\} := \{x: \delta(x)=0\}$ als Annahmebereich und $\{\delta=1\} := \{x: \delta(x)=1\}$ als Ablehnungsbereich oder kritischer Bereich des Tests δ bezeichnet.
Der Zusammenhang zwischen einem statistischen Test und der vorgegebenen Irrtumswahrscheinlichkeit α wird durch die folgende Definition hergestellt:

<u>Def.:</u> Sei $0<\alpha<1$. Ein Test δ heißt <u>Test zum Niveau α</u> (bezüglich H_0) falls gilt:

$$\forall\, \theta\in H_0 \quad P_\theta(\delta=1):=P_\theta(\{\delta=1\}) \leq \alpha \quad . \qquad (4)$$

D.h. die Wahrscheinlichkeit dafür, daß ein derartiger Test die Nullhypothese (fälschlicherweise) ablehnt, ist höchstens gleich α. Das Auftreten eines solchen Ereignisses wird auch als Fehler 1.Art bezeichnet. (Von einem Fehler 2.Art spricht man, wenn die Nullhypothese falsch ist, aber dennoch nicht abgelehnt wird.)
Ist nun ein Niveau-α-Test auf einer Statistik T (wie in (3)) aufgebaut, so kann man an Stelle von (4) auch schreiben:

$$\forall\, \theta\in H_0 \quad P_\theta(T\geq c) \leq \alpha \quad .$$

Eine weitere Möglichkeit ergibt sich durch die Verwendung der p-Funktion:

<u>Def.:</u> Sei $\theta\in\Gamma$ beliebig. Unter der <u>p-Funktion von θ</u> versteht man die durch $p_\theta(x):=P_\theta(T\geq T(x))$ definierte Funktion $p_\theta:S\to[0,1]$.

($p_\theta(x)$ bezeichnet man dann auch als Überschreitungswahrscheinlichkeit von x bzw. als p-Wert von x)
Setzt man jetzt zusätzlich voraus, daß T stetig ist, so läßt sich zeigen (siehe z.B. Sonnemann(1981)), daß die beiden folgenden Aussagen äquivalent sind:

a) $\forall\, \theta\in H_0 \quad P_\theta(T\geq c_\alpha) = \alpha$
mit $c_\alpha:=\inf\{c\in\mathbb{R}: \ \theta\in H_0 \quad P_\theta(T\geq c) \leq \alpha\}$

b) $\forall\, \theta\in H_0 \quad P_\theta(p_\theta\leq\alpha) = \alpha$.

In den folgenden Kapiteln wird vor allem die Aussage b) verwendet werden.
Nach dieser kurzen Einführung in die Grundlagen einfacher Testprobleme folgen nun einige Begriffsbildungen für multiple Testprobleme.

2. Multiple Testprobleme.

Im Gegensatz zum einfachen Testproblem sind beim multiplen Testproblem endlich viele Nullhypothesen $H_{01},\dots,H_{0n}$ $(n\geq 2)$ vorgegeben und bezüglich jeder einzelnen Hypothese H_{0i} ist eine entsprechende Entscheidung zu treffen. Die Nullhypothesen sollen dabei folgende Voraussetzungen erfüllen:

$$H_{0i}\subset\Gamma \qquad i=1,\dots,n$$
$$H_{0i}\neq\emptyset,\ H_{0i}\neq\Gamma \qquad i=1,\dots,n$$
$$H_{0i}\neq H_{0j} \qquad \text{für } i\neq j$$

Die Anwendung mehrerer Tests auf ein und dieselbe Nullhypothese wird hier also ausgeschlossen.
Definiert man nun $H_{1i}:=\Gamma-H_{0i}$, so läßt sich das multiple Testproblem als eine Zusammenfassung von n einfachen Testproblemen schreiben:

$$\begin{array}{c} H_{01} : H_{11} \\ H_{02} : H_{12} \\ \vdots \\ H_{0n} : H_{1n} \end{array}$$

(H_{0i},H_{1i}) stellt dann wieder eine Zerlegung des Parameterraumes Γ in zwei disjunkte Teilmengen dar.
Aufgabe eines multiplen Tests ist es, die einzelnen Nullhypothesen "simultan" zu testen.

Def.: Ein multipler Test bzw. eine multiple Testprozedur ist eine Abbildung $\Phi:S\to\{0,1\}^n$.

(Dabei ist $\{0,1\}^n=\{(x_1,\dots,x_n):\ x_i\in\{0,1\}\}$)

Offensichtlich läßt sich jeder multiple Test Φ als Zusammenfassung von n einfachen Tests auffassen (und umgekehrt):

$$\Phi=(\delta_1,\dots,\delta_n):S\to\{0,1\}^n$$
$$\text{mit } \Phi(x)=(\delta_1(x),\dots,\delta_n))$$

Insbesondere ist δ_i ein einfacher Test für H_{0i} .
Von entscheidender Bedeutung ist nun die Frage nach einer

geeigneten "Niveau-α-Test"-Definition für multiple Testprozeduren. Dazu erscheint es sinnvoll, noch einmal die entsprechende Situation in einem einfachen Testproblem zu betrachten:

Sei H_0 die zu überprüfende Nullhypothese und δ ein exakter Niveau-α-Test (d.h. $\forall \theta \in H_0 \quad P_\theta(\delta=1)=\alpha$!).
Für den "wahren" Parameter θ lassen sich dann folgende Fälle unterscheiden:

	Wahrscheinlichkeit für einen Fehler 1.Art	
$\theta \notin H_0$	-	(ein Fehler 1.Art kann nicht auftreten)
$\theta \in H_0$	α	

Seien jetzt H_{01} und H_{02} die einzelnen Nullhypothesen eines multiplen Testproblems, mit $H_{01} \cap H_{02} \neq \emptyset$, und sei $\Phi=(\delta_1,\delta_2)$ ein multipler Test, dessen "Komponenten" δ_1 und δ_2 jeweils exakte Niveau-α-Tests seien, d.h.

$$\forall \theta \in H_{01} \quad P_\theta(\delta_1=1)=\alpha$$
$$\forall \theta \in H_{02} \quad P_\theta(\delta_2=1)=\alpha$$

Die entsprechende Fallunterscheidung für das multiple Testproblem lautet dann:

		Wahrscheinlichkeit für einen Fehler 1.Art
$\theta \notin H_{01}$	$\theta \notin H_{02}$	-
$\theta \in H_{01}$	$\theta \notin H_{02}$	α
$\theta \notin H_{01}$	$\theta \in H_{02}$	α
$\theta \in H_{01}$	$\theta \in H_{02}$	$P_\theta(\delta=1$ oder $\delta_2=1)$

Gilt $\theta \in H_{01}$ und $\theta \in H_{02}$, dann tritt ein Fehler 1.Art offensichtlich genau dann auf, wenn wenigstens eine der beiden Hypothesen durch ihren jeweiligen Test abgelehnt wird.
Die Wahrscheinlichkeit eines derartigen Ereignisses läßt sich dann aber schreiben als:

$$P_\theta(\delta_1=1 \text{ oder } \delta_2=1) \ .$$

Da $\alpha = P_\theta(\delta_1=1) \leq P_\theta(\delta_1=1$ oder $\delta_2=1)$ ist und es sich außerdem in vielen Fällen um eine strikte Ungleichung handeln wird, wird die Wahrscheinlichkeit für das Auftreten eines Fehlers 1.Art (bei

wenigstens einem der beiden Testprobleme) i.a. größer als das vorgegebene Signifikanzniveau α sein!
Geht man davon aus, daß ein multipler Test einen derartigen "Fehler" vermeiden sollte, so könnte zunächst die Forderung nach einem sog. "Globaltest" ein mögliches Kriterium eines geeigneten multiplen Tests darstellen.

<u>Def.:</u> Ein multipler Test $\Phi=(\delta_1,\ldots,\delta_n)$ heißt <u>Test zum globalen Niveau α</u>, $0<\alpha<1$, falls gilt:

$$\forall\, \theta \in H_0 := \bigcap_{i=1}^{n} H_{0i} \qquad P_\theta\Big(\bigcup_{i=1}^{n} \{\delta_i = 1\}\Big) \le \alpha \quad .$$

Ein Test zum globalen Niveau α begeht somit einen Fehler 1.Art höchstens mit Wahrscheinlichkeit α. Dies allerdings unter der einschränkenden Voraussetzung, daß sämtliche Nullhypothesen "wahr" sind.
Die folgende Definition berücksichtigt, daß der "wahre" Parameter θ möglicherweise nur in einigen der H_{0i} liegt.

<u>Def.:</u> Ein multipler Test $\Phi=(\delta_1,\ldots,\delta_n)$ heißt <u>Test zum multiplen Niveau α</u>, $0<\alpha<1$, falls gilt:

$$\forall\, \theta \in \Gamma \qquad P_\theta\Big(\bigcup_{i \in I(\theta)} \{\delta_i = 1\}\Big) \le \alpha$$

wobei $I(\theta) := \{i:\ \theta \in H_{0i}\}$ ist.

Da $I(\theta)$ gerade die Indexmenge der "wahren" Nullhypothesen ist, besitzt ein Test zum multiplen Niveau α die Eigenschaft, daß ein Fehler 1.Art (bei wenigstens einem der n Testprobleme) höchstens mit Wahrscheinlichkeit α auftritt und zwar unabhängig davon, wieviele und welche Nullhypothesen tatsächlich "wahr" sind!

<u>Satz:</u> Sei $\{H_{01},\ldots,H_{0n}\}$ eine Familie von Nullhypothesen eines multiplen Testproblems und $\Phi=(\delta_1,\ldots,\delta_n)$ ein zugehöriger multipler Test. Dann gelten für $0<\alpha<1$ folgende Aussagen:

(1) Φ ist genau dann ein Test zum multiplen Niveau α, falls gilt:

$$\forall\, J \subset \{1,\ldots,n\} \quad \forall\, \theta \in \bigcap_{j \in J} H_{0j} \qquad P_\theta\Big(\bigcup_{j \in J} \{\delta_j = 1\}\Big) \le \alpha \quad .$$

(2) Ist Φ ein Test zum multiplen Niveau α, dann ist Φ auch ein Test zum globalen Niveau α.

Beweis: (1) Sei $\Phi=(\delta_1,\dots,\delta_n)$ ein Test zum multiplen Niveau α, d.h.

$$\forall\theta\in\Gamma \quad P_\theta\Big(\bigcup_{i\in I(\theta)}\{\delta_i=1\}\Big)\le\alpha$$

mit $I(\theta)=\{i:\ \theta\in H_{0i}\}$.

Für $J=\emptyset$ gilt stets: $P_\theta\big(\bigcup_{j\in\emptyset}\{\delta_j=1\}\big) = P_\theta(\emptyset) = 0 \le \alpha$, und für $\bigcap_{j\in J} H_{0j} = \emptyset$ ist die Aussage $P_\theta\big(\bigcup_{j\in J}\{\delta_j=1\}\big)\le\alpha$ trivialerweise erfüllt.

Sei jetzt $J\subset\{1,\dots,n\}$, mit $J\neq\emptyset$ und $\bigcap_{j\in J} H_{0j}\neq\emptyset$, und sei $\theta\in\bigcap_{j\in J} H_{0j}$.

Dann folgt sofort $J\subset I(\theta)$. Damit ist $\bigcup_{j\in J}\{\delta_j=1\}\subset\bigcup_{i\in I(\theta)}\{\delta_i=1\}$ und

$$P_\theta\Big(\bigcup_{j\in J}\{\delta_j=1\}\Big)\le P_\theta\Big(\bigcup_{i\in I(\theta)}\{\delta_i=1\}\Big)\le\alpha.$$

Ist umgekehrt $\theta\in\Gamma$ beliebig, dann folgt wegen $I(\theta)\subset\{1,\dots,n\}$:

$$P_\theta\Big(\bigcup_{i\in I(\theta)}\{\delta_i=1\}\Big)\le\alpha \qquad \forall\theta\in\bigcap_{i\in I(\theta)} H_{0i} \quad .$$

(2) Sei Φ ein Test zum multiplen Niveau α und $J:=\{1,\dots,n\}$. Dann folgt wegen (1) :

$$\forall\theta\in\bigcap_{j\in J} H_{0j}=\bigcap_{j=1}^{n} H_{0j} \quad P_\theta\Big(\bigcup_{j\in J}\{\delta_j=1\}\Big) = P_\theta\Big(\bigcup_{j=1}^{n}\{\delta_j=1\}\Big)\le\alpha$$

und damit ist Φ ein Test zum globalen Niveau α.

Da somit jeder Test zum multiplen Niveau α stets auch ein Test zum globalen Niveau α ist und außerdem die Wahrscheinlichkeit für jedes mögliche Auftreten von Fehlern 1.Art höchstens α beträgt, gilt der Begriff des Tests zum multiplen Niveau α nun als entscheidender Ausgangspunkt für die Suche nach geeigneten multiplen Testprozeduren.
Bevor jedoch auf spezielle Testprozeduren eingegangen wird, folgt noch ein kurzer Abschnitt über die "Widerspruchsfreiheit" von Testentscheidungen.

3. Widerspruchsfreie Entscheidungen.

Ist δ ein statistischer Test für ein einfaches Testproblem von der Form $H_0 : H_1$ und gilt $\delta(x)=1$, dann wird man (üblicherweise) H_0 ablehnen und sich für die Alternativhypothese entscheiden. Bei der Anwendung von multiplen Tests verlangt man nun, daß die Entscheidungen der einzelnen Tests "konsistent", d.h. (untereinander) widerspruchsfrei sind. Die folgenden Definitionen sollen diese Problematik verdeutlichen.

Def.: Ein multipler Test $\Phi=(\delta_1,...,\delta_n)$ heißt kompatibel 1.Art oder widerspruchsfrei 1.Art (bezüglich des multiplen Testproblems $H_{0i} : H_{1i}$, $i=1,...,n$), falls gilt:

$$\forall\, x\in S \quad \bigcap_{i:\,\delta_i(x)=1} H_{1i} \neq \emptyset .$$

Die obige Définition verlangt also, daß die einzelnen Alternativhypothesen, für die man sich bei der Durchführung eines multiplen Tests entscheiden kann, nicht in Widerspruch zueinander stehen.
Diese auf Lehmann(1957b) zurückgehende Definition läßt sich übrigens zum Begriff der allgemeinen Widerspruchsfreiheit (Lehmann(1957a)) erweitern. Davon wird aber hier kein Gebrauch gemacht.

Von grundsätzlicher Bedeutung sind nun die von Gabriel(1969) stammenden Begriffe der Kohärenz und Konsonanz:

Def.: Ein multipler Test $\Phi=(\delta_1,...,\delta_n)$ heißt kohärent (bezüglich des multiplen Testproblems $H_{0i} : H_{1i}$, $i=1,...,n$), falls gilt:

Aus $H_{0i}\subset H_{0j}$ und $\delta_j(x)=1$ folgt stets: $\delta_i(x)=1$.

Die "Konsequenz" eines kohärenten Tests besteht darin, daß bei der Ablehnung einer Hypothese H_{0j} auch jede Teilhypothese $H_{0i}\subset H_{0j}$ abgelehnt wird.
Kohärente (multiple) Tests vermeiden z.B. Resultate folgender Art:

$$H_{01}:\ \beta_1=0 \qquad \delta_1(x)=1$$
$$H_{02}:\ \beta_1=\beta_2=0 \qquad \delta_2(x)=0 \quad .$$

In der Testsituation

$$H_{01}: \beta_1=0 \qquad\qquad H_{02}: \beta_2=0$$
$$\delta_1(x)=0 \qquad\qquad \delta_2(x)=0$$
$$H_{03}: \beta_1=\beta_2=0$$
$$\delta_3(x)=1$$

bewirkt umgekehrt die Ablehnung von H_{03} keine einzige Ablehnung bei den Hypothesen H_{01} und H_{02}! Dieser Fall läßt sich durch Verwendung von konsonanten Testprozeduren ausschließen:

Def.: Ein multipler Test $\Phi=(\delta_1,...,\delta_n)$ heißt konsonant (bezüglich des multiplen Testproblems $H_{0i} : H_{1i}$, $i=1,...,n$), falls für jede Hypothese H_{0i}, mit $H_{0i} = \bigcap_{j:H_{0j}\supsetneq H_{0i}} H_{0j}$, gilt:

$$\text{Aus } \delta_i(x)=1 \text{ folgt : } \exists H_{0j} \supsetneq H_{0i} \quad \text{mit } \delta_j(x)=1 \quad .$$

(Hier wird die in Sonnemann(1982) erwähnte "strengere" Fassung des Konsonanzbegriffs von Gabriel(1969) verwendet.)

Nach diesem Exkurs über widerspruchsfreie Entscheidungen wird im nächsten Kapitel gezeigt, wie man mit Hilfe des "Abschlußprinzips" auf relativ einfache Weise eine spezielle Klasse von kohärenten Tests, die sog. Abschlußtests, konstruieren kann.

III. MULTIPLE TESTPROZEDUREN UND DAS PRINZIP DES ABSCHLUSSTESTS.

Nachdem im letzten Kapitel der Begriff des Tests zum multiplen Niveau α eingeführt wurde, dessen charakteristische Eigenschaft eine grundsätzliche Anforderung an multiple Testprozeduren darstellt, soll in diesem Kapitel vor allem auf die Konstruktion derartiger Testprozeduren eingegangen werden. Von entscheidender Bedeutung wird dabei das bereits erwähnte "Abschlußprinzip" sein.

1. Theoretische Grundlagen.

Vorgegeben sei wieder ein statistischer Raum $(S, \mathcal{B}, \{P_\theta : \theta \in \Gamma\})$ und eine reelle Zahl α, mit $0<\alpha<1$.
Ein multiples Testproblem wird im folgenden auch einfach durch die Familie der Nullhypothesen $\{H_{01},...,H_{0n}\}$ gekennzeichnet. Dabei wird vorausgesetzt, daß $H_{01},...,H_{0n}$ die in Kapitel II, Abschnitt 2, erwähnten Bedingungen erfüllen.

Für die Darstellung des eigentlichen Prinzips des Abschlußtests (Abschlußprinzip) benötigt man zunächst den Begriff der durchschnittsabgeschlossenen Nullhypothesenfamilie.

<u>Def.:</u> Eine Familie $\{H_{01},...,H_{0n}\}$ von Nullhypothesen heißt <u>durchschnittsabgeschlossen</u>, falls für beliebige Indizes i und j, mit $1\leq i,j\leq n$, gilt:

$$H_{0i} \cap H_{0j} \in \{H_{01},...,H_{0n}\} \quad \text{oder} \quad H_{0i} \cap H_{0j} = \emptyset .$$

So sind z.B. im linearen Regressionsmodell

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 + u_t \qquad t=1,...,T$$

die Familien $\{H_{01},H_{02}\}$ und $\{H_{01},H_{03},H_{04}\}$ mit

$$H_{01}: \beta_1=0 \qquad H_{02}: \beta_1=1$$
$$H_{03}: \beta_2=0 \qquad H_{04}: \beta_1=\beta_2=0$$

durchschnittsabgeschlossen, die Familie $\{H_{01},H_{03}\}$ dagegen nicht.

Der Abschlußtest ist nun ein Test zum multiplen Niveau α, der auf durchschnittsabgeschlossene Nullhypothesenfamilien angewandt wird.

<u>Def.:</u> Sei $\{H_{01},\dots,H_{0n}\}$ eine durchschnittsabgeschlossene Familie von Nullhypothesen. Ein multipler Test $\Phi=(\delta_1,\dots,\delta_n)$ heißt <u>Abschlußtest</u>, falls gilt:

a) Φ ist ein Test zum multiplen Niveau α.
b) Φ ist kohärent.

Der folgende Satz stellt die zentrale Aussage dieses Kapitels dar. Er zeigt, auf welche Weise Abschlußtests und damit kohärente Tests zum multiplen Niveau α konstruiert werden können.

<u>Satz:</u> (Marcus/Peritz/Gabriel(1976),Sonnemann(1981,1982))

Sei $\{H_{01},\dots,H_{0n}\}$ durchschnittsabgeschlossen. Dann gilt:

(1) Ist $\Phi=(\delta_1,\dots,\delta_n)$ ein multipler Test, dessen Komponenten δ_i Niveau-α-Tests sind, dann ist der durch

$$\sigma_i(x) = \begin{cases} 1 & \forall H_{0j} \subset H_{0i} \quad \delta_j(x)=1 \\ 0 & \text{sonst} \end{cases}$$

definierte multiple Test $\Psi=(\sigma_1,\dots,\sigma_n)$ ein Abschlußtest.

(2) Ist umgekehrt $\Psi=(\sigma_1,\dots,\sigma_n)$ ein Abschlußtest, dann existiert ein multipler Test $\Phi=(\delta_1,\dots,\delta_n)$, dessen Komponenten δ_i Niveau-α-Tests sind mit der Eigenschaft

$$\sigma_i(x) = \begin{cases} 1 & \forall H_{0j} \subset H_{0i} \quad \delta_j(x)=1 \\ 0 & \text{sonst} \end{cases} \quad .$$

<u>Beweis:</u> (1) Es ist zu zeigen: Ψ ist ein kohärenter Test zum multiplen Niveau α.
Die Kohärenzeigenschaft folgt unmittelbar aus der Definition von Ψ.
Sei nun $\theta\in\Gamma$ mit $I(\theta)=\{i:\ \theta\in H_{0i}\} \neq \emptyset$. Dann ist auch $\bigcap_{i\in I(\theta)} H_{0i} \neq \emptyset$.
Aus der Durchschnittsabgeschlossenheit von $\{H_{01},\dots,H_{0n}\}$ folgt dann:

$$\exists\, j\in\{1,\dots,n\} \text{ mit } H_{0j} = \bigcap_{i\in I(\theta)} H_{0i} \quad .$$

Wegen $\theta \in \bigcap_{i\in I(\theta)} H_{0i}$ ist $j\in I(\theta)$ und damit

$$\{\sigma_j=1\} \subset \bigcup_{i\in I(\theta)} \{\sigma_i=1\} \quad .$$

Nach Voraussetzung ist δ_j ein Test zum Niveau α (bezüglich H_{0j}). Da $\theta \in H_{0j}$ ist und wegen der Kohärenz von Ψ erhält man schließlich:

$$P_\theta\Big(\bigcup_{i \in I(\theta)} \{\sigma_i = 1\}\Big) = P_\theta(\sigma_j = 1) \le P_\theta(\delta_j = 1) \le \alpha \quad .$$

Dabei ergibt sich die erste Ungleichung aus der Definition von Ψ:

$$\sigma_j(x) = 1 \;\Rightarrow\; \delta_j(x) = 1 \quad .$$

(2) Sei nun Ψ ein Abschlußtest. Da Ψ ein Test zum multiplen Niveau α ist, gilt:

$$\forall I \subset \{1,\dots,n\} \quad \forall \theta \in \bigcap_{i \in I} H_{0i} \qquad P_\theta\Big(\bigcup_{i \in I} \{\sigma_i = 1\}\Big) \le \alpha \quad .$$

Setzt man $I := \{i\}$, mit $i \in \{1,\dots,n\}$ beliebig, dann folgt:

$$\forall \theta \in H_{0i} \qquad P_\theta(\sigma_i = 1) \le \alpha \quad .$$

Damit ist Ψ ein multipler Test, dessen Komponenten σ_i Niveau-α-Tests sind. Aus der Kohärenzeigenschaft von Ψ folgt schließlich

$$\sigma_i(x) = \begin{cases} 1 & \forall H_{0j} \subset H_{0i} \quad \delta_j(x) = 1 \\ 0 & \text{sonst} \end{cases}$$

und damit ist (2) bewiesen.

Das in (1) beschriebene Abschlußprinzip sei hier noch einmal formuliert:

Sei $\{H_{01},\dots,H_{0n}\}$ eine durchschnittsabgeschlossene Familie von Nullhypothesen und $\Phi = (\delta_1,\dots,\delta_n\}$ ein multipler Test, dessen Komponenten $\delta_1,\dots,\delta_n$ Niveau-α-Tests bezüglich der Hypothesen $H_{01},\dots,H_{0n}$ sind.
Die Entscheidungsregel des Abschlußtests in Bezug auf eine beliebige Hypothese H_{0i} lautet dann:

Lehne H_{0i} genau dann ab, wenn jede Teilhypothese $H_{0j} \subset H_{0i}$ (also auch H_{0i} selbst) durch ihren zugehörigen Niveau-α-Test δ_j abgelehnt wird.

Die Aussage (2) des Satzes besagt dann, daß allein durch Anwendung des Abschlußprinzips jeder mögliche Abschlußtest konstruiert werden kann.

Obwohl ein Abschlußtest i.a. nicht konsonant ist, läßt sich diese Eigenschaft durch Anwendung eines zusätzlichen speziellen Abschlußtests erzwingen. Allerdings erhält man dadurch meist eine

konservativere Testprozedur, in dem Sinne, daß die Wahrscheinlichkeiten bezüglich der Ablehnung etwaiger falscher Nullhypothesen sich nicht unbeträchtlich verringern können (im Verhältnis zum urspünglichen Abschlußtest).
Im folgenden wird daher auf diesen konsonanten Abschlußtest nicht weiter eingegangen (siehe dazu auch die Anmerkungen in Sonnemann(1982)).
Ein weiterer Aspekt des Abschlußtests, die Forderung nach Abgeschlossenheit hinsichtlich der Durchschnittsbildung, erscheint wohl zunächst als eine wesentliche Einschränkung im Hinblick auf eine allgemeine Anwendbarkeit des Abschlußtests.
Verwendet man aber z.B. als Nullhypothesen sog. "Elementarhypothesen" (dies ist bei praktischen Anwendungen sehr häufig der Fall), dann läßt sich sogar zeigen, daß jedem Test zum multiplen Niveau α ein entsprechender Abschlußtest zugeordnet werden kann (und umgekehrt), sodaß die Entscheidungen beider Testprozeduren bezüglich der ("Ziel")-Nullhypothesen übereinstimmen.
Somit kann die Forderung nach einer durchschnittsabgeschlossenen Nullhypothesenfamilie (für diesen Fall) gewissermaßen als eine Voraussetzung "ohne Beschränkung der Allgemeinheit" aufgefaßt werden.
Dieser Sachverhalt soll im folgenden präzisiert werden.

<u>Def.:</u> Sei $\{H_{01},\dots,H_{0n}\}$ eine beliebige Familie von Nullhypothesen und $\Phi=(\delta_1,\dots,\delta_n)$ ein zugehöriger multipler Test. Ist I eine nichtleere Teilmenge von $\{1,\dots,n\}$, mit $\bigcap_{i\in I} H_{0i} \neq \emptyset$, dann nennt man den durch

$$\delta_I(x) = \begin{cases} 1 & \max_{i\in I} \delta_i(x) = 1 \\ 0 & \text{sonst} \end{cases}$$

definierten Test <u>Vereinigungs-Durchschnitts-Test</u> bezüglich der Hypothese $H_{0I} = \bigcap_{i\in I} H_{0i}$.

<u>Bem.:</u> Ist z.B. Φ ein Test zum multiplen Niveau α und berücksichtigt man, daß

$$\{x: \max_{i\in I} \delta_i(x) = 1\} = \bigcup_{i\in I} \{x: \delta_i(x) = 1\}$$

ist, dann sieht man sofort, daß δ_I ein Test zum Niveau α bezüglich H_{0I} ist.

<u>Def.:</u> Sei $\{H_{01},\dots,H_{0n}\}$ eine beliebige Familie von Nullhypothesen. Eine Hypothese H_{0i}, $1\le i\le n$, heißt <u>Elementarhypothese</u>, falls gilt:

$$\forall j\neq i \qquad H_{0j} \not\subset H_{0i} \quad .$$

Die beiden letzten Begriffe sind nun für die folgende Aussage von grundlegender Bedeutung.

<u>Satz:</u> Sei $\{H_{01},\dots,H_{0n}\}$ eine Familie von Elementarhypothesen und $\Phi=(\delta_1,\dots,\delta_n)$ ein beliebiger multipler Test. Dann sind die folgenden Aussagen äquivalent:

(1) Φ ist ein Test zum multiplen Niveau α.

(2) Es existiert eine durchschnittsabgeschlossene Nullhypothesenfamilie $\{H_{01},\dots,H_{0n},H_{0n+1},\dots,H_{0m}\}$ und ein entsprechender Abschlußtest $\Psi=(\sigma_1,\dots,\sigma_n,\sigma_{n+1},\dots,\sigma_m\}$, sodaß gilt:

$$\sigma_i=\delta_i \qquad i=1,\dots n \quad .$$

<u>Beweis:</u> (1) $\Rightarrow$ (2)

Ist $\{H_{01},\dots,H_{0n}\}$ bereits durchschnittsabgeschlossen, dann ist (2) trivialerweise erfüllt mit $\Psi=\Phi$.

Ist $\{H_{01},\dots,H_{0n}\}$ nicht durchschnittsabgeschlossen, dann muß zunächst eine entsprechende durchschnittsabgeschlossene Familie konstruiert werden:

Für jede Teilmenge $I\subset\{1,\dots,n\}$, mit $|I|\ge 2$ und $\bigcap_{i\in I} H_{0i}\neq\emptyset$, für die außerdem gilt:

$$H_{0j}\supset\bigcap_{i\in I} H_{0i} \Rightarrow j\in I \qquad j\in\{1,\dots,n\}$$

definiere man die Menge

$$H_{0I}:=\bigcap_{i\in I} H_{0i} \quad .$$

Bezeichnet man dann diese (endlich vielen) Mengen H_{0I} mit $H_{0n+1},\dots,H_{0m}$, dann sieht man sofort, daß $\{H_{01},\dots,H_{0n}, H_{0n+1},\dots,$ $\dots,H_{0m}\}$ durchschnittsabgeschlossen ist. Außerdem erfüllen $H_{01},\dots,H_{0m}$ die in Kapitel II, Abschnitt 2, geforderten Eigenschaften von Nullhypothesen.

Insbesondere gilt: $\qquad H_{0i}\neq H_{0j} \qquad i\neq j$.

Verwendet man nun für $H_{01},\dots,H_{0n}$ die Tests $\delta_1,\dots,\delta_n$ und definiert ansonsten für $i=n+1,\ldots,m$:

$$\delta_i(x) := \begin{cases} 1 & \max_{j\in I} \delta_j(x) = 1 \\ 0 & \text{sonst} \end{cases}$$

wobei I aus der obigen Definition stammt, mit $H_{0I} = \bigcap_{i\in I} H_{0i}$, dann erhält man einen multiplen Test, dessen Komponenten $\delta_1,\ldots,\delta_n$ Niveau-α-Tests sind, weil $\Phi=\{\delta_1,\ldots,\delta_n\}$ ein Test zum multiplen Niveau α ist. Folglich sind auch $\delta_{n+1},\ldots,\delta_m$ Niveau-α-Tests (Vereinigungs-Durchschnitts-Tests). Durch Anwendung des Abschlußprinzips erhält man dann einen zugehörigen Abschlußtest. $\Psi=(\sigma_1,\ldots,\sigma_n,\sigma_{n+1},\ldots,\sigma_m\}$.

Zu zeigen ist jetzt noch:

$$\delta_i = \sigma_i \qquad i=1,\ldots,n$$

Sei dazu $i\in\{1,\ldots,n\}$ beliebig.
<u>Fall 1:</u> $\delta_i(x)=1$.
Sei $H_{0k} \subset H_{0i}$ beliebig, mit $k\neq i$. Dann ist $k\geq n+1$ und es gilt:

$$H_{0k} = \bigcap_{j\in I} H_{0j}$$

für ein eindeutig bestimmtes I (gemäß obiger Definition). Daraus folgt $i\in I$ und somit $\delta_k(x)=1$. Da $H_{0k} \subset H_{0i}$ beliebig war, ergibt sich schließlich

$$\sigma_i(x)=1 \quad .$$

<u>Fall 2:</u> $\delta_i(x)=0$.
Aus der Definition des Abschlußtests folgt sofort

$$\sigma_i(x)=0 \quad .$$

Insgesamt ergibt sich damit:

$$\delta_i(x)=\sigma_i(x) \qquad \text{für } i=1,\ldots n \quad .$$

(2) $\Rightarrow$ (1)
Da der Abschlußtest ein Test zum multiplen Niveau α ist (bezüglich $\{H_{01},\ldots,H_{0n}\}$), gilt natürlich:

$$\forall I \subset \{1,\ldots,n\} \quad \forall \theta\in \bigcap_{i\in I} H_{0i} \quad P_\theta\left(\bigcup_{i\in I} \{\delta_i=1\}\right) \leq \alpha \quad .$$

Damit ist aber $\Phi=(\delta_1,\ldots,\delta_n)$ ein Test zum multiplen Niveau α (bezüglich $\{H_{01},\ldots,H_{0n}\}$).

2. Beispiele von Abschlußtests.

Im folgenden werden drei Beispiele von Testprozeduren vorgestellt, die multiples Niveau α besitzen: der klassische Bonferroni-Test, der Holm-Test und der exakte Abschlußtest.
Dabei soll vor allem gezeigt werden, daß sich der Bonferroni-Test und der Holm-Test i.a. als spezielle Abschlußtests darstellen lassen.
Obwohl man die oben genannten Testprozeduren unter recht allgemeinen Voraussetzungen anwenden kann, erscheint es sinnvoll, im Hinblick auf entsprechende Anwendungen im nächsten Kapitel und um einige Resultate des letzten Abschnitts verwenden zu können, für die folgenden Ausführungen eine Reihe von Annahmen zu treffen:

(1) Die zu überprüfenden ("Ziel"-) Nullhypothesen $H_{01},\ldots,H_{0n}$ seien Elementarhypothesen.

(2) Jedem Test δ_i liege eine entsprechende Teststatistik T_i zugrunde.

(3) Eine Nullhypothese H_{0i} werde durch den Test δ_i genau dann abgelehnt, falls für die Teststatistik T_i gilt:

$$T_i(x) \geq c_{i\alpha} \quad ,$$

wobei $c_{i\alpha}$ ein von α abhängender "kritischer" Wert ist.

(4) Die Teststatistiken $T_1,\ldots,T_n$ seien stetige Zufallsvariablen.

Diese Annahmen ermöglichen insbesondere die Verwendung der entsprechenden p-Funktionen $p_1,\ldots,p_n$ (Siehe S.10).
Für jedes $\alpha\in(0,1)$ gilt dann:

$$\forall \theta\in H_{0i} \qquad P_\theta(p_i\leq\alpha) = \alpha \qquad\qquad i=1,\ldots,n.$$

Es folgen die Beschreibungen der einzelnen Testprozeduren.

2.1. Bonferroni-Test.

Def.: Ein multipler Test $\Phi=(\delta_1,\ldots,\delta_n)$ heißt Bonferroni-Test, falls $\delta_1,\ldots,\delta_n$ jeweils Tests zum Niveau α/n sind (bezüglich $H_{01},\ldots,H_{0n}$), d.h. falls gilt:

$$\delta_i(x)=\begin{cases}1 & p_i(x)\le\alpha/n\\ 0 & \text{sonst}\end{cases}\qquad i=1,\ldots,n\ .$$

Es wird jetzt gezeigt, daß sich jeder Bonferroni-Test (unter den gegebenen Annahmen) als Abschlußtest darstellen läßt und daher auf Grund des letzten Satzes ein Test zum multiplen Niveau α ist.
Entsprechend der dortigen Beweisführung wird zunächst eine durchschnittsabgeschlossene Nullhypothesenfamilie konstruiert (es sei denn, $\{H_{01},\ldots,H_{0n}\}$ ist bereits durchschnittsabgeschlossen).
Dabei wählt man wieder genau diejenigen $H_{0I}=\bigcap_{i\in I}H_{0i}$, für die gilt:

$$H_{0j}\supset\bigcap_{i\in I}H_{0i}\quad\Rightarrow\quad j\in I.$$

Die Familie der H_{0I} ist durchschnittsabgeschlossen (die Hypothesen $H_{01},\ldots,H_{0n}$ gehören natürlich dazu) und auf Grund der eindeutigen Darstellung der H_{0I} sind diese Mengen auch paarweise voneinander verschieden.
Unter Verwendung der p-Funktionen werden dann die Tests δ_I wie folgt definiert:

$$\delta_I(x)=\begin{cases}1 & \min\limits_{i\in I}p_i(x)\le\alpha/n\\ 0 & \text{sonst}\end{cases}\ .$$

Wegen $\quad P_\theta(\delta_I=1)=P_\theta(\min\limits_{i\in I}p_i\le\alpha/n)=P_\theta(\bigcup\limits_{i\in I}\{p_i\le\alpha/n\})\le$

$$\le\sum_{i\in I}P_\theta(p_i\le\alpha/n)=|I|\cdot\alpha/n\le n\cdot\alpha/n=\alpha$$

ist δ_I ein Test zum Niveau α.
Da außerdem gilt: $\quad\delta_{\{i\}}=\delta_i\qquad i=1,\ldots,n$
ergibt sich somit für den zugehörigen Abschlußtest $\Psi=(\sigma_{\{1\}},\ldots,$ $,\ldots\sigma_{\{n\}},\ \ldots\ ,\sigma_{\{1,\ldots,n\}})$:

$$\sigma_{\{i\}}=\delta_i\qquad i=1,\ldots,n$$

und daher ist $\Phi=(\delta_1,\ldots,\delta_n)$ ein Test zum multiplen Niveau α.

Der Bonferroni-Test ist somit eine verkürzte Darstellung des zugehörigen Abschlußtests.
Diese Tatsache soll an Hand eines einfachen Beispiels veranschaulicht werden.
Seien H_{01}, H_{02} und H_{03} die zu überprüfenden Elementarhypothesen. Dabei werde angenommen, daß sämtliche Durchschnitte dieser Hypothesen nichtleer und paarweise voneinander verschieden sind. Zusammen mit den zugehörigen p-Funktionen p_1, p_2 und p_3 ergibt sich dann die folgende Testsituation:

H_{01}	H_{02}	H_{03}
$p_1 \leq \alpha/3$	$p_2 \leq \alpha/3$	$p_3 \leq \alpha/3$
$H_{01} \cap H_{02}$	$H_{01} \cap H_{03}$	$H_{02} \cap H_{03}$
$\min(p_1, p_2) \leq \alpha/3$	$\min(p_1, p_3) \leq \alpha/3$	$\min(p_2, p_3) \leq \alpha/3$
	$H_{01} \cap H_{02} \cap H_{03}$	
	$\min(p_1, p_2, p_3) \leq \alpha/3$	

wobei für jede einzelne Hypothese die Bedingung angegeben ist, unter der sie abgelehnt wird.
Man sieht sofort, daß, unabhängig davon, ob man nun den Abschlußtest durchführt oder nur die Hypothesen H_{01}, H_{02} und H_{03} überprüft (Bonferroni-Test), die Entscheidungen beider Test-prozeduren hinsichtlich dieser Hypothesen in jedem Fall übereinstimmen.

2.2. Holm-Test.

Dieser von Holm(1979) vorgeschlagene multiple Test wird entsprechend der (allgemeineren) Darstellung in Sonnemann(1982) direkt als Abschlußtest eingeführt.

Def.: Verwendet man für jede Hypothese $H_{0I} = \bigcap_{i \in I} H_{0i}$ (siehe Abschnitt 2.1.) den Test

$$\delta_I(x) = \begin{cases} 1 & \min_{i \in I} p_i(x) \leq \alpha/|I| \\ 0 & \text{sonst} \end{cases}$$

dann nennt man den durch

$$\sigma_I(x) = \begin{cases} 1 & \forall H_{0J} \subset H_{0I} \quad \delta_J(x) = 1 \\ 0 & \text{sonst} \end{cases}$$

definierten Abschlußtest einen <u>Holm-Test</u>.

Dabei ist offensichtlich δ_I ein Test zum Niveau α, sodaß der Satz über die Konstruktion von Abschlußtests unmittelbar angewandt werden kann. Als Abschlußtest ist der Holm-Test natürlich ein Test zum multiplen Niveau α.

Unter der Voraussetzung, daß sämtliche Durchschnitte der Hypothesen $H_{01},\dots,H_{0n}$ nichtleer und paarweise voneinander verschieden sind, kann statt des oben beschriebenen Holm-Tests auch eine "abgekürzte" Version dieses Tests verwendet werden.
Dazu seien o.B.d.A. die p-Werte bezüglich der Hypothesen $H_{01},\dots,H_{0n}$ der Größe nach angeordnet:

$$p_1(x) \le p_2(x) \le \dots \le p_n(x) \quad .$$

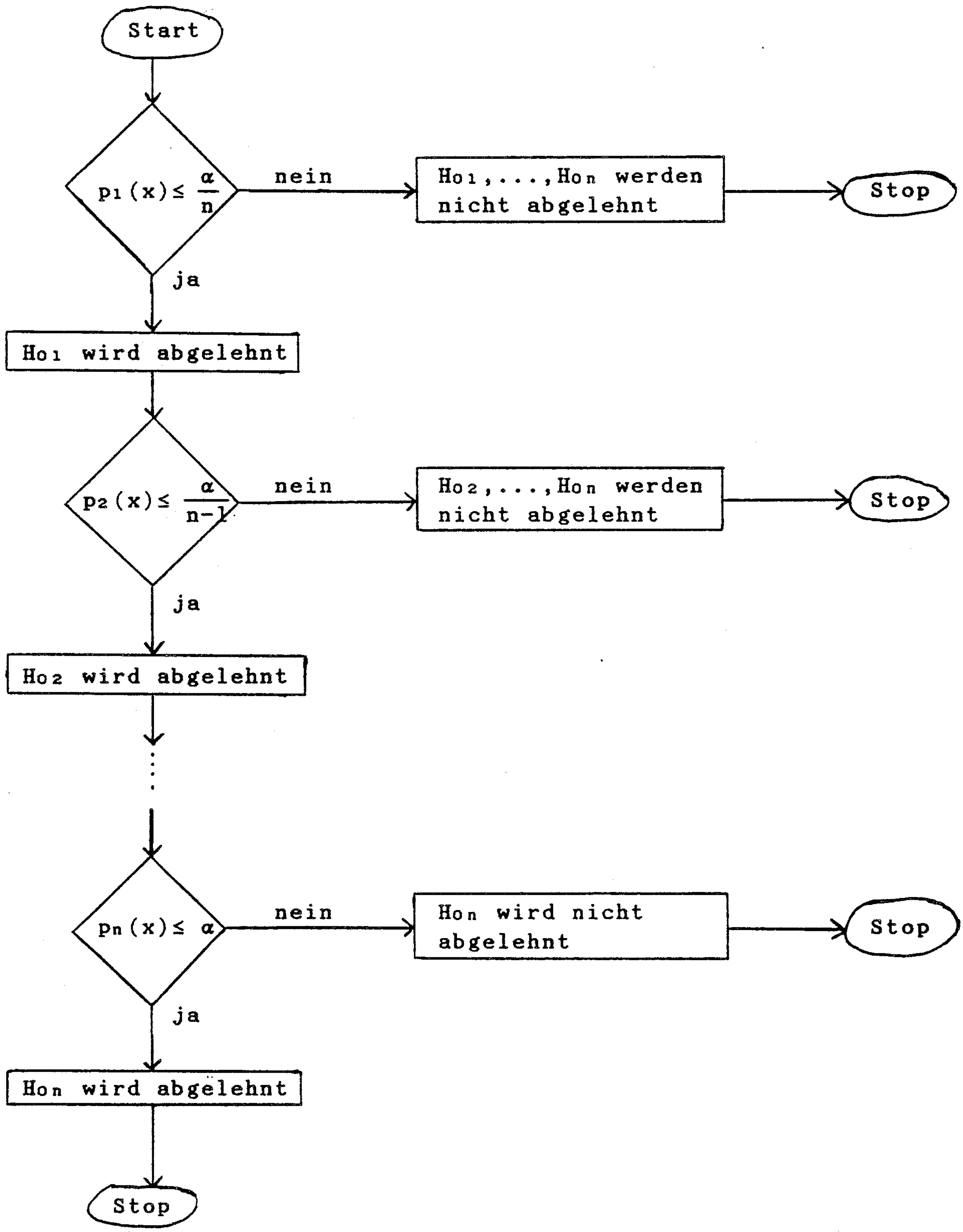
Der Test $\Phi = (\delta_1,\dots,\delta_n)$ überprüft dann die Hypothesen $H_{01},\dots,H_{0n}$ in Form eines (streng) sequentiellen Verfahrens:

$$\delta_1(x) = \begin{cases} 1 & p_1(x) \le \alpha/n \\ 0 & \text{sonst} \end{cases}$$

$$\delta_i(x) = \begin{cases} 1 & \delta_{i-1}(x)=1 \text{ und } p_i(x) \le \alpha/(n-i+1) \\ 0 & \text{sonst} \end{cases}$$

für $i=2,\dots,n$.

Dieses Verfahren läßt sich auch einfach mit Hilfe eines Flußdiagramms angeben (siehe nächste Seite).



(Dies ist die ursprüngliche Darstellung der Testprozedur von Holm(1979))

Daß beide Testprozeduren (hinsichtlich der Hypothesen $H_{01},\dots,H_{0n}$) tatsächlich zu denselben Entscheidungen führen, läßt sich wie folgt zeigen:

Sei $\delta_i(x)=1$. Dann ist $p_i(x)\le \frac{\alpha}{n-i+1}$ und außerdem gilt:

$$\delta_1(x)=\delta_2(x)=\dots=\delta_{i-1}(x)=1$$

bzw.

$$p_1(x)\le\frac{\alpha}{n}\ ,\ p_2(x)\le\frac{\alpha}{n-1}\ ,\ \dots\ ,\ p_{i-1}(x)\le\frac{\alpha}{n-i+1}\ .$$

Ist nun $H_{0I}=\bigcap_{j\in I} H_{0j}$ eine beliebige Teilhypothese von H_{0i}, d.h.

$H_{0I}\subset H_{0i}$, dann folgt aus der eindeutigen Darstellung von H_{0I}:

$$i\in I.$$

Ausgehend von der Indexmenge I lassen sich jetzt zwei Fälle unterscheiden:

Fall 1: $|I|\le n-i+1$

Daraus folgt:

$$\min_{j\in I} p_j(x)\ \le\ \frac{\alpha}{n-i+1}\ \le\ \frac{\alpha}{|I|}$$

und damit ist $\delta_I(x)=1$.

Fall 2: $|I|>n-i+1$

Setzt man $k:=\min I$, dann folgt wegen $|I|\le n-k+1$:

$$\frac{\alpha}{n-k+1}\ \le\ \frac{\alpha}{|I|}$$

Da $1\le k\le i-1$ ist, gilt somit

$$p_k(x)\ \le\ \frac{\alpha}{n-k+2}\ <\ \frac{\alpha}{n-k+1}\ \le\ \frac{\alpha}{|I|}$$

und daher ebenfalls $\delta_I(x)=1$.

Da H_{0I} beliebig gewählt war, folgt also:

$$\sigma_i(x):=\sigma_{\{i\}}(x)=1$$

Aus $\sigma_{\{i\}}(x)=1$ folgt aber auch umgekehrt $\delta_i(x)=1$.

Nach Voraussetzung gilt nämlich

$$\forall\, H_{0J}\subset H_{0i}\qquad \delta_J(x)=1\quad .$$

Damit ist

$$p_1(x) = \min_{1 \le j \le n} p_j(x) \le \frac{\alpha}{n}$$

$$p_2(x) = \min_{2 \le j \le n} p_j(x) \le \frac{\alpha}{n-1}$$

$$\vdots$$

$$p_i(x) = \min_{i \le j \le n} p_j(x) \le \frac{\alpha}{n-i+1}$$

und somit $\delta_i(x)=1$.

Die Signifikanzniveaus der Tests $\delta_1,\ldots,\delta_n$ können übrigens entsprechend "adjustiert" werden, falls nicht jede Anzahl m ($1\le m\le n$) von "wahren" Nullhypothesen möglich ist. (Näheres dazu findet man z.B. in Hommel(1985)).
Den Holm-Test kann man offensichtlich immer dann anwenden, wenn für die Durchschnittshypothesen H_{0I} keine "spezifischen" Tests zur Verfügung stehen. Die dabei verwendeten Vereinigungs-Durchschnitts-Tests δ_I haben allerdings den Nachteil, daß sie i.a. das jeweils vorgegebene Signifikanzniveau α nicht voll ausschöpfen können.
In jedem Fall aber stellt der Holm-Test eine "gleichmäßige" Verbesserung des klassischen Bonferroni-Tests dar.

Dies soll kurz begründet werden.

Sei H_{0i} eine beliebige (Ziel-) Nullhypothese, die beim Bonferroni-Test abgelehnt wird.

Wegen $p_i(x)\le \frac{\alpha}{n}$ folgt dann für jede Teilhypothese $H_{0I} = \bigcap_{j\in I} H_{0j} \subset H_{0i}$

$$\min_{j\in I} p_j(x) \le \frac{\alpha}{n} \le \frac{\alpha}{|I|} \qquad (i\in I\ !)$$

und damit wird H_{0i} auch unter dem Holm-Test abgelehnt.
Da aber $\forall I \neq \{1,\ldots,n\}$ gilt

$$\frac{\alpha}{n} < \frac{\alpha}{|I|}$$

lehnt der Holm-Test außerdem alle diejenigen Hypothesen H_{0I} ab, für die

$$\frac{\alpha}{n} < \min_{j \in I} p_j(x) < \frac{\alpha}{|I|}$$

ist.

Natürlich können unter den Nullhypothesen, die dadurch zusätzlich abgelehnt werden können, auch "wahre" Hypothesen vorkommen. Da der Holm-Test aber stets (!) das multiple Niveau α einhält, wird hier das höhere Risiko für das Auftreten von Fehlern 1.Art eingegangen, um mit entsprechend höherer Wahrscheinlichkeit (im Vergleich zum Bonferroni-Test!) falsche Nullhypothesen abzulehnen. An Hand des Beispiels aus Abschnitt 2.1. läßt sich diese "Verbesserung" durch den Holm-Test recht anschaulich darstellen:

H_{01} $p_1 \leq \alpha$	H_{02} $p_2 \leq \alpha$	H_{03} $p_3 \leq \alpha$
$H_{01} \cap H_{02}$ $\min(p_1, p_2) \leq \alpha/2$	$H_{01} \cap H_{03}$ $\min(p_1, p_3) \leq \alpha/2$	$H_{02} \cap H_{03}$ $\min(p_2, p_3) \leq \alpha/2$
	$H_{01} \cap H_{02} \cap H_{03}$ $\min(p_1, p_2, p_3) \leq \alpha/3$.	

2.3. Exakter Abschlußtest.

Falls (zusätzlich zu den bisherigen Annahmen) für jede Hypothese $H_{0I} = \bigcap_{i \in I} H_{0i}$ ein spezifischer Test zur Verfügung steht, der z.B. auf einer stetigen Teststatistik aufgebaut ist, dann können die beim Holm-Test verwendeten Vereinigungs-Durchschnitts-Tests natürlich durch diese spezifischen Tests ersetzt werden.

Def.: Existiert zu jeder Hypothese $H_{0I} = \bigcap_{i \in I} H_{0i}$ ein Test δ_I mit der Eigenschaft

$$\forall \theta \in H_{0I} \quad P_\theta(\delta_I = 1) = \alpha \qquad (1)$$

dann bezeichnet man den dazugehörigen Abschlußtest als exakten Abschlußtest.

Verwendet man für den Test δ_I die p-Funktion p_I, so kann (1) auch in der Form

$$\forall \theta \in H_{0I} \quad P_\theta(p_I \leq \alpha) = \alpha$$

geschrieben werden.
Da zur Anwendung des Abschlußprinzips für jede Hypothese H_{0I} ein Niveau-α-Test erforderlich ist, ist beim exakten Abschlußtest diese Voraussetzung (in gewisser Hinsicht) "optimal" erfüllt.

Aus der Tatsache, daß jeder Test δ_I das vorgegebene Signifikanzniveau α ausschöpft, folgt i.a. aber nicht, daß dann auch das multiple Niveau α (d.h. die Wahrscheinlichkeit dafür, daß wenigstens eine der "wahren" Nullhypothesen abgelehnt wird) ausgeschöpft wird.

Die Testsituation aus Abschnitt 2.1. besitzt bei der Anwendung des exakten Abschlußtests folgendes Aussehen:

H_{01} $p_1 \leq \alpha$	H_{02} $p_2 \leq \alpha$	H_{03} $p_3 \leq \alpha$
$H_{01} \cap H_{02}$ $p_{12} \leq \alpha$	$H_{01} \cap H_{03}$ $p_{13} \leq \alpha$	$H_{02} \cap H_{03}$ $p_{23} \leq \alpha$
	$H_{01} \cap H_{02} \cap H_{03}$ $p_{123} \leq \alpha$	

Dabei bedeuten p_{12}, p_{13}, p_{23} und p_{123} die zu den spezifischen Tests gehörenden p-Funktionen.

IV. ANWENDUNG VON MULTIPLEN TESTPROZEDUREN AUF ELEMENTARE LINEARE HYPOTHESEN.

Ausgehend von einem einfachen linearen Regressionsmodell

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_{K-1} x_{tK-1} + \beta_K + u_t \qquad t=1,\ldots,T$$

soll in diesem Kapitel gezeigt werden, auf welche Weise Elementarhypothesen von der Form

$$H_{0i}: \beta_i = 0 \qquad i=1,\ldots,K-1$$

mit Hilfe multipler Testprozeduren überprüft werden können. Als Beispiele werden dazu insgesamt sechs verschiedene Testprozeduren herangezogen:

(1) Bonferroni-Test
(2) Holm-Test
(3) Exakter Abschlußtest
(4) Scheffé-Test
(5) LSD-Test
(6) Multipler t-Test .

Jeder einzelne dieser multiplen Tests kann zur Überprüfung der oben angegebenen Hypothesen verwendet werden.
Die drei letztgenannten Tests, auf die in Abschnitt 1 noch eingegangen wird, dienten vor allem zu Vergleichszwecken bei einer Reihe von Monte Carlo Experimenten. Eine Auswahl der Ergebnisse dieser Experimente ist in Abschnitt 2 sowie im Anhang dieser Arbeit angegeben.

Zunächst folgt aber eine Darstellung der wichtigsten Voraussetzungen und eine kurze Beschreibung der einzelnen Testprozeduren.

1. Grundlagen und Beispiele.

Vorgegeben sei ein einfaches lineares Regressionsmodell, das sich (unter Berücksichtigung der Matrixschreibweise) auch in der Form

$$y = X\beta + u$$

darstellen läßt. Dabei bedeutet y ein Tx1 Spaltenvektor mit den

Werten der abhängigen Variablen, X eine nichtstochastische TxK Matrix (T>K) mit den Werten der unabhängigen Variablen (einschließlich Absolutglied), β ein (unbekannter) Kx1 Parametervektor und u ein normalverteilter Tx1 Störvektor mit E(u)=0 und $E(uu')=\sigma^2 I$ (I - TxT Einheitsmatrix, σ^2 unbekannt). Außerdem soll gelten: rg(X)=K .
(Dies sind die üblichen Annahmen des klassischen linearen Regressionsmodells.)
Auf der Grundlage des Stichprobenraumes $S=\mathbb{R}^T$ und des Parameterraumes $\Gamma=\{(\beta_1,\ldots,\beta_K,\sigma^2):\ \beta_i\in\mathbb{R},\ i=1,\ldots,K;\ \sigma^2>0\}$ sollen nun die Hypothesen

$$H_{0i}:\ \beta_i=0 \qquad i=1,\ldots,K-1$$

getestet werden.
(Dabei wird sinnvollerweise $K\geq 3$ vorausgesetzt. Auf die Überprüfung des Koeffizienten β_K wird hier verzichtet.)

Es folgt eine kurze Beschreibung der einzelnen Testprozeduren.

(1) Bonferroni-Test

Jede Hypothese H_{0i} wird mit dem gewöhnlichen t-Test zum Niveau $\alpha/(K-1)$ überprüft.

Bekanntlich beruht hier der t-Test auf der Teststatistik

$$\frac{\hat\beta_i}{(\hat\sigma^2\cdot a_{ii})^{\frac{1}{2}}} \quad ,$$

die t-verteilt ist mit T-K Freiheitsgraden. Dabei bedeuten $\hat\beta_i$ die i-te Komponente des Vektors

$$\hat\beta = (X'X)^{-1}X'y \quad ,$$

$\hat\sigma^2$ die erwartungstreue Schätzung für σ^2, mit

$$\hat\sigma^2 = \frac{\hat u'\hat u}{T-K} \qquad (\hat u = y - X\beta)$$

und a_{ii} das i-te Diagonalelement der Matrix $\sigma^2(X'X)^{-1}$.

(Der Bonferroni-Test ist ein Test zum multiplen Niveau α)

(2) Holm-Test

Ausgehend von den obigen t-Statistiken werden zunächst die entsprechenden p-Werte bezüglich der Hypothesen H_{0i}, $i=1,\ldots,K-1$, berechnet. Dann kann die in Kapitel III beschriebene "Kurzversion" des Holm-Tests angewendet werden.

(Der Holm-Test ist ein Test zum multiplen Niveau α)

(3) Exakter Abschlußtest

Für jede Hypothese H_{0I}, mit $I\subset\{1,\ldots,K-1\}$ und

$$H_{0I} = \bigcap_{i\in I} H_{0i}$$

wird der übliche F-Test mit $(|I|,T-K)$ Freiheitsgraden zum Niveau α (!) verwendet.
(Wegen der Teststatistik siehe z.B. Johnston (1984)).
Anschließend wird der in Kapitel III beschriebene Abschlußtest durchgeführt.

(Der exakte Abschlußtest ist ein Test zum multiplen Niveau α)

(4) Scheffè-Test

Beim Scheffè-Test handelt es sich um einen multiplen Test

$$\Phi=(\delta_1,\ldots,\delta_{K-1})$$

der definiert ist durch

$$\delta_i(x) = \begin{cases} 1 & \left|\dfrac{\beta_i}{(\sigma^2\cdot a_{ii})^{\frac{1}{2}}}\right| \geq S \\ 0 & \text{sonst} \end{cases}$$

mit $S = ((K-1)\cdot F_\alpha(K-1,T-k))^{\frac{1}{2}}$, wobei $F_\alpha(K-1,T-K)$ der kritische Wert (zum Niveau α) einer F-Verteilung mit $(K-1,T-K)$ Freiheitsgraden ist.
Zu einer allgemeineren Darstellung des Scheffè-Tests siehe z.B. Savin(1984).

(Der Scheffè-Test ist ein Test zum multiplen Niveau α. Auf den genauen Nachweis dieser Eigenschaft wird allerdings verzichtet. Ähnliches gilt für die beiden folgenden Testprozeduren.)

(5) LSD-Test

Beim LSD-Test wird zunächst die "globale" Nullhypothese

$$H_0: \beta_1=\beta_2=...=\beta_{K-1}=0$$

zum Niveau α getestet (F-Test). Wird H_0 nicht abgelehnt, gilt keine der Hypothesen H_{0i} als abgelehnt. Ansonsten wird jede Hypothese H_{0i} mit einem t-Test zum Niveau α überprüft.

(Der LSD-Test ist ein Test zum globalen Niveau α, aber für $K-1>2$ kein Test zum multiplen Niveau α !)

(6) Multipler t-Test

Jede Hypothese H_{0i} wird mit einem einfachen t-Test zum Niveau α (!) getestet.

(Der multiple t-Test ist weder ein Test zum globalen noch zum multiplen Niveau α !!!)

Faßt man die jeweiligen Eigenschaften dieser Testprozeduren zusammen, so ergibt sich folgendes Bild:

- Der Bonferroni-Test, der Holm-Test, der exakte Abschlußtest und der Scheffè-Test sind Tests zum multiplen Niveau α (und damit auch Tests zum globalen Niveau α).
- Der LSD-Test ist ein Test zum globalen Niveau α, aber i.a. kein Test zum multiplen Niveau α.
- Der multiple t-Test ist weder ein Test zum multiplen noch zum globalen Niveau α.

Für konkrete Anwendungen wäre es jetzt wünschenswert, unter diesen sechs Testprozeduren die am besten geeignete auszuwählen.
Betrachtet man aber insbesondere ökonometrische Anwendungen, so wird die Eigenschaft 'Test zum multiplen Niveau α' allein sicherlich nicht ausreichen, um eine interessante Alternative zu dem dort üblicherweise verwendeten multiplen t-Test darzustellen.
Eine derartige Testprozedur sollte auch in der Lage sein, falsche Nullhypothesen mit relativ hoher Wahrscheinlichkeit abzulehnen.
Konservative Testprozeduren, wie z.B. der Bonferroni-Test, wird man daher wohl kaum ernsthaft in Erwägung ziehen.

Während nun für den Scheffè-Test (und auch für den Bonferroni-

Test) eine ganze Reihe von "Power"-Untersuchungen durchgeführt wurden (siehe z.B. Savin(1984)), scheint dies beim Holm-Test und beim exakten Abschlußtest (zumindest was ökonometrische Anwendungen anbelangt) nicht bzw. kaum der Fall zu sein, obwohl die Konstruktion dieser beiden Testprozeduren durchaus eine wesentliche Verbesserung, etwa gegenüber dem Bonferroni-Test, erwarten läßt.
Im nächsten Abschnitt soll daher, an Hand einiger Monte Carlo Simulationen (die bei Untersuchungen multipler Tests wohl angebracht sind) der Versuch eines "empirischen Vergleichs" der oben beschriebenen Testprozeduren unternommen werden.

2. Empirischer Vergleich verschiedener Testprozeduren.

Am Beginn derartiger Untersuchungen steht zunächst die Frage nach den jeweiligen "Größen", die mit Hilfe der Monte Carlo Methode simuliert werden sollen. Dazu wurden den einzelnen Experimenten insgesamt fünf "Vergleichskriterien" zugrunde gelegt:

a) die Wahrscheinlichkeit dafür, daß die Hypothese H_{0i}: $\beta_i = 0$, $1 \leq i \leq K-1$, abgelehnt wird,
b) die Wahrscheinlichkeit dafür, daß wenigstens eine der "wahren" Nullhypothesen abgelehnt wird,
c) die Wahrscheinlichkeit dafür, daß wenigstens eine der "falschen" Nullhypothesen abgelehnt wird,
d) die Wahrscheinlichkeit dafür, daß alle "falschen" Nullhypothesen abgelehnt werden,
e) die Wahrscheinlichkeit dafür, daß weder ein Fehler 1.Art noch ein Fehler 2.Art auftritt.

(Siehe dazu auch Hommel(1986))
Bei den folgenden drei Simulationsbeispielen konnten aus Platzgründen bzw. wegen zu erwartender großer Rechenzeiten allerdings nur einige dieser Kriterien berücksichtigt werden.

Beispiel 1:

Um das Design bei den einzelnen Simulationsbeispielen möglichst realistisch zu gestalten, wurde bei der Wahl der unabhängigen Variablen ausschließlich auf reale ökonomische Zeitreihen zurückgegriffen. So bildet den Ausgangspunkt für Beispiel 1 die Modellgleichung

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 \mathbb{1} + u,$$

bei der die Variablen (Vektoren) $x_1, \ldots, x_7$ einer (Konsum-) Gleichung der reduzierten Form des Klein-I-Modells (USA,1921-1941) entnommen wurden. ($\mathbb{1}$ ist hier ein Tx1 Spaltenvektor von Einsen.) Die einzelnen Variablen haben folgende Bedeutung:

x_1 - Government wage bill
x_2 - Taxes
x_3 - Government nonwage expenditures
x_4 - Time measured as year-1931
x_5 - Profits lagged one year
x_6 - Capital stock lagged one year
x_7 - Total production of private industry

(Die zugehörigen Daten finden man z.B. in Theil(1971). Bezüglich der obigen Gleichung siehe auch das Beispiel 2 von Savin(1984).)

Die Designmatrix X (T=21) hat somit folgendes Aussehen:

$$X = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, \mathbb{1}] \quad .$$

Insgesamt wurden 1000 Simulationen durchgeführt, bei denen die X-Matrix nicht verändert wurde. X kann daher als eine nicht-stochastische Matrix von unabhängigen Variablen angesehen werden. Für den Störvektor u wurde angenommen:

$$u = (u_1, \ldots, u_{21})' \sim N(0, \sigma^2 I) \ , \quad \text{mit } \sigma^2 = 1.0 \quad .$$

Beim Koeffizientenvektor $\beta = (\beta_1, \ldots, \beta_8)'$ wurde zunächst die Konstante β_8 mit $\beta_8 = 10.0$ festgesetzt. Wegen zu erwartender großer Rechenzeiten, insbesondere für den exakten Abschlußtest, wurde für die übrigen Koeffizienten $\beta_1, \ldots, \beta_7$ nur eine einzige Parameterkombination zugrundegelegt, und zwar

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 \quad .$$

Hierdurch wird gerade der Fall betrachtet, bei dem sämtliche Nullhypothesen

$$H_{0i}: \ \beta_i = 0 \qquad i = 1, \ldots, 7$$

"wahr" sind.

Auf der Grundlage der durchgeführten Simulationen konnten damit sämtliche einfachen bzw. multiplen (empirischen) Signifikanzniveaus berechnet werden. Diese beiden Größen (die Wahrscheinlichkeit dafür, eine "wahre" Nullhypothese abzulehnen bzw. die Wahrscheinlichkeit dafür, wenigstens eine der "wahren" Nullhypothesen abzulehnen) sind wohl für jede Testprozedur von wesentlicher Bedeutung. Besonderes Interesse sollte daher auch dem multiplen t-Test gelten, der bekanntlich kein multiples Niveau α besitzt und auf Grund der gegebenen Parameterkombination ein nicht unbeträchtliches multiples Signifikanzniveau aufweisen dürfte.
Obwohl dies bereits aus dem bisher Gesagten hervorgehen sollte, sei hier noch einmal darauf hingewiesen, daß bei jeder einzelnen Simulation jede Hypothese $H_{0i}: \beta_i=0$, $1\leq i\leq 7$, durch jede der sechs Testprozeduren getestet wurde.

In den folgenden Tabellen sind nun zunächst die (einfachen) empirischen Signifikanzniveaus angegeben (α=0.01,0.05,0.10). Anschließend findet man auch eine Zusammenstellung der empirischen multiplen Signifikanzniveaus.

<u>Tabelle 1:</u> Empirische Signifikanzniveaus (α=0.01)

Testprozedur	Hypothesen						
	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	H_{06}	H_{07}
Bonferroni-Test	.000	.003	.001	.000	.002	.000	.001
Holm-Test	.000	.003	.001	.000	.002	.000	.001
Exakter Abschlußtest	.000	.001	.000	.000	.000	.000	.000
Scheffè-Test	.000	.000	.000	.000	.000	.000	.000
LSD-Test	.000	.001	.001	.001	.003	.000	.002
Multipler t-Test	.006	.008	.010	.008	.011	.005	.009

Tabelle 2: Empirische Signifikanzniveaus (α=0.05)

Testprozedur	Hypothesen						
	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	H_{06}	H_{07}
Bonferroni-Test	.005	.007	.009	.006	.006	.002	.005
Holm-Test	.005	.007	.009	.007	.008	.002	.006
Exakter Abschlußtest	.004	.004	.006	.003	.005	.000	.004
Scheffè-Test	.000	.001	.000	.000	.001	.000	.000
LSD-Test	.011	.012	.015	.012	.010	.009	.010
Multipler t-Test	.055	.054	.048	.051	.042	.044	.046

Tabelle 3: Empirische Signifikanzniveaus (α=0.10)

Testprozedur	Hypothesen						
	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	H_{06}	H_{07}
Bonferroni-Test	.012	.010	.014	.012	.015	.010	.011
Holm-Test	.014	.013	.014	.015	.017	.010	.014
Exakter Abschlußtest	.007	.006	.012	.007	.009	.008	.009
Scheffè-Test	.000	.003	.001	.000	.002	.000	.002
LSD-Test	.035	.036	.034	.035	.031	.039	.027
Multipler t-Test	.112	.109	.103	.101	.098	.110	.090

Tabelle 4: Empirische multiple Signifikanzniveaus

Testprozedur	α		
	0.01	0.05	0.10
Bonferroni-Test	.005	.029	.053
Holm-Test	.005	.029	.053
Exakter Abschlußtest	.001	.016	.031
Scheffè-Test	.000	.002	.006
LSD-Test	.005	.035	.090
Multipler t-Test	.040	.193	.361

Die Ergebnisse der Tabellen 1-3 entsprechen im wesentlichen den Erwartungen. So liegen z.B. die empirischen Signifikanzniveaus des multiplen t-Tests relativ nahe beim vorgegebenen Wert α, während sich die entsprechenden Ergebnisse des Bonferroni-Tests dem (erwarteten) Wert $\alpha/7$ annähern. Die Unterschiede zum Holm-Test fallen hier nur sehr gering aus, was sich aber bei zunehmender Abweichung von den Nullhypothesen sehr rasch ändern dürfte.
Auffallend ist vielleicht die Tatsache, daß die Signifikanzniveaus des Scheffè-Tests hier praktisch den Wert Null haben.
An den Ergebnissen für den LSD-Test kann man (im Vergleich zum multiplen t-Test) ablesen, wie weit sich die "Vorschaltung" eines globalen F-Tests vor den jeweiligen t-Tests auf die einzelnen Signifikanzniveaus auswirken kann.
Die entsprechenden Werte für den exakten Abschlußtest fallen natürlich noch geringer aus, da zur Ablehnung einer Hypothese H_{0i} ja eine ganze Folge von F-Tests durchlaufen werden muß.
Etwas interessanter ist die Situation bei den multiplen Signifikanzniveaus:
Im Falle des multiplen t-Tests wird z.B. bei einem vorgegebenen Signifikanzniveau von $\alpha=0.05$ bei fast 20% der durchgeführten Simulationen wenigstens eine "wahre" Nullhypothese abgelehnt. Eine derartige Fehlerwahrscheinlichkeit kann bei den anderen Testprozeduren nicht auftreten, da diese zumindest globales Niveau α besitzen. Allerdings fällt auf, daß hier das multiple Signifikanzniveau in recht unterschiedlichem Maße ausgeschöpft wird.

<u>Beispiel 2:</u>

Bei der zweiten Modellgleichung, die die Form

$$y = \beta_1 \ln x_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 \ln x_6 + \beta_7 x_7 + \beta_8 1 + u$$

besitzt, wurden diesmal österreichische Wirtschaftsdaten aus den Jahren 1956-1985 (Quartalsdaten, T=120) verwendet. Die dabei auftretenden Variablen haben folgende Bedeutung:

- x_1 - Privater Konsum, verzögert (lag 1), saisonbereinigt
- x_2 - Disponibles Einkommen, saisonbereinigt
- x_3 - Disponibles Einkommen, verzögert (lag 1), saisonbereinigt
- x_4 - Inflationsrate
- x_5 - Rendite der Neuemissionen
- x_6 - Geldvolumen
- x_7 - Trend für Quartalsdaten (Januar 1956 = 9)

(Quelle: Datenbank/Institut für Höhere Studien, 1986)

Es wurden wieder 1000 Simulationen durchgeführt, bei denen die X-Matrix stets konstant gehalten wurde.
Abgesehen von einer Änderung der Varianz, nämlich $\sigma^2=2.0$, galten für den Störvektor u dieselben Voraussetzungen wie in Beispiel 1. Die Konstante β_8 wurde diesmal mit $\beta_8=1.0$ und die übrigen Koeffizienten wieder mit

$$\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=\beta_6=\beta_7=0$$

festgesetzt.
Bei diesem Simulationsbeispiel wurden noch einmal dieselben Kriterien wie in Beispiel 1 zugrunde gelegt. Die Ergebnisse sind in den folgenden Tabellen enthalten.

Tabelle 5: Empirische Signifikanzniveaus (α=0.01)

Testprozedur	Hypothesen						
	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	H_{06}	H_{07}
Bonferroni-Test	.003	.001	.000	.001	.003	.001	.001
Holm-Test	.003	.001	.000	.001	.003	.001	.001
Exakter Abschlußtest	.001	.000	.000	.000	.000	.000	.000
Scheffè-Test	.000	.000	.000	.000	.000	.000	.000
LSD-Test	.001	.000	.000	.000	.002	.002	.000
Multipler t-Test	.010	.006	.006	.009	.014	.011	.010

Tabelle 6: Empirische Signifikanzniveaus (α=0.05)

Testprozedur	Hypothesen						
	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	H_{06}	H_{07}
Bonferroni-Test	.009	.002	.004	.005	.013	.007	.006
Holm-Test	.009	.002	.004	.005	.013	.008	.006
Exakter Abschlußtest	.004	.001	.001	.001	.003	.002	.002
Scheffè-Test	.001	.000	.000	.000	.000	.000	.001
LSD-Test	.009	.011	.012	.015	.021	.011	.011
Multipler t-Test	.044	.047	.051	.050	.056	.059	.045

Tabelle 7: Empirische Signifikanzniveaus (α=0.10)

Testprozedur	Hypothesen						
	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	H_{06}	H_{07}
Bonferroni-Test	.013	.013	.011	.012	.018	.014	.012
Holm-Test	.014	.014	.011	.012	.018	.014	.013
Exakter Abschlußtest	.007	.004	.004	.003	.013	.005	.008
Scheffè-Test	.001	.000	.000	.000	.000	.001	.001
LSD-Test	.021	.027	.036	.033	.036	.032	.028
Multipler t-Test	.092	.094	.105	.105	.110	.105	.084

Tabelle 8: Empirische multiple Signifikanzniveaus

Testprozedur	α		
	0.01	0.05	0.10
Bonferroni-Test	.010	.044	.086
Holm-Test	.010	.044	.086
Exakter Abschlußtest	.001	.014	.040
Scheffè-Test	.000	.002	.003
LSD-Test	.005	.052	.094
Multipler t-Test	.062	.226	.456

Berücksichtigt man die jeweiligen Schwankungen durch den zugrundeliegenden Zufallsmechanismus, so liefern die Tabellen 5-7 ein ähnliches Bild wie die Tabellen 1-3, sodaß weitere Erläuterungen wohl nicht notwendig sind.
Aus Tabelle 8 ersieht man, daß die multiplen Signifikanzniveaus gegenüber Beispiel 1 etwas zugenommen haben (Multipler t-Test!!), was vermutlich vor allem auf die unterschiedliche Abhängigkeit der x-Regressoren zurückzuführen sein dürfte.

Beispiel 3:

An Hand dieses Beispiels soll nun untersucht werden, wie die einzelnen Testprozeduren auf Änderungen verschiedener Regressionskoeffizienten reagieren. Ausgangspunkt ist diesmal die Gleichung

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 1 + u$$

mit den Variablen

x_1 - Nichtdauerhafter Konsum, verzögert(lag 1), saisonbereinigt
x_2 - Disponibles Einkommen, saisonbereinigt
x_3 - disponibles Einkommen, verzögert(lag 1), saisonbereinigt
x_4 - Reale Geldmenge [M3], saisonbereinigt (dividiert durch 10000)

Dabei wurden wieder österreichische Wirtschaftsdaten (Quartalsdaten) verwendet, diesmal für den Zeitraum 1976-1985. (Quelle: Datenbank/Institut für Höhere Studien, 1986)

Die Varianz der Störvariablen u wurde bei diesem Beispiel mit $\sigma^2=5.0$ festgesetzt. Ansonsten galt natürlich:

$$u \sim N(0, \sigma^2 I) \quad .$$

Während der Koeffizient β_5 bei den hier durchgeführten Simulationen stets den Wert $\beta_5=20.0$ besaß, wurde bezüglich der Koeffizienten $\beta_1, \beta_2, \beta_3$ und β_4 folgende Vorgangsweise gewählt:
Für jeweils 1000 Simulationen wurde jedem dieser 4 Koeffizienten genau einer der Werte

$$0.0, 0.25, 0.50$$

zugeordnet.
Die dabei auftretende Anzahl von insgesamt $3^4=81$ verschiedenen Parameterkombinationen dürfte wohl eine recht interessante Vergleichsbasis für die zu untersuchenden Testprozeduren darstellen, insbesondere durch das z.T. gleichzeitige Auftreten von "wahren" bzw. "falschen" Nullhypothesen ($\beta_i=0$ bzw. $\beta_i \neq 0$).
Mit Ausnahme der Kombination $\beta_1=\beta_2=\beta_3=\beta_4=0.0$ wurde daher jedesmal zusätzlich die 'Wahrscheinlichkeit, wenigstens eine der "falschen" Nullhypothesen abzulehnen', berechnet.
Aus Platzgründen kann in dieser Arbeit allerdings nur ein Teil der Ergebnisse präsentiert werden. Die dargestellten Tabellen (siehe Anhang) beschränken sich daher auf den Fall

$$\beta_4=0.25 \quad .$$

Dazu einige Erläuterungen:

Auf jeder Seite (des Anhangs) sind die Ergebnisse von jeweils drei Kombinationen zusammengefaßt ($\alpha=0.05$). Dabei sind in den obersten drei Tabellen die empirischen Ablehnungswahrscheinlichkeiten bezüglich der Nullhypothesen

$$H_{0i}: \beta_i = 0 \quad i=1,2,3,4$$

angegeben (bei gegebener Kombination und entsprechender Testprozedur). Bei jeder einzelnen Tabelle ist unterhalb der Überschrift ('Empirische Ablehnungswahrscheinlichkeiten') die jeweils zugrundeliegende Kombination der Regressionskoeffizienten angeführt.
Anschließend folgen noch die Ergebnisse bezüglich der 'Wahrscheinlichkeit, wenigstens eine der "falschen" Nullhypothesen abzulehnen' (jeweils für drei Kombinationen).

Die Ergebnisse dieser Simulationsstudie lassen sich wie folgt zusammenfassen:

Betrachtet man die Gruppe der Testprozeduren, die multiples Niveau α besitzen (Bonferroni-Test, Holm-Test, Exakter Abschlußtest, Scheffè-Test), so zeigt sich bei fast jeder der untersuchten Kombinationen, daß die empirischen Ablehnungswahrscheinlichkeiten des Bonferroni-Tests, des Holm-Tests und des Scheffè-Tests durch die entsprechenden Werte des exakten Abschlußtests (z.T. erheblich) übertroffen werden. Dies gilt insbesondere bei Vorliegen mehrerer "falscher" Nullhypothesen. So betragen etwa die Unterschiede gegenüber dem Holm-Test in manchen Fällen bis zu 15 Prozentpunkten.
Die Ablehnungswahrscheinlichkeiten des Holm-Tests ihrerseits sind in allen Fällen (erwartungsgemäß) mindestens ebenso groß wie die des Bonferroni-Tests. Die geringste "Power" bei der Ablehnung "falscher" Nullhypothesen besitzt schließlich der Scheffè-Test.
Daß die Ergebnisse für den LSD-Test und den multiplen t-Test übereinstimmen, läßt sich wohl mit dem relativ hohen Wert des Koeffizienten $\beta_4 = 0.25$ erklären. Der globale F-Test stellt in diesem Fall praktisch keine wirksame Schranke mehr dar.
Die beiden letztgenannten Testprozeduren lassen sich allerdings nur sehr bedingt mit den anderen vergleichen, da sie bekanntlich kein multiples Niveau α haben.
Um diesen Umstand noch hervorzuheben, sind im folgenden für einige ausgewählte Kombinationen die empirischen multiplen Signifikanzniveaus angegeben.

Tabelle 9: Empirische multiple Signifikanzniveaus ($\beta_1=0.0, \beta_2=0.0, \beta_3=0.0, \beta_4=0.0$)

Testprozedur	α 0.01	0.05	0.10
Bonferroni-Test	.004	.041	.074
Holm-Test	.004	.041	.074
Exakter Abschlußtest	.002	.025	.059
Scheffè-Test	.001	.004	.012
LSD-Test	.005	.041	.083
Multipler t-Test	.028	.125	.255

Tabelle 10: Empirische multiple Signifikanzniveaus ($\beta_1=0.0, \beta_2=0.0, \beta_3=0.0, \beta_4=0.5$)

Testprozedur	α 0.01	0.05	0.10
Bonferroni-Test	.006	.046	.074
Holm-Test	.008	.053	.097
Exakter Abschlußtest	.007	.050	.090
Scheffè-Test	.002	.006	.016
LSD-Test	.037	.131	.237
Multipler t-Test	.037	.131	.237

Tabelle 11: Empirische multiple Signifikanzniveaus ($\beta_1=0.0, \beta_2=0.25, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	α 0.01	0.05	0.10
Bonferroni-Test	.003	.027	.054
Holm-Test	.004	.038	.062
Exakter Abschlußtest	.007	.040	.070
Scheffè-Test	.001	.003	.013
LSD-Test	.019	.090	.174
Multipler t-Test	.019	.090	.174

(Die zuvor gemachten Aussagen über das Abschneiden des Bonferroni-Tests, des Holm-Tests, des exakten Abschlußtests und des Scheffè-Tests gelten übrigens auch für den Fall $\beta_4=0.0$ bzw. $\beta_4=0.5$).

Insgesamt kann wohl (in Bezug auf ökonometrische Anwendungen) der vorläufige Schluß gezogen werden, daß mit dem exakten Abschlußtest eine recht interessante und vielversprechende Testprozedur zur Verfügung steht, deren spezielle Eigenschaften aber noch genauer untersucht werden sollten.

V. MULTIPLE TESTPROBLEME UND SPEZIFIKATIONSTESTS.

Multiple Testprobleme treten nicht nur beim Testen linearer Hypothesen auf, sondern vor allem auch bei der Anwendung von Spezifikationstests zur Überprüfung der Annahmen des linearen Regressionsmodells (Homoskedastizität, Konstanz der Regressionskoeffizienten, Unabhängigkeit, Linearität, Normalität u.s.w.).
In diesem Kapitel soll nun ein sequentielles Verfahren beschrieben werden, mit dem eine entsprechend angeordnete Folge von "Spezifikationshypothesen" getestet werden kann. Dieses Verfahren, das eine Reihe von Anwendungsmöglichkeiten in der Ökonometrie besitzt, wurde zunächst von Anderson(1962,1971) zur Bestimmung des Grades einer polynomialen Regressionsfunktion entwickelt.
Mittlerweile wurde es aber auch von Mizon(1977), Phillips(1984), Kiviet/Phillips(1985) und Phillips/McCabe(1985) zur Überprüfung von Spezifikationshypothesen herangezogen.
Ein einfaches Beispiel für die Anwendung der Anderson-Prozedur ist in Kapitel I, S. 5/6, angegeben. Das dabei auftretende Testproblem

$$
\begin{aligned}
H_{01}&: \varrho = 0\\
H_{02}&: \varrho = 0,\ \sigma_1^2 = \sigma_2^2\\
H_{03}&: \varrho = 0,\ \sigma_1^2 = \sigma_2^2,\ \beta_1 = \beta_2
\end{aligned}
$$

läßt sehr anschaulich eine der wesentlichen Voraussetzungen dieses Verfahrens erkennen: Die zu überprüfenden Nullhypothesen müssen in einer auf- bzw. absteigenden Folge angeordnet sein.
Allgemeiner kann man diese Voraussetzungen wie folgt formulieren:
Sei $(S,\mathcal{B},\{P_\theta:\ \theta\in\Gamma\})$ ein statistischer Raum und seien $H_1,\ldots,H_n$ Teilmengen von Γ mit

$$
\begin{aligned}
H_i &\neq \Gamma && i=1,\ldots,n\\
H_1\cap\ldots\cap H_n &\neq \emptyset\\
H_1\cap\ldots\cap H_{i+1} &\subset H_1\cap\ldots\cap H_i && i=1,\ldots,n-1\ .
\end{aligned}
$$

Definiert man dann als Nullhypothesen die Mengen

$$
\begin{aligned}
H_{01} &= H_1\\
H_{02} &= H_1\cap H_2\\
&\ \ \vdots\\
H_{0n} &= H_1\cap\ldots\cap H_n
\end{aligned}
$$

so besitzen $H_{01},\dots,H_{0n}$ die Eigenschaft:

$$H_{01} \subset H_{02} \quad \dots \subset H_{0n} \quad .$$

(In diesem Zusammenhang spricht man auch von hierarchisch angeordneten Hypothesen ("nested hypotheses").)

Eine durchaus mögliche Vorgangsweise zur Überprüfung von $H_{01},\dots,$ $\dots,H_{0n}$, auf die im folgenden aber nicht weiter eingegangen werden soll, sei hier nur kurz erwähnt:
Da die Familie $\{H_{01},\dots,H_{0n}\}$ offensichtlich durchschnittsabgeschlossen ist (für $i<j$ ist ja $H_{0i} \cap H_{0j} = H_{0j}$), läßt sich, unter der Voraussetzung, daß für jede Hypothese H_{0i} ein spezifischer Test δ_i (zum Niveau α) zur Verfügung steht, ein entsprechender Abschlußtest durchführen.
Dieser Abschlußtest lehnt dabei die Hypothese H_{0i} genau dann ab, falls gilt:

$$\delta_j(x) = 1 \qquad \text{für } i \le j \le n \quad .$$

Die Erfordernis eines spezifischen Tests δ_i zur Überprüfung der Hypothese $H_{0i} = H_1 \cap \dots \cap H_i$ dürfte sich allerdings bei manchen praktischen Anwendungen als recht restriktiv erweisen.
Diese Voraussetzung wird daher in den folgenden Abschnitten, in denen vor allem die Anderson-Prozedur sowie die verbesserte Variante von Bauer/Hackl(1986) behandelt werden, entscheidend modifiziert.
Dabei wird sich insbesondere herausstellen, daß sich die Anderson-Prozedur als spezieller Abschlußtest darstellen läßt.

1. Streng hierarchische Teststrukturen.

Zunächst seien noch einmal die oben genannten Voraussetzungen bezüglich der Mengen $H_1,\dots,H_n$ angegeben:
$(S,\mathcal{B},\{P_\theta\colon \theta\in\Gamma\})$ sei ein statistischer Raum und $H_1,\dots,H_n$ seien Teilmengen von Γ mit

$$\begin{aligned} &H_i \neq \Gamma && \text{für } i=1,\dots,n \\ &H_1\cap\dots\cap H_n \neq \emptyset && \\ \text{und}\quad &H_1\cap\dots\cap H_{i+1} \subset H_1\cap\dots\cap H_i && \text{für } i=1,\dots,n-1 \ . \end{aligned}$$

Die weiteren Voraussetzungen der Anderson-Prozedur lauten dann wie folgt:
Sei $(S,\mathcal{B},\{P_\theta\colon \theta\in\Gamma_i\})$, $i=1,..,n$, eine endliche Folge von statistischen Räumen mit den zugehörigen Parameterräumen

$$\Gamma_1 := \Gamma$$
$$\text{und} \qquad \Gamma_i := H_1 \cap \ldots \cap H_{i-1} \qquad i=2,\ldots,n\ .$$

Für jedes i=1,...,n sei außerdem

$$H_{0i}^* \ :\ \Gamma_i \setminus H_{0i}^* \qquad \text{mit } H_{0i}^* = H_1 \cap \ldots \cap H_i\ ,$$

ein einfaches Testproblem bezüglich $(S, \mathcal{B}, \{P_\theta\colon \theta \in \Gamma_i\})$ und δ_i ein Test zum Niveau α_i $(0<\alpha_i<1)$ bezüglich H_{0i}^*, d.h. es gilt

$$\forall \theta \in H_{0i}^* \quad P_\theta(\delta_i = 1) \le \alpha_i \ .$$

(Auf Grund der besonderen Definition der einzelnen Parameterräume kann man hier auch von einer <u>streng hierarchischen Teststruktur</u> sprechen.)

Mit Hilfe dieser Voraussetzungen läßt sich jetzt der folgende Satz formulieren:

<u>Satz:</u> Sei $0<\alpha<1$. Setzt man $\alpha_1 = \ldots = \alpha_n := \alpha/n$ und $H_{0i} := H_{01}^* \cap \ldots \cap H_{0i}^*$ für i=1,...,n, so ist der durch

$$\sigma_i(x) = \begin{cases} 1 & \max\limits_{1\le j\le i} \delta_j(x) = 1 \\ 0 & \text{sonst} \end{cases} \qquad i=1,\ldots,n$$

definierte multiple Test $\psi = (\sigma_1,\ldots,\sigma_n)$ ein Abschlußtest bezüglich $\{H_{01},\ldots,H_{0n}\}$.

Beweis: Sei $\theta \in H_{0i}$. Dann gilt:

$$P_\theta(\sigma_i = 1) = P_\theta(\max_{1\le j\le i} \delta_j = 1) = P_\theta(\bigcup_{j=1}^{i} \{\delta_j = 1\}) \le$$

$$\le \sum_{j=1}^{i} P_\theta(\delta_j = 1) \le i\cdot\alpha/n \le \alpha \ .$$

Somit ist σ_i ein Test zum Niveau α bezüglich H_{0i}.
Da $\{H_{01},\ldots,H_{0n}\}$ durchschnittsabgeschlossen ist, kann der entsprechende Abschlußtest konstruiert werden.
Wegen

$$\sigma_i(x) = 1 \iff \forall H_{0j} \subset H_{0i} \quad \sigma_j(x) = 1$$

ist aber gerade $\psi = (\sigma_1,\ldots,\sigma_n)$ dieser Abschlußtest.

<u>Bem.:</u> Der Abschlußtest $\Psi=(\sigma_1,\ldots,\sigma_n)$ bezieht sich natürlich auf den zugrundeliegenden Raum $(S,\mathcal{B},\{P_\theta: \theta\in\Gamma\})$. Durch die Definition von H_{0i} und σ_i, $1\leq i\leq n$, werden gerade sämtliche einfachen Testprobleme der zugehörigen statistischen Räume zu einem einzigen multiplen Testproblem zusammengefaßt. Dabei ist besonders zu beachten, daß für die Überprüfung der Hypothese $H_{0i}=H_{01}^*\cap\ldots\cap H_{0i}^*$ die Gültigkeit von H_{0j}^*, $i<j\leq n$, nicht vorausgesetzt werden kann!

Der im obigen Satz definierte Abschlußtest wird hier auch als Anderson-Prozedur bezeichnet, obwohl Anderson(1962,1971) für sein Verfahren eigentlich ganz spezielle unabhängige Tests verwendete, für die eine Reihe von Optimalitätsaussagen gemacht werden konnten.
(Siehe dazu auch die Ausführungen in Phillips/McCabe(1985).)
Bei Unabhängigkeit der δ_j, $1\leq j\leq n$, lassen sich übrigens die Voraussetzungen des Satzes ohne weiteres modifizieren:
Man setze dazu $\alpha_1=\ldots=\alpha_n:=1-\sqrt[n]{1-\alpha}$ und für $\theta\in H_{0i}$ gilt dann:

$$P_\theta\left(\bigcup_{j=1}^{i}\{\delta_j=1\}\right) = 1-P_\theta\left(\bigcap_{j=1}^{i}\{\delta_j=0\}\right) = 1-\prod_{j=1}^{i} P_\theta(\delta_j=0) \leq$$

$$\leq 1-(\sqrt[n]{1-\alpha})^i \leq 1-(\sqrt[n]{1-\alpha})^n = \alpha \quad ,$$

wobei die Unabhängigkeit der δ_j verwendet wurde. Die weitere Beweisführung erfolgt dann analog wie oben.
Vergleicht man die beiden Signifikanzschranken α/n und $1-\sqrt[n]{1-\alpha}$, so gilt zwar auf Grund der Bernoullischen Ungleichung stets

$$\frac{\alpha}{n} \leq 1-\sqrt[n]{1-\alpha} \quad ,$$

aber für kleine Werte von α sind die Unterschiede derart gering, daß bei Vorliegen unabhängiger Tests ohne weiteres die Schranken α/n verwendet werden können.

Es folgt eine kurze Anleitung zur Durchführung der Anderson-Prozedur:

<u>1. Schritt:</u> Man teste zunächst H_{01}^* (mit δ_1) zum Niveau α/n. Wird H_{01}^* abgelehnt, dann werden auch $H_{01},\ldots,H_{0n}$ abgelehnt und das Verfahren ist beendet. Wird H_{01}^* nicht abgelehnt (und somit auch nicht H_{01}) → 2.Schritt.

<u>2.Schritt:</u> Man teste H_{02}^* (mit δ_2) zum Niveau α/n. Wird H_{02}^*

abgelehnt, dann werden auch $H_{02},\dots,H_{0n}$ abgelehnt und das Verfahren ist beendet. Wird H_{02}^* nicht abgelehnt (und somit auch nicht H_{02}) → 3.Schritt.

3.Schritt: .....

.

.

.

n.Schritt: Teste H_{0n}^* zum Niveau α/n. Wird H_{0n}^* abgelehnt, dann wird auch H_{0n} abgelehnt und das Verfahren ist beendet. Wird H_{0n}^* nicht abgelehnt, dann gilt auch keine der Hypothesen $H_{01},\dots,H_{0n}$ als abgelehnt und das Verfahren ist ebenfalls beendet.

Das zu Beginn dieses Kapitels angegebene Testproblem läßt sich somit ohne weiteres mit Hilfe der obigen Anleitung "behandeln". Die zugehörigen Teststatistiken findet man in Phillips(1984).

Von Bauer/Hackl(1986) stammt nun die folgende (sequentiell verwerfende) Variante der Anderson-Prozedur. Es handelt sich dabei im wesentlichen um die Anwendung des Holm-Tests auf die vorliegende (streng hierarchische) Teststruktur. Dabei wird lediglich die zusätzliche Voraussetzung gemacht , das für die einzelnen Hypothesen H_{0i}^* entsprechende p-Funktionen p_i zur Verfügung stehen. Insbesondere werden also stetige Teststatistiken vorausgesetzt.

Satz: (Bauer/Hackl(1986))

Seien $H_{01},\dots,H_{0n}$ wie bisher definiert. Dann ist der multiple Test $\gamma=(\sigma_1,\dots\sigma_n)$ mit

$$\sigma_i(x) = \begin{cases} 1 & \delta_j(x)=1 \qquad i\le j\le n \\ 0 & \text{sonst} \end{cases}$$

und $$\delta_j(x) = \begin{cases} 1 & \min\limits_{1\le k\le j} p_k(x) \le \dfrac{\alpha}{j} \\ 0 & \text{sonst} \end{cases}$$

ein Abschlußtest bezüglich $\{H_{01},\dots,H_{0n}\}$.

Beweis: Auf Grund seiner Konstruktion ist δ_j natürlich ein Test zum Niveau α bezüglich H_{0i}. Für $\theta\in H_{0i}$ gilt nämlich:

$$P_\theta(\delta_j=1) = P_\theta(\min_{1\le k\le j} p_k \le \frac{\alpha}{j}) = P_\theta(\bigcup_{k=1}^{j}\{p_k\le\frac{\alpha}{j}\}) \le$$

$$\le \sum_{k=1}^{j} P_\theta(p_k\le\frac{\alpha}{j}) = j\cdot\frac{\alpha}{j} = \alpha \quad .$$

Die Anwendung des Abschlußprinzips führt dann zum Abschlußtest

$$\psi = (\sigma_1,\ldots,\sigma_n).$$

Ähnlich wie beim Verhältnis zwischen dem Bonferroni-Test und dem Holm-Test (Siehe Kapitel III), so stellt auch die Testprozedur von Bauer/Hackl eine gleichmäßige Verbesserung gegenüber der Anderson-Prozedur dar. Insbesondere lehnt die Testprozedur von Bauer/Hackl jede Nullhypothese ab, die bereits bei der Anderson-Prozedur abgelehnt wurde.

Die praktische Durchführung dieser Testprozedur verläuft wie folgt:

<u>1.Schritt:</u> Sei $p_k(x)=\min_{1\le j\le n} p_j(x)$. Ist $p_k(x)>\frac{\alpha}{n}$, dann wird keine der Hypothesen $H_{01},\ldots,H_{0n}$ abgelehnt und das Verfahren ist beendet.

Ist $p_k(x)\le\frac{\alpha}{n}$, dann werden alle Hypothesen H_{0j}, mit $k\le j\le n$, abgelehnt. → 2.Schritt.

<u>2.Schritt.</u> Sei $p_l(x)=\min_{1\le j\le k-1} p_j(x)$. Ist $p_l(x)>\frac{\alpha}{k-1}$, dann werden keine weiteren Hypothesen abgelehnt und das Verfahren ist beendet.

Ist $p_l(x)\le\frac{\alpha}{k-1}$, dann werden alle Hypothesen H_{0j}, mit $l\le j\le k-1$, abgelehnt. Im Falle $l=1$ ist das Verfahren damit beendet, ansonsten vertausche man k und l. → Wiederholung des 2.Schritts.

2. <u>Verallgemeinerte hierarchische Teststrukturen.</u>

Betrachtet man die bisher untersuchten Testprobleme, so ergeben sich wohl eine Reihe von Überlegungen, die insbesondere eine Verallgemeinerung des Konzepts der streng hierarchischen Teststruktur als wünschenswert erscheinen lassen.

(1) Im letzten Abschnitt wurde das Konzept der Durchschnittsabgeschlossenheit auf streng hierarchische Teststrukturen angewandt. Diese Verbindung führte allerdings zu keinen neuen Hypothesenstrukturen. Die Anordung der zu überprüfenden Null-

hypothesen war vorher bereits bekannt und stellt lediglich einen Spezialfall einer durchschnittsabgeschlossenen Familie dar.

(2) Auf jeder einzelnen "Hierarchiestufe" wird nur eine einzige Nullhypothese (H_{0i}^*) untersucht. Es wäre sinnvoll, wenn man auch das Auftreten mehrerer Hypothesen zulassen würde.

(3) Sämtliche bisher behandelten Testprobleme (auch diejenigen früherer Kapitel) sollten aus einer einheitlichen Teststruktur abgeleitet werden können.

(4) Die Erweiterung einer gegebenen Teststruktur sollte u.a. auch durch entsprechende Anwendungsbeispiele gerechtfertigt werden.

Mit Hilfe einer einfachen Modifizierung der in Abschnitt 1 gemachten Voraussetzungen, läßt sich nun eine "verallgemeinerte" hierarchische Teststruktur konstruieren, die durchaus die obigen Überlegungen berücksichtigt:

An Stelle von H_{0i}^* werden jetzt für jeden Parameterraum Γ_i, $i=1,\ldots,n$, endlich viele Mengen $H_{0ij} \subset \Gamma_i$, $j=1,\ldots,n_i$, vorausgesetzt, die die folgenden Bedingungen erfüllen sollen:

a) $$H_{0ij} \neq \emptyset \;,\; H_{0ij} \neq \Gamma_i \qquad j=1,\ldots,n_i$$
$$H_{0ij} \neq H_{0ik} \qquad j \neq k$$

b) Die Familie $\{H_{0ij},\ j=1,\ldots,n_i\}$ ist durchschnitts-abgeschlossen.

c) $$\bigcap_{j=1}^{n_i} H_{0ij} = H_1 \cap \ldots \cap H_i \quad (=\Gamma_{i+1})$$

Ausgehend von dieser neuen Teststruktur läßt sich jetzt wiederum ein Abschlußtest konstruieren:

<u>Satz:</u> Sei $0<\alpha<1$. Für $i=1,\ldots,n$ definiere man die Hypothesen:

$$H_{01j_1} \cap H_{02j_2} \cap \ldots \cap H_{0ij_i} \qquad \begin{array}{l} j_1=1,\ldots,n_1 \\ j_2=1,\ldots,n_2 \\ \vdots \\ j_i=1,\ldots,n_i \end{array} .$$

Dann gilt:

(a) Die Familie aller dieser Hypothesen ist durchschnitts-abgeschlossen.

(b) Existiert zu jeder einzelnen Durchschnittshypothese $H_{01j_1} \cap \ldots \cap H_{0ij_i}$ ein entsprechender Niveau-α-Test, dann erhält man durch Anwendung des Abschlußprinzips einen Abschlußtest.

Beweis: Es ist lediglich die Durchschnittsabgeschlossenheit zu zeigen.
Seien dazu $H_{01j_1} \cap \ldots \cap H_{0ij_i}$ und $H_{01k_1} \cap \ldots \cap H_{01k_l}$ zwei beliebige Durchschnittshypothesen (o.B.d.A. sei $i \leq l$). Bildet man dann den Durchschnitt

$$(H_{01j_1} \cap \ldots \cap H_{0ij_i}) \cap (H_{01k_1} \cap \ldots \cap H_{01k_l}) =$$
$$= (H_{01j_1} \cap H_{01k_1}) \cap \ldots \cap (H_{0ij_i} \cap H_{0ik_i}) \cap \ldots \cap H_{01k_l} \quad ,$$

so erhält man auf Grund der Tatsache, daß jede einzelne Familie $\{H_{0mj},\ j=1,\ldots,n_m\}$, $m=1,\ldots,n$, durchschnittsabgeschlossen ist, eine Durchschnittshypothese gemäß der im obigen Satz angegebenen Konstruktion. Damit ist aber die Familie aller dieser Hypothesen durchschnittsabgeschlossen.

Der obige Satz stellt gerade die Anwendung des Abschlußprinzips auf verallgemeinerte hierarchische Strukturen dar. Die bisher behandelten Testprozeduren ergeben sich daher unmittelbar als Spezialfälle:
So läßt sich z.B. im Falle $n_i=1$, $i=1,\ldots,n$ (eine einzelne Nullhypothese ist trivialerweise durchschnittsabgeschlossen!), unter Verwendung der entsprechenden Niveau-α-Tests, sowohl die Anderson-Prozedur als auch die Testprozedur von Bauer/Hackl ableiten. Für $n=1$ können dann weitere Testprozeduren, wie etwa der Bonferroni-Test, der Holm-Test und der exakte Abschlußtest hergeleitet werden.
Im folgenden soll ein einfaches Beispiel betrachtet werden, das insbesondere für ökonometrische Anwendungen von Interesse ist.

Gegeben sei ein Modell von der Form

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 + u_t \qquad t=1,\ldots.,T$$
$$\text{mit } u_t = \rho\, u_{t-1} + \varepsilon_t \quad \text{und } \varepsilon_t \underset{\sim}{\text{iid}} N(0,\sigma^2).$$

Definiert man $\Gamma = \{(\beta_1,\beta_2,\beta_3,\sigma^2,\rho):\ \beta_i \in \mathbb{R},\ i=1,2,3,\ \sigma^2>0,\ |\rho|<1\}$, so läßt sich folgende hierarchische Teststruktur konstruieren:

$\underline{\Gamma_1=\Gamma:}$ $\quad H_{011}:\ \rho=0$

$\underline{\Gamma_2=H_{011}:}$ $\quad H_{021}:\ \beta_1=0 \qquad H_{022}:\ \beta_2=0$

$\qquad H_{023}:\ \beta_1=\beta_2=0$.

Die durchschnittsabgeschlossene Familie besteht dann aus den Hypothesen

$$H_{01}:\ \varrho=0$$

$$H_{02}:\ \varrho=0,\ \beta_1=0 \qquad\qquad H_{03}:\ \varrho=0, \beta_2=0$$

$$H_{04}:\ \varrho=0,\ \beta_1=\beta_2=0 \qquad .$$

Verwendet man für H_{011} die p-Funktion des Durbin-Watson Tests (p_1) und für die übrigen Hypothesen $H_{021}, H_{022}, H_{023}$ die entsprechenden p-Funktionen des t- bzw. F-Tests (p_1, p_2, p_3), dann sind mit

$$\delta_1(x) = \begin{cases} 1 & p_1(x) \le \alpha \\ 0 & \text{sonst} \end{cases}$$

$$\delta_i(x) = \begin{cases} 1 & \min(p_1(x), p_i(x)) \le \dfrac{\alpha}{2} \\ 0 & \text{sonst} \end{cases} \qquad i=2,3,4$$

sämtliche Voraussetzungen des obigen Satzes erfüllt und der zugehörige Abschlußtest kann durchgeführt werden.

3. Empirischer Vergleich verschiedener Testprozeduren.

Auf der Grundlage einer einfachen Modellgleichung von der Form

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

mit den Annahmen ($i=1,2$)

- X_i – T_ixK Matrix (nichtstochastisch, mit den Werten der unabhängigen Variablen), $T_i>K$, $rg(X_i)=K$, $T_1+T_2=:T$
- β_i – unbekannte Kx1 Parametervektoren
- y_i – T_ix1 Spaltenvektoren (mit den Werten der abhängigen Variablen)
- u_i – T_ix1 Störvektoren, mit $u_i \sim N(0, \sigma_i^2 I_{T_i})$ und $E(u_1 u_2')=0$

wurden eine Reihe von Simulationsexperimenten durchgeführt, deren Ergebnisse am Ende dieses Abschnitts in einer repräsentativen Auswahl zusammengestellt sind.

Die bei diesem Modell interessierenden Hypothesen

$$H_{01}^*:\ \sigma_1^2=\sigma_2^2$$

$$H_{02}^*:\ \beta_1=\beta 2 \qquad (\text{gegeben } \sigma_1^2=\sigma_2^2\,!)$$

wurden jeweils mit Hilfe der Anderson-Prozedur und der Prozedur von Bauer/Hackl getestet.

Die dabei zugrundeliegenden Parameterräume lauten

$$\Gamma_1 = \{(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2): \beta_i \in R^{T_i}, \sigma_i^2 > 0, i=1,2\}$$
$$\Gamma_2 = \{(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2): \beta_i \in R^{T_i}, \sigma_i^2 > 0, i=1,2, \underline{\text{und}}\ \sigma_1^2 = \sigma_2^2\} .$$

Somit gilt:

$$H_{01} = H_{01}^*: \sigma_1^2 = \sigma_2^2$$
$$H_{02} = H_{01}^* \cap H_{02}^*: \sigma_1^2 = \sigma_2^2, \beta_1 = {}_2 \quad .$$

Phillips(1984) verwendet für die Hypothese H_{01} die Teststatistik

$$F_1 = \frac{RSS_2}{RSS_1} \cdot \frac{T_1 - K}{T_2 - K} \qquad F(T_2 - K, T_1 - K)$$

und für die Hypothese H_{02} die Teststatistik

$$F_2 = \frac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \cdot \frac{T_1 + T_2 - 2K}{K} \qquad F(K, T_1 + T_2 - 2K)$$

wobei RSS_1, RSS_2 und RSS die jeweiligen Residuenquadratsummen bedeuten.

Phillips/McCabe(1983) zeigen, daß diese beiden Teststatistiken unabhängig sind und daher hängt die Verteilung von F_1 von der Gültigkeit der Hypothese H_{02} nicht ab!
Damit sind sämtliche Voraussetzungen für die Anwendung beider Testprozeduren erfüllt. (Zur konkreten Durchführung siehe die Anleitungen in Abschnitt 1).

Es folgen die Angaben für das Simulationsdesign.

Die Anzahl der "Beobachtungen" wurde mit T=40 bzw. $T_1 = T_2 = 20$, die Zahl der Variablen mit K=4 (einschließlich Absolutglied) festgesetzt. Für $X_1 = [x_{11}, x_{12}, x_{13}, 1]$ und $X_2 = [x_{21}, x_{22}, x_{23}, 1]$ wurden Variablen aus Kapitel IV. verwendet, und zwar

x_{i1} – Nichtdauerhafter Konsum, verzögert(lag 1), saisonbereinigt
x_{i2} – Disponibles Einkommen, saisonbereinigt
x_{i3} – Disponibles Einkommen, verzögert(lag 1), saisonbereinigt
i=1,2 .

Die Koeffizienten $\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}$ und $\beta_{21}, \beta_{23}, \beta_{24}$ der Vektoren $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14})'$ und $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23}, \beta_{24})'$ wurden wie folgt definiert:

$$\beta_{11}=\beta_{21}=0.4$$
$$\beta_{12}=\beta_{22}=0.1$$
$$\beta_{13}=0.2$$
$$\beta_{14}=\beta_{24}=10.0 \quad .$$

Die Werte für β_{23} wurden variiert (β_{23}=0.21,0.22,0.23,0.24,0.25). Die Varianz σ_1^2 betrug σ_1^2=2.0, während σ_2^2 wiederum verschiedene Werte annehmen konnte (σ_2^2=2.0,3.0,4.0,5.0,6.0,7.0,8.0,9.0,10.0). Zu jeder der 45 Kombinationsmöglichkeiten wurden 1000 Simulationen durchgeführt (α=0.05).

Natürlich stand fest, daß die Ablehnungswahrscheinlichkeiten der Testprozedur von Bauer/Hackl mindestens so groß ausfallen würden wie die entsprechenden Werte der Anderson-Prozedur. Dennoch sind die folgenden Ergebnisse recht interessant, weil sie zeigen, wie groß die jeweiligen Unterschiede u.U. werden können (bei entsprechender Variation von β_{23} und σ_2^2).
(Daß die Ergebnisse für H_{02} übereinstimmen ergibt sich unmittelbar aus der Konstruktion der beiden Testprozeduren.)

Tabelle 12: Empirische Ablehnungswahrscheinlichkeiten (σ_2^2=2.0)

β_{23}	Anderson-Prozedur		Bauer/Hackl-Prozedur	
	H_{01}	H_{02}	H_{01}	H_{02}
0.21	.024	.104	.025	.104
0.22	.016	.342	.029	.342
0.23	.023	.756	.046	.756
0.24	.029	.977	.053	.977
0.25	.018	1.000	.037	1.000

Tabelle 13: Empirische Ablehnungswahrscheinlichkeiten ($\sigma_2^2=3.0$)

β_{23}	Anderson-Prozedur		Bauer/Hackl-Prozedur	
	H_{01}	H_{02}	H_{01}	H_{02}
0.21	.110	.156	.112	.156
0.22	.109	.354	.124	.354
0.23	.117	.681	.170	.681
0.24	.099	.918	.172	.918
0.25	.129	.993	.204	.993

Tabelle 14: Empirische Ablehnungswahrscheinlichkeiten ($\sigma_2^2=4.0$)

β_{23}	Anderson-Prozedur		Bauer/Hackl-Prozedur	
	H_{01}	H_{02}	H_{01}	H_{02}
0.21	.260	.317	.267	.317
0.22	.277	.463	.301	.463
0.23	.248	.686	.319	.686
0.24	.257	.914	.337	.914
0.25	.293	.982	.402	.982

Tabelle 15: Empirische Ablehnungswahrscheinlichkeiten ($\sigma_2^2=5.0$)

β_{23}	Anderson-Prozedur		Bauer/Hackl-Prozedur	
	H_{01}	H_{02}	H_{01}	H_{02}
0.21	.443	.481	.449	.481
0.22	.431	.543	.449	.543
0.23	.405	.723	.476	.723
0.24	.449	.921	.584	.921
0.25	.448	.986	.575	.986

Tabelle 16: Empirische Ablehnungswahrscheinlichkeiten ($\sigma_2^2=6.0$)

β_{23}	Anderson-Prozedur		Bauer/Hackl-Prozedur	
	H_{01}	H_{02}	H_{01}	H_{02}
0.21	.583	.614	.591	.614
0.22	.555	.646	.571	.646
0.23	.542	.761	.602	.761
0.24	.539	.911	.655	.911
0.25	.558	.981	.666	.981

Tabelle 17: Empirische Ablehnungswahrscheinlichkeiten ($\sigma_2^2=7.0$)

β_{23}	Anderson-Prozedur		Bauer/Hackl-Prozedur	
	H_{01}	H_{02}	H_{01}	H_{02}
0.21	.687	.706	.692	.706
0.22	.673	.731	.691	.731
0.23	.694	.844	.738	.844
0.24	.672	.916	.750	.916
0.25	.686	.984	.765	.984

Tabelle 18: Empirische Ablehnungswahrscheinlichkeiten ($\sigma_2^2=8.0$)

β_{23}	Anderson-Prozedur		Bauer/Hackl-Prozedur	
	H_{01}	H_{02}	H_{01}	H_{02}
0.21	.753	.784	.758	.784
0.22	.752	.821	.778	.821
0.23	.807	.880	.841	.880
0.24	.767	.948	.842	.948
0.25	.773	.981	.845	.981

Tabelle 19: Empirische Ablehnungswahrscheinlichkeiten ($\sigma_2^2=9.0$)

β_{23}	Anderson-Prozedur H_{01}	Anderson-Prozedur H_{02}	Bauer/Hackl-Prozedur H_{01}	Bauer/Hackl-Prozedur H_{02}
0.21	.814	.835	.823	.835
0.22	.827	.868	.842	.868
0.23	.827	.904	.855	.904
0.24	.842	.963	.888	.963
0.25	.829	.979	.886	.979

Tabelle 20: Empirische Ablehnungswahrscheinlichkeiten ($\sigma_2^2=10.0$)

β_{23}	Anderson-Prozedur H_{01}	Anderson-Prozedur H_{02}	Bauer/Hackl-Prozedur H_{01}	Bauer/Hackl-Prozedur H_{02}
0.21	.867	.881	.872	.881
0.22	.874	.895	.885	.895
0.23	.881	.928	.897	.928
0.24	.870	.963	.915	.963
0.25	.879	.984	.929	.984

ANHANG

Tabelle 21: Empirische Ablehnungswahrscheinlichkeiten (1)
($\beta_1=0.0, \beta_2=0.0, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.006	.009	.013	.605
Holm-Test	.008	.012	.015	.605
Exakter Abschlußtest	.012	.009	.012	.634
Scheffè-Test	.003	.003	.000	.379
LSD-Test	.035	.041	.048	.816
Multipler t-Test	.035	.041	.048	.816

Tabelle 22: Empirische Ablehnungswahrscheinlichkeiten (2)
($\beta_1=0.0, \beta_2=0.0, \beta_3=0.25, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.016	.010	.055	.578
Holm-Test	.022	.015	.062	.579
Exakter Abschlußtest	.023	.014	.068	.610
Scheffè-Test	.003	.002	.012	.356
LSD-Test	.060	.044	.152	.783
Multipler t-Test	.060	.044	.152	.783

Tabelle 23: Empirische Ablehnungswahrscheinlichkeiten (3)
($\beta_1=0.0, \beta_2=0.0, \beta_3=0.5, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.009	.011	.259	.630
Holm-Test	.017	.022	.290	.642
Exakter Abschlußtest	.020	.026	.334	.678
Scheffè-Test	.002	.004	.110	.391
LSD-Test	.039	.047	.461	.798
Multipler t-Test	.039	.047	.461	.798

Tabelle 24: Empirische Wahrscheinlichkeiten für die Ablehnung wenigstens einer falschen Nullhypothesen

Testprozedur	Kombination		
	(1)	(2)	(3)
Bonferroni-Test	.605	.584	.693
Holm-Test	.605	.584	.693
Exakter Abschlußtest	.634	.623	.745
Scheffè-Test	.379	.357	.431
LSD-Test	.816	.799	.864
Multipler t-Test	.816	.799	.864

Tabelle 25: Empirische Ablehnungswahrscheinlichkeiten (4)
($\beta_1=0.0, \beta_2=0.25, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.018	.061	.015	.619
Holm-Test	.027	.069	.021	.622
Exakter Abschlußtest	.029	.070	.025	.688
Scheffè-Test	.027	.069	.021	.622
LSD-Test	.064	.163	.047	.794
Multipler t-Test	.064	.163	.047	.794

Tabelle 26: Empirische Ablehnungswahrscheinlichkeiten (5)
($\beta_1=0.0, \beta_2=0.25, \beta_3=0.25, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.013	.071	.055	.603
Holm-Test	.020	.077	.071	.607
Exakter Abschlußtest	.028	.097	.086	.649
Scheffè-Test	.004	.023	.022	.373
LSD-Test	.063	.172	.140	.790
Multipler t-Test	.063	.172	.140	.790

Tabelle 27: Empirische Ablehnungswahrscheinlichkeiten (6)
($\beta_1=0.0, \beta_2=0.25, \beta_3=0.5, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.022	.082	.226	.605
Holm-Test	.028	.092	.262	.616
Exakter Abschlußtest	.037	.113	.349	.705
Scheffè-Test	.004	.024	.078	.373
LSD-Test	.053	.181	.438	.801
Multipler t-Test	.053	.181	.438	.801

Tabelle 28: Empirische Wahrscheinlichkeiten für die Ablehnung wenigstens einer "falschen" Nullhypothese

Testprozedur	Kombination		
	(4)	(5)	(6)
Bonferroni-Test	.660	.651	.702
Holm-Test	.662	.651	.704
Exakter Abschlußtest	.730	.709	.813
Scheffè-Test	.417	.397	.427
LSD-Test	.852	.862	.903
Multipler t-Test	.852	.862	.903

Tabelle 29: Empirische Ablehnungswahrscheinlichkeiten (7)
($\beta_1=0.0, \beta_2=0.5, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.010	.315	.009	.620
Holm-Test	.015	.337	.014	.643
Exakter Abschlußtest	.020	.384	.023	.718
Scheffè-Test	.003	.146	.003	.356
LSD-Test	.049	.545	.051	.816
Multipler t-Test	.049	.545	.051	.816

Tabelle 30: Empirische Ablehnungswahrscheinlichkeiten (8)
($\beta_1=0.0, \beta_2=0.5, \beta_3=0.25, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.007	.335	.071	.618
Holm-Test	.019	.366	.087	.638
Exakter Abschlußtest	.027	.432	.106	.693
Scheffè-Test	.000	.155	.022	.392
LSD-Test	.052	.554	.155	.800
Multipler t-Test	.052	.554	.155	.800

Tabelle 31: Empirische Ablehnungswahrscheinlichkeiten (9)
($\beta_1=0.0, \beta_2=0.5, \beta_3=0.5, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.011	.335	.193	.575
Holm-Test	.024	.370	.231	.602
Exakter Abschlußtest	.041	.422	.319	.692
Scheffè-Test	.001	.158	.082	.362
LSD-Test	.051	.557	.408	.791
Multipler t-Test	.051	.557	.408	.791

Tabelle 32: Empirische Wahrscheinlichkeiten für die Ablehnung wenigstens einer "falschen" Nullhypothese

Testprozedur	Kombination		
	(7)	(8)	(9)
Bonferroni-Test	.793	.792	.794
Holm-Test	.793	.792	.795
Exakter Abschlußtest	.860	.862	.905
Scheffè-Test	.472	.511	.509
LSD-Test	.947	.949	.972
Multipler t-Test	.947	.949	.972

Tabelle 33: Empirische Ablehnungswahrscheinlichkeiten (10)
($\beta_1=0.25, \beta_2=0.0, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.033	.016	.012	.584
Holm-Test	.035	.022	.021	.589
Exakter Abschlußtest	.041	.020	.018	.624
Scheffè-Test	.005	.001	.001	.378
LSD-Test	.100	.050	.049	.790
Multipler t-Test	.100	.050	.049	.790

Tabelle 34: Empirische Ablehnungswahrscheinlichkeiten (11)
($\beta_1=0.25, \beta_2=0.0, \beta_3=0.25, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.040	.013	.046	.606
Holm-Test	.045	.017	.053	.610
Exakter Abschlußtest	.067	.020	.068	.635
Scheffè-Test	.009	.001	.011	.363
LSD-Test	.121	.048	.134	.799
Multipler t-Test	.121	.048	.134	.799

Tabelle 35: Empirische Ablehnungswahrscheinlichkeiten (12)
($\beta_1=0.25, \beta_2=0.0, \beta_3=0.5, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.038	.011	.205	.592
Holm-Test	.038	.018	.227	.600
Exakter Abschlußtest	.053	.022	.310	.655
Scheffè-Test	.012	.002	.092	.349
LSD-Test	.098	.051	.398	.800
Multipler t-Test	.098	.051	.398	.800

Tabelle 36: Empirische Wahrscheinlichkeiten für die Ablehnung wenigstens einer "falschen" Nullhypothese

Testprozedur	Kombination		
	(10)	(11)	(12)
Bonferroni-Test	.609	.652	.675
Holm-Test	.613	.653	.675
Exakter Abschlußtest	.654	.695	.759
Scheffè-Test	.383	.377	.410
LSD-Test	.845	.881	.903
Multipler t-Test	.845	.881	903

Tabelle 37: Empirische Ablehnungswahrscheinlichkeiten (13)
($\beta_1=0.25, \beta_2=0.25, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.032	.061	.014	.621
Holm-Test	.038	.072	.017	.627
Exakter Abschlußtest	.039	.081	.020	.703
Scheffè-Test	.007	.018	.002	.387
LSD-Test	.094	.171	.042	.814
Multipler t-Test	.094	.171	.042	.814

Tabelle 38: Empirische Ablehnungswahrscheinlichkeiten (14)
($\beta_1=0.25, \beta_2=0.25, \beta_3=0.25, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.032	.073	.063	.591
Holm-Test	.034	.081	.071	.591
Exakter Abschlußtest	.060	.101	.099	.660
Scheffè-Test	.010	.021	.018	.396
LSD-Test	.112	.173	.133	.799
Multipler t-Test	.112	.173	.133	.799

Tabelle 39: Empirische Ablehnungswahrscheinlichkeiten (15)
($\beta_1=0.25, \beta_2=0.25, \beta_3=0.5, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.045	.072	.231	.603
Holm-Test	.048	.085	.252	.614
Exakter Abschlußtest	.073	.118	.391	.705
Scheffè-Test	.008	.028	.100	.355
LSD-Test	.109	.177	.432	.808
Multipler t-Test	.109	.177	.432	.808

Tabelle 40: Empirische Wahrscheinlichkeiten für die Ablehnung wenigstens einer "falschen" Nullhypothese

Testprozedur	Kombination		
	(13)	(14)	(15)
Bonferroni-Test	.679	.666	.726
Holm-Test	.680	.666	.726
Exakter Abschlußtest	.756	.759	.849
Scheffè-Test	.409	.428	.416
LSD-Test	.899	.914	.942
Multipler t-Test	.899	.914	.942

Tabelle 41: Empirische Ablehnungswahrscheinlichkeiten (16)
($\beta_1=0.25, \beta_2=0.5, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.040	.313	.008	.584
Holm-Test	.046	.341	.021	.598
Exakter Abschlußtest	.060	.377	.032	.690
Scheffè-Test	.013	.142	.002	.363
LSD-Test	.111	.527	.052	.796
Multipler t-Test	.111	.527	.052	.796

Tabelle 42: Empirische Ablehnungswahrscheinlichkeiten (17)
($\beta_1=0.25, \beta_2=0.5, \beta_3=0.25, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.034	.334	.038	.579
Holm-Test	.040	.363	.052	.597
Exakter Abschlußtest	.085	.426	.096	.656
Scheffè-Test	.009	.178	.008	.365
LSD-Test	.119	.579	.131	.774
Multipler t-Test	.119	.579	.131	.774

Tabelle 43: Empirische Ablehnungswahrscheinlichkeiten (18)
($\beta_1=0.25, \beta_2=0.5, \beta_3=0.5, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.032	.358	.240	.610
Holm-Test	.041	.390	.274	.642
Exakter Abschlußtest	.079	.450	.388	.729
Scheffè-Test	.011	.168	.103	.399
LSD-Test	.097	.581	.419	.801
Multipler t-Test	.097	.581	.419	.801

Tabelle 44: Empirische Wahrscheinlichkeiten für die Ablehnung wenigstens einer "falschen" Nullhypothese

Testprozedur	Kombination		
	(16)	(17)	(18)
Bonferroni-Test	.767	.766	.848
Holm-Test	.767	.766	.848
Exakter Abschlußtest	.853	.869	.944
Scheffè-Test	.473	.502	.560
LSD-Test	.956	.968	.984
Multipler t-Test	.956	.968	.984

Tabelle 45: Empirische Ablehnungswahrscheinlichkeiten (19)
($\beta_1=0.5, \beta_2=0.0, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.121	.013	.006	.603
Holm-Test	.126	.015	.014	.612
Exakter Abschlußtest	.150	.017	.013	.637
Scheffè-Test	.043	.002	.000	.364
LSD-Test	.284	.051	.037	.818
Multipler t-Test	.284	.051	.037	.818

Tabelle 46: Empirische Ablehnungswahrscheinlichkeiten (20)
($\beta_1=0.5, \beta_2=0.0, \beta_3=0.25, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.142	.015	.059	.585
Holm-Test	.150	.020	.067	.589
Exakter Abschlußtest	.202	.027	.091	.622
Scheffè-Test	.049	.002	.017	.389
LSD-Test	.306	.047	.154	.801
Multipler t-Test	.306	.047	.154	.801

Tabelle 47: Empirische Ablehnungswahrscheinlichkeiten (21)
($\beta_1=0.5, \beta_2=0.0, \beta_3=0.5, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.133	.009	.198	.623
Holm-Test	.141	.017	.222	.641
Exakter Abschlußtest	.172	.025	.322	.672
Scheffè-Test	.041	.001	.087	.377
LSD-Test	.280	.043	.414	.806
Multipler t-Test	.280	.043	.414	.806

Tabelle 48: Empirische Wahrscheinlichkeiten für die Ablehnung wenigstens einer "falschen" Nullhypothese

Testprozedur	Kombination		
	(19)	(20)	(21)
Bonferroni-Test	.688	.702	.769
Holm-Test	.690	.702	.769
Exakter Abschlußtest	.735	.768	.842
Scheffè-Test	.403	.435	.462
LSD-Test	.936	.939	.957
Multipler t-Test	.936	.939	.957

Tabelle 49: Empirische Ablehnungswahrscheinlichkeiten (22)
($\beta_1=0.5, \beta_2=0.25, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.110	.085	.004	.591
Holm-Test	.120	.093	.010	.602
Exakter Abschlußtest	.148	.108	.019	.679
Scheffè-Test	.042	.020	.001	.348
LSD-Test	.259	.202	.034	.795
Multipler t-Test	.259	.202	.034	.795

Tabelle 50: Empirische Ablehnungswahrscheinlichkeiten (23)
($\beta_1=0.5, \beta_2=0.25, \beta_3=0.25, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.129	.052	.057	.589
Holm-Test	.137	.063	.070	.598
Exakter Abschlußtest	.186	.105	.128	.682
Scheffè-Test	.045	.017	.013	.371
LSD-Test	.265	.153	.151	.805
Multipler t-Test	.265	.153	.151	.805

Tabelle 51: Empirische Ablehnungswahrscheinlichkeiten (24)
($\beta_1=0.5, \beta_2=0.25, \beta_3=0.5, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.129	.082	.224	.593
Holm-Test	.145	.096	.249	.609
Exakter Abschlußtest	.215	.144	.377	.681
Scheffè-Test	.051	.023	.102	.345
LSD-Test	.277	.189	.416	.783
Multipler t-Test	.277	.189	.416	.783

Tabelle 52: Empirische Wahrscheinlichkeiten für die Ablehnung wenigstens einer "falschen" Nullhypothese

Testprozedur	Kombination		
	(22)	(23)	(24)
Bonferroni-Test	.724	.716	.774
Holm-Test	.724	.716	.774
Exakter Abschlußtest	.810	.839	.908
Scheffè-Test	.403	.423	.451
LSD-Test	.937	.942	.971
Multipler t-Test	.937	.941	.971

Tabelle 53: Empirische Ablehnungswahrscheinlichkeiten (25)
($\beta_1=0.5, \beta_2=0.5, \beta_3=0.0, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.118	.309	.011	.626
Holm-Test	.134	.345	.020	.643
Exakter Abschlußtest	.191	.394	.035	.717
Scheffè-Test	.041	.143	.004	.400
LSD-Test	.258	.542	.044	.813
Multipler t-Test	.258	.542	.044	.813

Tabelle 54: Empirische Ablehnungswahrscheinlichkeiten (26)
($\beta_1=0.5, \beta_2=0.5, \beta_3=0.25, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.116	.311	.063	.608
Holm-Test	.134	.345	.069	.634
Exakter Abschlußtest	.224	.443	.128	.689
Scheffè-Test	.039	.145	.017	.378
LSD-Test	.271	.551	.137	.818
Multipler t-Test	.271	.551	.137	.818

Tabelle 55: Empirische Ablehnungswahrscheinlichkeiten (27)
($\beta_1=0.5, \beta_2=0.5, \beta_3=0.5, \beta_4=0.25$)

Testprozedur	Nullhypothesen			
	H_{01}	H_{02}	H_{03}	H_{04}
Bonferroni-Test	.139	.329	.229	.587
Holm-Test	.155	.369	.262	.617
Exakter Abschlußtest	.253	.473	.426	.718
Scheffè-Test	.055	.161	.108	.365
LSD-Test	.283	.559	.440	.800
Multipler t-Test	.283	.559	.440	.800

Tabelle 56: Empirische Wahrscheinlichkeiten für die Ablehnung wenigstens einer "falschen" Nullhypothese

Testprozedur	Kombination		
	(25)	(26)	(27)
Bonferroni-Test	.818	.820	.860
Holm-Test	.819	.820	.860
Exakter Abschlußtest	.925	.937	.975
Scheffè-Test	.528	.521	.570
LSD-Test	.985	.984	.995
Multipler t-Test	.985	.984	.995

LITERATUR

Anderson, T.W.(1962), "The Choice of Degree of a Polynomial Regression as a Multiple Decision Problem", Annals of Mathematical Statistics 33, 255-265.

Anderson, T.W.(1971), *The Statistical Analysis of Time Series.* New York: John Wiley & Sons.

Bauer,P./Hackl P.(1986), "Multiple Testing in a Set of Nested Hypotheses" (erscheint in Statistics 17 (1987)).

Dhrymes, P.J., Howrey, E.P., Hymans, S.H., Kmenta, J., Leamer, E.E., Quandt, R.E., Ramsey, J.B., Shapiro, H.T. and Zanowitz, V. (1972), "Criteria for the Evaluation of Econometric Models", Annals of Economic and Social Measurement 1/3,291-324.

Einot,I./Gabriel, K.R.(1975), "A Study of the Powers of Several Methods of Multiple Comparisons", Journal of the American Statistical Association 70, 574-583.

Gabriel,K.R.(1969), "Simultaneous Test Procedures - Some Theory of Multiple Comparisons", Annals of Mathematical Statistics 40, 224-250.

Holm,St.(1979), "A Simple Sequentially Rejective Multiple Test Procedure", Scandinavian Journal of Statistics 6, 65-70.

Hommel,G.(1984), " Multiple Test Procedures for Arbitrary Dependence Structures", (erscheint in Metrika).

Hommel,G.(1985), "Multiple Vergleiche mittels Rangtests - Alle Paarvergleiche", in: *Neuere Verfahren der nichtparametrischen Statistik*. (Reihe: Medizinische Informatik und Statistik) Hsg. G.Pflug., Berlin: Springer-Verlag.

Hommel,G.(1986), "Grundlagen Multipler Testprobleme", Vortrag vor dem Ausschuss 'Technische Statistik', Deutsche Statistische Gesellschaft, Bamberg.

Johnston,J.(1984), *Econometric Methods* . McGraw Hill, 3rd Edition.

Kiviet,J.F./Phillips,G.D.A.(1985), "Testing Strategies for Model Specification", Faculty of Actuarial Science & Econometrics, University of Amsterdam.

Lehmann,E.L.(1957a), "A Theory of Some Multiple Decision Problems I.", Annals of Mathematical Statistics 28, 1-25.

Lehmann,E.L.(1957b), "A Theory of Some Multiple Decision Problems II.", Annals of Mathematical Statistics 28, 547-572.

Marcus,R./Peritz,E./Gabriel,K.R.(1976), " On Closed Testing Procedures with Special Reference to Ordered Analysis of Variance", Biometrika 63, 655-660.

Miller,R.G.jr.(1966,1981), *Simultaneous Statistical Inference*. New York: McGraw-Hill.(Neuauflage:1981,Berlin, Springer Verlag).

Mizon,G.E.(1977), "Inferential Procedures in Non-Linear Models: An Application in a U.K. Industrial Cross Section Study of Factor Substitution and Returns to Scale", Econometrica 45, 1221-1242.

Naik,U.D.(1975), "Some Selection Rules for Comparing p Processes with a Standard", Communications in Statistics 4, 519-535.

Phillips,G.D.A.(1984), "Some Recent Developments in Econometric Test Methodology", Research Memorandum No. 205 , Institut für Höhere Studien, Wien.

Phillips,G.D.A./McCabe,B.P.M.(1983), "The Independence of Tests for Structural Change in Regression Models", Economic Letters 12, 283-287.

Phillips,G.D.A./McCabe,B.P.M.(1985), "A Sequential Approach to Testing for Misspecification in Simultaneous Equation Models", Econometric Society 5th World Congress, MIT, Cambridge, Massachusetts,USA.

Savin,N.E.(1984), "Multiple Hypotheses Testing", in: *Handbook of Econometrics*, Griliches,Z./Intriligator,M.D.(Hsg.), Amsterdam, North-Holland Publishing.

Sonnemann,E.(1981), "Tests zum Multiplen Niveau α", ROeS-Seminar, Bad Ischl.

Sonnemann,E.(1982), "Allgemeine Lösungen Multipler Testprobleme", EDV in Medizin und Biologie 13, 120-128.

Theil,H.(1971), *Principles of Econometrics*, New York:Wiley & Sons.