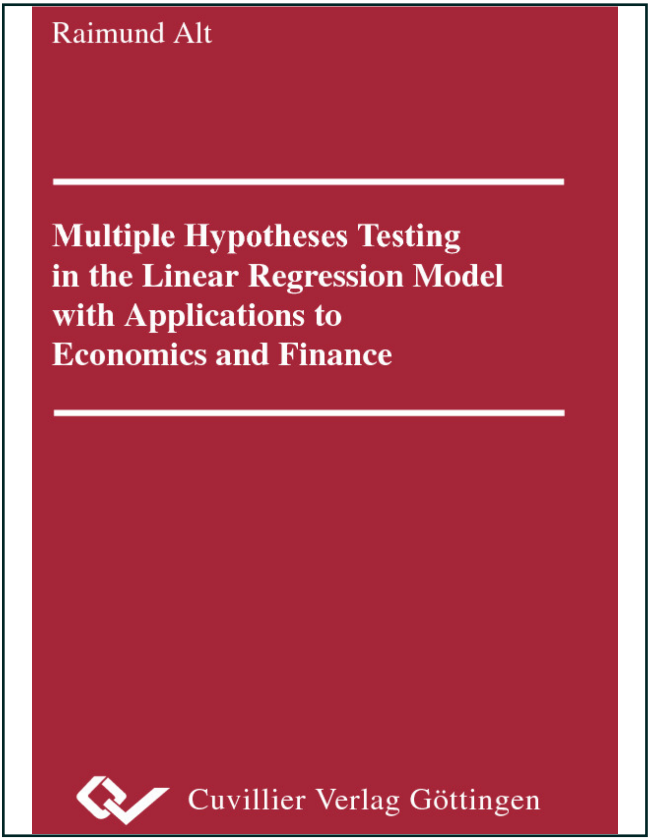
Raimund Alt (Autor)

Multiple Hypotheses Testing in the Linear Regression Model with Applications to Economics and Finance



https://cuvillier.de/de/shop/publications/2694

Chapter 1

Introduction

Multiple hypotheses testing is a branch of statistics which is concerned with specific issues arising when two or more null hypotheses are tested simultaneously. While multiple test procedures are applied in a number of fields, it seems that they are mostly used in biometrics and medical statistics. Incidentally, one of the earliest contributions to the subject is due to the eminent biometrician Fisher (1935[27]) who proposed the application of two now classical test procedures – the Bonferroni procedure and the LSD procedure – in the context of pairwise multiple comparisons of means.

One of the most important issues in multiple hypotheses testing is the adequate control of type I error probabilities. As is well known, a type I error is committed, whenever a true null hypothesis is rejected. Controlling type I error probabilities in a multiple test situation is a nontrivial problem, because the occurrence of *multiple* type I errors has to be taken into account.

To illustrate this so-called 'multiplicity effect', let us consider the case of testing two null hypotheses H_{01} and H_{02} against the alternative hypotheses H_{11} and H_{12}, respectively. We assume that for each null hypothesis there is an exact level α test, i.e. each test has the property, that the probability of falsely rejecting its respective null hypothesis is equal to some prespecified level α. Provided that all combinations of the two null hypotheses are logically possible, the type I error probabilities for this multiple test situation are as follows:

Table 1.1 Type I error probabilities when testing H_{01} and H_{02}

State of the world	Type I error probability
(1) H_{01} true, H_{02} true	$P(H_{01}$ is rejected or H_{02} is rejected$)$
(2) H_{01} true, H_{02} false	α
(3) H_{01} false, H_{02} true	α
(4) H_{01} false, H_{02} false	0

While the type I error probabilities in the states (2) and (3) are bounded by α, which is a direct consequence of the way the tests were defined, this does not hold for state (1). If both null hypotheses are true, it is not guaranteed that the probability of committing any type I error is bounded by α. On the contrary, the type I error probability for state (1) satisfies the inequality

$$\alpha \leq P(H_{01} \text{ is rejected or } H_{02} \text{ is rejected}) \leq 2\alpha.$$

In general we are not able to calculate the exact probability, unless the dependence structure of the tests is explicitly known.

The situation becomes somewhat dramatic, if we consider the testing of three null hypotheses H_{01}, H_{02} and H_{03} against the alternative hypotheses H_{11}, H_{12} and H_{13}, respectively. Again we assume that for each null hypothesis there is an exact level α test. Provided that all combinations of the three null hypotheses are logically possible, we get the following type I error probabilities:

Table 1.2 Type I error probabilities when testing H_{01}, H_{02} and H_{03}

State of the world	Type I error probability
(1) H_{01} true, H_{02} true, H_{03} true	$P(H_{01}$ is rejected or H_{02} is rejected or H_{03} is rejected)
(2) H_{01} true, H_{02} true, H_{03} false	$P(H_{01}$ is rejected or H_{02} is rejected)
(3) H_{01} true, H_{02} false, H_{03} true	$P(H_{01}$ is rejected or H_{03} is rejected
(4) H_{01} false, H_{02} true, H_{03} true	$P(H_{02}$ is rejected or H_{03} is rejected)
(5) H_{01} true, H_{02} false, H_{03} false	α
(6) H_{01} false, H_{02} true, H_{03} false	α
(7) H_{01} false, H_{02} false, H_{03} true	α
(8) H_{01} false, H_{02} false, H_{03} false	0

It turns out that in four out of eight states the probability of committing any type I error is not bounded by α. If we consider the general case of testing n null hypotheses $H_{01}, \dots, H_{0n}$, implying 2^n possible states, the type I error probability is under control only in n states. In all other states, ignoring the state where all null hypotheses are false, the type I error probability may be substantially larger than the given level α. The reason for this 'multiplicity effect' is simple. The textbook-like approach described above does not take into account the specific nature of multiple test situations. As a consequence the inflated occurrence of type I errors results in spurious significance in the employed tests.

To get an intuition of how large the type I error probabilities may be, let us consider the general case of testing n null hypotheses under the additional assumption, that the corresponding tests are independent, given that all null hypotheses are true. In this case we are able to derive a simple formula for the so-called global type I error probability:

$$
\begin{aligned}
& P(\text{at least one of the true null hypotheses is rejected}) \\
= \ & 1 - P(\text{none of the true null hypotheses is rejected}) \\
= \ & 1 - \prod_{i=1}^{n} (1-\alpha) \\
= \ & 1 - (1-\alpha)^n.
\end{aligned}
$$

Below we list some values of the formula $1-(1-\alpha)^n$ for $\alpha = 0.05$.

Table 1.3 Global type I error probability

n	$1-(1-\alpha)^n$
2	0.0975
3	0.1426
4	0.1855
5	0.2262
6	0.2649
7	0.3017
8	0.3366
9	0.3698
10	0.4013

It turns out that the global type I error probability, i.e. the probability of committing any type I error, given that all null hypotheses are true, can become quite large, even for small values of n. For example, the case $n = 2$ yields a global type I error probability of 0.0975, which comes rather close to the upper bound $2\alpha = 0.10$.

Compared with their established position in biometrics and medical statistics, multiple test procedures are only of marginal importance in fields like economics or finance. This is somewhat surprising since the relevant literature abounds of examples, where the testing of several null hypotheses is far more the rule than the exception. But, it is a matter of fact that empirical researchers working in these fields usually do not adjust for multiplicity when two or more null hypotheses are tested within a given model.

Nevertheless there are a number of references emphasizing the importance of multiple test procedures for applied research. Some of them are given in the following chapters, two of them are mentioned here.

The first reference is Epstein (1987[25]), who made an interesting remark in his book '*A History of Econometrics*', when he referred to the early work of the Cowles Commission: " *The problem of multiple hypotheses was a primary objection by mathematical statisticians in* *the 1940's to the Cowles Commission methodology. Marschak* [...] *interpreted this problem very narrowly as the proper adjustment to the size of significance tests when the same data are used repeatedly for discriminating among hypotheses*". Epstein added: "*This problem* [multiple hypotheses, R.A.] *tended to be neglected by the next generation of practitioners but urgently needs more careful attention for future modeling efforts*".

The second reference is Savin (1984[80]), who wrote a survey article entitled '*Multiple Hypothesis Testing*' for one of the volumes of the internationally renowned series '*Handbooks of Economics*'. The article presented an overview about the properties of the Bonferroni and the Scheffé procedure within the linear regression model. Unfortunately the author seemed not to be aware of the path-breaking developments that have taken place in the field of multiple hypotheses testing during the 1970's. In particular he did not mention the influential papers by Marcus, Peritz and Gabriel (1976[66]) and Holm (1979[43]).

Marcus, Peritz and Gabriel (1976[66]) introduced the concept of closed test procedures, the application of which results in the construction of multi-step or stepwise procedures. These procedures have become very popular in biometrics and medical statistics, which is due to the fact, that they are in general more powerful than traditional single-step procedures. Holm (1979[43]) suggested a sequentially rejective version of a closed test procedure, which is at least as powerful as the Bonferroni procedure (but in general more powerful), while still controlling the type I error probability for each combination of true null hypotheses. As a consequence, Holm's version, also called Bonferroni-Holm procedure, has more and more replaced the classical Bonferroni procedure.

Meanwhile there are several modern textbooks available, e.g. Hochberg and Tamhane (1987[39]) or Hsu (1996[47]), replacing Miller's (1981[69]) classic but outdated monograph. A survey article covering the latest developments in the field of multiple hypotheses testing was recently published by Pigeot (2000[76]).

Unfortunately closed test procedures are virtually unknown in the economics and financial literature, aside from a few exceptions like Neusser (1991[71]), Madlener and Alt (1996[65])