

**MULTIPLE TEST PROCEDURES
AND THE CLOSURE PRINCIPLE**

A new look at multiple hypotheses testing
in the linear regression model

Raimund ALT

Forschungsbericht/
Research Memorandum No. 278

March 1991

Presented at the 6th World Congress of the Econometric Society, Barcelona 1990, and at
the WAS'90 Seminar on Multivariate Statistics, Lodz 1990.

Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich des Autors/der Autorin (der Autoren/Autorinnen) und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden. Nachdruck nur auszugsweise und mit genauer Quellenangabe gestattet.

All contributions are to be regarded as preliminary and should not be quoted without consent of the respective author(s). All contributions are personal and any opinions expressed should never be regarded as opinion of the Institute for Advanced Studies.

This series contains investigations by the members of the Institute's staff, visiting professors, and others working in collaboration with our departments.

ZUSAMMENFASSUNG

In dieser Arbeit wird gezeigt, wie mit Hilfe des Abschlußprinzips lineare Hypothesen im Rahmen des klassischen Regressionsmodells getestet werden können. Das von Marcus/Peritz/Gabriel (1976) entwickelte Abschlußprinzip führt zur Konstruktion von Testprozeduren, die im allgemeinen erheblich mächtiger sein können als die herkömmlichen Testprozeduren wie die Bonferroni- oder die Scheffe-Prozedur. Eine Simulationsstudie zeigt anschaulich die empirische Überlegenheit der abgeschlossenen Testprozeduren gegenüber herkömmlichen Verfahren.

SUMMARY

In this paper we show how to apply the closure test principle in case of testing linear hypotheses within the classical linear regression model. The closure test principle which was introduced by Marcus/Peritz/Gabriel (1976) results in the construction of test procedures which are in general much more powerful than conventional test procedures like the Bonferroni procedure or the Scheffe procedure. A small simulation study provides some evidence of the superiority of closed test procedures compared to classical test procedures.

1. Introduction

In biometrics and medical statistics many researchers advocate the use of multiple comparison procedures (multiple test procedures) in case of testing two or more null hypotheses. While testing of multiple hypotheses plays an essential role in econometrics as well, multiple comparison methods are rarely mentioned in econometric literature. One exception, for example, is the survey article from Savin (1984), where he describes different features of two classical multiple test procedures, namely the Bonferroni procedure and the Scheffe procedure.

One disadvantage of these two test procedures is that they are often quite conservative, i.e. the probability of rejecting at least one of the true null hypotheses is in general relatively small compared to the given level α . For this reason it might be useful to consider other test procedures like the so-called closed test procedures, which have been developed by Marcus, Peritz, Gabriel (1976) and which are generally much more powerful. These test procedures seem to be quite unknown in econometrics. In particular, Savin (1984) did not mention them.

The presentation of the paper is now as follows:

After introducing some basic definitions for multiple test procedures it will be shown how to apply the fundamental theorem for closed test procedures (closure principle) to linear hypotheses, for example, that the regression coefficients are equal to zero. Then it will be demonstrated how to apply this principle to a set of general linear hypotheses. Then a small simulation study is presented which provides some evidence of the empirical performance of closed test procedures compared to that of the Bonferroni procedure or the Scheffe procedure.

2. The closure principle: Theoretical foundations

Let H_{01}, \dots, H_{0n} be our interesting null hypotheses, where H_{01}, \dots, H_{0n} are subsets of a parameter space Γ with $H_{0i} \neq H_{0j}$ for $i \neq j$, and let $\bar{\Phi} = (\varphi_1, \dots, \varphi_n)$ be a corresponding test procedure. Now the following definition describes a type I error concept which seems to be quite appropriate in the area of multiple hypotheses testing.

Definition 2.1: Let $0 < \alpha < 1$ and let C_1, \dots, C_n be the critical regions of the tests $\varphi_1, \dots, \varphi_n$. $\bar{\Phi} = (\varphi_1, \dots, \varphi_n)$ controls the multiple level α , if for every non-empty index set $I \subset \{1, \dots, n\}$

$$P\left(\bigcup_{i \in I} C_i \mid H_{0i} \text{ true, } i \in I\right) \leq \alpha.$$

Clearly a test procedure which controls the multiple level α also controls the overall level α , since $I = \{1, \dots, n\}$ is a special subset of $\{1, \dots, n\}$ and so we have

$$P\left(\bigcup_{i=1}^n C_i \mid H_{01}, \dots, H_{0n} \text{ true}\right) \leq \alpha.$$

But the multiple-level- α concept is of course much more general. It guarantees that the probability of rejecting at least one of the true null hypotheses is always smaller or equal than α for every combination of true null hypotheses.

The closure test concept makes also use of the following two definitions.

Definition 2.2: A set $\{H_{01}, \dots, H_{0n}\}$ of null hypotheses is *closed under intersection* if for any two indices $i \neq j$

$$H_{0i} \cap H_{0j} \in \{H_{01}, \dots, H_{0n}\}$$

Definition 2.3: The test procedure $\bar{\Phi} = (\varphi_1, \dots, \varphi_n)$ is called *coherent*, if $H_{0j} \subset H_{0i}$ implies $C_i \subset C_j$, i.e. the rejection of any null hypothesis H_{0i} implies the rejection of every subhypothesis $H_{0j} \subset H_{0i}$.

The closure test theorem is now as follows:

Theorem 2.4: (Marcus, Peritz, Gabriel (1976))

Let H_{01}, \dots, H_{0n} be closed under intersection and let C_i be the critical region of the test φ_i with $P(C_i | H_{0i} \text{ true}) \leq \alpha$, $1 \leq i \leq n$. Then the test rule

Reject H_{0i} , if every subhypothesis $H_{0j} \subset H_{0i}$ is rejected by its level- α -test φ_j

defines a coherent test procedure $\Psi = (\gamma_1, \dots, \gamma_n)$ which controls the multiple level α .

Such a test procedure is called a closed test procedure or a closure test.

For a proof see for example Marcus, Peritz, Gabriel (1976), Sonnemann (1982) or Alt (1988).

In the next section we will demonstrate how to apply this theorem for testing linear hypotheses in the classical regression model.

3. Linear hypotheses and closed test procedures

Let us consider the classical linear regression model

$$y = X\beta + u$$

where u is a $T \times 1$ disturbance vector with $u \sim N(0, \sigma^2 I)$,

$\beta' = (\beta_0, \beta_1, \dots, \beta_K)$ is an unknown $1 \times (K+1)$ parameter vector and X is a $T \times (K+1)$ nonstochastic matrix of rank $K+1$.

We will assume that we are interested in testing the following null hypotheses:

$$H_{01}: \beta_1 = 0$$

.

.

.

$$H_{0K}: \beta_K = 0$$

For the sake of simplicity we will restrict ourselves to the case $K=3$ and in the following subsections we will demonstrate how to construct closed test procedures, i.e. coherent test procedures which control the multiple level α .

3.1. The closed Bonferroni procedure

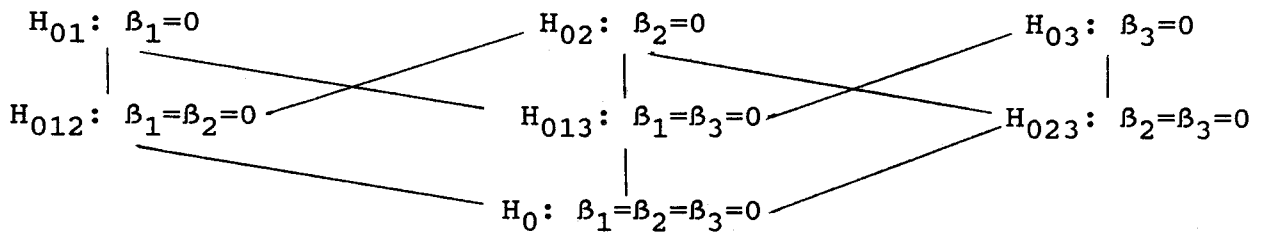
Our interesting null hypotheses are now

$$H_{01}: \beta_1=0$$

$$H_{02}: \beta_2=0$$

$$H_{03}: \beta_3=0$$

Clearly this set of hypotheses is not closed under intersection. In order to obtain this property we have to add all possible intersection hypotheses:



It is easy to verify that the set $\{H_{01}, \dots, H_{023}, H_0\}$ is closed under intersection which means that now the first assumption of the closure theorem is fulfilled.

It is well known that the test statistic

$$Q_i = \frac{\hat{\beta}_i^2}{\hat{\sigma}^2 a_{ii}}$$

is distributed as $F(1, T-K-1)$ under $H_{0i}: \beta_i=0$, where $\hat{\beta}_i$ is the ols estimator of β_i , $\hat{\sigma}^2$ is the usual unbiased estimator of σ^2 and a_{ii} is the i -th diagonal element of the matrix $(X'X)^{-1}$.

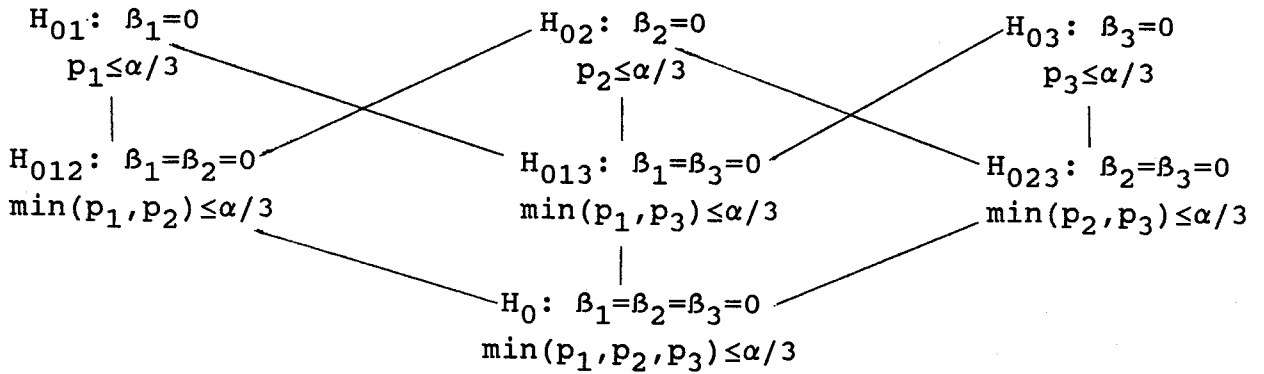
Now let p_1 , p_2 and p_3 be the corresponding p-values, i.e.

$$p_i := P(Q_i \geq Q_i(Y) | H_{0i} \text{ true}) \quad i=1,2,3$$

where $Q_i(y)$ is the value of the test statistic Q_i evaluated at the observed vector y .

Now we will construct a level- α -test for every null hypothesis in $\{H_{01}, \dots, H_{023}, H_0\}$ for a given level α .

Let us consider the following test scheme:



where below each null hypothesis a rejection region is given which defines a level- α -test for this hypothesis, i.e.

$$P(\text{Reject } H | H \text{ true}) \leq \alpha$$

For example, given $H_{01}: \beta_1=0$, we have

$$P(p_1 \leq \alpha/3) = \alpha/3 \leq \alpha$$

or, given $H_{012}: \beta_1=\beta_2=0$, we have

$$\begin{aligned}
 P(\min(p_1, p_2) \leq \alpha/3) &= P(p_1 \leq \alpha/3 \text{ or } p_2 \leq \alpha/3) \leq P(p_1 \leq \alpha/3) + P(p_2 \leq \alpha/3) = \\
 &= \alpha/3 + \alpha/3 = 2\alpha/3 \leq \alpha
 \end{aligned}$$

and so on.

This means that the second assumption of the closure theorem is fulfilled and we can now apply the test rule which results in a coherent test procedure controlling the multiple level α .

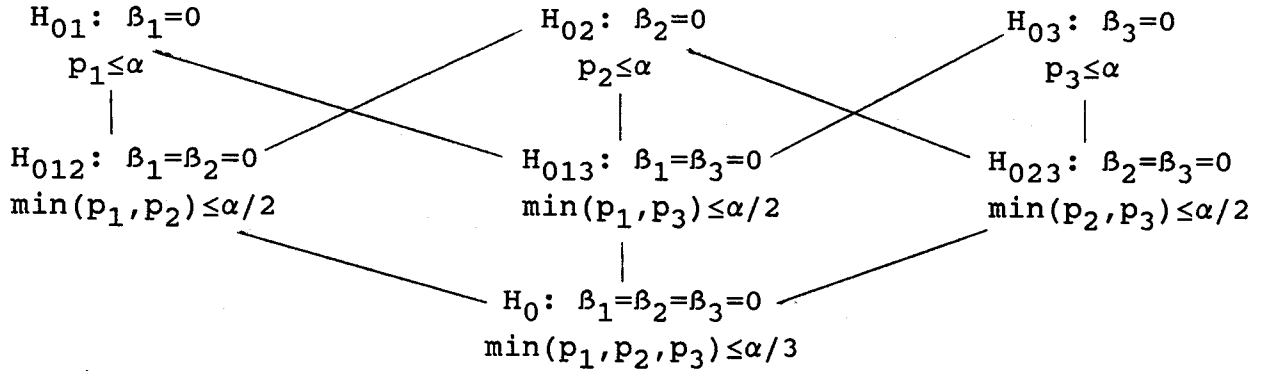
It turns out that if we are only interested in testing H_{01} , H_{02} and H_{03} it is not necessary to consider the intersection hypotheses. If, for example, $p_1 \leq \alpha/3$, then all subhypotheses are rejected by their level- α -tests.

What we have constructed is nothing else than the closed version of the classical Bonferroni procedure on the basis of three null

hypotheses H_{01} , H_{02} and H_{03} . But now we are able to show how to improve the classical Bonferroni procedure.

3.2. The closed Holm procedure

Let us modify our test scheme as follows:



where at the first stage we have replaced $\alpha/3$ by α and at the second stage $\alpha/3$ by $\alpha/2$.

Again each null hypotheses is connected with a critical region defining a level- α -test for this null hypothesis.

For example, given $H_{01}: \beta_1=0$, we have now

$$P(p_1 \leq \alpha) = \alpha$$

or, given $H_{012}: \beta_1=\beta_2=0$, we have

$$\begin{aligned} P(\min(p_1, p_2) \leq \alpha/2) &= P(p_1 \leq \alpha/2 \text{ or } p_2 \leq \alpha/2) \leq P(p_1 \leq \alpha/2) + P(p_2 \leq \alpha/2) \\ &= \alpha/2 + \alpha/2 = \alpha \end{aligned}$$

Applying the closure test rule we again get a test procedure which controls the multiple level α . We will call it the closed Holm procedure, because there is a shortcut version of this procedure which was developed by Holm (1979).

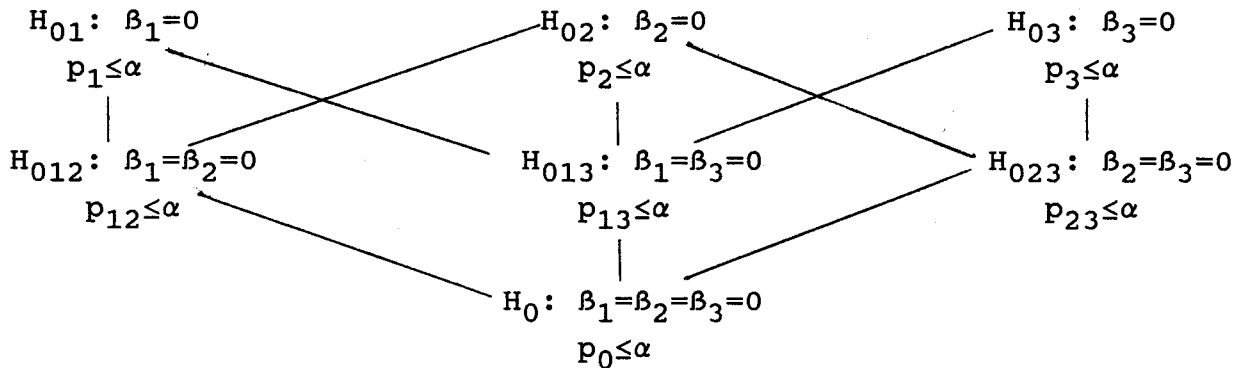
It is easy to see that every null hypothesis which is rejected by the closed Bonferroni procedure is also rejected by the closed Holm procedure. But the latter procedure is able to reject more null hypotheses since the critical values at the first two stages are greater than the ones used for the closed Bonferroni procedure.

Together with the fact that the closed Holm procedure does not exceed the multiple level α , it turns out that the closed Holm procedure is uniformly better than the closed Bonferroni procedure!

3.3. The closed LSD-procedure

There is still another interesting modification which is called the closed LSD procedure because of its relation to Fisher's Least Significant Difference test, Fisher (1935), which was developed in the context of multiple comparisons of several means.

In this case our testing scheme is as follows:



where p_{12} , p_{13} , p_{23} and p_0 are the p-values of the corresponding direct F-tests. Now we have for each null hypothesis of our scheme an exact level- α -test. The application of the closure test rule is straightforward and so we get another example of a closure test. This test procedure seems to be quite attractive, though there is no such simple implication between this test procedure and the two mentioned before. But the results of a simulation study presented in section 4 indicate that the closed LSD procedure might be quite an attractive alternative compared to other multiple test procedures.

3.4. Testing general linear hypotheses

Suppose that we are interested in testing the general null hypotheses

$$H_{01}: C_1\beta = c_1$$

.

.

.

$$H_{0q}: C_q\beta = c_q$$

where C_1, \dots, C_q and c_1, \dots, c_q are known with $C' = [C_1', \dots, C_q']$ having full column rank $= q = \text{rank}(C \ c)$.

Then the application of the closure principle is straightforward. We only need to construct all possible intersection hypotheses by stacking the C_i -vectors. This results in a set of hypotheses based on submatrices of C . The assumptions of the closure principle are fulfilled by defining a level- α -test for each null hypothesis. Then the test rule can be applied and we get a corresponding test procedure controlling the multiple level α .

4. A simulation study

In this section we will present the results of a small simulation study designed to compare the empirical performance of four multiple test procedures, namely the Bonferroni procedure, the Holm procedure, the closed LSD procedure and another classical test procedure, the Scheffe procedure.

A simple regression model of the form

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + u_t \quad t=1, \dots, 120$$

was simulated, where the x_t 's were generated once from the relation

$$x_t = 0.7x_{t-1} + e_t \quad e_t \stackrel{i.i.d.}{\sim} N(0, 1)$$

and remained constant for all experiments. The u_t 's were independent drawings from a $N(0,10.0)$ distributed random variable. The coefficient β_0 was 2.0 while β_1 and β_2 were varied using the values 0.0, 0.2, 0.4, 0.6, 0.8, 1.0. So there were 36 different parameter combinations of β_1 and β_2 .

We considered here the testing of two null hypotheses

$$H_{01}: \beta_1=0$$

$$H_{02}: \beta_2=0$$

where the level α was given by 0.05.

The Bonferroni procedure was carried out by using the p-values p_1 and p_2 , each at level 0.025. The Holm procedure was carried out at level 0.025 for the intersection hypothesis $H_{012}: \beta_1=\beta_2=0$ and 0.05 for the single hypotheses H_{01} and H_{02} . The application of the closed LSD procedure is straightforward while the Scheffe procedure employs the use of the statistics Q_1 and Q_2 , each of which was compared with the critical value $S^2 = 2F_{0.05}(2,117)$, where $F_{0.05}(2,117)$ is the upper 0.05 significance point of an F distribution with 2 and 117 degrees of freedom.

10000 replications were performed for each combination of β_1 and β_2 . The following tables show the estimated rejection probabilities for the Bonferroni (B) procedure, the Holm (H) procedure, the closed LSD (CLSD) procedure and the Scheffe (S) procedure.

Table 1: Estimated Rejection Probabilities
 $H_{01}: \beta_1=0$

β_1	β_2	S	B	H	CLSD
0.0	0.0	.0169	.0259	.0284	.0268
	0.2	.0141	.0236	.0282	.0287
	0.4	.0156	.0264	.0339	.0398
	0.6	.0151	.0267	.0376	.0465
	0.8	.0144	.0258	.0409	.0505
	1.0	.0124	.0247	.0415	.0502
0.2	0.0	.0362	.0530	.0551	.0545
	0.2	.0362	.0573	.0590	.0748
	0.4	.0347	.0529	.0557	.0880
	0.6	.0379	.0561	.0626	.0983
	0.8	.0360	.0544	.0690	.0986
	1.0	.0400	.0607	.0855	.1019
0.4	0.0	.1230	.1664	.1672	.1809
	0.2	.1203	.1686	.1700	.2154
	0.4	.1194	.1645	.1687	.2361
	0.6	.1201	.1651	.1824	.2437
	0.8	.1190	.1636	.1982	.2464
	1.0	.1223	.1687	.2281	.2528
0.6	0.0	.3010	.3748	.3759	.4233
	0.2	.2923	.3622	.3643	.4508
	0.4	.2959	.3739	.3852	.4804
	0.6	.2999	.3705	.4073	.4840
	0.8	.3066	.3718	.4396	.4838
	1.0	.2963	.3712	.4684	.4864
0.8	0.0	.5429	.6188	.6191	.6870
	0.2	.5368	.6209	.6264	.7090
	0.4	.5420	.6221	.6433	.7230
	0.6	.5424	.6254	.6710	.7257
	0.8	.5462	.6247	.6984	.7247
	1.0	.5390	.6161	.7097	.7184
1.0	0.0	.7693	.8259	.8270	.8788
	0.2	.7644	.8234	.8317	.8880
	0.4	.7729	.8304	.8511	.8934
	0.6	.7736	.8295	.8689	.8929
	0.8	.7729	.8276	.8849	.8952
	1.0	.7699	.8259	.8845	.8866

Table 2: Estimated Rejection Probabilities
 $H_{02}: \beta_2=0$

β_1	β_2	S	B	H	CLSD
0.0	0.0	.0159	.0270	.0295	.0260
	0.2	.0378	.0575	.0601	.0617
	0.4	.1190	.1623	.1636	.1752
	0.6	.2975	.3653	.3661	.4103
	0.8	.5490	.6321	.6329	.6913
	1.0	.7644	.8218	.8235	.8748
0.2	0.0	.0129	.0235	.0270	.0270
	0.2	.0336	.0542	.0558	.0704
	0.4	.1213	.1633	.1640	.2081
	0.6	.2960	.3728	.3758	.4540
	0.8	.5388	.6172	.6225	.7112
	1.0	.7714	.8276	.8350	.8864
0.4	0.0	.0155	.0245	.0325	.0381
	0.2	.0346	.0526	.0549	.0827
	0.4	.1242	.1720	.1750	.2423
	0.6	.2913	.3667	.3759	.4674
	0.8	.5451	.6244	.6441	.7261
	1.0	.7668	.8273	.8499	.8958
0.6	0.0	.0138	.0244	.0366	.0447
	0.2	.0367	.0578	.0635	.0954
	0.4	.1233	.1699	.1844	.2519
	0.6	.2994	.3673	.4024	.4851
	0.8	.5496	.6298	.6782	.7365
	1.0	.7745	.8310	.8687	.8912
0.8	0.0	.0125	.0225	.0360	.0468
	0.2	.0379	.0585	.0727	.1016
	0.4	.1236	.1682	.2030	.2528
	0.6	.2927	.3688	.4373	.4827
	0.8	.5421	.6230	.6977	.7244
	1.0	.7711	.8282	.8832	.8942
1.0	0.0	.0134	.0239	.0393	.0481
	0.2	.0346	.0554	.0834	.0996
	0.4	.1202	.1645	.2226	.2458
	0.6	.3044	.3810	.4716	.4885
	0.8	.5326	.6161	.7167	.7245
	1.0	.7788	.8366	.8948	.8971

The reported results clearly indicate the superiority of the Holm procedure and the closed LSD procedure compared with the Bonferroni procedure and the Scheffe procedure. In particular the differences between the closed LSD procedure and the Bonferroni/Scheffe procedures are in general quite substantial.

5. Conclusions

This paper deals with the application of closed test procedures to linear hypotheses within the classical linear regression model. It was shown via the closure test theorem how to construct such test procedures which always control the multiple level α . A specific closure test, namely the closed LSD procedure, seems to be quite attractive compared with the Bonferroni/Scheffe procedures.

References:

- Alt, R. (1988), "Hierarchical test problems and the closure principle" in: Bauer, P./Hommel, G./Sonnemann, E. (ed.) - Multiple Hypotheses Testing-, Reihe: Medizinische Informatik und Statistik, Springer Verlag.
- Fisher, R.A. (1935), The Design of experiments. Oliver and Boyd Ltd, Edinburgh and London.
- Holm, S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure", Scandinavian Journal of Statistics 6, 65-70.
- Marcus, R./Peritz, E./Gabriel, K.R. (1976), "On closed testing procedures with special reference to ordered analysis of variance", Biometrika, 63, 3, 655-660.
- Savin, N.E. (1984) "Multiple Hypothesis Testing" in: Handbook of Econometrics, Vol.2, Griliches, Z./Intriligator, M.D.(ed.), Amsterdam, North-Holland Publishing.
- Scheffe, H. (1953) "A method for judging all contrasts in the analysis of variance", Biometrika 40, 87-104.
- Sonnemann, E. (1982) "Allgemeine Lösungen Multipler Testprobleme", EDV in Medizin und Biologie 13, 120-128.