ECONOMETRIC INSTITUTE

TESTING NONNESTED LINEAR HYPOTHESES: A BAYESIAN APPROACH BASED ON INCOMPLETELY SPECIFIED PRIOR DISTRIBUTIONS

J. BOUMAN

REPORT 8534/A

Erasmus

ERASMUS UNIVERSITY ROTTERDAM - P.O. BOX 1738 - 3000 DR ROTTERDAM - THE NETHERLANDS

Testing nonnested linear hypotheses:

A Bayesian approach based on incompletely specified prior distributions.

Drs. J. Bouman

Contents.

1. Introduction

This paper is concerned with the problem of testing two linear statistical models, i.e.,

(1.1) $H_0: y = X\beta + u$ vs. $H_1: y = Z\gamma + u$,

where y is a n-dimensional vector of observations, X is a n×k matrix of constants of rank k, Z is a n×ℓ matrix of constants of rank ℓ, $\beta \in \mathbb{R}^k$ and $\gamma \in \mathbb{R}^\ell$ are unknown parameter vectors and u is an unobservable n-dimensional random vector which is assumed to be generated by the following multivariate normal distribution

(1.2) $u \sim n(0, \sigma^2 I)$,

where $\sigma > 0$ is an unknown parameter.
Under the assumption (1.2) it is seen from (1.1) that

(1.3) $y \sim n(\mu, \sigma^2 I)$,

where $\mu = E(y) = X\beta$ or $Z\gamma$.
This shows that (1.1) can be formulated as the following problem of testing linear hypotheses: Given that the observable random vector y is generated by a $n(\mu, \sigma^2 I)$ distribution we want to test

(1.4) $H_0: \mu = X\beta$ vs. $H_1: \mu = Z\gamma \neq X\beta$.

In general, the problem of finding good tests has been attacked in a variety of ways.
Two major approaches can be described as follows:

(I) Restricted classes of tests.

Since (except in rare cases) there does not exist a test which is uniformly best, narrow classes of procedures has been proposed using criteria such as
a) level of significance
b) invariance or symmetry
c) unbiasedness.

Researches have then sought procedures which improve (in terms of the power of the test) all others within the restricted class.

(II) Minimax and Bayes procedures.

After specifying a loss function for the problem, the tests are compared on the basis of their risk functions. Since there does not exist a test with uniformly minimum risk, we compare the procedures by global criteria such as

d) the maximum of the risk function

e) the weighted average of the risk function.

Then we choose the test which minimizes the maximum risk (Minimax) or the average risk (Bayes), respectively.

For our problem of testing linear hypotheses approach (I) is only succesfull in the case of nested linear models. That is, when the linear hypotheses are nested, there exists a uniformly most powerful (UMP) test in the restricted class of invariant tests with level α (conditions a) en b)). This test coincides with the classical F test for the nested problem. Moreover, if $|\ell-k| = 1$ this test turns out to be UMP unbiased (condition c)) and equivalent to the well-known t test for this special nested case.

On the other hand, when the linear hypotheses are nonnested the approach (I) does not yield a UMP test. In particular, it is shown by Bouman [2] that in the nonnested case invariance considerations do not reduce the problem sufficiently far for the existence of a UMP invariant level α test.

Since we are mainly concerned with the nonnested case and since approach (I) is not succesfull for this problem, it is reasonable to apply approach (II) to the problem of testing linear hypotheses.

This approach requires the specification of a loss function. Usually, a simple loss function is chosen by specifying $\ell_i > 0$ as the loss associated with the wrong decision of rejecting H_i when this hypothesis is true ($i = 0, 1$) and by assuming that the loss of a correct decision is zero.

With this loss structure it can be shown that every unbiased test with size $\alpha = \ell_1/(\ell_0 + \ell_1)$ is a minimax test for our problem.

In particular, this implies that the (trivial) test which rejects H_0 with probability $\alpha = \ell_1/(\ell_0 + \ell_1)$, regardless of the observations, is a minimax test. It should be noted that the latter test is a purely randomized test which can be performed through an auxiliary random experiment. In other words, the minimax approach does not lead to a satisfactory solution for our problem.

Let us next consider the Bayesian approach. As is shown by Zellner [12] and Gaver and Geisel [5] (see also Judge et.al. [7]), with the above loss structure and suitable prior distributions (weight functions) for the parameters it is possible to derive the Bayes test, i.e., the test which minimizes the expected (average) risk, for the problem of testing linear hypotheses (nested as well as nonnested). It should be emphasized that this approach requires a completely specified loss function and completely specified prior distributions. That is, in order to apply this principle it is necessary to assume not only that the loss function is known and that the parameters are random variables but also that the loss function has a very simple form and that the prior distributions are completely known.

Since these assumptions are usually not warranted in applications of the type we consider, in this study we choose a different approach in order to find a test for our problem. This approach, which can be thought of as a combination of the approaches (I) and (II), does not require a loss function nor completely specified priors (or weight functions). To be more specific, instead of expected (average) risk we concentrate on the expected (average) power function of a test and moreover we assume that the parameters possess certain incompletely specified prior distributions (i.e., distributions containing unknown parameters). Within this framework we try to find the test which maximizes the expected power under H_1, in the class of tests whose expected power under H_0 does not exceed α (a preassigned level).

However, since we work with incompletely specified prior distributions, the expected power function will depend on the unknown prior parameters and there does not exist a test which maximizes the expected power uniformly, i.e., for all values of the prior parameters. Since the problem of testing linear hypotheses remains invariant under a certain group of transformations, a possible way out is to restrict further attention to the invariant tests and then to try to solve the above problem within this restricted class.

The outline of this study is as follows.
In Section 2 we derive the Bayes test for our problem under the assumption of a completely specified loss function and completely known prior distributions. We also show that the Bayesian approach can be considered as a possible way of reducing the original problem to a more simple form.
In Section 3 we introduce certain imcompletely specified prior distributions and derive the corresponding reduced problem.
Section 4 is concerned with the problem of finding the UMP invariant test for the reduced hypotheses.
The probability distribution of the test statistic is derived in Section 5.
Although we are primarily interested in the problem of testing nonnested linear hypotheses, our approach is applicable to the nested as well as the nonnested case, and in Section 6 we reconsider the nested case.
In Section 7 we show that the test possesses the above stated optimum property. A summary of the results and a description of the required computations are given in Section 8. Finally, in the appendices some special results are proved.

2. The Bayesian approach to testing linear hypotheses

As we saw in the foregoing section the problem of testing linear hypotheses can be formulated as the problem of testing

$$(2.1) \qquad H_0: \mu = X\beta \quad \text{vs.} \quad H_1: \mu = Z\gamma \neq X\beta$$

on the basis of the observable random vector y with probability distribution

$$(2.2) \qquad y \sim n(\mu, \sigma^2 I),$$

where $\beta \in \mathbb{R}^k$, $\gamma \in \mathbb{R}^\ell$ and $\sigma > 0$ are unknown parameters and X and Z are given nonstochastic (regressor) matrices.
The matrix X is of the order $n \times k$ with rank k and Z is of the order $n \times \ell$ with rank ℓ, where $k < n$ and $\ell < n$.
Before considering the Bayesian approach we shall rewrite the above problem in a slightly different form.
In general, let $M(A)$ denote the linear (vector-)subspace (of $\mathbb{R}^n$) spanned by the columnvectors of the $n \times m$ matrix A. Using this notation, the above problem can be reformulated as follows:
On the basis of the observable random vector y with a probability distribution as given in (2.2) we want to test

$$(2.3) \qquad H_0: \mu \in M(X) \quad \text{vs.} \quad H_1: \mu \in M(Z) \setminus M(X).$$

Note that $\sigma > 0$ is an unknown nuisance parameter.
In the Bayesian approach to this problem the parameters presented by μ and σ are considered as unobservable random variables with a known probability distribution, which reflects our prior knowledge (ideas).
The distribution (2.2) is now considered as the conditional distribution of y given μ and σ. That is, if $f(y|\mu, \sigma)$ denotes the conditional probability density function (p.d.f.) of y given μ and σ, we have:

$$(2.4) \qquad f(y|\mu, \sigma) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\{-\frac{1}{2\sigma^2}(y-\mu)'(y-\mu)\}.$$

Often it is convenient to represent the prior information in such a way that the required computations can be easily performed and that the functional form of the posterior distribution (the probability distribution of the parameters given the sample y) is the same as that for the prior distribution. Such a prior is called natural conjugate and it treats prior information as if it were a previous sample of the same process.

In our case the natural conjugate family is formed by the so-called normal-inverted gamma-2 distributions. The joint prior distribution of μ and σ is defined by specifying the marginal distribution of σ and the conditional distribution of μ given σ.

For the marginal distribution of σ we choose the inverted gamma-2 distribution (see Raiffa and Schlaifer [11], p. 228), which is defined through the following p.d.f.:

$$(2.5)\qquad h(\sigma) = \frac{2}{\Gamma(\frac{v}{2})}\left(\frac{v\omega^2}{2}\right)^{v/2}\sigma^{-v-1}e^{-\frac{v\omega^2}{2\sigma^2}},\ \sigma > 0$$

$$= 0 \text{ elsewhere,}$$

where the parameters $v > 0$ and $\omega > 0$ are supposed to be given.

The first two moments of the distribution are:

$$(2.6)\qquad \begin{aligned} E(\sigma) &= \omega\sqrt{\tfrac{1}{2}v}\,\frac{\Gamma(\frac{v-1}{2})}{\Gamma(\frac{v}{2})},\ v > 1 \\ \mathrm{Var}(\sigma) &= \omega^2\frac{v}{v-2} - [E(\sigma)]^2,\ v > 2. \end{aligned}$$

When $v \to \infty$, the limiting distribution is degenerate with all probability mass concentrated in the point $\sigma = \omega$.

Next we consider the specification of the conditional distribution of μ given σ. Here we have to distinguish between H_0 and H_1. Under H_0 we know that $\mu \in M(X)$ and if we treat μ as an outcome of a random vector, it is necessary to require that the total mass of the distribution of μ given σ belongs to $M(X)$ and not to any linear subset of dimension less than k.

A necessary and sufficient condition for this being the case is that

$$E(\mu|\sigma) = X\eta$$

$$Cov(\mu|\sigma) = X\Omega X',$$

where $\eta \in \mathbb{R}^k$ and Ω is a symmetric positive-definite matrix of the order $k{\times}k$ and where it is assumed that the moments exist.
The vector η and the matrix Ω may depend on σ. In view of this a natural candidate for the conditional prior distribution of μ given σ (under H_0) is the following singular normal distribution

(2.7) $$H_0\colon \mu|\sigma \sim n(X\eta, \sigma^2 X\Omega X'),$$

where $\eta \in \mathbb{R}^k$ is a given vector and Ω is a given symmetric, positive-definite $k{\times}k$ matrix.
Under H_1, similar considerations lead to the prior

(2.8) $$H_1\colon \mu|\sigma \sim n(Z\xi, \sigma^2 Z\Lambda Z'),$$

where $\xi \in \mathbb{R}^\ell$ is a given vector and Λ is a given symmetric, positive-definite $\ell{\times}\ell$ matrix.
It should be emphasized that the singular normal distribution has no density function (with respect to Lebesque measure in $\mathbb{R}^n$). This distribution is defined in terms of its characteristic function. That is, if P is the probability measure of a $n(\theta, \Gamma)$ distribution, P is uniquely determined by specifying

$$\int_{\mathbb{R}^n} e^{it'x} dP(x) = \exp\{it'\theta - \tfrac{1}{2}t'\Gamma t\},$$

where $t \in \mathbb{R}^n$ and where i denotes the imaginary unit. Although this definition also works in the case of a nonsingular normal distribution, the latter distribution is usually defined directly in terms of the well-known density function.
The product of the distributions in (2.7) and (2.8) with the distribution as specified by (2.5) yields the joint prior distributions of (μ, σ) under H_0 and H_1, respectively.

Note that these prior distributions are completely specified, i.e., they do not contain unknown parameters.
If we think of the unknown parameters as (unknown) outcomes of random variables possessing a (known) prior distribution, the sample y can be considered as the result of a two-stage process. Under H_0 for instance, first (μ, σ) is selected according to the joint prior distribution given by the product of the distributions specified in (2.5) and (2.7). Then, given the outcome (μ, σ), the sample y is selected according to the distribution specified by the p.d.f. $f(y|\mu, \sigma)$ as given in (2.4). Combining these two stages, we can think of y as an outcome form the marginal distribution of y under H_0. (This distribution is often called the predictive distribution under H_0.)
The marginal distribution of y under H_0 can be computed from the joint distribution of y and (μ, σ) under H_0, where the latter distribution can be found by taking the product of the prior distribution of (μ, σ) under H_0 and the conditional distribution of y given (μ, σ).
Since the joint distribution of y and (μ, σ) does not have a density function, the p.d.f. of the marginal distribution of y under H_0 cannot be derived in the usual way through integration on μ and σ.
However, by making use of the fact that every probability distribution has a unique characteristic function, we shall derive the conditional p.d.f. of y given σ (under H_0). The latter p.d.f. together with (2.5) then yields the p.d.f. of the marginal distribution of y under H_0.
Let i be the imaginary unit and t be a vector in $\mathbb{R}^n$, then the characteristic function of the distribution of y given σ is defined by $E(e^{it'y}|\sigma)$.
According to the double-expectation theorem we have:

$$(2.9) \qquad E(e^{it'y}|\sigma) = E(E(e^{it'y}|\mu, \sigma)),$$

where the inner expectation is taken with respect to the conditional distribution of y given (μ, σ) and the outer expectation with respect to the conditional distribution of μ given σ under H_0.
From (2.4) it follows that

$$(2.10) \qquad E(e^{it'y}|\mu, \sigma) = \exp\{it'\mu - \tfrac{1}{2}\sigma^2 t't\}.$$

Substitution of (2.10) into (2.9) yields

$$(2.11)\quad E(e^{it'y}|\sigma) = E(\exp\{it'\mu - \tfrac{1}{2}\sigma^2 t't\})$$

$$= e^{-\frac{1}{2}\sigma^2 t't} E(e^{it'\mu}|\sigma).$$

As we saw in (2.7), under H_0, the conditional distribution of μ given σ is $n(X\eta, \sigma^2 X\Omega X')$, which implies that

$$(2.12)\quad E(e^{it'\mu}|\sigma) = \exp\{it'X\eta - \tfrac{1}{2}\sigma^2 t'X\Omega X't\}.$$

Substitution of (2.12) into (2.11) gives

$$(2.13)\quad E(e^{it'y}|\sigma) = \exp\{-\tfrac{1}{2}\sigma^2 t't\} \exp\{it'X\eta - \tfrac{1}{2}\sigma^2 t'X\Omega X't\}$$

$$= \exp\{it'\bar{\mu}_0 - \tfrac{1}{2}\sigma^2 t'V_0 t\},$$

where

$$(2.14)\quad \bar{\mu}_0 = X\eta$$

and

$$(2.15)\quad V_0 = I + X\Omega X'.$$

It is not difficult to verify that V_0 is symmetric and positive definite.

The result (2.13) shows that the distribution of y given σ (under H_0) is a nonsingular $n(\bar{\mu}_0, \sigma^2 V_0)$ distribution. Hence, the p.d.f. $k_0(y)$ of the marginal (predictive) distribution of y under H_0 can be found from

$$(2.16)\quad k_0(y) = \int_0^\infty (2\pi)^{-\frac{n}{2}} \sigma^{-n} [\det(V_0)]^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}(y - \bar{\mu}_0)'V_0^{-1}(y - \bar{\mu}_0)\}\, h(\sigma)d\sigma,$$

where $h(\sigma)$ is defined in (2.5).

Substitution of $\sigma = z^{-\frac{1}{2}}$ and integration on z yields

$$(2.17)\qquad k_0(y) = \frac{\Gamma(\frac{n+v}{2})\, v^{\frac{v}{2}}\, \omega^{-n}}{\pi^{\frac{n}{2}}\, \Gamma(\frac{v}{2})\, \sqrt{\det(V_0)}} (v + Q_0(y))^{-\frac{n+v}{2}},$$

where

$$(2.18)\qquad Q_0(y) = (y - \bar{\mu}_0)'V_0^{-1}(y - \bar{\mu}_0)/\omega^2.$$

The density in (2.17) is the p.d.f. of a multivariate t distribution, see Raiffa and Schlaifer [11], pp. 256-259. In general, when the n-dimensional random vector x has the p.d.f.

$$(2.19)\qquad f(x) = \frac{\Gamma(\frac{n+r}{2})\, r^{\frac{r}{2}}}{\pi^{\frac{n}{2}}\, \Gamma(\frac{r}{2})\, \sqrt{\det(V)}} (r + (x-\theta)'V^{-1}(x-\theta))^{-\frac{n+r}{2}},$$

we say that x has a multivariate (n-dimensional) t distribution with r degrees of freedom and parameters θ and V, where $r > 0$, $\theta \in \mathbb{R}^n$ and V is a symmetric, positive-definite $n \times n$ matrix. We use the notation $x \sim t_r(\theta, V)$. The first two moments of x are

$$(2.20)\qquad \begin{aligned} E(x) &= \theta,\ r > 1 \\ Cov(x) &= \frac{r}{r-2} V,\ r > 2. \end{aligned}$$

If $r \to \infty$, the limiting distribution of x is $n(\theta, V)$.
It follows from (2.17) and (2.18) that the marginal distribution of y under H_0 is a multivariate t distribution with v degrees of freedom and parameters $\bar{\mu}_0$ and $\omega^2 V_0$, i.e.,

$$(2.21)\qquad H_0\colon y \sim t_v(\bar{\mu}_0, \omega^2 V_0).$$

In a similar way, under H_1 we can think of the sample y as an outcome from the marginal (predictive) distribution under H_1 with the following p.d.f.

$$(2.22)\qquad k_1(y) = \frac{\Gamma(\frac{n+v}{2})\, v^{\frac{v}{2}}\, \omega^{-n}}{\pi^{\frac{n}{2}}\, \Gamma(\frac{v}{2})\, \sqrt{\det(V_1)}} (v + Q_1(y))^{-\frac{n+v}{2}},$$

where

$$(2.23)\qquad Q_1(y) = (y - \bar{\mu}_1)' V_1^{-1} (y - \bar{\mu}_1)/\omega^2.$$

Here we have

$$(2.24)\qquad \bar{\mu}_1 = Z\xi$$

and

$$(2.25)\qquad V_1 = I + Z\Lambda Z'.$$

Again the marginal distribution of y is multivariate t, that is

$$(2.26)\qquad H_1\colon y \sim t_v(\bar{\mu}_1, \omega^2 V_1).$$

Since $\bar{\mu}_0$, V_0, $\bar{\mu}_1$, V_1, v and ω are given, it is seen from (2.17) and (2.22) that the marginal distributions of y under H_0 and H_1 are completely specified.

In other words, by considering μ and σ as random variables with known prior distributions under H_0 and H_1, the problem is reduced to that of testing the following simple hypotheses:

$$(2.27)\qquad H_0'\colon y \sim t_v(\bar{\mu}_0, \omega^2 V_0) \text{ against } H_1'\colon y \sim t_v(\bar{\mu}_1, \omega^2 V_1).$$

When the p.d.f. $k_i(y)$ is interpreted as the conditional p.d.f. of y given H_i and if p_i is a prior probability assigned to H_i, i = 0, 1 ($p_0 + p_1 = 1$), we can derive the Bayes test for our problem.

Let $\ell_i > 0$ be the loss associated with the wrong decision of rejecting H_i when this hypotheses is true (i = 0, 1) and let the loss of a correct decision be zero, then the expected risk or Bayes risk $r(\phi)$ of a test (critical function) ϕ is equal to:

(2.28) $\quad r(\phi) = \ell_0 p_0 E_0(E(\phi(y)|\mu, \sigma)) + \ell_1 p_1 E_1(E(1 - \phi(y)|\mu, \sigma)),$

where E_i is the expectation taken with respect to the prior distribution of (μ, σ) under H_i, $i = 0, 1$.
From the double-expectation theorem it follows that

(2.29) $$r(\phi) = \ell_0 p_0 E(\phi(y); H_0') + \ell_1 p_1 E(1 - \phi(y); H_1')$$

$$= \ell_1 p_1 + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\ell_0 p_0 k_0(y) - \ell_1 p_1 k_1(y)] \, \phi(y) dy_1 \ldots dy_n,$$

where $y' = (y_1 \; y_2 \; \ldots \; y_n)$ and where $k_0(y)$ and $k_1(y)$ are as defined in (2.17) and (2.22), respectively.
Hence, the Bayes test, i.e., the test ϕ which minimizes $r(\phi)$ has the form

(2.30) $$\phi(y) = 1 \text{ if } \ell_0 p_0 k_0(y) - \ell_1 p_1 k_1(y) < 0$$

$$= 0 \text{ otherwise.}$$

That is, the Bayes test rejects $H_0(\phi(y) = 1)$ if

(2.31) $$\frac{k_1(y)}{k_0(y)} > \frac{p_0 \ell_0}{p_1 \ell_1}$$

and accepts $H_0(\phi(y) = 0)$ otherwise.
It easily follows from (2.17) and (2.22) that the Bayes test rejects H_0 when

(2.32) $$\frac{v + Q_1(y)}{v + Q_0(y)} < [\frac{\det(V_0)}{\det(V_1)}]^{\frac{1}{n+v}} (\frac{p_1 \ell_1}{p_0 \ell_0})^{\frac{2}{n+v}}.$$

Further it is not difficult to verify that the test (2.32) is also the Bayes test for the reduced problem (2.27), provided that the same prior probabilities and losses are assigned to the latter problem.
Moreover, it follows from the lemma of Neyman and Pearson that this test is most powerful (of its own size) for testing (2.27).
If one finds it difficult to specify p_i and ℓ_i, $i = 0, 1$, it is also possible to compute the most powerful test of size α (where α is a

preassigned significance level) for problem (2.27). Again it easily follows from the Neyman-Pearson lemma that this test has critical region

$$\frac{v + Q_1(y)}{v + Q_0(y)} < c, \tag{2.33}$$

where the critical value of c has to be computed from

$$P\left(\frac{v + Q_1(y)}{v + Q_0(y)} < c;\ H_0'\right) = \alpha.$$

Although the distribution of y under H_0' is known, the critical value c is very difficult to compute.
An advantage of the Bayesian approach is that, through complete specification of the prior distributions, the original problem is reduced to a simple problem of testing two simple hypotheses. On the other hand, however, it can be expected that this method is sensitive (at least for relatively small samples) for the specific choice of the priors or the parameters of the prior distributions.
If we consider the Bayesian approach as a possible way of reducing a problem, it seems reasonable to choose incompletely specified prior distributions (i.e., distributions which contain unknown parameters). In this manner we do not obtain a problem of testing two simple hypotheses, but it turns out that for a suitable choice of the prior distributions the reduced problem can be solved by applying invariance considerations to it. The resulting test is uniformly most powerful (UMP) among the invariant tests and can also be obtained by means of the generalized likelihood-ratio (GLR) criterion, as we shall see in the next sections.
For a slightly different derivation of the Bayes test (2.32) and a discussion of the Bayesian approach to testing linear models we refer to Gaver and Geisel [5] and Zellner [12].

3. Incompletely specified prior distributions

In this section we shall consider incompletely specified prior distributions for the parameters μ and σ, i.e., prior distributions which contain unknown parameters.
To be more specific, we again assume that μ and σ are random variables with prior distributions belonging to the class of normal-inverted gamma-2 distributions, but we no longer assume that the parameters of these prior distributions are completely known.
We recall that the original problem can be formulated as:

(3.1) Given that the observable random vector y has a $n(\mu, \sigma^2 I)$ distribution, we want to test

$$H_0: \mu \in \mathcal{M}(X) \quad \text{vs.} \quad H_1: \mu \in \mathcal{M}(Z) \setminus \mathcal{M}(X),$$

where $\sigma > 0$ is unknown.

As before, we consider the $n(\mu, \sigma^2 I)$ distribution as the conditional distribution of y given (μ, σ) and we treat σ as an (unknown) outcome of a random variable with the following p.d.f.:

$$h(\sigma) = \frac{2}{\Gamma(\frac{v}{2})} \left(\frac{v\omega^2}{2}\right)^{v/2} \sigma^{-v-1} e^{-\frac{v\omega^2}{2\sigma^2}}, \quad \sigma > 0 \tag{3.2}$$

$$= 0 \text{ elsewhere,}$$

where $v > 0$ and $\omega > 0$.

Moreover, given σ, we think of μ as an (unknown) outcome of a random vector with the following distributions under H_0 and H_1, respectively:

$$\begin{aligned} &H_0: \mu|\sigma \sim n(X\eta, \sigma^2 X\Omega X') \\ &H_1: \mu|\sigma \sim n(Z\xi, \sigma^2 Z\Lambda Z'), \end{aligned} \tag{3.3}$$

where $\eta \in \mathbb{R}^k$, $\xi \in \mathbb{R}^\ell$, Ω is a symmetric, positive-definite $k \times k$ matrix and Λ is a symmetric, positive-definite $\ell \times \ell$ matrix.

In Section 2, we saw that treating μ and σ as random variables with prior distributions as stated in (3.2) and (3.3) is equivalent to saying that y is a sample from either a $t_v(\bar{\mu}_0, \omega^2 V_0)$ or a $t_v(\bar{\mu}_1, \omega^2 V_1)$ distribution. That is, by treating μ and σ as random variables the original problem is "reduced" to the problem of testing

(3.4) $$H_0': y \sim t_v(\bar{\mu}_0, \omega^2 V_0) \quad \text{vs.} \quad H_1': y \sim t_v(\bar{\mu}_1, \omega^2 V_1),$$

where

(3.5) $$\bar{\mu}_0 = X\eta$$
$$\bar{\mu}_1 = Z\xi$$
$$V_0 = I + X\Omega X'$$
$$V_1 = I + Z\Lambda Z'.$$

When no further restrictions are placed on v, ω, η, ξ, Ω and Λ, we cannot say that problem (3.4) is a reduction of the original problem (3.1) and problem (3.4) turns out to be too complicated for a solution. On the other hand, as we saw in Section 2, by completely specifying the prior parameters v, ω, η, ξ, Ω and Λ the reduced problem (3.4) becomes a problem of testing two simple hypotheses. However, in this case we are risking the possibility that our guesses of the prior parameters are wrong and that the resulting most powerful test (see (2.33)) is bad. We are looking for a situation somewhere in between these two extreme cases. The original problem (3.1) itself can be seen as a limiting case of such a situation. This follows from the fact that for $v\to\infty$, $\Omega\to 0$ and $\Lambda\to 0$ we have $t_v(\bar{\mu}_0, \omega^2 V_0) \to n(\bar{\mu}_0, \omega^2 I)$ and $t_v(\bar{\mu}_1, \omega^2 V_1) \to n(\bar{\mu}_1, \omega^2 I)$. Hence, if $v\to\infty$, $\Omega\to 0$ and $\Lambda\to 0$, and if we consider $\eta \in \mathbb{R}^k$, $\xi \in \mathbb{R}^\ell$ and $\omega > 0$ as unknown, the reduced problem (3.4) becomes:

(3.6) $$H_0': y \sim n(X\eta, \omega^2 I) \quad \text{vs.} \quad H_1': y \sim n(Z\xi, \omega^2 I),$$

which is equivalent to problem (3.1).

In order to find suitable restrictions to be placed on the prior parameters, we first observe that the original problem (3.1) is invariant under the transformations

$$(3.7) \qquad X \to XA \quad \text{and} \quad Z \to ZB,$$

for all nonsingular k×k matrices A and ℓ×ℓ matrices B. This is easily seen from the fact that $M(XA) = M(X)$ and $M(ZB) = M(Z)$ for any nonsingular A and B.

The transformations (3.7) can be interpreted as changes of the coordinate system in which the explanatory variables are expressed. Since the original problem (3.1) is independent of the particular coordinate system chosen, it is natural to require that the reduced problem (3.4) satisfies the same property. That is, we require that the $t_v(\bar{\mu}_0, \omega^2 V_0)$ and the $t_v(\bar{\mu}_1, \omega^2 V_1)$ distribution remain unchanged after the transformations (3.7).

As is seen from (3.6) this can be achieved in a simple way by specifying η, Ω, ξ and Λ as follows:

$$(3.8) \qquad \begin{aligned} \eta &= (X'X)^{-1}X'q_0 \\ \Omega &= (X'X)^{-1} \\ \xi &= (Z'Z)^{-1}Z'q_1 \\ \Lambda &= (Z'Z)^{-1}, \end{aligned}$$

where $q_0 \in \mathbb{R}^n$ and $q_1 \in \mathbb{R}^n$ are specified independently of X and Z. With this choice of the prior parameters, $\bar{\mu}_0$, $\bar{\mu}_1$, V_0 and V_1 become

$$(3.9) \qquad \begin{aligned} \bar{\mu}_0 &= X(X'X)^{-1}X'q_0 \\ \bar{\mu}_1 &= Z(Z'Z)^{-1}Z'q_1 \\ V_0 &= I + X(X'X)^{-1}X' \\ V_1 &= I + Z(Z'Z)^{-1}Z', \end{aligned}$$

which clearly are invariant under (3.7).
In the second place we consider the choice of $\bar{\mu}_0$ and $\bar{\mu}_1$, or equivalently, q_0 and q_1. Since the original hypotheses H_0 and H_1 provide no information concerning q_0 and q_1 and since H_0' and H_1' in (3.4) take the place of H_0 and H_1 for the purpose of testing problem (3.1), knowledge about q_0 and q_1 should provide as little help for this task as possible. We might expect that the problem of testing H_0' against H_1' becomes more difficult when H_0' is "close" to H_1'. For this reason we try to specify q_0 and q_1 in such a way that H_0' is as close as possible to H_1'.
What we need is a measure of "distance" between a distribution under H_0' and a distribution under H_1'. That is, we are looking for a nonnegative number d which measures the distance between a $t_v(\bar{\mu}_0, \omega^2 V_0)$ and a $t_v(\bar{\mu}_1, \omega^2 V_1)$ distribution, where $\bar{\mu}_0$, $\bar{\mu}_1$, V_0 and V_1 are as specified in (3.9).
We first consider the general case. Let P_0 and P_1 be probability distributions in $\mathbb{R}^n$ and suppose that these distributions have density functions (with respect to Lebesque measure) $f_0(x)$ and $f_1(x)$, respectively ($x \in \mathbb{R}^n$).
Let the function, $\mu(z; \lambda)$ be defined as

$$(3.10) \qquad \mu(z;\lambda) = \frac{z^\lambda - 1}{\lambda}, \; z \geq 0, \; \lambda \neq 0,$$

where by definition

$$(3.11) \qquad \mu(z;0) = \lim_{\lambda \to 0} \mu(z;\lambda) = \ln z.$$

Then the generalized Kullback-Leibler "distance" from P_0 to P_1 is defined as:

$$(3.12) \qquad d_\lambda = E_0\left[\mu\left(\frac{f_0(X)}{f_1(X)}; \lambda\right)\right], \; \lambda \geq 0,$$

where $X \in \mathbb{R}^n$ is a stochastic vector and where the expectation E_0 is taken with respect to P_0.
It may be shown that for any $\lambda \geq 0$ the number d_λ is always defined though possible infinite. Moreover we have for any $\lambda \in [0,\infty)$:

(i) $d_\lambda \geq 0$

(ii) $d_\lambda = 0$ if and only if $P_0 = P_1$.

Note that for $\lambda = 0$ we get

$$d_0 = E_0[\ln(\frac{f_0(X)}{f_1(X)})],$$

which is the well-known Kullback-Leibler number, see Kullback [9] and Bickel and Doksum [1], p. 226.

In the case of measuring the "distance" between two n-dimensional t distributions with the same number of degrees of freedom v, a convenient choice of λ turns out to be $\lambda = \frac{2}{n+v}$. For this choice of λ the required computations are easily performed and moreover for any n and any $v > 0$ the number d_λ takes on a finite value.

As is shown in Appendix D, when $P_0 = t_v(\bar{\mu}_0, \omega^2 V_0)$, $P_1 = t_v(\bar{\mu}_1, \omega^2 V_1)$, where $\bar{\mu}_0$, $\bar{\mu}_1$, V_0 and V_1 are as given in (3.9), and f_0 and f_1 are the corresponding p.d.f.'s, the "distance" $d = d_{\frac{2}{n+v}}$ from P_0 to P_1 is equal to:

$$(3.13) \qquad d = \tfrac{1}{2}[\det(V_0 V_1^{-1})]^{-\frac{1}{n+v}}[v + tr(V_0 V_1^{-1}) + \frac{1}{\omega^2}(\bar{\mu}_1 - \bar{\mu}_0)' V_1^{-1} (\bar{\mu}_1 - \bar{\mu}_0)] - \frac{n+v}{2}.$$

When we consider d as a function of q_0 and q_1 (through $\bar{\mu}_0$ and $\bar{\mu}_1$, see (3.9)) it is easily seen that d has the minimal value

$$(3.14) \qquad \bar{d} = \tfrac{1}{2}[\det(V_0 V_1^{-1})]^{-\frac{1}{n+v}}[v + tr(V_0 V_1^{-1})] - \frac{n+v}{2}$$

for all points q_0 and q_1 which satisfy $\bar{\mu}_0 = \bar{\mu}_1$.

Hence if we restrict attention to the points q_0 and q_1 with the property that $\bar{\mu}_0 = \bar{\mu}_1$, the hypothesis H_0' appears to be "closest" to H_1'.

Since we know from (3.9) that $\bar{\mu}_0 = X(X'X)^{-1}X'q_0 \in M(X)$ and $\bar{\mu}_1 = Z(Z'Z)^{-1}Z'q_1 \in M(Z)$, the restriction $\bar{\mu}_0 = \bar{\mu}_1$ implies that

$$(3.15) \qquad \bar{\mu}_0 = \bar{\mu}_1 = \bar{\mu} \in M(X) \cap M(Z).$$

The linear subspace $M(X) \cap M(Z)$ plays an important rôle in the testing problem. Note that in the original problem (3.1) the points (μ, σ) with $\mu \in M(X) \cap M(Z)$ and $\sigma > 0$, which belong to H_0, are limit points of H_1. That is, these points form the boundary between the hypotheses H_0 and H_1.
If we suppose that $M(X) \cap M(Z)$ has dimension p, it is not difficult to verify that (see Bouman [2])

(3.16) $\quad 0 \leq p \leq r \leq \min(k, \ell),$

where $r = \text{rank}(X'Z)$.
In general, hypotheses H_0 and H_1 are said to be nested if and only if every point in H_0 is a limit point of H_1 or vice versa. Otherwise the hypotheses are called nonnested. We speak of strictly separate hypotheses if H_0 and H_1 have no common limit points. This means that the hypotheses in problem (3.1) are nested if and only if $M(X) \subset M(Z)$ or $M(Z) \subset M(X)$, which is equivalent to $p = \dim(M(X) \cap M(Z)) = \min(k,\ell)$. Further it follows that our problem is nonnested if and only if $p < \min(k,\ell)$.
Since the point $\mu = 0$ always belongs to $M(X) \cap M(Z)$ the hypotheses in problem (3.1) are never strictly separate (the points $(0, \sigma)$ with $\sigma > 0$ are always common limit points of H_0 and H_1).

The above considerations show that if $\bar{\mu}_0 = \bar{\mu}_1$ is the only restriction placed on the prior means, the common prior mean $\bar{\mu}$ belongs to the boundary between $M(X)$ and $M(Z)$ but is otherwise unknown.
In other words, if $p > 0$ we have

(3.17) $\quad \bar{\mu} = C\delta,$

where $\delta \in \mathbb{R}^p$ is supposed to be unknown and where C is a n×p matrix, the columns of which form an arbitrary basis for the p-dimensional linear subspace $M(X) \cap M(Z)$. When $p = 0$ we have $M(X) \cap M(Z) = \{0\}$, which implies that

(3.18) $\quad \bar{\mu} = 0.$

If no further restrictions are placed on the prior parameters $v > 0$ and $\omega > 0$, the above specification of η, ξ, Ω and Λ yields the following prior distributions of μ and σ:

The prior distribution of σ has p.d.f.

$$(3.19)\qquad h(\sigma) = \frac{2}{\Gamma(\frac{v}{2})}\left(\frac{v\omega^2}{2}\right)^{v/2}\sigma^{-v-1}e^{-\frac{v\omega^2}{2\sigma^2}},\ \sigma > 0$$

$$= 0 \text{ elsewhere,}$$

where $v > 0$ and ω are unknown parameters. When $p > 0$, the conditional distributions of μ given σ are

$$(3.20)\qquad \begin{aligned} &H_0\colon \mu|\sigma \sim n(C\delta,\ \sigma^2X(X'X)^{-1}X') \\ &H_1\colon \mu|\sigma \sim n(C\delta,\ \sigma^2Z(Z'Z)^{-1}Z'), \end{aligned}$$

where $\delta \in \mathbb{R}^p$ is unknown.
In the case $p = 0$ we have

$$(3.21)\qquad \begin{aligned} &H_0\colon \mu|\sigma \sim n(0,\ \sigma^2X(X'X)^{-1}X') \\ &H_1\colon \mu|\sigma \sim n(0,\ \sigma^2Z(Z'Z)^{-1}Z') \end{aligned}$$

Note that the prior distribution of μ and σ are incompletely specified.
As we saw above, by treating μ and σ as random variables with the stated prior distributions the original testing problem is reduced to

$$(3.22)\qquad H_0'\colon y \sim t_v(C\delta,\ \omega^2V_0) \quad \text{vs.} \quad H_1'\colon y \sim t_v(C\delta,\ \omega^2V_1)$$

for the case $p > 0$, where $v > 0$, $\omega > 0$ and $\delta \in \mathbb{R}^p$ are unknown parameters.
When $p = 0$ we get

$$(3.23)\qquad H_0'\colon y \sim t_v(0,\ \omega^2V_0) \quad \text{vs.} \quad H_1'\colon y \sim t_v(0,\ \omega^2V_1)$$

with $v > 0$ and $\omega > 0$ unknown.
The matrices V_0 and V_1 are given by

$$(3.24)\qquad \begin{aligned} V_0 &= I + X(X'X)^{-1}X' \\ V_1 &= I + Z(Z'Z)^{-1}Z'. \end{aligned}$$

The columns of the n×p matrix C form a basis for the linear subspace $\mathcal{M}(X) \cap \mathcal{M}(Z)$. It will be clear that C can always be constructed from the given matrices X and Z.

The problems (3.22) and (3.23) are invariant under the transformations (3.7). This follows from the fact that $\mathcal{M}(XA) \cap \mathcal{M}(ZB) = \mathcal{M}(X) \cap \mathcal{M}(Z)$. When $p = 0$ we obtain $\mathcal{M}(XA) \cap \mathcal{M}(ZB) = \{0\}$. In the case $p > 0$ the construction of a basis for $\mathcal{M}(XA) \cap \mathcal{M}(ZB)$ from the matrices XA and ZB yields the n×p matrix C_* which satisfies $C_* = CD$ for some nonsingular p×p matrix D. Hence after a transformation of the type (3.7) the vector $C\delta$ is represented by $C_*\delta_*$ with $\delta_* = D^{-1}\delta \in \mathbb{R}^p$. Together with the fact that V_0 and V_1 in (3.24) remain unchanged under (3.7), this shows the stated invariance property.

The hypotheses H_0' and H_1' in (3.22) and (3.23) are composite and, as is typically the case in such a situation, no UMP test exists for the reduced problem. However, if we restrict attention to the invariant tests a UMP test can be found within this restricted class as will be shown in the next sections.

Now it remains to show how the matrix C can be found from the given matrices X and Z (in the case $p > 0$). As is shown in Bouman [2] the dimension of $\mathcal{M}(X) \cap \mathcal{M}(Z)$ is equal to p if and only if the matrix $(X'X)^{-1}X'Z(Z'Z)^{-1}Z'X$ (or equivalently, $(Z'Z)^{-1}Z'X(X'X)^{-1}X'Z$) has an eigenvalue 1 with multiplicity p. Suppose that the value of p is given, then in most applications the matrices X and Z have precisely p columnvectors in common.

These common vectors can be taken as the matrix C, which shows that in most cases it is very easy to find the matrix C.

It occasionally happens that X and Z have less than p columns in common. In this case we can always find a n×p submatrix X_{1*} of X and a n×(k−p) submatrix X_{2*} of X such that:

$$\text{rank}([X_{1*} \vdots X_{2*}]) = k \text{ and } \text{rank}([X_{2*} \vdots Z]) = k+\ell-p.$$

Then there exist matrices G_1 and G_2 with $G_2 \neq 0$ such that

$$(3.25) \qquad X_{1*} = X_{2*}G_1 + ZG_2.$$

Hence if we can find G_1 we can take

$$(3.26) \qquad C = X_{1*} - X_{2*}G_1.$$

The matrix G_1 is obtained from

(3.27) $$G = (F'F)^{-1}F'X_{1*},$$

where $G = \begin{bmatrix} G_1 \\ \cdots \\ G_2 \end{bmatrix}$ and $F = [X_{2*} \vdots Z]$.

Obviously, the above procedure also works in the case where X and Z have precisely p columns in common.
Finally we shall consider the number $\bar{d}$ from (3.14) which measures the "closeness" of H_0' to H_1' in (3.22) and (3.23). When the original hypotheses H_0 and H_1 are nested we might expect a small value of $\bar{d}$, i.e., we might expect that H_0' is close to H_1'. On the other hand, when H_0 and H_1 are nonnested we might expect a large value of $\bar{d}$.
We shall now show that this is precisely what happens. From (3.14) and (3.24) it is seen that $\bar{d}$ depends on X and Z through V_0 and V_1 and in order to investigate this dependence we first compute $\det(V_0)$, $\det(V_1)$ and V_1^{-1}.
It can easily be verified that

(3.28) $$\det(V_0) = 2^k$$
$$\det(V_1) = 2^\ell$$
$$V_1^{-1} = I - \tfrac{1}{2}Z(Z'Z)^{-1}Z'.$$

Substitution of (3.28) into (3.14) yields

(3.29) $$\bar{d} = 2^{\frac{\ell-k}{n+v}}\left[\frac{n+v}{2} - \frac{1}{4}\ell + \frac{1}{2}k - \frac{1}{4}t\right] - \frac{n+v}{2},$$

where t is defined as

(3.30) $$t = \mathrm{tr}[X(X'X)^{-1}X'Z(Z'Z)^{-1}Z'].$$

The number t can be considered as a measure of the degree of "nonnestedness" of the original hypotheses H_0 and H_1.
For given values of k and ℓ we always have

(3.31) $$0 \leq t \leq \min(k, \ell).$$

Further it can be shown that H_0 and H_1 are nested, i.e., $\mathcal{M}(X) \subset \mathcal{M}(Z)$ or $\mathcal{M}(Z) \subset \mathcal{M}(X)$, if and only if $t = \min(k, \ell)$. Also, H_0 and H_1 are nonnested if and only if $t < \min(k, \ell)$. It follows from (3.29) that $\bar{d}$ attains a minimum value (for given k and ℓ) when H_0 and H_1 are nested. The more H_0 and H_1 are nonnested, i.e., the smaller the value of t, the larger the value of $\bar{d}$. For given k and ℓ it is seen that $\bar{d}$ attains a maximum value for $t = 0$, which corresponds to the most extreme nonnested case $\mathcal{M}(X) \perp \mathcal{M}(Z)$, i.e., $X'Z = 0$. The trivial nested case $\mathcal{M}(X) = \mathcal{M}(Z)$ which occurs if and only if $t=k=\ell$ yields $\bar{d} = 0$, that is, in this case we have $H_0' = H_1'$. Note that $\mathcal{M}(X) = \mathcal{M}(Z)$ if and only if $V_0 = V_1$ and that the converse of the above statement is also true, i.e., $\bar{d} = 0$ implies that $\mathcal{M}(X) = \mathcal{M}(Z)$.

4. A UMP invariant test for the reduced problem

This section is concerned with the problem of finding a solution for the reduced testing problem as defined in the foregoing section.
That is, if $p = \dim[\mathcal{M}(X) \cap \mathcal{M}(Z)] > 0$, we consider the problem of testing

$$(4.1) \qquad H_0': y \sim t_v(C\delta, \omega^2 V_0) \quad \text{vs.} \quad H_1': y \sim t_v(C\delta, \omega^2 V_1)$$

and if $p = 0$ we consider

$$(4.2) \qquad H_0': y \sim t_v(0, \omega^2 V_0) \quad \text{vs.} \quad H_1': y \sim t_v(0, \omega^2 V_1),$$

where C, V_0 and V_1 are given matrices. C is of the order $n \times p$ with rank p, $V_0 = I + X(X'X)^{-1}X'$ and $V_1 = I + Z(Z'Z)^{-1}Z'$. On the other hand, $v > 0$, $\omega > 0$ and $\delta \in \mathbb{R}^p$ are unknown parameters.
As is typically the case in a situation of testing two composite hypotheses, no UMP test exists for the problems (4.1) and (4.2).
Then the usual procedure is to restrict attention to a certain subclass of tests and solve the problem of finding the UMP test within this restricted class. When a problem remains invariant under a certain group of transformations of the sample space onto itself, it is natural to restrict attention to tests which exhibit the same property. That is, we only consider tests (functions of y) that are invariant with respect to these transformations.
The transformations can be interpreted as changes of the coordinate system in which the data (y) are expressed. When a problem is independent of the particular coordinate system chosen, it is reasonable to restrict attention to tests which satisfy the same property, since otherwise the acceptance or rejection of the hypothesis under consideration would depend on the choice of the coordinates, which is quite arbitrary and has no bearing on the problem.
Within the class of invariant tests we try to find a UMP test, which (if it exists) is called the UMP invariant test.
A discussion of this type of reduction on symmetry grounds or invariance can be found in Lehmann [10], Chapter 6.
In order to show that the problems (4.1) and (4.2) remain invariant under a certain group of transformations we first reformulate these problems in a different way.

Let F be the family of even n-dimensional p.d.f.'s, that is, $f \in F$ satisfies

$$f(x) \geq 0, \quad x \in \mathbb{R}^n$$

$$(4.3) \qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x)dx_1 \ldots dx_n = 1$$

$$f(-x) = f(x), \quad x \in \mathbb{R}^n.$$

Let $f_{iv} \in F$ be defined by

$$(4.4) \qquad f_{iv}(x) = \frac{\Gamma(\frac{n+v}{2})v^{v/2}}{\pi^{n/2}\Gamma(\frac{v}{2})\sqrt{\det(V_i)}}(v + x'V_i^{-1}x)^{-\frac{n+v}{2}},$$

where $v > 0$ and $i = 0, 1$.

It follows that the p.d.f. of a $t_v(C\delta, \omega^2 V_i)$ distribution can be written as

$$(4.5) \qquad \omega^{-n} f_{iv}(\frac{x-C\delta}{\omega}), \quad i = 0, 1.$$

Hence problem (4.1) can be reformulated as:

Given that the sample y has the p.d.f. $\omega^{-n} f(\frac{y-C\delta}{\omega})$, $f \in F$, we want to test H_0': $f = f_{0v}$ vs. H_1': $f = f_{1v}$.

In a similar way problem (4.2) can be written as:

Given that the sample y has the p.d.f. $\omega^{-n} f(\frac{y}{\omega})$, $f \in F$, we want to test H_0': $f = f_{0v}$ vs. H_1': $f = f_{1v}$.

It should be emphasized that the parameters $v > 0$, $\omega > 0$ and $\delta \in \mathbb{R}^p$ are unknown.

Next we will show that problem (4.1) remains invariant under the group of transformation given by

$$(4.6) \qquad G_1: g(y) = ay + C\alpha,$$

for all $a \in \mathbb{R}^1$, $a \neq 0$ and all $\alpha \in \mathbb{R}^p$.

Let $g \in G_1$ and consider $z = g(y) = ay + C\alpha$. The inverse transformation is $y = g^{-1}(z) = (z-C\alpha)/a$ and if y has the p.d.f. $\omega^{-n} f((y-C\delta)/\omega)$ it is easily seen that the p.d.f. of z becomes

$$(4.7)\qquad |a|^{-n}\omega^{-n}f\left(\left(\frac{z-C\alpha}{a}-C\delta\right)/\omega\right) = (|a|\omega)^{-n}f\left(\frac{z-C(a\delta+\alpha)}{|a|\omega}\right) =$$

$$\omega_*^{-n}f\left(\frac{z-C\delta_*}{\omega_*}\right),$$

where $\omega_* = |a|\,\omega > 0$ and $\delta_* = a\delta + \alpha \in \mathbb{R}^p$.
This shows that after the transformation $z = g(y)$ the problem becomes:
Given that the sample z has the p.d.f. $\omega_*^{-n}f\left(\frac{z-C\delta_*}{\omega_*}\right)$, $f \in F$, we want to test H_0': $f = f_{0v}$ vs. H_1': $f = f_{1v}$, where $v > 0$, $\omega_* > 0$ and $\delta_* \in \mathbb{R}^p$ are unknown.
Since the problem in terms of z is exactly the same as that in terms of y, it follows that the transformations (4.6) leave problem (4.1) invariant.
In a similar way it is seen that problem (4.2) remains invariant under the group of transformations

$$(4.8)\qquad G_2\colon\ g(y) = ay,\ a \in \mathbb{R}^1,\ a \neq 0.$$

Note that G_2 is a subgroup of G_1.*
In the case of problem (4.1), through invariance considerations, we restrict attention to tests (critical functions) $\phi(y)$ which are invariant with respect to the group (4.6), i.e.,

$$(4.9)\qquad \phi(ay+C\alpha) = \phi(y),$$

for all $a \in \mathbb{R}^1$, $a \neq 0$ and all $\alpha \in \mathbb{R}^p$.
For problem (4.2) we only consider tests which satisfy $\phi(ay) = \phi(y)$, for all $a \in \mathbb{R}^1$, $a \neq 0$.
We shall first derive a UMP invariant test for problem (4.1), that is, a test which is UMP among the invariant tests (4.9).
Suppose that v is fixed at an arbitrary level, then a UMP invariant test for problem (4.1) exists, as is shown in Appendix A. This UMP invariant test turns out to be independent of v and is therefore also UMP invariant for the more general problem with v unknown.
In order to derive the UMP invariant test we need some preliminary results.

* It is not difficult to see that the original testing problem as defined in (3.1) of Section 3 also remains invariant under G_1 or G_2.

Let the columns of the n×p matrix R_1 be an orthonormal basis for the p-dimensional linear subspace $\mathcal{M}(C)$ and let the columns of the n × (n-p) matrix R_2 be an orthonormal basis for the orthogonal complement $\mathcal{M}(C)^{\perp}$ of $\mathcal{M}(C)$.
For $x \in \mathbb{R}^{n-p}$ we define the function $\ell_i(x)$, i = 0, 1, as

$$(4.10) \qquad \ell_i(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{iv}(R_1 z + R_2 x)dz_1 \cdots dz_p,$$

where $z' = (z_1\ z_2\ \cdots\ z_p)$ and where f_{iv} is as given in (4.4).
Further let $w_2 = R_2'y$ and $w_2' = (W_{21}\ W_{22}\ \cdots\ W_{2(n-p)})$.
As is shown in Appendix A, see (A.17), the UMP invariant test for problem (4.1) with fixed v rejects ($\phi(y) = 1$) if

$$(4.11) \qquad \frac{\int_0^{\infty} \eta^{n-p-1} \ell_1(\frac{\eta}{W_{2(n-p)}} w_2)d\eta}{\int_0^{\infty} \eta^{n-p-1} \ell_0(\frac{\eta}{W_{2(n-p)}} w_2)d\eta} > c.$$

By making use of the property that the marginal distributions of a multivariate t distribution are again multivariate t, the substitution of (4.4) into (4.10) yields:

$$(4.12) \qquad \ell_i(x) = \frac{\Gamma(\frac{n-p+v}{2})v^{v/2}}{\pi^{\frac{n-p}{2}}\Gamma(\frac{v}{2})\sqrt{\det(\bar{V}_i)}} (v+x'\bar{V}_i^{-1}x)^{-\frac{n-p+v}{2}},$$

where $\bar{V}_i = R_2'V_iR_2$ (i = 0, 1).

If we define

$$(4.13) \qquad A_i = w_2'\bar{V}_i^{-1}w_2/W_{2(n-p)}^2, \quad i = 0, 1$$

and make the transformation

$$z = \frac{1}{1 + \frac{A_i}{v}\eta^2},$$

the integrals in (4.11) become

(4.14) $$\int_0^\infty \eta^{n-p-1} \ell_i\left(\frac{\eta}{W_{2(n-p)}} w_2\right) d\eta =$$

$$\frac{1}{2} \frac{\Gamma(\frac{n-p+v}{2})\, A_i^{-\frac{n-p}{2}}}{\pi^{\frac{n-p}{2}}\, \Gamma(\frac{v}{2}) \sqrt{\det(\bar{V}_i)}} \int_0^1 z^{\frac{v}{2}-1} (1-z)^{\frac{n-p}{2}-1} dz =$$

$$\frac{1}{2} \frac{\Gamma(\frac{n-p}{2})}{\pi^{\frac{n-p}{2}} \sqrt{\det(\bar{V}_i)}} A_i^{-\frac{n-p}{2}}, \quad i = 0, 1.$$

Subsitution of (4.14) into (4.11) yields the following rejection region of the UMP invariant test

(4.15) $$\frac{w_2' \bar{V}_1^{-1} w_2}{w_2' \bar{V}_0^{-1} w_2} < c_*.$$

Note that this rejection region does not depend on v. Therefore, in order to prove that the UMP invariant test (4.15) does not depend on v, it is sufficient to show that the probability distribution under H_0' of the test statistic

(4.16) $$S = \frac{w_2' \bar{V}_1^{-1} w_2}{w_2' \bar{V}_0^{-1} w_2}$$

does not depend on v.

As a matter of fact we will show that the distribution of the above test statistic does not depend on v, ω and δ under H_0' as well as H_1'.
Let the random vector T be defined by $T' = (T_1\ T_2\ \cdots\ T_{n-p-1})$, where

$$T_i = \frac{W_{2i}}{W_{2(n-p)}}, \quad i = 1, 2, \ldots, n-p-1,$$

then S can be written as a function of T,

(4.17) $$S = \frac{T_*' \bar{V}_1^{-1} T_*}{T_*' \bar{V}_0^{-1} T_*},$$

where $T_*' = (T_1\ T_2\ \dots\ T_{n-p-1}\ 1)$.
With the aid of the result (A.15) of Appendix A and (4.14), it is not difficult to see that the p.d.f. $k_i(t_1, t_2, \dots, t_{n-p-1})$ of T under H_i', i = 0, 1, is equal to

$$(4.18)\qquad k_i(t_1, t_2, \dots, t_{n-p-1}) = \frac{\Gamma(\frac{n-p}{2})}{\pi^{\frac{n-p}{2}}\sqrt{\det(\bar{V}_i)}}\,(t_*'\bar{V}_i^{-1}t_*)^{-\frac{n-p}{2}},$$

i = 0, 1, where $t_*' = (t_1\ t_2\ \dots\ t_{n-p-1}\ 1)$.
Since the distribution of T (see (4.18)) does not depend on v, ω and δ under H_0' and H_1', it follows that S and therefore also the UMP invariant test (4.15) does not depend on v, ω and δ under H_0' and H_1'. This shows that the test (4.15) is UMP invariant for problem (4.1) with v > 0 unknown.
In a similar way it can be shown that the test with critical region

$$(4.19)\qquad S = \frac{y'V_1^{-1}y}{y'V_0^{-1}y} < c_*$$

is UMP invariant for problem (4.2) (the case p = 0).
The critical value c_* in (4.15) and (4.19) has to be chosen in such a way that the size of the test becomes equal to a preassigned significance level α.
Throughout the above derivation it was assumed that n-p > 1 (for p > 0 as well as p = 0). When $n-p \leq 1$ the UMP invariant test has critical function $\phi(y) \equiv \alpha$.
This is a purely randomized test which rejects with probability α regardless of the observations.

In applications it is not necessary to compute R_2 and $w_2 = R_2'y$ in order to find the test statistic S. With the aid of the relations

(i) $\qquad w_2 = R_2'y$

(ii) $\qquad R_1R_1' + R_2R_2' = I$

(iii) $\qquad \bar{V}_i^{-1} = (R_2'V_iR_2)^{-1} = R_2'V_i^{-1}R_2 - R_2'V_i^{-1}R_1(R_1'V_i^{-1}R_1)^{-1}R_1'V_i^{-1}R_2,\ i = 0, 1$

(which can be derived from $(R'V_iR)^{-1} = R'V_i^{-1}R$, where $R = [R_1 \vdots R_2]$)

(iv) $[V_i^{-1} - V_i^{-1}R_1(R_1'V_i^{-1}R_1)^{-1}R_1'V_i^{-1}]\, R_1 = 0,\ i = 0,\ 1,$

we get the following expression for $w_2'\overline{V}_i^{-1}w_2$

$$w_2'\overline{V}_i^{-1}w_2 = y'[V_i^{-1} - V_i^{-1}R_1(R_1'V_i^{-1}R_1)^{-1}R_1'V_i^{-1}]y,\ i = 0,\ 1. \tag{4.20}$$

Since $R_1 = CA$ for some nonsingular p×p matrix A, we have

$$R_1(R_1'V_i^{-1}R_1)^{-1}R_1' = C(C'V_i^{-1}C)^{-1}C',\ i = 0,\ 1, \tag{4.21}$$

substitution of which into (4.20) yields

$$w_2'\overline{V}_i^{-1}w_2 = y'[V_i^{-1} - V_i^{-1}C(C'V_i^{-1}C)^{-1}C'V_i^{-1}]y,\ i = 0,\ 1. \tag{4.22}$$

The right-hand side of (4.22) can be expressed in terms of the maximum likelihood estimator of the parameter vector δ under H_i', i.e., under the assumption that $y \sim t_v(C\delta,\ \omega^2V_i)$, $i = 0,\ 1$. In order to see this, let $\hat{\delta}_i$ be the maximum likelihood estimator of δ under H_i', then we have

$$\hat{\delta}_i = (C'V_i^{-1}C)^{-1}C'V_i^{-1}y,\ i = 0,\ 1. \tag{4.23}$$

If we define the residual vector $\hat{u}_i$ as

$$\hat{u}_i = y - C\hat{\delta}_i,\ i = 0,\ 1 \tag{4.24}$$

it follows from (4.22) that

$$w_2'\overline{V}_i^{-1}w_2 = \hat{u}_i'V_i^{-1}\hat{u}_i,\ i = 0,\ 1. \tag{4.25}$$

Hence the UMP invariant test for problem (4.1) rejects when

$$S = \frac{\hat{u}_1'V_1^{-1}\hat{u}_1}{\hat{u}_0'V_0^{-1}\hat{u}_0} < c_*. \tag{4.26}$$

The above results also show that the test (4.26) can be obtained by applying the GLR criterion to problem (4.1). Similarly, the test (4.19) turns out to be the GLR test for problem (4.2).

Until now we did not use the fact that the matrices C, V_0 and V_1 have a special structure.
Upon using

$$\begin{aligned} \mathcal{M}(C) &= \mathcal{M}(X) \cap \mathcal{M}(Z) \\ V_0 &= I + X(X'X)^{-1}X' \qquad (4.27) \\ V_1 &= I + Z(Z'Z)^{-1}Z', \end{aligned}$$

the tests (4.19) and (4.26) can be considerably simplified.
From (4.27) it is easily seen that

$$\begin{aligned} V_0^{-1} &= I - \tfrac{1}{2}X(X'X)^{-1}X' \\ V_1^{-1} &= I - \tfrac{1}{2}Z(Z'Z)^{-1}Z', \end{aligned} \qquad (4.28)$$

If M_X and M_Z are defined by

$$\begin{aligned} M_X &= I - X(X'X)^{-1}X' \\ M_Z &= I - Z(Z'Z)^{-1}Z', \end{aligned} \qquad (4.29)$$

it follows from (4.28) that

$$\begin{aligned} V_0^{-1} &= \tfrac{1}{2}(I + M_X) \\ V_1^{-1} &= \tfrac{1}{2}(I + M_Z). \end{aligned} \qquad (4.30)$$

Further it is seen from $\mathcal{M}(C) \subset \mathcal{M}(X)$ and $\mathcal{M}(C) \subset \mathcal{M}(Z)$ that $M_XC = M_ZC = 0$, which together with (4.30) implies that

$$V_i^{-1}C = \tfrac{1}{2}C, \; i = 0, 1. \qquad (4.31)$$

The maximum likelihood estimators $\hat{\delta}_i$ become

$$\hat{\delta}_i = (C'C)^{-1}C'y, \; i = 0, 1, \qquad (4.32)$$

see (4.23), which shows that $\hat{\delta}_i$ is equal to the ordinary least-squares estimator of δ after the regression from y on C.

Therefore, if the corresponding least-squares residual vector is denoted by $\hat{u}_C$, i.e.,

$$(4.33) \qquad \hat{u}_C = y - C\hat{\delta} = M_C y$$

where

$$(4.34) \qquad \begin{aligned} \hat{\delta} &= (C'C)^{-1}C'y \\ M_C &= I - C(C'C)^{-1}C', \end{aligned}$$

it follows from (4.24) that

$$(4.35) \qquad \hat{u}_i = \hat{u}_C = M_C y, \; i = 0, 1.$$

By making use of (4.30) and (4.35) it is seen that the test statistic S from (4.26) can be written as

$$(4.36) \qquad S = \frac{y'M_C(I + M_Z)M_C y}{y'M_C(I + M_X)M_C y}.$$

Let $\hat{u}_X$ and $\hat{u}_Z$ be the least-squares residual vectors after regression from y on X and Z, respectively, i.e.,

$$(4.37) \qquad \begin{aligned} \hat{u}_X &= y - X\hat{\beta} = M_X y \\ \hat{u}_Z &= y - Z\hat{\gamma} = M_Z y, \end{aligned}$$

where $\hat{\beta} = (X'X)^{-1}X'y$ and $\hat{\gamma} = (Z'Z)^{-1}Z'y$. Then it follows from (4.36) that S takes the form

$$(4.38) \qquad S = \frac{\hat{u}_C'\hat{u}_C + \hat{u}_Z'\hat{u}_Z}{\hat{u}_C'\hat{u}_C + \hat{u}_X'\hat{u}_X},$$

where use has been made of $M_C^2 = M_C$, $M_C M_X M_C = M_X$, $M_C M_Z M_C = M_Z$, $y'M_C y = \hat{u}_C'\hat{u}_C$, $y'M_X y = \hat{u}_X'\hat{u}_X$ and $y'M_Z y = \hat{u}_Z'\hat{u}_Z$.

That is, the UMP invariant test for problem (4.1) has the rejection region

$$(4.39) \qquad S = \frac{\hat{u}_C'\hat{u}_C + \hat{u}_Z'\hat{u}_Z}{\hat{u}_C'\hat{u}_C + \hat{u}_X'\hat{u}_X} < c_*.$$

In the same way it can be shown that the UMP invariant test (4.19) for problem (4.2) rejects when

$$(4.40) \qquad S = \frac{y'y + \hat{u}_Z'\hat{u}_Z}{y'y + \hat{u}_X'\hat{u}_X} < c_*.$$

It is not difficult to verify from (4.39) and (4.40) that with probability 1 we have $\frac{1}{2} < S < 2$.

The above results show that the test statistic S can be very easily computed from y, C, X and Z. Now it remains to find the critical value c_*. This problem shall be discussed in the next section.

5. The distribution of the test statistic

In this section we shall derive the probability distribution of the test statistic under the hypothesis H_0'. That is, we shall derive the distribution function of S as given in (4.38) of Section 4 under the assumption that $y \sim t_v(C\delta, \omega^2 V_0)$ (the case $p > 0$) and the distribution function of S as given in (4.40) of Section 4 under the assumption that $y \sim t_v(0, \omega^2 V_0)$ (the case $p = 0$). We shall consider the former case in detail, the case $p = 0$ can be treated in a similar way.
In order to derive the probability distribution of S under H_0' it is convenient to use the equivalent form of S as given in (4.16), i.e.,

$$(5.1) \qquad S = \frac{w_2' \bar{V}_1^{-1} w_2}{w_2' \bar{V}_0^{-1} w_2},$$

where $\bar{V}_i = R_2' V_i R_2$, $i = 0, 1$ and $w_2 = R_2' y$, the columns of the $n \times (n-p)$ matrix R_2 being an orthonormal basis for the linear subspace $M(C)^{\perp}$. Since under H_0' we have $y \sim t_v(C\delta, \omega^2 V_0)$ it easily follows that

$$(5.2) \qquad w_2 \sim t_v(0, \omega^2 \bar{V}_0).$$

Hence if $u = \frac{w_2}{\omega}$, we have

$$(5.3) \qquad S = \frac{u' \bar{V}_1^{-1} u}{u' \bar{V}_0^{-1} u}, \quad u \sim t_v(0, \bar{V}_0).$$

Now we shall show that S has the same distribution as the ratio $x' \bar{V}_1^{-1} x / x' \bar{V}_0^{-1} x$, where the random vector x has a $n(0, \bar{V}_0)$ distribution. Let $g_0(u)$ be the p.d.f. of $u \sim t_v(0, \bar{V}_0)$, then it is not difficult to verify that

$$(5.4) \qquad g_0(u) = \int_0^\infty k_0(u|z)\, h(z) dz,$$

where $k_0(u|z)$ is the conditional p.d.f. of u given $Z = z$ and $h(z)$ is the marginal p.d.f. of the random variable Z, $k_0(u|z)$ and $h(z)$ being equal to

$$k_0(u|z) = (2\pi)^{-\frac{n-p}{2}} z^{-(n-p)} [\det(\bar{V}_0)]^{-\frac{1}{2}} \exp\{-\frac{1}{2z^2} u'\bar{V}_0^{-1}u\}$$

(5.5)

$$h(z) = \frac{2}{\Gamma(\frac{v}{2})} (\frac{v}{2})^{v/2} z^{-v-1} e^{-\frac{v}{2z^2}}, \quad z > 0$$

$$= 0 \text{ elsewhere.}$$

That is, the conditional distribution of u given Z = z is a $n(0, z^2\bar{V}_0)$ distribution and the marginal distribution of Z is a inverted gamma-2 distribution with parameters v and 1.

Let the (n-p)-dimensional random vector x be defined by

$$x = \frac{u}{Z}, \tag{5.6}$$

then it follows that the conditional distribution of x given Z = z is a $n(0, \bar{V}_0)$ distribution. Since the latter distribution does not depend on z, it is seen that x and Z are stochastically independent and that the unconditional (marginal) distribution of x is a $n(0, \bar{V}_0)$ distribution. Substitution of (5.6), i.e., $u = Zx$, into (5.3) yields

$$S = \frac{(Zx)'\bar{V}_1^{-1}(Zx)}{(Zx)'\bar{V}_0^{-1}(Zx)} = \frac{Z^2x'\bar{V}_1^{-1}x}{Z^2x'\bar{V}_0^{-1}x} = \frac{x'\bar{V}_1^{-1}x}{x'\bar{V}_0^{-1}x}, \tag{5.7}$$

where $x \sim n(0, \bar{V}_0)$ and this proves the stated property. This result also shows that the distribution of S under H_0' does not depend on the unknown parameters v, ω and δ, a fact which was already observed in Section 4. The fact that the test statistic can be written as

$$S = \frac{x'\bar{V}_1^{-1}x}{x'\bar{V}_0^{-1}x} \text{ with } x \sim n(0, \bar{V}_0),$$

enables us to find the distribution of S under H_0', i.e.,

$$F_0(s) = P(S \leq s;\ H_0'). \tag{5.8}$$

Let $\tau_1, \tau_2, \ldots, \tau_M$ be the M different eigenvalues of the matrix $\overline{V}_1^{-1}\overline{V}_0$ and let $m_1, m_2, \ldots, m_M$ be the corresponding multiplicities. It is not difficult to see that τ_j is real and $\tau_j > 0$ for $j = 1, 2, \ldots, M$.
We also have $\sum_{j=1}^{M} m_j = n-p$. As is shown in Appendix B (see (B.15)) the distribution function $F_0(s)$ can be written as (provided that $M > 1$):

$$(5.9) \qquad F_0(s) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin \varepsilon(u;s)}{u\,\gamma(u;s)}du,$$

where

$$(5.10) \qquad \gamma(u;s) = \prod_{j=1}^{M} (1 + (\tau_j - s)^2 u^2)^{\frac{m_j}{4}}$$

and

$$(5.11) \qquad \varepsilon(u;s) = \sum_{j=1}^{M} \frac{m_j}{2} \operatorname{arctg}((\tau_j - s)u).$$

The value of the integrand in (5.9) at $u = 0$ is defined by

$$(5.12) \qquad \left.\frac{\sin \varepsilon(u;s)}{u\,\gamma(u;s)}\right|_{u=0} = \lim_{u\downarrow 0} \frac{\sin \varepsilon(u;s)}{u\,\gamma(u;s)} = \tfrac{1}{2}\sum_{j=1}^{M} \tau_j m_j - \tfrac{1}{2}(n-p)s.$$

Next consider the case $p = 0$, that is, we want to derive the distribution function of the test statistic S as given in (4.40) under the hypothesis H_0': $y \sim t_v(0, \omega^2 V_0)$.
If we express S in the equivalent form (4.19), i.e.,

$$(5.13) \qquad S = \frac{y'V_1^{-1}y}{y'V_0^{-1}y},$$

it can be shown in a similar way that $F_0(s) = P(S \leq s; H_0')$ takes the form (5.9), where now $\tau_1, \tau_2, \ldots, \tau_M$ are the M different eigenvalues of the matrix $V_1^{-1}V_0$ with corresponding multiplicities $m_1, m_2, \ldots, m_M$. In this case we have $\sum_{j=1}^{M} m_j = n$ and (5.12) becomes:

(5.14) $$\left.\frac{\sin \varepsilon(u;s)}{u\,\gamma(u;s)}\right|_{u=0} = \tfrac{1}{2}\sum_{j=1}^{M} \tau_j m_j - \tfrac{1}{2}ns.$$

The above results show that $F_0(s)$ can be computed for any s through numerical integration of (5.9). We refer to Appendix B for more details on this numerical integration.

Here we are more interested in finding the critical value c of the UMP invariant test with critical region

(5.15) $$S < c.$$

That is, we want to find c such that in the case $p > 0$

(5.16) $$\sup_{(v,\omega,\delta)} P(S < c;\, H_0') = \alpha,$$

where α is a preassigned significance level.

When $p = 0$, we have to compute c such that

(5.17) $$\sup_{(v,\omega)} P(S < c;\, H_0') = \alpha.$$

However, as we have seen above the probability distribution of S does not depend on the unknown parameters and it follows that the critical value c can be found by solving the equation $F_0(c) = \alpha$ or

(5.18) $$\frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin \varepsilon(u;c)}{u\,\gamma(u;c)}du = \alpha.$$

The solution of (5.18) not only requires numerical integration but also an iteration procedure. This would not be an objection if the value of c could be determined once and for all, that is, if the distribution of S under H_0' does not depend on the particular testing problem (the matrices X and Z), but only on n, k and ℓ. In the latter situation we can tabulate the critical values of c and use these tables for any problem of testing linear hypotheses. However, as we shall see below, only in a few special cases tabulation is possible and in general the distribution of S under H_0' depends on the particular matrices X and Z under consideration through the eigenvalues $\tau_1, \tau_2, \ldots, \tau_M$ and multiplicities $m_1, m_2, \ldots, m_M$.

In view of this it seems more attractive to report the p-value of the test (also called the observed size), instead of computing the critical value c.

Since the p-value of our test is defined by $F_0(S)$, where S is the test statistic, we have

$$(5.19)\qquad F_0(S) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin \varepsilon(u;S)}{u\,\gamma(u;S)}du,$$

which shows that the computation of the p-value only requires numerical integration and no iteration procedure. The procedures of first computing c (for given α) and then rejecting H_0' if $S < c$ is equivalent to computing the p-value $F_0(S)$ and rejecting H_0' if $F_0(S) < \alpha$.

In other words, we can think of $F_0(S)$ as a standardized test statistic and the critical region becomes $F_0(S) < \alpha$, since under H_0' the random variable $F_0(S)$ has an uniform distribution on the interval (0, 1).

The use of $F_0(S)$ also means that we are very flexible in our choice of α.

If we are interested in the power of the UMP invariant test, we first have to compute the critical value c, since the power is defined as $P(S < c; H_1')$. The probability distribution of S under H_1' again does not depend on the unknown parameters (v, ω and δ in the case $p > 0$ and v, ω in the case $p = 0$) and, as indicated in Appendix B, in a way similar to the case H_0' we can derive the distribution function of S under H_1'. This results into the following expression for the power of the test

$$(5.20)\qquad P(S < c; H_1') = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin \varepsilon_*(u;\frac{1}{c})}{u\,\gamma_*(u;\frac{1}{c})}du,$$

where

$$(5.21)\qquad \gamma_*(u;s) = \prod_{j=1}^{M} \left(1 + \left(\frac{1}{\tau_j} - s\right)^2 u^2\right)^{\frac{m_j}{4}}$$

and

$$(5.22)\qquad \varepsilon_*(u;s) = \sum_{j=1}^{M} \frac{m_j}{2} \operatorname{arctg}\left(\left(\frac{1}{\tau_j} - s\right)u\right).$$

The value of the integrand in (5.20) at u = 0 now becomes

$$\left.\frac{\sin \varepsilon_*(u;s)}{u\,\gamma_*(u;s)}\right|_{u=0} = \tfrac{1}{2} \sum_{j=1}^{M} \frac{m_j}{\tau_j} - \tfrac{1}{2}(n-p)s$$

for $p \geq 0$.

When we want to compute the p-value $F_0(S)$, the critical value c or the power $P(S < c;\ H_1')$, we always need the eigenvalues $\tau_1, \tau_2, \ldots, \tau_M$ and corresponding multiplicities $m_1, m_2, \ldots, m_M$ of the matrix $\bar{V}_1^{-1}\bar{V}_0$ in the case $p > 0$ or of the matrix $V_1^{-1}V_0$ in the case $p = 0$.

In order to compute the eigenvalues of $\bar{V}_1^{-1}\bar{V}_0$ in the case $p > 0$ it is not necessary to compute the matrix $\bar{V}_1^{-1}\bar{V}_0 = (R_2'V_1R_2)^{-1}R_2'V_0R_2$, where the columns of R_2 are an orthonormal basis for $M(C)^{\perp}$.

We shall show that the eigenvalues of $\bar{V}_1^{-1}\bar{V}_0$ are equal to the nonzero eigenvalues of the n×n matrix

$$(5.23) \qquad P = [V_1^{-1} - V_1^{-1}C(C'V_1^{-1}C)^{-1}C'V_1^{-1}]V_0.$$

The matrix $\bar{V}_1^{-1}\bar{V}_0$ can be written as

$$\bar{V}_1^{-1}\bar{V}_0 = (R_2'V_1R_2)^{-1}R_2'V_0R_2 = R_2'AR_2R_2'V_0R_2,$$

where $A = V_1^{-1} - V_1^{-1}R_1(R_1'V_1^{-1}R_1)^{-1}R_1'V_1^{-1}$ and where the columns of R_1 form an orthonormal basis for $M(C)$ (see formula (iii) of Section 4).

Since $R_2R_2' = I - R_1R_1'$ and $AR_1 = 0$ it follows that $AR_2R_2' = A$ and this yields

$$(5.24) \qquad \bar{V}_1^{-1}\bar{V}_0 = R_2'AV_0R_2.$$

Let λ be an eigenvalue of $\bar{V}_1^{-1}\bar{V}_0$ with corresponding eigenvector x, then $\bar{V}_1^{-1}\bar{V}_0x = \lambda x$, $\lambda > 0$ and therefore

$$(5.25) \qquad R_2'AV_0R_2x = \lambda x.$$

This gives $R_2R_2'AV_0R_2x = \lambda R_2x$ and since $R_2R_2'A = A$ it is seen that

(5.26) $\quad AV_0y = \lambda y,$

where $y = R_2x$.
From (5.26) it follows that $\lambda > 0$ is an eigenvalue of AV_0 with eigenvector y.
Now we always have $R_1 = CB$ for some nonsingular $p\times p$ matrix B and this shows that

$$A = V_1^{-1} - V_1^{-1}C(C'V_1^{-1}C)^{-1}C'V_1^{-1}.$$

Hence $AV_0 = P$ and we have

(5.27) $\quad Py = \lambda y.$

That is, if λ is an eigenvalue of $\overline{V}_1^{-1}\overline{V}_0$ it follows that λ is an eigenvalue of P.
Conversely, if λ is an nonzero eigenvalue of P it can be shown in a similar way that λ is an eigenvalue of $\overline{V}_1^{-1}\overline{V}_0$ and this proves the above statement.
If we define the matrix P for the case $p = 0$ as

(5.28) $\quad P = V_1^{-1}V_0,$

it follows form the above discussion that the probability distribution of the test statistic S under H_0' and H_1' (for $p > 0$ as well as $p = 0$) is given by (5.9) and (5.20), respectively, where $\tau_1, \tau_2, \ldots, \tau_M$ are the nonzero eigenvalues of the matrix P and $m_1, m_2, \ldots, m_M$ the corresponding multiplicities.
In order to apply the UMP invariant test to a certain problem we therefore need the nonzero eigenvalues of the $n\times n$ matrix P. These eigenvalues can be numerically computed for any given C, V_0 and V_1, but this can be a rather time-consuming process. At this point it should be emphasized that in the above discussion nowhere did we use the fact that for our problem the matrices C, V_0 and V_1 have a special structure, i.e.,

$$\mathcal{M}(C) = \mathcal{M}(X) \cap \mathcal{M}(Z)$$

$$V_0 = I + X(X'X)^{-1}X' \tag{5.29}$$

$$V_1 = I + Z(Z'Z)^{-1}Z'.$$

When we no longer consider the problem

$$H_0': y \sim t_v(C\delta,\ \omega^2 V_0) \quad \text{vs.} \quad H_1': y \sim t_v(C\delta,\ \omega^2 V_1) \tag{5.30}$$

as a reduction of the original problem of testing two linear hypotheses, that is, if we consider the problem (5.30) for any given matrices C, V_0, V_1 (where C is of the order n×p with rank p and V_0 and V_1 are symmetric and positive definite), it was shown in Section 4 that the UMP invariant test (with respect to the transformation $y \to ay + C\alpha$) for this general problem has rejection region

$$S = \frac{\hat{u}_1' V_1^{-1} \hat{u}_1}{\hat{u}_0' V_0^{-1} \hat{u}_0} < c, \tag{5.31}$$

where $\hat{u}_i = y - C\hat{\delta}_i$ and $\hat{\delta}_i = (C'V_i^{-1}C)^{-1}C'V_i^{-1}y$, i = 0, 1.
Moreover, as we saw above, the distribution function of S under H_0' is given by (5.9) and this function depends on C, V_0 and V_1 through the nonzero eigenvalues of the matrix P as defined in (5.23), for any given C, V_0 and V_1.
In the above discussion it was assumed that $p > 0$, of course similar remarks can be made for the case $p = 0$.
As is shown in Section 4, by making use of the fact that in our problem the matrices C, V_0 and V_1 have the special form as given in (5.29), the computation of the general test statistic in (5.31) can be considerably simplified (see formula (4.38)). Now we shall see that the same holds true for the computation of the eigenvalues of the matrix P.
Substitution of (5.29) into (5.23) yields for the case $p > 0$:

$$P = I + X(X'X)^{-1}X' - \tfrac{1}{2}Z(Z'Z)^{-1}Z' - \tfrac{1}{2}Z(Z'Z)^{-1}Z'X(X'X)^{-1}X' - C(C'C)^{-1}C'. \tag{5.32}$$

If the columns of the n×k matrix $\bar{X}$ form an orthonormal basis for $M(X)$, the columns of the n×ℓ matrix $\bar{Z}$ form an orthonormal basis for $M(Z)$ and the columns of the n×p matrix $\bar{C}$ form an orthonormal basis for $M(C)$, the expression (5.32) can be written as

$$(5.33) \qquad P = I + \bar{X}\bar{X}' - \tfrac{1}{2}\bar{Z}\bar{Z}' - \tfrac{1}{2}\bar{Z}\bar{Z}'\bar{X}\bar{X}' - \bar{C}\bar{C}'.$$

When p = 0 we get in a similar way

$$(5.34) \qquad P = I + \bar{X}\bar{X}' - \tfrac{1}{2}\bar{Z}\bar{Z}' - \tfrac{1}{2}\bar{Z}\bar{Z}'\bar{X}\bar{X}'.$$

Now it is shown in Appendix C that the eigenvalues of the n×n matrix P as given in (5.33) or (5.34) can be deduced from the eigenvalues of the k×k matrix $\bar{X}'\bar{Z}\bar{Z}'\bar{X}$ (or equivalently, the eigenvalues of the ℓ×ℓ matrix $\bar{Z}'\bar{X}\bar{X}'\bar{Z}$).

The eigenvalues of $\bar{X}'\bar{Z}\bar{Z}'\bar{X}$ always lie between 0 and 1.

If $p = \dim(M(X) \cap M(Z))$ and $r = \text{rank}(X'Z)$, it is not difficult to verify that $\bar{X}'\bar{Z}\bar{Z}'\bar{X}$ has an eigenvalue 0 with multiplicity k-r and an eigenvalue 1 with multiplicity p and vice versa. In other words, the values of p and r can always be concluded from the eigenvalues of $\bar{X}'\bar{Z}\bar{Z}'\bar{X}$. See Bouman [2], pp. 26-29 and 94-96.

Suppose that $\bar{X}'\bar{Z}\bar{Z}'\bar{X}$ had R different eigenvalues, say, $\rho_1, \rho_2, \ldots, \rho_R$ with $0 < \rho_j < 1$, $j = 1, 2, \ldots, R$ and let $r_1, r_2, \ldots, r_R$ be the corresponding multiplicities.

Note that $r = p + \sum_{j=1}^{R} r_j$ and that the matrix $\bar{Z}'\bar{X}\bar{X}'\bar{Z}$ has precisely the same eigenvalues and multiplicities as $\bar{X}'\bar{Z}\bar{Z}'\bar{X}$, except for the eigenvalue 0, which has a multiplicity ℓ-r.

As is shown in Appendix C, the matrix P has the following eigenvalues, where

$$(5.35) \qquad \alpha_j = \frac{1}{4}(5-\rho_j) - \frac{1}{4}\sqrt{(1-\rho_j)(9-\rho_j)}, \; j = 1, 2, \ldots, R,$$

and where we assume that $0 \leq p < r < \min(k, \ell)$.

Table 1. Eigenvalues > 0 of the matrix P.

Eigenvalue τ_i	Multiplicity m_i
$\tau_1 = \frac{1}{2}$	$m_1 = \ell - r$
$\tau_2 = \alpha_1$	$m_2 = r_1$
$\tau_3 = \alpha_2$	$m_3 = r_2$
$\vdots$	$\vdots$
$\tau_{R+1} = \alpha_R$	$m_{R+1} = r_R$
$\tau_{R+2} = 1$	$m_{R+2} = n+p-k-\ell$
$\tau_{R+3} = \alpha_R^{-1}$	$m_{R+3} = r_R$
$\tau_{R+4} = \alpha_{R-1}^{-1}$	$m_{R+4} = r_{R-1}$
$\vdots$	$\vdots$
$\tau_{2R+2} = \alpha_1^{-1}$	$m_{2R+2} = r_1$
$\tau_{2R+3} = 2$	$m_{2R+3} = k-r$
$M = 2R+3$	$\sum_{i=1}^{n} m_i = n-p$

It should be noted that M is the number of different eigenvalues > 0 of P, $\frac{1}{2} < \alpha_j < 1$, $j = 1, 2, \ldots, R$ and that P has an eigenvalue 0 with multiplicity p.

In Appendix B it was shown that the test statistic S can be written as

$$S = \frac{\sum_{i=1}^{M} \tau_i \eta_i}{\sum_{i=1}^{M} \eta_i},$$

where $\eta_1, \eta_2, \ldots, \eta_M$ are random variables with $P(\eta_i > 0) = 1$.
This shows that $\min_i(\tau_i) < S < \max_i(\tau_i)$ and it is seen from Table 1 that we always have $\frac{1}{2} < S < 2$, a fact which was already concluded at the end of Section 4.

The above results show that the eigenvalues of the n×n matrix P can be found from the eigenvalues of the k×k matrix $\bar{X}'\bar{Z}\bar{Z}'\bar{X}$. In order to find these latter eigenvalues it is not necessary to compute the matrices $\bar{X}$ and $\bar{Z}$. This follows from the fact that the matrices $\bar{X}'\bar{Z}\bar{Z}'\bar{X}$ and $(X'X)^{-1}X'Z(Z'Z)^{-1}Z'X$ have the same nonzero eigenvalues (see Bouman [2], p. 19).

Hence, the eigenvalues of P can be found from the eigenvalues of the k×k matrix $(X'X)^{-1}X'Z(Z'Z)^{-1}Z'X$ or, equivalently, from the eigenvalues of the $\ell\times\ell$ matrix $(Z'Z)^{-1}Z'X(X'X)^{-1}X'Z$.

The number M, being the number of different eigenvalues > 0 of P, does not always take on the value 2R+3. As a matter of fact for given R the maximum value of M is equal to 2R+3 and this value is only attained in the case $0 \leq p < r < \min(k, \ell)$.

For instance, when $r = \ell < k$ the eigenvalues $\tau_1 = \frac{1}{2}$ vanishes and $M = 2R+2$. Similarly, if $r = k < \ell$ the eigenvalue $\tau_{2R+3} = 2$ vanishes and again $M = 2R+2$. Moreover, if $r = k = \ell$ both $\tau_1 = \frac{1}{2}$ and $\tau_{2R+3} = 2$ vanish and we have $M = 2R+1$.

Another case occurs if $0 \leq p = r < \min(k, \ell)$.

Since $p = r$ if and only if $R = 0$, if follows that $\tau_2 = \alpha_1, \ldots, \tau_{R+1} = \alpha_R$ and $\tau_{R+3} = \alpha_R^{-1}, \ldots, \tau_{2R+2} = \alpha_1^{-1}$ vanish and we have $M = 3$.

Let us next consider the case of nested linear hypotheses. As we saw in Section 3 the problem of testing linear hypotheses can be divided into two categories:

(i) The case of nested linear hypotheses, i.e., $\mathcal{M}(X) \subset \mathcal{M}(Z)$ or $\mathcal{M}(Z) \subset \mathcal{M}(X)$.

(ii) The case of nonnested linear hypotheses.

The case (i) occurs if and only if $p = \min(k, \ell)$ and can be subdivided into:

(a) The trivial case $\mathcal{M}(X) = \mathcal{M}(Z)$, i.e., $p = k = \ell$.

(b) The nontrivial case $\mathcal{M}(X) \subset \mathcal{M}(Z)$ and $\mathcal{M}(X) \neq \mathcal{M}(Z)$, i.e., $p = k < \ell$.

(c) The nontrivial case $\mathcal{M}(Z) \subset \mathcal{M}(X)$ and $\mathcal{M}(X) \neq \mathcal{M}(Z)$, i.e., $p = \ell < k$.

For these three cases we get the following eigenvalues > 0 and multiplicities of P (after renumbering the τ_i and m_i):

(a) In this case $p = r = k = \ell$ and therefore $\ell-r = 0$, $R = 0$, $k-r = 0$ and $M = 1$. We have $\tau_1 = 1$ and $m_1 = n-k = n-\ell = n-p$.

(b) Now $p = r = k < \ell$ and it is seen that $k-r = 0$, $R = 0$ and $M = 2$. We get $\tau_1 = \frac{1}{2}$, $\tau_2 = 1$, $m_1 = \ell-k$ and $m_2 = n-\ell$.

(c) In this case $p = r = \ell < k$ and therefore $\ell-r = 0$, $R = 0$ and $M = 2$. Now we have $\tau_1 = 1$, $\tau_2 = 2$, $m_1 = n-k$ and $m_2 = k-\ell$.

The trivial case (a) yields a reduced problem with $H_0' = H_1'$. It follows that $S = 1$ and that the UMP invariant test with size α rejects with probability α regardless of the observations. More interesting are the nontrivial nested cases (b) and (c). It follows that in these cases the distribution of S only depends on X and Z through k and ℓ.
Therefore, in these cases it is possible to tabulate the critical value c for different values of α and $\ell-k$, $n-\ell$ or $k-\ell$, $n-k$, respectively. However, as will be shown in the next section, tabulation is not necessary since in the cases (b) and (c) we can use the tables of the F distribution in order to find the value of c.
On the other hand, if we are testing nonnested linear hypotheses, tabulation of c is no longer possible since the distribution of S depends on the particular X and Z matrices.
Finally, we shall consider a large sample approximation to the distribution of the test statistic S. From the results of this section and Appendix B it follows that under H_0' we have

$$S = \frac{\sum_{i=1}^{n-p} \lambda_i \xi_i}{\sum_{i=1}^{n-p} \xi_i}, \tag{5.36}$$

where $\frac{1}{2} \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{n-p} \leq 2$ are the nonzero eigenvalues of the matrix P and $\xi_1, \xi_2, \ldots, \xi_{n-p}$ are mutually independent random variables with $\xi_i \sim \chi^2(1)$, $i = 1, 2, \ldots, n-p$.
We know that there are M different λ_i's, i.e., $\frac{1}{2} \leq \tau_1 < \tau_2 < \ldots < \tau_M \leq 2$ with multiplicities $m_1, m_2, \ldots, m_M$, as shown in Table 1. Note that $\sum_{j=1}^{M} m_j = n-p$ and $P(\tau_1 \leq S \leq \tau_M) = 1$. We only consider the case $M > 1$, since in the trivial case $M = 1$ we always have $P(S = \tau_1) = 1$.

If $F_0(s) = P(S \leq s; H_0')$, it is seen from (5.36) that

$$F_0(s) = P(Q_s \leq 0), \tag{5.37}$$

where

$$Q_s = \sum_{i=1}^{n-p} (\lambda_i - s)\xi_i. \tag{5.38}$$

Hence, for any s the distribution function $F_0(s)$ can be found from the distribution function of the random variable Q_s.

As is shown in Appendix E, if $s \neq 1$, we have

$$\frac{Q_s - (n-p)(a_n - s)}{\sqrt{2(n-p)[(a_n - s)^2 + b_n]}} \to n(0, 1) \tag{5.39}$$

in distribution when $n \to \infty$.

Here a_n and b_n are defined by

$$\begin{aligned} a_n &= \frac{1}{n-p} \sum_{i=1}^{n-p} \lambda_i = \frac{1}{n-p} \sum_{j=1}^{M} m_j \tau_j \\ b_n &= \frac{1}{n-p} \sum_{i=1}^{n-p} \lambda_i^2 - a_n^2 = \frac{1}{n-p} \sum_{j=1}^{M} m_j \tau_j^2 - a_n^2. \end{aligned} \tag{5.40}$$

From (5.37) and (5.39) it follows that for $s \neq 1$ and large n, we approximately have

$$F_0(s) \approx N\left(\frac{\sqrt{n-p}\,(s - a_n)}{\sqrt{2[(s - a_n)^2 + b_n]}}\right), \tag{5.41}$$

where $N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$.

Note that a_n and b_n can easily be calculated from Table 1.

The result (5.41) shows that for large n the p-value $F_0(S)$ as given in (5.19) can be approximated by

(5.42) $$F_0(S) \approx N\left(\frac{\sqrt{n-p}\,(S - a_n)}{\sqrt{2[(S - a_n)^2 + b_n]}}\right).$$

The large sample approximation (5.41) also enables us to approximate the level α critical value c of the test, i.e., the value of c which satisfies $F_0(c) = \alpha$.

When $0 < \alpha \leq \frac{1}{2}$, it is shown in Appendix E that, for large n, the critical value c can be approximated by

(5.43) $$c \approx c_A = a_n - \sqrt{q_n(\alpha)b_n}\ ,$$

where

(5.44) $$q_n(\alpha) = \frac{2t_\alpha^2}{n-p - 2t_\alpha^2}\ ,$$

t_α being the (100α)th percentile of the n(0, 1) distribution.

Since we know in advance that $c > \tau_1$, the approximation c_A in (5.43) makes no sense if $c_A \leq \tau_1$.

In Appendix E it is shown that a necessary and sufficient condition for $c_A > \tau_1$ is given by

(5.45) $$\alpha > N(d_1),$$

where

(5.46) $$d_1 = \frac{\sqrt{n-p}\,(\tau_1 - a_n)}{\sqrt{2[(\tau_1 - a_n)^2 + b_n]}}.$$

Note that $d_1 < 0$.

Except for the case $\tau_1 = 1$, it can be shown that $d_1 \to -\infty$ if $n\to\infty$, and therefore, $N(d_1) \to 0$ if $n\to\infty$. Hence, except for the case $\tau_1 = 1$, the condition (5.45) is no restriction on the level α for large n.

The case $\tau_1 = 1$ occurs if and only if $(Z) \subset (X)$ and $(Z) \neq (X)$, i.e., the nested case (c) with $p = \ell < k$.

In the latter case we have $M = 2$, $\tau_1 = 1$, $\tau_2 = 2$, $m_1 = n-k$ and $m_2 = k-\ell$.

This gives:

$$a_n = 1 + \frac{k-\ell}{n-\ell}$$

$$b_n = \frac{k-\ell}{n-\ell} - \left(\frac{k-\ell}{n-\ell}\right)^2,$$

which yields $d_1 = -\sqrt{(k-\ell)/2} \not\to -\infty$ if $n\to\infty$.
Therefore, in the nested case (c), the condition $\alpha > N(d_1) = N(-\sqrt{(k-\ell)/2})$ may be a restriction on the choice of α, even when n is large. However, in this nested case there is no need to approximate c, since we can find the exact value of c with the aid of the F distribution.
In order to compute the values of a_n and b_n it is often more easy to use the formulae

$$a_n = \frac{1}{n-p}\,\mathrm{tr}(P)$$

$$b_n = \frac{1}{n-p}\,\mathrm{tr}(P^2) - a_n^2,$$

which yield

(5.47)
$$a_n = \frac{1}{n-p}\,(n+k - \tfrac{1}{2}\ell - p - \tfrac{1}{2}\mathrm{tr}(AB))$$
$$b_n = \frac{1}{n-p}\,(n+3k - \tfrac{3}{4}\ell - p - \tfrac{5}{2}\mathrm{tr}(AB) + \tfrac{1}{4}\,\mathrm{tr}[(AB)^2]) - a_n^2,$$

where $A = (X'X)^{-1}X'Z$ and $B = (Z'Z)^{-1}Z'X$.
Once the value of c is computed or approximated, it can be shown in a similar way (see Appendix E) that, for $c \neq 1$ and large n, the power of the test $P(S < c;\, H_1')$ as given in (5.20) can be approximated by

(5.48) $$P(S < c;\, H_1') \approx N\left(\frac{\sqrt{n-p}\,(a_n^* - \frac{1}{c})}{\sqrt{2[(a_n^* - \frac{1}{c})^2 + b_n^*]}}\right),$$

where

$$a_n^* = \frac{1}{n-p}\sum_{j=1}^{M}\frac{m_j}{\tau_j}$$

$$b_n^* = \frac{1}{n-p}\sum_{j=1}^{M}\frac{m_j}{\tau_j^2} - (a_n^*)^2.$$

6. A reconsideration of the nested case

In this section we reconsider the nontrivial nested cases:
(i) $M(X) \subset M(Z)$ and $M(X) \neq M(Z)$
(ii) $M(Z) \subset M(X)$ and $M(X) \neq M(Z)$.
As is well known, in the case of nested linear hypotheses we can use the classical F test.
It will be shown in this section that in case (i) the general (Bayesian-) approach, which we have developed in the foregoing sections, exactly reproduces the above mentioned F test. However, in case (ii) the situation is different and it turns out that the two approaches lead to different tests.
Moreover, it will be shown that a number of special (nested) cases can easily be incorporated into our general approach.
As said before, when the linear hypotheses are nested we can apply the classical F test to the (original) problem of testing

$$H_0\colon \mu \in M(X) \text{ against } H_1\colon \mu \in M(Z)\setminus M(X).$$

Since in case (ii) we get $M(Z)\setminus M(X) = \emptyset$, the hypotheses for this case are reformulated as

$$H_0\colon \mu \in M(X)\setminus M(Z) \text{ against } H_1\colon \mu \in M(Z).$$

The F test can be obtained by applying the GLR principle to the original problem. Moreover, it is well known (see Lehmann [10], Chapter 7, pp. 265-272) that the F test is UMP invariant with respect to a certain group of transformations G. This group differs from the group G_1 considered in (4.6) of Section 4. As a matter of fact G contains G_1 as a subgroup. It should be emphasized that in case of testing nonnested linear hypotheses no F test is obtained by applying the GLR principle or invariance considerations to the original problem.
Before applying our approach to the nested cases (i) and (ii) we shall briefly discuss the F tests.
In case (i) the F test has the following rejection region

$$(6.1) \qquad F = \frac{\hat{u}_X'\hat{u}_X - \hat{u}_Z'\hat{u}_Z}{\hat{u}_Z'\hat{u}_Z} \cdot \frac{n-\ell}{\ell-k} > c^*,$$

where $\hat{u}_X = y - X\hat{\beta}$, $\hat{u}_Z = y - Z\hat{\gamma}$ with $\hat{\beta} = (X'X)^{-1}X'y$ and $\hat{\gamma} = (Z'Z)^{-1}Z'y$.
Writing

$$(6.2) \qquad F = \frac{y'(M_X - M_Z)y}{y'M_Z y} \cdot \frac{n-\ell}{\ell-k},$$

where $M_X = I - X(X'X)^{-1}X'$ and $M_Z = I - Z(Z'Z)^{-1}Z'$, it is easily verified that under H_0: $y \sim n(X\beta, \sigma^2 I)$ we have:

$$(6.3) \qquad F \sim F(\ell-k, n-\ell).$$

From (6.3) it follows that the F test with level α rejects when

$$(6.4) \qquad F = \frac{\hat{u}_X'\hat{u}_X - \hat{u}_Z'\hat{u}_Z}{\hat{u}_Z'\hat{u}_Z} \cdot \frac{n-\ell}{\ell-k} > f_{1-\alpha},$$

where $f_{1-\alpha}$ is the $(100(1-\alpha))$th percentile of the $F(\ell-k, n-\ell)$ distribution.
Under H_1 we have $y \sim n(Z\gamma, \sigma^2 I)$ and it is seen that

$$(6.5) \qquad F \sim F(\ell-k, n-\ell, \theta_1), \quad \theta_1 > 0,$$

where $\theta_1 = \gamma'Z'M_X Z\gamma/\sigma^2$, $\gamma \in \mathbb{R}^\ell$, $\sigma > 0$.
With the aid of (6.5) we can compute the power of the test $P(F > f_{1-\alpha}; H_1)$ for different values of γ and σ.

Next consider case (ii). Now the F test has rejection region

$$(6.6) \qquad F = \frac{\hat{u}_Z'\hat{u}_Z - \hat{u}_X'\hat{u}_X}{\hat{u}_X'\hat{u}_X} \cdot \frac{n-k}{k-\ell} < c^*.$$

In this case F can be written as

$$(6.7) \qquad F = \frac{y'(M_Z - M_X)y}{y'M_X y} \cdot \frac{n-k}{k-\ell}.$$

Under H_0: $y \sim n(X\beta, \sigma^2 I)$ we get

$$F \sim F(k-\ell, n-k, \theta_2),\ \theta_2 > 0,$$

where $\theta_2 = \beta'X'M_Z X\beta/\sigma^2$, $\beta \in \mathbb{R}^k$, $\sigma > 0$ $(X\beta \notin M(Z))$.
Let $G(x, \theta_2)$, $\theta_2 \geq 0$ be the distribution function of a $F(k-\ell, n-k, \theta_2)$ distribution, then

$$\sup_{\theta_2>0} G(x, \theta_2) = G(x, 0) \text{ for all } x.$$

Now we have to choose c^* in such a way that

$$\sup_{(\beta,\sigma)} P(F < c^*; H_0) = \alpha.$$

Since $P(F < c^*, H_0) = G(c^*, \theta_2)$, $\theta_2 > 0$, it follows that

$$\sup_{(\beta,\sigma)} P(F < c^*; H_0) = \sup_{\theta_2>0} G(c^*, \theta_2) = G(c^*, 0).$$

That is, c^* must satisfy $G(c^*, 0) = \alpha$.
Now $G(x,0)$ is the distribution function of a $F(k-\ell, n-k)$ distribution and it is seen that the rejection region of the F test in case (ii) becomes

$$(6.8) \qquad F = \frac{\hat{u}_Z'\hat{u}_Z - \hat{u}_X'\hat{u}_X}{\hat{u}_X'\hat{u}_X} \cdot \frac{n-k}{k-\ell} < f_\alpha',$$

where f_α' is the (100α)th percentile of the $F(k-\ell, n-k)$ distribution.
Under H_1: $y \sim n(Z\gamma, \sigma^2 I)$ we have

$$(6.9) \qquad F \sim F(k-\ell, n-k),$$

and therefore the power of the test is equal to

$$P(F < f_\alpha'; H_1) = G(f_\alpha', 0) = \alpha \text{ for all } \gamma \in \mathbb{R}^\ell \text{ and } \sigma > 0.$$

Let us next apply the method proposed in the foregoing sections to the nested cases (i) and (ii).
That is, we assume that μ and σ are (unobservable) random variables possessing the following incompletely specified prior distributions:
The marginal prior distribution of σ is an inverted gamma-2 distribution with unknown parameters $v > 0$ and $\omega > 0$. Under H_0 the conditional prior distribution of μ given σ is a $n(C\delta, \sigma^2 X(X'X)^{-1}X')$ distribution and under H_1 a $n(C\delta, \sigma^2 Z(Z'Z)^{-1}Z')$ distribution, where the columns of the $n\times p$ matrix C form a basis for the p-dimensional linear subspace $M(X) \cap M(Z)$ and where $\delta \in \mathbb{R}^p$ is unknown.
We shall first consider case (i). Then we have $M(X) \subset M(Z)$ and $M(X) \neq M(Z)$, i.e., $p = k < \ell$. This implies that $C = X$ and the conditional priors of μ given σ become

$$(6.10)\qquad \begin{aligned} &H_0\colon \mu|\sigma \sim n(X\delta, \sigma^2 X(X'X)^{-1}X') \\ &H_1\colon \mu|\sigma \sim n(X\delta, \sigma^2 Z(Z'Z)^{-1}Z'), \end{aligned}$$

where $\delta \in \mathbb{R}^k$ is unknown.
This results into the following reduced problem

$$(6.11)\qquad H_0'\colon y \sim t_v(X\delta, \omega^2 V_0) \text{ vs. } H_1'\colon y \sim t_v(X\delta, \omega^2 V_1),$$

where as before $V_0 = I + X(X'X)^{-1}X'$ and $V_1 = I + Z(Z'Z)^{-1}Z'$.
From Section 4, see (4.39), we know that the UMP invariant test for problem (6.11) rejects if

$$(6.12)\qquad S = \frac{\hat{u}_X'\hat{u}_X + \hat{u}_Z'\hat{u}_Z}{2\hat{u}_X'\hat{u}_X} < c.$$

It is not difficult to verify that

$$(6.13)\qquad S = \tfrac{1}{2} + \tfrac{1}{2}\left(1 + \frac{\ell-k}{n-\ell} F\right)^{-1},$$

where F is as defined in (6.1).

The inverse transformation of (6.13) is

$$F = \frac{1-S}{S-\frac{1}{2}} \cdot \frac{n-\ell}{\ell-k},$$

and it follows that the critical region $S < c$ is equivalent to $F > c^*$. This shows that the UMP invariant test for problem (6.11) has rejection region

$$(6.14) \qquad F = \frac{\hat{u}_X'\hat{u}_X - \hat{u}_Z'\hat{u}_Z}{\hat{u}_Z'\hat{u}_Z} \cdot \frac{n-\ell}{\ell-k} > c^*.$$

In order to find the critical value c^* we derive the distribution of F under H_0': $y \sim t_v(X\delta, \omega^2 V_0)$.
We first write F in the equivalent form

$$(6.15) \qquad F = \frac{y'(M_X - M_Z)y}{y'M_Z y} \cdot \frac{n-\ell}{\ell-k}.$$

Let $u = (y-X\delta)/\omega$, then $u \sim t_v(0, V_0)$.
Substitution of $y = \omega u + X\delta$ into (6.15) yields

$$(6.16) \qquad F = \frac{u'(M_X - M_Z)u}{u'M_Z u} \cdot \frac{n-\ell}{\ell-k}, \text{ with } u \sim t_v(0, V_0),$$

where use has been made of $M_X X = M_Z X = 0$.
By using the argument applied to the test statistic S in Section 5, it is seen that the random variable F in (6.16) can be rewritten as

$$(6.17) \qquad F = \frac{w'(M_X - M_Z)w}{w'M_Z w} \cdot \frac{n-\ell}{\ell-k}, \text{ where } w \sim n(0, V_0),$$

which also shows that the distribution of F under H_0' does not depend on v, ω and δ.
The numerator as well as the denominator in (6.17) are quadratic forms in normally distributed random variables.
In general we have:

(a) If $x \sim n(\mu, V)$, then $x'Ax \sim \chi^2(m, \theta)$, with $m = \text{rank}(A)$ and $\theta = \mu'A\mu$ if and only if AV is idempotent.

(b) If $x \sim n(\mu, V)$, then $x'Ax$ and $x'Bx$ are stochastically independent if and only if $AVB = 0$.

Since $(M_X - M_Z)V_0 = (M_X - M_Z)(I + X(X'X)^{-1}X') = M_X - M_Z$ and $(M_X - M_Z)^2 = M_X - M_Z$ it follows from (a) that

$$(6.18) \qquad w'(M_X - M_Z)w \sim \chi^2(\ell-k),$$

where also use has been made of rank $(M_X - M_Z) = tr(M_X - M_Z) = tr(M_X) - tr(M_Z) = n-k - (n-\ell) = \ell-k$.
In a similar way it is seen from $M_Z V_0 = M_Z(I + X(X'X)^{-1}X') = M_Z$, $M_Z^2 = M_Z$ and $rank(M_Z) = tr(M_Z) = n-\ell$ that

$$(6.19) \qquad w'M_Z w \sim \chi^2(n-\ell).$$

Moreover, $(M_X - M_Z)V_0 M_Z = (M_X - M_Z)M_Z = M_Z - M_Z = 0$ and it follows from (b) that $w'(M_X - M_Z)w$ and $w'M_Z w$ are stochastically independent.
This independence together with (6.18) and (6.19) implies that under H_0':

$$(6.20) \qquad F \sim F(\ell-k, n-\ell).$$

The results (6.14) and (6.20) show that the UMP invariant test with level α for testing the reduced problem (6.11) has critical region

$$(6.21) \qquad F = \frac{\hat{u}_X'\hat{u}_X - \hat{u}_Z'\hat{u}_Z}{\hat{u}_Z'\hat{u}_Z} \cdot \frac{n-\ell}{\ell-k} > f_{1-\alpha},$$

where $f_{1-\alpha}$ is the $(100(1-\alpha))$th percentile of the $F(\ell-k, n-\ell)$ distribution.
The test (6.21) is exactly the same as the classical F test (6.4). In other words, in the case of testing nested linear hypotheses with $\mathcal{M}(X) \subset \mathcal{M}(Z)$ and $\mathcal{M}(X) \neq \mathcal{M}(Z)$ our approach reproduces the classical F test.
It should be noted that the random variable

$$F = \frac{\hat{u}_X'\hat{u}_X - \hat{u}_Z'\hat{u}_Z}{\hat{u}_Z'\hat{u}_Z} \cdot \frac{n-\ell}{\ell-k}$$

has a $F(\ell-k, n-\ell)$ distribution under H_0: $y \sim n(X\beta, \sigma^2 I)$ as well as under H_0': $y \sim t_v(X\delta, \omega^2 V_0)$.

This is no longer the case under H_1 and H_1'. The distribution of F under H_1: $y \sim n(Z\gamma, \sigma^2 I)$ is given in (6.5). On the other hand, in a way similar to the derivation of the distribution of F under H_0' it can be shown that under H_1': $y \sim t_v(X\delta, \omega^2 V_1)$ we have:

(6.22) $\frac{1}{2}F \sim F(\ell-k, n-\ell)$.

Hence, the power of the test (6.21) under H_1' becomes

(6.23) $P(F > f_{1-\alpha}; H_1') = 1 - F(\frac{1}{2}f_{1-\alpha})$,

where F(x) is the distribution function of a $F(\ell-k, n-\ell)$ distribution. It is easily verified that $P(F > f_{1-\alpha}; H_1') > \alpha$ for $0 < \alpha < 1$.

In the second place we consider the nested case (ii), that is, $M(Z) \subset M(X)$ and $M(X) \neq M(Z)$. Now we have $p = \ell < k$ and $C = Z$. This yields the following prior distribution of μ and σ:

$$(6.24)\quad \begin{aligned} H_0&: \mu|\sigma \sim n(Z\delta, \sigma^2 X(X'X)^{-1}X') \\ H_1&: \mu|\sigma \sim n(Z\delta, \sigma^2 Z(Z'Z)^{-1}Z'). \end{aligned}$$

The reduced problem becomes:

(6.25) H_0': $y \sim t_v(Z\delta, \omega^2 V_0)$ vs. H_1': $y \sim t_v(Z\delta, \omega^2 V_1)$,

where v, ω and δ are unknown and V_0 and V_1 are as before. In this case the UMP invariant test for problem (6.25) has critical region

$$(6.26)\quad S = \frac{2\,\hat{u}_Z'\hat{u}_Z}{\hat{u}_Z'\hat{u}_Z + \hat{u}_X'\hat{u}_X} < c.$$

Let F be as defined in (6.6), then we have

$$(6.27)\quad S = \frac{2}{1 + (1 + \frac{k-\ell}{n-k}F)^{-1}}$$

and the inverse transformation of (6.27) is

$$F = \frac{2(S-1)}{2-S} \cdot \frac{n-k}{k-\ell}.$$

It follows that the region $S < c$ is equivalent to $F < c^*$.
Hence, the UMP invariant test for problem (6.25) has rejection region

$$(6.28) \qquad F = \frac{\hat{u}_Z'\hat{u}_Z - \hat{u}_X'\hat{u}_X}{\hat{u}_X'\hat{u}_X} \cdot \frac{n-k}{k-\ell} < c^*.$$

The critical value c^* can be found from the distribution of F under H_0': $y \sim t_v(Z\delta, \omega^2 V_0)$. By using similar arguments as before it can be shown that under H_0' we have

$$(6.29) \qquad \tfrac{1}{2}F \sim F(k-\ell, n-k).$$

Then it is seen from $\alpha = P(F < c^*; H_0') = P(\frac{1}{2}F < \frac{1}{2}c^*; H_0')$ that $\frac{1}{2}c^* = f_\alpha'$, which implies $c^* = 2f_\alpha'$, where as before f_α' is the (100α)th percentile of the $F(k-\ell, n-k)$ distribution.
These results show that the UMP invariant test with level α for testing the reduced problem (6.25) has rejection region

$$(6.30) \qquad F = \frac{\hat{u}_Z'\hat{u}_Z - \hat{u}_X'\hat{u}_X}{\hat{u}_X'\hat{u}_X} \cdot \frac{n-k}{k-\ell} < 2f_\alpha'.$$

The test (6.30) differs from the classical F test for the nested case (ii) as given in (6.8). That is, our approach does not reproduce the classical F test in the nested case with $M(Z) \subset M(X)$ and $M(X) \neq M(Z)$. Note that in case (ii) the random variable

$$F = \frac{\hat{u}_Z'\hat{u}_Z - \hat{u}_X'\hat{u}_X}{\hat{u}_X'\hat{u}_X} \cdot \frac{n-k}{k-\ell}$$

does not have the same distribution under H_0 and H_0'. On the other hand, in this case it can be shown that F has the same distribution under H_1: $y \sim n(Z\gamma, \sigma^2 I)$ and H_1': $y \sim t_v(Z\delta, \omega^2 V_1)$, i.e., (6.31) $F \sim F(k-\ell, n-k)$ under H_1'.

From (6.31) it follows that the power of the test (6.30) under H_1' becomes equal to

(6.32) $P(F < 2\, f_\alpha';\, H_1') = G(2f_\alpha')$,

where $G(x)$ is the distribution function of a $F(k-\ell, n-k)$ distribution. Again it is easily verified that $P(F < f_\alpha';\, H_1') > \alpha$ for $0 < \alpha < 1$.

The nested cases (i) and (ii) are not the only cases where a F distribution can be used. It turns out that besides the nested cases also in a few nonnested cases the test $S < c$ is equivalent to $W > c^*$, where under H_0' the statistic W has a F distribution. In order to see this we write the test statistic S in the form (see Appendix B)

$$(6.33) \qquad S = \frac{\sum_{i=1}^{M} \tau_i \eta_i}{\sum_{i=1}^{M} \eta_i},$$

where M is the number of different eigenvalues $\tau_1, \tau_2, \ldots, \tau_M$ with multiplicities $m_1, m_2, \ldots, m_M$ as given in Table 1 of the foregoing section.

Under H_0' the random variables $\eta_1, \eta_2, \ldots, \eta_M$ are mutually independent with $\eta_i \sim \chi^2(m_i)$, $i = 1, 2, \ldots, M$.

Suppose that $M = 2$ with $\tau_1 < \tau_2$, then

$$(6.34) \qquad S = \frac{\tau_1 \eta_1 + \tau_2 \eta_2}{\eta_1 + \eta_2}.$$

Let W be defined by

$$(6.35) \qquad W = \frac{\eta_1/m_1}{\eta_2/m_2},$$

then under H_0' we have

(6.36) $W \sim F(m_1, m_2)$.

Writing S in terms of W we get

$$(6.37) \qquad S = \frac{\tau_1 \frac{m_1}{m_2} W + \tau_2}{\frac{m_1}{m_2} W + 1},$$

which has the inverse transformation

$$(6.38) \qquad W = \frac{\tau_2 - S}{S - \tau_1} \cdot \frac{m_2}{m_1}.$$

It follows that the region $S < c$ is equivalent to $W > c^*$ and that the UMP invariant level α test for the reduced problem in the case $M = 2$ has critical region

$$(6.39) \qquad W = \frac{\tau_2 - S}{S - \tau_1} \cdot \frac{m_2}{m_1} > f^*_{1-\alpha},$$

where $f^*_{1-\alpha}$ is the $(100(1-\alpha))$th percentile of the $F(m_1, m_2)$ distribution.

The most important cases with $M = 2$ are again the nested cases (i) and (ii). First consider case (i), then we have $\tau_1 = \frac{1}{2}$ and $\tau_2 = 1$, $m_1 = \ell - k$ and $m_2 = n - \ell$. The test (6.39) becomes

$$(6.40) \qquad W = \frac{1-S}{S-\frac{1}{2}} \cdot \frac{n-\ell}{\ell-k} > f_{1-\alpha}.$$

It is easily verified that (see (6.13))

$$W = F = \frac{\hat{u}_X'\hat{u}_X - \hat{u}_Z'\hat{u}_Z}{\hat{u}_Z'\hat{u}_Z} \cdot \frac{n-\ell}{\ell-k}$$

and it follows that the test (6.40) is equal to the classical F test as derived in (6.21) and (6.4).

In the second place consider the nested case (ii). Now we have $\tau_1 = 1$, $\tau_2 = 2$, $m_1 = n-k$ and $m_2 = k-\ell$. The test (6.39) becomes

$$(6.41)\qquad W = \frac{2-S}{S-1}\cdot\frac{k-\ell}{n-k} > f^*_{1-\alpha}.$$

From (6.27) it is seen that $W = \frac{2}{F}$, where

$$F = \frac{\hat{u}_Z'\hat{u}_Z - \hat{u}_X'\hat{u}_X}{\hat{u}_X'\hat{u}_X}\cdot\frac{n-k}{k-\ell}.$$

Hence (6.41) is equivalent to

$$(6.42)\qquad F < \frac{2}{f^*_{1-\alpha}} = 2f'_\alpha,$$

where $\frac{1}{2}F = \frac{1}{W} \sim F(k-\ell,\ n-k)$ under H_0'.
This is precisely the test (6.30) derived above for the nested case (ii).
All other cases with $M = 2$, i.e., cases where the F test (6.39) can be used, are nonnested.
From Table 1 of Section 5 it follows that in all these cases we have $n = k+\ell-p$ and either $k > r$, $\ell > r$, $R = 0$ or $k = \ell = r$, $R = 1$.
Finally we shall consider the cases where $k = 0$ or $\ell = 0$ and $k = n$ or $\ell = n$. The analysis in the foregoing sections was carried out under the assumption that $0 < k < n$ and $0 < \ell < n$. Therefore the cases where k or ℓ take on a boundary value do not follow at once from the general case. However, as we shall see below, these boundary cases can easily be incorporated into the general case.
First consider the situation where $k = 0$ and $0 < \ell < n$ (the case $0 < k < n$ and $\ell = 0$ can be derived in a similar way). When $k = 0$ we have $\mathcal{M}(X) = \{0\}$ and the original problem has the form

$$(6.43)\qquad H_0\colon \mu = 0 \text{ vs. } H_1\colon \mu \in \mathcal{M}(Z)\setminus\{0\},$$

where $y \sim n(\mu,\ \sigma^2 I)$.
Since we have $p = k = 0 < \ell$ it follows that (6.43) is a subcase of the nested case (i).

If we specify the conditional prior distributions of μ given σ as follows

$$(6.44)\qquad \begin{aligned} H_0&: \mu|\sigma \sim n(0,\ 0) \\ H_1&: \mu|\sigma \sim n(0,\ \sigma^2 Z(Z'Z)^{-1}Z'), \end{aligned}$$

it is easily seen that the reduced problem becomes

$$(6.45)\qquad H_0': y \sim t_v(0,\ \omega^2 I) \text{ vs. } H_1': y \sim t_v(0,\ \omega^2 V_1),$$

where as before $V_1 = I + Z(Z'Z)^{-1}Z'$ and v and ω are unknown.
Note that the conditional prior distribution of μ given σ under H_0 is a degenerate distribution with all probability mass concentrated at the single point $\mu = 0$.
From the result (4.19) of Section 4 (with $V_0 = I$) it follows that the UMP invariant test for problem (6.45) rejects when

$$(6.46)\qquad S = \frac{y'V_1^{-1}y}{y'y} < c.$$

Since $V_1^{-1} = I - \frac{1}{2}Z(Z'Z)^{-1}Z'$, the test statistic S can be written as

$$(6.47)\qquad S = \frac{y'y + \hat{u}_Z'\hat{u}_Z}{2y'y}.$$

As would be expected the result (6.47) can be obtained from the general case (4.40) through the substitution of $\hat{u}_X'\hat{u}_X = y'y$ when $M(X) = \{0\}$.
The probability distribution of S under H_0' can be found by using the method described in Section 5. In this case we have $M = 2$, $\tau_1 = \frac{1}{2}$, $\tau_2 = 1$, $m_1 = \ell$ and $m_2 = n-\ell$.
Since we have to do with a case of nested models it follows from the results of the first part of this section that the test $S < c$ is equivalent to the classical F test for problem (6.43), i.e.,

$$(6.48)\qquad F = \frac{y'y - \hat{u}_Z'\hat{u}_Z}{\hat{u}_Z'\hat{u}_Z} \cdot \frac{n-\ell}{\ell} > f_{1-\alpha}$$

where $f_{1-\alpha}$ is the $(100(1-\alpha))$th percentile of the $F(\ell,\ n-\ell)$ distribution.

Usually (6.48) is written in the equivalent form

$$(6.49)\qquad F = \frac{\hat{\gamma}'Z'Z\hat{\gamma}}{\hat{u}_Z'\hat{u}_Z} \cdot \frac{n-\ell}{\ell} > f_{1-\alpha}, \text{ where } \hat{\gamma} = (Z'Z)^{-1}Z'y,$$

which is precisely the classical F test for the problem of testing H_0: $\gamma = 0$ against H_1: $\gamma \neq 0$ in the linear model $y = Z\gamma + u$ with $u \sim n(0, \sigma^2 I)$. Of course, this latter problem is equivalent to problem (6.43).

From the above discussion we conclude that the case $k = 0$ (or similarly $\ell = 0$) can be incorporated without any difficulty into our general approach.

A very interesting particular case occurs if besides $k = 0$ we have $\ell = 1$ and $Z = \iota$, where $\iota' = (1\ 1\ \dots\ 1)$.

In this particular situation the original problem (6.43) becomes equivalent to the problem of testing

$$(6.50)\qquad H_0\text{: } \theta = 0 \quad \text{against} \quad H_1\text{: } \theta \neq 0$$

on the basis of a random sample $y' = (Y_1\ Y_2\ \dots\ Y_n)$ from a $n(\theta, \sigma^2)$ distribution.

The reduced problem has the form

$$(6.51)\qquad H_0'\text{: } y \sim t_v(0, \omega^2 I) \quad \text{against} \quad H_1'\text{: } y \sim t_v(0, \omega^2(I + \iota(\iota'\iota)^{-1}\iota')).$$

The UMP invariant test for problem (6.51) rejects when (see (6.47))

$$(6.52)\qquad S = \frac{\sum_{i=1}^{n} Y_i^2 + \sum_{i=1}^{n} (Y_i - \bar{Y})^2}{2 \sum_{i=1}^{n} Y_i^2} < c,$$

where $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$.

Since (6.52) is equivalent to

$$(6.53)\qquad F = \frac{n\bar{Y}^2}{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2} > f_{1-\alpha},$$

where $f_{1-\alpha}$ is the $(100(1-\alpha))$th percentile of the $F(1, n-1)$ distribution, which in turn is equivalent to

$$(6.54) \qquad |T| > t_{1-\frac{1}{2}\alpha},$$

where

$$T = \frac{\sqrt{n}\ \bar{Y}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

and $t_{1-\frac{1}{2}\alpha}$ is the $(100(1-\frac{1}{2}\alpha))$th percentile of the $t(n-1)$ distribution, it follows that our general approach reproduces the classical t test for problem (6.50).

In the second place consider the case where $0 \leq k < n$ and $\ell = n$ (again the case $k = n$ and $0 \leq \ell < n$ can be handled in a similar way). When $\ell = n$ we have $M(Z) = \mathbb{R}^n$ and the original problem can be written as

$$(6.55) \qquad H_0: \mu \in M(X) \quad \text{vs.} \quad H_1: \mu \in \mathbb{R}^n \backslash M(X).$$

Since $p = k < \ell = n$ it follows that (6.55) is a subcase of (i). The conditional prior distributions of μ given σ are specified by

$$(6.56) \qquad \begin{aligned} &H_0: \mu|\sigma \sim n(X\delta, \sigma^2 X(X'X)^{-1}X') \\ &H_1: \mu|\sigma \sim n(X\delta, \sigma^2 I), \end{aligned}$$

where use has been made of $Z(Z'Z)^{-1}Z' = I$.

Note that in this case the conditional distribution of μ given σ under H_1 is a nonsingular normal distribution. The reduced problem becomes

$$(6.57) \qquad H_0': y \sim t_v(X\delta, \omega^2 V_0) \quad \text{vs.} \quad H_1': y \sim t_v(X\delta, \omega^2(2I)),$$

where $V_0 = I + X(X'X)^{-1}X'$ and where v, ω and δ are unknown. From (4.26) it is seen that the UMP invariant test for problem (6.57) rejects when

$$(6.58) \qquad S = \frac{\frac{1}{2}\ \hat{u}_1'\hat{u}_1}{\hat{u}_0'V_0^{-1}\hat{u}_0} < c,$$

where $\hat{u}_i = y - C(C'V_i^{-1}C)^{-1}C'V_i^{-1}y$, $i = 0, 1$ with $C = X$ and $V_1 = 2I$.
Since $\hat{u}_0 = \hat{u}_1 = \hat{u}_X = M_X y$ we get

(6.59) $$S = \frac{\hat{u}_X'\hat{u}_X}{2\hat{u}_X'\hat{u}_X} = \tfrac{1}{2}.$$

As would be expected this result can be obtained form the general case (4.38) through the substitution of $\hat{u}_C'\hat{u}_C = \hat{u}_X'\hat{u}_X$ and $\hat{u}_Z'\hat{u}_Z = 0$ when $M(X) \subset M(Z) = \mathbb{R}^n$.
With the aid of the results of Appendix A it is seen that the UMP invariant level α test for problem (6.57) rejects with probability α regardless of the observations.

Note that in this case the F statistic $\frac{\hat{u}_X'\hat{u}_X - \hat{u}_Z'\hat{u}_Z}{\hat{u}_Z'\hat{u}_Z} \cdot \frac{n-\ell}{\ell-k}$ is not defined since $\hat{u}_Z'\hat{u}_Z = 0$ and $n-\ell = 0$.
The above results show that the observations are of no use in testing a problem of the type (6.55). This is not surprising since in this case the alternative hypothesis H_1 is left almost entirely unspecified.
In the foregoing derivation it was assumed that $k > 0$, when $k = 0$ a similar result is obtained.
From Section 5 it is seen that in the case $\ell = n$ we have $M = 1$, $\tau_1 = \frac{1}{2}$ and $m_1 = \ell-k = n-k$, which shows that $P(S = \frac{1}{2}) = 1$, as would be expected and it follows that the boundary case $\ell = n$ can easily be incorporated into our general approach.

7. An optimum property of the test

In the foregoing sections we have derived a test for the problem of testing linear hypotheses which has the property of being UMP invariant for the so-called reduced problem.
In this section we shall show that this test is also optimal in a certain sense for the original problem. We no longer consider the parameters μ and σ as random and we interprete the prior distributions as weight functions which express the importance the experimenter attaches to the various values of the parameters.
Throughout this section we use the following notation:

$$\theta = (\mu, \sigma),\ \Theta_0 = \{(\mu, \sigma) \mid \mu \in M(X),\ \sigma > 0\},$$

$$(7.1) \qquad \Theta_1 = \{(\mu, \sigma) \mid \mu \in M(Z)\backslash M(X),\ \sigma > 0\},\ \Theta = \Theta_0 \cup \Theta_1,$$

$$f(x;\theta) = (2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\{-\frac{1}{2\sigma^2}(x-\mu)'(x-\mu)\},$$

where $x \in \mathbb{R}^n$ and $\theta \in \Theta$.
Then the problem of testing two linear hypotheses can be formulated as:

(7.2) Given that the sample $y \sim f(x;\theta)$, $\theta \in \Theta$, we want to test $H_0\colon \theta \in \Theta_0$ vs. $H_1\colon \theta \in \Theta_1$.

Now θ is treated as an unknown parameter and not as a random vector, which means that we consider $f(x;\theta)$ as the unconditional p.d.f. of y.
Suppose that we assign a weight function W to the various subsets of the parameterspace Θ.
We first assume that this weight function is completely specified and expresses the importance attached to the various subsets of Θ.
The mathematical expression for a weight function is a probability measure defined on a suitable class of subsets of Θ (a nonnegative set function with $W(\Theta) = 1$).
In a testing problem of the form (7.2) usually the weight function W is specified in steps. That is, we first specify the weight functions W_0 and W_1 on Θ_0 and Θ_1, respectively and then assign (positive) weights w_0

and w_1 to H_0 and H_1, where $w_1 = 1 - w_0$. The weight function W_i is a completely specified probability measure defined on a suitable class of subsets of Θ_i, i = 0, 1. Note that $W_i(\Theta_i) = 1$ for i = 0, 1.
Now the weight function W on Θ is defined by

(7.3) $$W = w_0W_0 + w_1W_1,$$

that is, a (measurable) subset A of Θ has weight $W(A) = w_0W_0(A) + w_1W_1(A)$. It is easily verified that W is a completely specified probability measure.
Note that $W(\Theta) = w_0 + w_1 = 1$ and $W(\Theta_i) = w_i$, i = 0, 1. For subsets of A_i of Θ_i we have:

(7.4) $$W(A_i) = w_iW_i(A_i),\ i = 0, 1.$$

Problem (7.2) can be considered as a statistical decision problem with two possible decisions: d_0 = "accept H_0" and d_1 = "reject H_0". Let D be a stochastic variable with possible outcomes d_0 and d_1, then a test $\phi(y)$ (critical function) is defined as the conditional probability that D takes on the value d_1 given that the sample outcome is y, i.e.,

(7.5) $$\phi(y) = P(D = d_1|y),\ y \in \mathbb{R}^n.$$

The power function $\pi(\theta, \phi)$ of the test ϕ is

(7.6) $$\pi(\theta, \phi) = P(D = d_1) =$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(D = d_1|y)\ f(y;\ \theta)dy_1 \ldots dy_n =$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)\ f(y;\ \theta)dy_1 \ldots dy_n,\ \theta \in \Theta,$$

where $y' = (y_1\ y_2 \ldots y_n)$.
We also have

(7.7) $$1 - \pi(\theta, \phi) = P(D = d_0).$$

Further we specify the following loss function $\ell(\theta, d) \geq 0$, $d = d_0, d_1$, $\theta \in \Theta$ for the above decision problem:

$$(7.8) \qquad \begin{aligned} \ell(\theta, d) &= 0 \text{ if } d = d_0,\ \theta \in \Theta_0 \\ &= 0 \text{ if } d = d_1,\ \theta \in \Theta_1 \\ &= \ell_0 \text{ if } d = d_1,\ \theta \in \Theta_0 \\ &= \ell_1 \text{ if } d = d_0,\ \theta \in \Theta_1, \end{aligned}$$

where $\ell_i > 0$ ($i = 0, 1$) is a given number.
The risk (expected loss) of a test ϕ is given by

$$(7.9) \qquad R(\theta, \phi) = E(\ell(\theta, D)) = \ell(\theta, d_0)\, P(D = d_0) + \ell(\theta, d_1)\, P(D = d_1),\ \theta \in \Theta.$$

From (7.8) it follows that

$$(7.10) \qquad \begin{aligned} R(\theta, \phi) &= \ell_0 P(D = d_1),\ \theta \in \Theta_0 \\ &= \ell_1 P(D = d_0),\ \theta \in \Theta_1 \end{aligned}$$

and by making use of (7.6) and (7.7) we get

$$(7.11) \qquad \begin{aligned} R(\theta, \phi) &= \ell_0 \pi(\theta, \phi),\ \theta \in \Theta_0 \\ &= \ell_1 - \ell_1 \pi(\theta, \phi),\ \theta \in \Theta_1. \end{aligned}$$

Usually $\pi(\theta, \phi)$, $\theta \in \Theta_0$ is called the probability of a type I error and $\pi(\theta, \phi)$, $\theta \in \Theta_1$ is called the power of the test.
As is well known there does not exist a test ϕ^* which minimizes the risk $R(\theta, \phi)$ for all $\theta \in \Theta$ and all ϕ, i.e., a test ϕ^* with $R(\theta, \phi^*) \leq R(\theta, \phi)$ for all $r \in \Theta$ and all ϕ. In view of this a natural procedure is to consider the (weighted) average risk of a test ϕ with respect to weight function W as given in (7.3), i.e.,

$$(7.12) \qquad r(\phi) = \int_{\Theta} R(\theta, \phi)\, dW(\theta),$$

and minimizing this average risk among all tests.
That is, we try to find a test ϕ^* with $r(\phi^*) \leq r(\phi)$ for all ϕ.

With the aid of (7.3) and (7.11) we have

$$(7.13)\qquad r(\phi) = \int_{\Theta} R(\theta, \phi)dW(\theta) =$$

$$\int_{\Theta_0} R(\theta, \phi)dW(\theta) + \int_{\Theta_1} R(\theta, \phi)dW(\theta) =$$

$$w_0 \int_{\Theta_0} R(\theta, \phi)dW_0(\theta) + w_1 \int_{\Theta_1} R(\theta, \phi)dW_1(\theta) =$$

$$w_0\ell_0 \int_{\Theta_0} \pi(\theta, \phi)dW_0(\theta) + w_1\ell_1 - w_1\ell_1 \int_{\Theta_1} \pi(\theta, \phi)dW_1(\theta)$$

$$= w_1\ell_1 + w_0\ell_0\tilde{\pi}_0(\phi) - w_1\ell_1\tilde{\pi}(\phi),$$

where the average probability of a type I error $\tilde{\pi}_0(\phi)$ and the average power $\tilde{\pi}_1(\phi)$ are defined as

$$\tilde{\pi}_0(\phi) = \int_{\Theta_0} \pi(\theta, \phi)dW_0(\theta)$$

$$(7.14)$$

$$\tilde{\pi}_1(\phi) = \int_{\Theta_1} \pi(\theta, \phi)dW_1(\theta).$$

Let the function k_i be defined by

$$(7.15)\qquad k_i(x) = \int_{\Theta_i} f(x;\ \theta)dW_i(\theta),\ i = 0,\ 1,$$

where $f(x;\ \theta)$ is as given above.
Then $k_i(x)$ ($i = 0, 1$) satisfies the requirements of a p.d.f. and we get:

$$(7.16)\qquad \tilde{\pi}_0(\phi) = \int_{\Theta_0} \pi(\theta, \phi)dW_0(\theta) =$$

$$\int_{\Theta_0} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)f(y;\ \theta)dy_1 \ldots dy_n dW_0(\theta)$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y) \int_{\Theta_0} f(y;\ \theta)dW_0(\theta)dy_1 \ldots dy_n =$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)k_0(y)dy_1 \ldots dy_n,$$

where use has been made of (7.6).
Similarly, we have

$$(7.17) \qquad \tilde{\pi}_1(\phi) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)k_1(y)dy_1 \cdots dy_n.$$

Upon substituting (7.16) and (7.17) into (7.13) we obtain

$$(7.18) \qquad r(\phi) = w_1\ell_1 + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)[w_0\ell_0k_0(y) - w_1\ell_1k_1(y)]dy_1 \cdots dy_n.$$

It easily follows that the test ϕ_1^* with miminum average risk becomes:

$$(7.19) \qquad \phi_1^*(y) = 1 \text{ if } w_0\ell_0k_0(y) - w_1\ell_1k_1(y) < 0$$

$$= 0 \text{ if } \quad " \qquad " \quad \geq 0.$$

In other words, the test ϕ_1^* rejects $H_0(\phi_1^*(y) = 1)$ if

$$(7.20) \qquad \frac{k_1(y)}{k_0(y)} > \frac{w_0\ell_0}{w_1\ell_1}.$$

Usually the test ϕ_1^* in (7.19) is called the Bayes test for problem (7.2) and it is seen that this test depends on the weight function W (specified by w_0, w_1, W_0 and W_1) and on the loss function $\ell(\theta, d)$. Now we take the weight functions W_0 and W_1 equal to the probability measures corresponding to the prior distributions of θ under H_0 and H_1, respectively, as given in (2.5), (2.7) and (2.8) of Section 2 and moreover, we take the weights w_0 and w_1 as considered in Section 2. Then it is easily seen that $k_i(x)$ becomes

$$(7.21) \qquad k_i(x) = \frac{\Gamma(\frac{n+v}{2})v^{\frac{v}{2}}\omega^{-n}}{\pi^{n/2}\Gamma(\frac{v}{2})\sqrt{\det(V_i)}} (v + Q_i(x))^{-\frac{n+v}{2}}, \; i = 0, 1,$$

where

$$(7.22) \qquad Q_i(x) = (x - \bar{\mu}_i)'V_i^{-1}(x - \bar{\mu}_i)/\omega^2, \; i = 0, 1$$

and

$$\bar{\mu}_0 = X\eta$$

$$\bar{\mu}_1 = Z\xi$$

(7.23)

$$V_0 = I + X\Omega X'$$

$$V_1 = I + Z\Lambda Z'.$$

Since the parameters v, ω, η, ξ, Ω and Λ are given it follows that the weight function $W = w_0 W_0 + w_1 W_1$ is completely specified. Substitution of (7.21) and $w_i = p_i$, i = 0, 1 into (7.20) shows that the test ϕ_1^* which minimizes the average risk $r(\phi) = \int_\Theta R(\theta, \phi)dW(\theta)$ has rejection region

$$\text{(7.24)} \qquad \frac{v + Q_1(y)}{v + Q_0(y)} < [\frac{\det(V_0)}{\det(V_1)}]^{\frac{1}{n+v}} (\frac{p_1 \ell_1}{p_0 \ell_0})^{\frac{2}{n+v}} = c,$$

which is precisely the Bayes test (2.32) derived in Section 2 under the assumption that θ is a random vector.

From the mathematical point of view it does not matter whether we consider θ as nonrandom and W as a weight function or θ as random and W as a prior probability measure. The solution is the same only the interpretation differs.

A difficulty with the above procedure is that a complete specification of the weight function W and the loss function $\ell(\theta, d)$ is required. If one is not able to (or willing to) specify a weight function and a loss function, it is natural to concentrate on the power function $\pi(\theta, \phi)$.

Again there does not exist a test ϕ^* which minimizes the probability of a type I error for all $\theta \in \Theta_1$ and all ϕ and at the same time maximizes the power of the test for all $\theta \in \Theta_1$ and all ϕ. Then the usual procedure is to restrict attention to tests which satisfy $\pi(\theta, \phi) \leq \alpha$ for all $\theta \in \Theta_0$, i.e., the probability of a type I error does not exceed the preassigned significance level α, and to attempt to maximize the power $\pi(\theta, \phi)$ for all $\theta \in \Theta_1$ subject to the above condition. When such an optimal test exists it is called uniformly most powerful (UMP) of level α.

For our problem (7.2) no UMP test exists and it is customary to narrow the class of tests still further and to try to find the UMP test within this smaller class. Well-known criteria for narrowing the class of level α tests are unbiasedness and invariance. However, except for the case of nested models, reduction through unbiasedness or invariance considerations does not lead to a solution for problem (7.2).
A possible way out is to concentrate on the average probability of type I error $\tilde{\pi}_0(\phi)$ and the average power of the test $\tilde{\pi}_1(\phi)$ as given in (7.14). In this approach no loss function is specified and the weight function W is only partially specified, that is, the weight functions W_0 and W_1 are completely specified but the weights w_0 and w_1 are considered as unknown.
Now we try to find a test ϕ^* which maximizes the average power $\tilde{\pi}_1(\phi)$ among all tests with an average probability of a type I error $\tilde{\pi}_0(\phi)$ not exceeding the level α. In other words we try to find a test ϕ^* which satisfies

$$\begin{aligned} &\tilde{\pi}_0(\phi^*) \leq \alpha. \\ (7.25)\qquad &\tilde{\pi}_1(\phi^*) \geq \tilde{\pi}_1(\phi), \text{ for all } \phi \text{ with } \tilde{\pi}_0(\phi) \leq \alpha. \end{aligned}$$

If such a test ϕ^* exists it will be called a test with best average power for problem (7.2). From (7.16) and (7.17) it follows that problem (7.25) is equivalent to: Find the test ϕ^* which satisfies

$$\begin{aligned} &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi^*(y)k_0(y)dy_1 \ldots dy_n \leq \alpha. \\ (7.26)\qquad &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi^*(y)k_1(y)dy_1 \ldots dy_n \geq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)k_1(y)dy_1 \ldots dy_n, \end{aligned}$$

for all ϕ with $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)k_0(y)dy_1 \ldots dy_n \leq \alpha$, where the average densities $k_i(x)$ $(i = 0, 1)$ are as given in (7.15). Since the latter problem is precisely the problem of finding the most powerful level α test for

$$(7.27)\qquad H_0': y \sim k_0(x) \quad \text{vs.} \quad H_1': y \sim k_1(x),$$

where k_0 and k_1 are completely specified, it follows from the lemma of Neyman and Pearson that, apart from trivial cases, the solution for problem (7.26) becomes

$$(7.28) \qquad \phi_2^*(y) = 1 \text{ if } \frac{k_1(y)}{k_0(y)} > c$$

$$= 0 \text{ if } \quad " \quad \leq c,$$

where c has to be determined from

$$\tilde{\pi}_0(\phi_2^*) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_2^*(y) k_0(y) dy_1 \cdots dy_n = P\left(\frac{k_1(y)}{k_0(y)} > c; H_0' \right) = \alpha.$$

Hence, the test with best average power for the original problem (7.2) is given by (7.28) and the average power of ϕ_2^* can be found from

$$\tilde{\pi}_1(\phi_2^*) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_2^*(y) k_1(y) dy_1 \cdots dy_n = P\left(\frac{k_1(y)}{k_0(y)} > c; H_1' \right).$$

If we again take the weight functions W_0 and W_1 equal to the probability measures induced by the prior distributions of θ under H_0 and H_1, respectively, as given in (2.5), (2.7) and (2.8) of Section 2, it is easily verified from (7.21) and (7.28) that test ϕ_2^* with best average power for the problem of testing two linear hypotheses (7.2) rejects when ($\phi_2^*(y) = 1$):

$$(7.29) \qquad \frac{v + Q_1(y)}{v + Q_0(y)} < c,$$

where c has to be chosen such that

$$P\left(\frac{v + Q_1(y)}{v + Q_0(y)} < c; H_0' \right) = \alpha,$$

and where $Q_i(y)$, $i = 0, 1$ is defined in (7.22) and (7.23). The test (7.29) is exactly equal to the test (2.33) of Section 2, which was derived under the assumption that θ is a random vector.

Although the above procedure does not require the specification of a loss function $\ell(\theta, d)$ and prior weights w_0 and w_1, it remains a difficulty that we need completely specified weight functions W_0 and W_1. For instance, in order to apply the test (7.29) we have to choose the real numbers $v > 0$, $\omega > 0$, the vectors $\eta \in \mathbb{R}^k$, $\xi \in \mathbb{R}^\ell$ and the matrices Ω and Λ.

It would be an improvement if we could find a test which maximizes the average power for a whole class of weight functions. We shall show below that under certain conditions such a test indeed exists.

Let $\{W_{0\psi};\ \psi \in \Psi\}$ and $\{W_{1\psi};\ \psi \in \Psi\}$ be families of weight functions defined on suitable classes of subsets of Θ_0 and Θ_1, respectively. The vector ψ which labels the weight functions belongs to a given space Ψ. That is, $W_{i\psi}$ $(i = 0, 1)$ is a weight function for every $\psi \in \Psi$.

If ϕ is a test for the problem (7.2) and if $\pi(\theta, \phi)$ denotes the power function of this test, i.e.,

$$(7.30) \qquad \pi(\theta, \phi) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y) f(y;\ \theta) dy_1 \ldots dy_n,\ \theta \in \Theta,$$

it follows that the average probability of a type I error and the average power of the test depend on the vector ψ which labels the weight functions, that is,

$$(7.31) \qquad \begin{aligned} \tilde{\pi}_0(\psi, \phi) &= \int_{\Theta_0} \pi(\theta, \phi) dW_{0\psi}(\theta) \\ \tilde{\pi}_1(\psi, \phi) &= \int_{\Theta_1} \pi(\theta, \phi) dW_{1\psi}(\theta) \end{aligned}$$

where $\psi \in \Psi$.

Now we are looking for a test ϕ^* which maximizes the average power $\tilde{\pi}_1(\psi, \phi)$ for all $\psi \in \Psi$ among all tests with an average probability of a type I error $\tilde{\pi}_0(\psi, \phi)$ not exceeding α. That is, we try to find a test ϕ^* which satisfies

$$(7.32) \qquad \begin{aligned} &\tilde{\pi}_0(\psi, \phi^*) \leq \alpha,\ \psi \in \Psi. \\ &\tilde{\pi}_1(\psi, \phi^*) \geq \tilde{\pi}_1(\psi, \phi), \text{ for all } \psi \in \Psi \text{ and all } \phi \\ &\text{with } \tilde{\pi}_0(\psi, \phi) \leq \alpha,\ \psi \in \Psi. \end{aligned}$$

If such as optimal test ϕ^* exists, it will be called a test with uniformly best average power for problem (7.2).

Upon substituting (7.30) into (7.31) we get

$$(7.33)\qquad \tilde{\pi}_i(\psi, \phi) = \int_{\Theta_i} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)f(y;\ \theta)dy_1 \ldots dy_n dW_{i\psi}(\theta)$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y) \int_{\Theta_i} f(y;\ \theta)dW_{i\psi}(\theta)dy_1 \ldots dy_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)k_i(y;\ \psi)dy_1 \ldots dy_n,\ i = 0,\ 1,$$

where the functions $k_i(x;\ \psi)$ are defined by

$$(7.34)\qquad k_i(x;\ \psi) = \int_{\Theta_i} f(x;\ \theta)dW_{i\psi}(\theta),\ i = 0,\ 1,$$

with $x \in \mathbb{R}^n$ and $\psi \in \Psi$.

It should be noted that the function $k_i(x;\ \psi)$ $(i = 0,\ 1)$ satisfies the requirements of a p.d.f. for every $\psi \in \Psi$. With the aid of (7.33) it is easily verified that problem (7.32) is equivalent to the problem of finding a test ϕ^* satisfying

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi^*(y)k_0(y;\ \psi)dy_1 \ldots dy_n \leq \alpha,\ \psi \in \Psi.$$

$$(7.35)\qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi^*(y)k_1(y;\ \psi)dy_1 \ldots dy_n \geq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)k_1(y;\ \psi)dy_1 \ldots dy_n,$$

for all $\psi \in \Psi$ and all ϕ with $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)k_0(y;\ \psi)dy_1 \ldots dy_n \leq \alpha$,

$\psi \in \Psi$.

Since (7.35) is precisely the problem of finding the UMP level α test for

$$(7.36)\qquad H_0'\colon y \sim k_0(x;\ \psi) \quad \text{vs.} \quad H_1'\colon y \sim k_1(x;\ \psi),$$

where $\psi \in \Psi$ is considered as an unknown parametervector, it follows that (subject to its existence) the test with uniformly best average power for problem (7.2) is equal to the UMP level α test for problem (7.36).

Unfortunately, however, no such test exists for problem (7.36).
Since problem (7.2) is invariant under the group of transformations

(7.37) $G: g(y) = ay + C\alpha$ for all $a \in \mathbb{R}^1$, $a \neq 0$ and $\alpha \in \mathbb{R}^p$,

where the columns of the $n \times p$ matrix C span the p-dimensional linear subspace $\mathcal{M}(X) \cap \mathcal{M}(Z)$ (we only consider the case $p > 0$, the case $p = 0$ can be treated in a similar way), it is natural to restrict attention to the invariant tests, i.e., tests ϕ which satisfy

(7.38) $\phi(g(y)) = \phi(y)$ for all $g \in G$.

Then, among the invariant tests we try to find the test with uniformly best average power.
In other words, we try to solve problem (7.32) or, equivalently, problem (7.36) subject to the extra restriction (7.38).
The transformation $g \in G$ induce a group of transformations $\bar{G}$ in the parameter space Θ given by:

(7.39) $\bar{G}: \bar{g}(\theta) = (a\mu + C\alpha, |a|\sigma)$ for all $a \in \mathbb{R}^1$,

$a \neq 0$ and $\alpha \in \mathbb{R}^p$, where $\theta = (\mu, \sigma) \in \Theta$.
Moreover, the group $\bar{G}$ induces a group $\tilde{G}$ of transformations $\tilde{g}$ in the space Ψ of points ψ which label the weight functions.
If the weight functions $W_{i\psi}$ are chosen in such a way that the families $\{W_{i\psi}; \psi \in \Psi\}$, $i = 0, 1$, remain invariant under the group of transformations $\bar{G}$, it follows that problem (7.36) remains invariant under the transformation $g \in G$.
Since problem (7.32) subject to the restriction of invariance is equivalent to the problem of finding the UMP level α invariant test for (7.36), it is seen that under the above conditions the test with uniformly best average power among the invariant tests for problem (7.2) is equal to the UMP invariant level α test for problem (7.36), provided that the latter test exists.
To be more specific, let $\psi = (\nu, \omega, \delta)$ and $\Psi = \{(\nu, \omega, \delta) | \nu > 0, \omega > 0, \delta \in \mathbb{R}^p\}$ and take the weight functions $W_{0\psi}$ and $W_{1\psi}$ equal to the probability measures corresponding to the prior distributions of θ under H_0 and H_1, respectively, as given in (3.19) and (3.20) of Section 3.

With this choice of $W_{i\psi}$ it is easily seen that the functions $k_i(x;\ \psi)$ becomes equal to:

$$(7.40)\qquad k_i(x;\ \psi) = \frac{\Gamma(\frac{n+v}{2})\ v^{\frac{v}{2}}\omega^{-n}}{\pi^{n/2}\Gamma(\frac{v}{2})\ \sqrt{\det(V_i)}}\left[v + \frac{1}{\omega^2}(x-C\delta)'V_i^{-1}(x-C\delta)\right]^{-\frac{n+v}{2}},$$

$i = 0,\ 1$, where $V_0 = I + X(X'X)^{-1}X'$ and $V_1 = I + Z(Z'Z)^{-1}Z'$.
The function $k_i(x;\ \psi)$ is the density of a multivariate t distribution and this shows that problem (7.36) can be written as:

$$(7.41)\qquad H_0':\ y \sim t_v(C\delta,\ \omega^2 V_0)\quad \text{vs.}\quad H_1':\ y \sim t_v(C\delta,\ \omega^2 V_1),$$

with $\psi = (v,\ \omega,\ \delta) \in \Psi$ unknown.
Moreover, the group $\tilde{G}$ of transformations $\tilde{g}$ induced by the group G in the space Ψ becomes:

$$(7.42)\qquad \tilde{G}:\ \tilde{g}(\psi) = (v,\ |a|\omega,\ a\delta + \alpha),$$

for all $a \in \mathbb{R}'$, $a \neq 0$ and $\alpha \in \mathbb{R}^p$, where $\psi = (v,\ \omega,\ \delta) \in \Psi$.
It easily follows that problem (7.41) remains invariant under the group G and the above argument shows that the test ϕ_3^* with uniformly best average power among the invariant tests for problem (7.2) is equal to the UMP invariant level α test for problem (7.41). The latter test can be obtained from the results of Appendix A and Section 4 and it is seen that the test ϕ_3^* with uniformly best average power among the invariant tests becomes:

$$(7.43)\qquad \phi_3^*(y) = 1 \text{ if } S = \frac{\hat{u}_C'\hat{u}_C + \hat{u}_Z'\hat{u}_Z}{\hat{u}_C'\hat{u}_C + \hat{u}_X'\hat{u}_X} < c$$

$$= 0 \text{ if } \qquad \text{"} \qquad \geq c,$$

where $\hat{u}_C$, $\hat{u}_X$ and $\hat{u}_Z$ are the vectors of residuals after least-squares regression of y on C, X and Z, respectively, and where the critical value c has to be determined from $P(S < c;\ H_0') = \alpha$.
This test was derived in Section 4, see (4.39), under the assumption that $\theta = (\mu,\ \sigma)$ is a random vector.

The average probability of a type I error $\tilde{\pi}_0(\psi, \phi_3^*)$ and the average power $\tilde{\pi}_1(\psi, \phi_3^*)$ of the test (7.43) turn out to be independent of $\psi \in \Psi$, i.e.,

$$\tilde{\pi}_0(\psi, \phi_3^*) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_3^*(y)k_0(y; \psi)dy_1 \ldots dy_n$$

$$= P(S < c; H_0') = \alpha \text{ for all } \psi \in \Psi,$$

and

$$\tilde{\pi}_1(\psi, \phi_3^*) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_3^*(y)k_1(y; \psi)dy_1 \ldots dy_n$$

$$= P(S < c; H_1') = \text{constant for all } \psi \in \Psi.$$

A final judgement of the tests ϕ_1^*, ϕ_2^* and ϕ_3^* as given in (7.24), (7.29) and (7.43) can be made by computing the true probability of a type I error and the true power of these tests, i.e., by computing the value of the power functions

$$\text{(7.44)} \qquad \pi(\theta, \phi_i^*) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_i^*(y)f(y; \theta)dy_1 \ldots dy_n; \; i = 1, 2, 3$$

for some relevant points $\theta \in \Theta_0$ and $\theta \in \Theta_1$, respectively, where $f(y; \theta)$ is the p.d.f. of a $n(\mu, \sigma^2 I)$ distribution.

In the above discussion three different approaches were considered in order to find a test for the problem of testing two linear hypotheses (problem (7.2)).

These approaches result into:

(i) A test with minimum average risk (ϕ_1^*).

(ii) A test with best average power (ϕ_2^*).

(iii) A test with uniformly best average power among the invariant tests (ϕ_3^*).

Although it is possible to use other weight functions than those considered above, the resulting procedures loose much of their simplicity when a different type of weight functions is chosen. Finally, we note that when θ is considered as a random vector and W_i or $W_{i\psi}$, $i = 0, 1$ as a prior probability distribution, only the interpretation differs but the results remain the same. That is, in this case we obtain a test ϕ_1^* with minimum expected risk, a test ϕ_2^* with best expected power and a test ϕ_3^* with uniformly best expected power among the invariant tests, respectively.

8. Summary of the results and conclusions

In this section we shall summarize the results from the foregoing sections and meanwhile we shall draw some conclusions. Moreover, we shall give a description of the required computations in order to use the test in practice.

The data set in a problem of testing linear hypotheses consists of (y, X, Z), where the sample y is considered as an empirical outcome of some n-dimensional random vector, X is a given nonstochastic $n \times k$ matrix with rank k and Z is a given nonstochastic $n \times \ell$ matrix with rank ℓ. Often X and Z are referred to as the regressor matrices.

Given the above data set, the problem of testing linear hypotheses (under the normality assumption) has the following form

$$(8.1) \qquad H_0: y \sim n(X\beta, \sigma^2 I) \quad \text{vs.} \quad H_1: y \sim n(Z\gamma, \sigma^2 I),$$

where $\beta \in \mathbb{R}^k$, $\gamma \in \mathbb{R}^\ell$ and $\sigma > 0$ are unknown parameters and where under H_1 the points γ with $Z\gamma = X\beta$, $\beta \in \mathbb{R}^k$ are excluded.

Let $p = \dim(M(X) \cap M(Z))$, suppose that $p > 0$ and let C be a $n \times p$ matrix, the columns of which span $M(X) \cap M(Z)$. As we saw above, if we treat the parameters $(\mu, \sigma) = (X\beta, \sigma)$ or $(Z\gamma, \sigma)$ as an (unobservable) outcome from a certain incompletely specified prior distribution and if we consider the distributions in (8.1) as conditional on (μ, σ), the sample y can be thought of as originating from a multivariate t distribution with v degrees of freedom (denoted by the symbol t_v).

In other words, under the above assumptions, problem (8.1) is equivalent to the reduced problem of testing

$$(8.2) \qquad H_0': y \sim t_v(C\delta, \omega^2 V_0) \quad \text{vs.} \quad H_1': y \sim t_v(C\delta, \omega^2 V_1),$$

where $v > 0$, $\omega > 0$ and $\delta \in \mathbb{R}^p$ are unknown and $V_0 = I + X(X'X)^{-1}X'$, $V_1 = I + Z(Z'Z)^{-1}Z'$.

Similarly, in the case $p = 0$ the reduced problem becomes

$$(8.3) \qquad H_0': y \sim t_v(0, \omega^2 V_0) \quad \text{vs.} \quad H_1': y \sim t_v(0, \omega^2 V_1).$$

Since (8.2) or (8.3) is equivalent to (8.1), it is natural to reject H_0 if and only if H_0' is rejected.

That is, we derive the best test for (8.2) or (8.3) and use this test for problem (8.1).
With the best test we mean the UMP invariant level α test for (8.2) or (8.3), where α is some preassigned significance level.
It follows that in the case $p > 0$ we reject H_0 if

$$(8.4) \qquad S = \frac{\hat{u}_C'\hat{u}_C + \hat{u}_Z'\hat{u}_Z}{\hat{u}_C'\hat{u}_C + \hat{u}_X'\hat{u}_X} < c,$$

where $\hat{u}_C$, $\hat{u}_X$ and $\hat{u}_Z$ are the residual vectors after least-squares regression of y on C, X and Z, respectively, and where the critical value c has to be chosen in such a way that

$$(8.5) \qquad P(S < c;\ H_0') = \alpha.$$

It should be noted that the probability distribution of S under H_0' does not depend on ν, ω and δ. In the case $p = 0$ we reject H_1 if

$$(8.6) \qquad S = \frac{y'y + \hat{u}_Z'\hat{u}_Z}{y'y + \hat{u}_X'\hat{u}_X} < c.$$

Again c is chosen in accordance with (8.5).
Since the computation of the distribution function $F_0(s)$ of S under H_0' requires numerical integration, it is more easy to use the p-value $F_0(S)$ of the test instead of the level α critical value c.
In terms of $F_0(S)$ we get the following decision rule, which is equivalent to $S < c$:

(8.7) Reject H_0 if $F_0(S) < \alpha$,

where S is as defined in (8.4) when $p > 0$, or (8.6) if $p = 0$ and where

$$(8.8) \qquad F_0(S) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin \varepsilon(u;\ S)}{u\ \gamma(u;\ S)}du.$$

For large n we can use the approximation

$$(8.9) \qquad F_0(S) \approx N\left(\frac{\sqrt{n-p}\,(S - a_n)}{\sqrt{2[(S - a_n)^2 + b_n]}}\right).$$

where $N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt.$

The functions $\gamma(u; s)$, $\varepsilon(u; s)$ in (8.8) and a_n, b_n in (8.9) are known. They depend on the matrices X and Z through the nonzero eigenvalues and corresponding multiplicities of the n×n matrix P defined by

$$(8.10) \qquad \begin{aligned} P &= I + X(X'X)^{-1}X' - \tfrac{1}{2}Z(Z'Z)^{-1}Z' - \tfrac{1}{2}Z(Z'Z)^{-1}Z'X(X'X)^{-1}X' \\ &\quad - C(C'C)^{-1}C', \text{ if } p > 0 \\ P &= I + X(X'X)^{-1}X' - \tfrac{1}{2}Z(Z'Z)^{-1}Z' - \tfrac{1}{2}Z(Z'Z)^{-1}Z'X(X'X)^{-1}X', \\ &\quad \text{if } p = 0. \end{aligned}$$

It should be emphasized that the eigenvalues of P can be found without first computing P.

As is shown in Appendix C and summarized in Table 1 of Section 5 the eigenvalues of P can be found from the eigenvalues of the k×k matrix $(X'X)^{-1}X'Z(Z'Z)^{-1}Z'X$ or, equivalently, from the eigenvalues of the $\ell\times\ell$ matrix $(Z'Z)^{-1}Z'X(X'X)^{-1}X'Z$.

In the special case of testing (8.1) when $\mathcal{M}(X) \subset \mathcal{M}(Z)$ and $\mathcal{M}(X) \neq \mathcal{M}(Z)$, the test (8.7) turns out to be equivalent to the classical F test with level α applied to this situation. In other words, our general approach reproduces the F test in the case of testing nested linear hypotheses.

An interesting subcase of the above situation occurs if we want to test:

$$(8.11) \qquad H_0\colon y \sim n(0, \sigma^2 I) \quad \text{vs.} \quad H_1\colon y \sim n(Z\gamma, \sigma^2 I).$$

The data set for this problem becomes (y, Z) (the matrix X vanishes) and if we set $\hat{u}_X = y$ in (8.6), i.e., if we take

$$(8.12) \qquad S = \frac{y'y + \hat{u}_Z'\hat{u}_Z}{2y'y},$$

again the test (8.7) can be applied and turns out to be equivalent to the F test for problem (8.11).

Our general approach can also be applied to a number of trivial cases which occur if $p = k = \ell$, $n-p \leq 1$, $k = n$ or $\ell = n$, respectively. In these cases we always get $S = c$ with $c = \frac{1}{2}$, 1, 2 and $P(S = c) = 1$, or S is not defined which happens in the case $n = p$. In all these situations the "best" test procedure is to reject H_0 with probability α regardless of the observations. That is, in the trivial cases the sample y is of no use in testing (8.1).

As was shown in Section 7, the above test uniformly maximizes the average (expected) power among all invariant tests whose average (expected) probability of a type I error does not exceed α.

In order to apply the test in practice a number of computations have to be carried out. We shall now describe how these required computations can be made from the given data set (y, X, Z).

We first assume that $k > 0$ and $\ell > 0$.

(i) Compute $A = (X'X)^{-1}X'Z$ and $B = (Z'Z)^{-1}Z'X$.
If $k < \ell$, compute AB and the eigenvalues of this matrix. When $\ell < k$, compute BA and the eigenvalues of BA. In the case $\ell = k$ it does not matter which of AB or BA is computed.
Suppose that AB (or BA) has an eigenvalue 1 with multiplicity $p \geq 0$ and $R \geq 0$ different eigenvalues $\rho_1, \rho_2, \ldots, \rho_R$ with $0 < \rho_j < 1$ and multiplicities $r_1, r_2, \ldots, r_R$, where $r_j > 0$.
Then it follows that $\dim(M(X) \cap M(Z)) = p$ and $r = \mathrm{rank}(X'Z) = p + \sum_{j=1}^{R} r_j$.

(ii) If $p > 0$, the matrices X and Z usually have precisely p common columnvectors. Then the matrix C is defined as the $n\times p$ matrix formed by these common vectors.
It may occasionally happen that the number of common columnvectors is smaller than p. In that case the matrix C can be computed as outlined in Section 3, see (3.26).

(iii) Compute $\hat{\beta} = (X'X)^{-1}X'y$, $\hat{\gamma} = (Z'Z)^{-1}Z'y$, $\hat{u}_X = y - X\hat{\beta}$, $\hat{u}_Z = y - Z\hat{\gamma}$ and if $p > 0$ also compute $\hat{\delta} = (C'C)^{-1}C'y$ and $\hat{u}_C = y - C\hat{\delta}$.

(iv) From (iii) we compute the test statistic:

$$S = \frac{\hat{u}_C'\hat{u}_C + \hat{u}_Z'\hat{u}_Z}{\hat{u}_C'\hat{u}_C + \hat{u}_X'\hat{u}_X} \text{ if } p > 0$$

$$S = \frac{y'y + \hat{u}_Z'\hat{u}_Z}{y'y + \hat{u}_X'\hat{u}_X} \quad \text{if } p = 0.$$

In the trivial cases $p = k = \ell$, $n-p = 1$, $k = n$ or $\ell = n$ we always get $S = c$, where $c = \frac{1}{2}$, 1 or 2 and $P(S = c) = 1$. The "best" procedure is to reject H_0 with probability α through an auxiliary random experiment.

(v) In the nontrivial cases, since we know n, k, ℓ and from (i), p, R, $\rho_1, \rho_2, \ldots, \rho_R$, $r_1, r_2, \ldots, r_R$ and r, we compute the eigenvalues $\tau_1, \tau_2, \ldots, \tau_M$ and multiplicities $m_1, m_2, \ldots, m_M$ as indicated in Table 1 of Section 5.

(vi) Compute the p-value of the test

$$F_0(S) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \varepsilon(u;\, S)}{u\, \gamma(u;\, S)} du,$$

through numerical integration.

The functions $\gamma(u;\, s)$ and $\varepsilon(u;\, s)$ are given by

$$\gamma(u;\, s) = \prod_{j=1}^{M} (1 + (\tau_j - s)^2 u^2)^{\frac{m_j}{4}}$$

$$\varepsilon(u;\, s) = \sum_{j=1}^{M} \frac{m_j}{2} \operatorname{arctg}((\tau_j - s)u),$$

where

$$\left.\frac{\sin \varepsilon(u;s)}{u\, \gamma(u;\, s)}\right|_{u=0} = \tfrac{1}{2} \sum_{j=1}^{M} \tau_j m_j - \tfrac{1}{2}(n-p)s.$$

(vii) Reject H_0 if $F_0(S) < \alpha$.

When $k = 0$ and $\ell > 0$, instead of (i) we get $p = k = 0$, $R = 0$ and $r = 0$. Step (iii) becomes: Compute $\hat{\gamma} = (Z'Z)^{-1}Z'y$, $\hat{u}_X = y$ and $\hat{u}_Z = y - Z\hat{\gamma}$. The remaining computations are unchanged. Similarly, if $k > 0$ and $\ell = 0$, step (i) becomes $p = \ell = 0$, $R = 0$ and $r = 0$, while under (iii) we compute $\hat{\beta} = (X'X)^{-1}X'y$, $\hat{u}_X = y - X\hat{\beta}$ and $\hat{u}_Z = y$. Again the other computations remain unchanged.

For large n the p-value $F_0(S)$ in (vi) can be approximated by

$$F_0(S) \approx N\left(\frac{\sqrt{n-p}\,(S - a_n)}{\sqrt{2[(S - a_n)^2 + b_n]}}\right),$$

where $N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$ and where a_n and b_n are given by

$$a_n = \frac{1}{n-p} \sum_{j=1}^{M} m_j \tau_j = \frac{1}{n-p} (n+k - \tfrac{1}{2}\ell - p - \tfrac{1}{2}\mathrm{tr}(AB))$$

$$b_n = \frac{1}{n-p} \sum_{j=1}^{M} m_j \tau_j^2 - a_n^2 = \frac{1}{n-p} (n+3k - \tfrac{3}{4}\ell - p - \tfrac{5}{2}\mathrm{tr}(AB) + \tfrac{1}{4}\mathrm{tr}[(AB)^2]) - a_n^2$$

References

[1] Bickel, P.J., and K.A. Doksum, "Mathematical Statistics, Basic Indeas and Selected Topics", Holden-Day, San Francisco (1977).

[2] Bouman, J., "Testing Nonnested Linear Hypotheses I: Reduction by Invariance Considerations", Report 8403/S of the Econometric Institute (1984), Erasmus University Rotterdam.

[3] Cramér, H., "Mathematical Methods of Statistics", Princeton University Press, Princeton (1971).

[4] Durbin, J., and G.S. Watson, "Testing for Serial Correlation in Least-Squares Regression III", Biometrika, 58 (1971), 1-19.

[5] Gaver, K.M., and M.S. Geisel, "Discriminating among Alternative Models: Bayesian and Non-Bayesian Methods", in Frontiers in Econometrics, ed. by P. Zarembka, New York: Academic Press (1974), 49-77.

[6] Imhof, P.J., "Computing the Distribution of Quadratic Forms in Normal Variables", Biometrika, 48 (1961), 419-426.

[7] Judge, G.G., W.E. Griffiths, R. Carter Hill, and Tsoung-Chao Lee, "The Theory and Practice of Econometrics", Wiley, New York (1980).

[8] Koerts, J. and A.P.J. Abrahamse, "On the Theory and Application of the General Linear Model", Rotterdam University Press, Rotterdam (1969).

[9] Kullback, S., "Information Theory and Statistics", Wiley, New York (1959).

[10] Lehmann, E.L., "Testing Statistical Hypotheses", Wiley New York (1959).

[11] Raiffa, H., and R. Schlaifer, "Applied Statistical Decision Theory", Boston (1961).

[12] Zellner, A., "An Introduction to Bayesian Inference in Econometrics", Wiley, New York (1971).

Appendix A

Uniformly most powerful invariant tests for a certain class of problems

In this appendix we shall derive a uniformly most powerful (UMP) invariant test for a certain class of testing problems.
The problems can be formulated as follows.
Let F be the class of even probability density functions on $\mathbb{R}^n$. That is, any probability density function (p.d.f.) $f \in F$ satisfies:

(i) $f(x) \geq 0$, $x = (x_1\ x_2\ \dots\ x_n)' \in \mathbb{R}^n$

(ii) $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x)dx_1 \dots dx_n = 1$

(iii) $f(-x) = f(x)$, $x \in \mathbb{R}^n$.

Some well-known examples are:

- $f(x) = (2\pi)^{-\frac{n}{2}}[\det(V)]^{-\frac{1}{2}}\exp\{-\frac{1}{2}x'V^{-1}x\}$,

  where V is a symmetric, positive definite n×n matrix (a multivariate normal distribution).

- $$f(x) = \frac{\Gamma(\frac{n+v}{2})\ v^{\frac{v}{2}}}{\pi^{\frac{n}{2}}\ \Gamma(\frac{v}{2})\ \sqrt{\det(V)}}(v + x'V^{-1}x)^{-\frac{n+v}{2}},$$

  where V is a symmetric, positive definite n×n matrix and $v > 0$ (a multivariate t distribution).

- $f(x) = 2^{-n}\exp\{-\sum_{i=1}^{n}|x_i|\}$.

Let $y = (y_1\ y_2\ \dots\ y_n)' \in \mathbb{R}^n$ be a vector of (observable) random variables with p.d.f.

$$\sigma^{-n}f(\frac{y - C\delta}{\sigma}).$$

where $\delta \in \mathbb{R}^p$, $\sigma > 0$, $f \in F$ and where C is a given nonstochastic n×p matrix with rank p.
In the case p = 0, the p.d.f. of y takes the form

$$\sigma^{-n} f(\frac{y}{\sigma}), \ \sigma > 0, \ f \in F.$$

Suppose that f_0, $f_1 \in F$ are given functions and that we want to test the hypotheses

(A.1) $\quad H_0\colon f = f_0$ against $H_1\colon f = f_1$

on the basis of the observation y.
Since (f_0, f_1) may be any pair of functions from F, we have a whole class of problems of the type (A.1). In problem (A.1) the parameters $\delta \in \mathbb{R}^p$ and $\sigma > 0$ are considered as unknown, which means that the hypotheses H_0 and H_1 are composite. As is typically the case in such situations, no UMP test exists. The usual approach is to restrict attention to a certain subclass of tests and then to try to find the UMP test within this restricted class. When a problem exhibits certain symmetry or invariance properties, it seems natural to restrict attention to those tests which satisfy the same invariance properties. Within this restricted class of invariant tests it is often possible to derive a UMP test.
It can easily be verified that problem (A.1) remains invariant under the following group G' of transformations g':

(A.2) $\quad G'\colon g'(y) = ay + C\alpha$

for all $a \in \mathbb{R}^1$, $a \neq 0$ and all $\alpha \in \mathbb{R}^p$.
When p = 0, we have: $g'(y) = ay$ for all $a \in \mathbb{R}^1$, $a \neq 0$.
The class of invariant tests then consists of all tests (critical functions) ϕ which satisfy

(A.3) $\quad \phi(g'(y)) = \phi(y)$ for all $g' \in G'$.

In this appendix we shall show that the test

(A.4)

$$\phi(y) = 1 \text{ if } \frac{\int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \sigma^{-(n+1)} f_1(\frac{y-C\delta}{\sigma}) d\delta_1 \cdots d\delta_p d\sigma}{\int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \sigma^{-(n+1)} f_0(\frac{y-C\delta}{\sigma}) d\delta_1 \cdots d\delta_p d\sigma} > c$$

$$= \gamma \text{ if } \qquad " \qquad = c$$

$$= 0 \text{ if } \qquad " \qquad < c$$

is UMP invariant for testing problem (A.1).

If $p = 0$, we get:

(A.5)

$$\phi(y) = 1 \text{ if } \frac{\int_0^\infty \sigma^{-(n+1)} f_1(\frac{y}{\sigma}) d\sigma}{\int_0^\infty \sigma^{-(n+1)} f_0(\frac{y}{\sigma}) d\sigma} > c$$

$$= \gamma \text{ if } \qquad " \qquad = c$$

$$= 0 \text{ if } \qquad " \qquad < c.$$

Problems of the type (A.1) are considered in Lehmann [10], Ch. 6, par. 6.3, pp. 218, 219 and 248, 249 (Ex. 5).
In order to derive the UMP invariant test it is often more easy to transform the original problem, through a suitable (1-1) transformation $w = h(y)$, into an equivalent problem with a more simple structure. After applying invariance considerations to this new problem the resulting UMP invariant test can then be expressed in terms of y through the substitution of $w = h(y)$. In our case we consider the linear transformation

(A.6) $\qquad w = R'y,$

where the orthogonal $n \times n$ matrix R is defined as follows

(A.7) $\qquad R = [R_1 \vdots R_2],$

where the columns of the n×p matrix R_1 form an orthonormal basis for the p-dimensional linear subspace $M(C)$ spanned by the columns of C. Consequently, the columns of the n×(n-p) matrix R_2 are an orthonormal basis for $M(C)^{\perp}$.

Since $y = Rw$ is the inverse transformation and $|\det(R)| = 1$, the p.d.f. of w becomes:

$$\sigma^{-n} f\left(\frac{Rw - C\delta}{\sigma}\right) = \sigma^{-n} f\left(\frac{R(w - R'C\delta)}{\sigma}\right).$$

By construction we have $R_2'C = 0$ and this gives

$$R'C\delta = \begin{bmatrix} R_1'C\delta \\ \cdots \\ R_2'C\delta \end{bmatrix} = \begin{bmatrix} R_1'C\delta \\ \cdots \\ 0 \end{bmatrix}.$$

Hence, after the reparameterization $\theta = R_1'C\delta$ and by making use of $w' = (w_1' \;\; w_2')$, where $w_i = R_i'y$, $i = 1, 2$, we get the following p.d.f. $h(w;\, \theta, \sigma)$ of w:

$$\text{(A.8)} \qquad h(w;\, \theta, \sigma) = \sigma^{-n} f\left(R_1\left(\frac{w_1 - \theta}{\sigma}\right) + R_2\left(\frac{w_2}{\sigma}\right)\right),$$

$\theta \in \mathbb{R}^p$, $\sigma > 0$ and $f \in F$.

After the transformation $w = R'y$ the problem becomes:

> On the basis of the vector of observations w with p.d.f. (A.8) we want to test $H_0\colon f = f_0$ vs. $H_1\colon f = f_1$.

In this problem $\theta \in \mathbb{R}^p$ and $\sigma > 0$ are unknown nuisance parameters. The above problem is invariant with respect to the group of transformation G generated by the following 2 subgroups:

$$G_1\colon \quad g_1(w) = \begin{bmatrix} w_1 + b \\ \cdots \\ w_2 \end{bmatrix} \text{ for all } b \in \mathbb{R}^p.$$

$$G_2\colon \quad g_2(w) = cw,\; c \in \mathbb{R}^1,\; c \neq 0.$$

This can be seen as follows. In the first place consider G_1.

Let $z = g_1(w)$ for $b \in \mathbb{R}^p$. Then the inverse transformation becomes

$$w = g_1^{-1}(z) = \begin{bmatrix} z_1 - b \\ \cdots \\ z_2 \end{bmatrix},$$

which shows that the p.d.f. of z takes the form

$$h(g_1^{-1}(z);\ \theta,\ \sigma) = \sigma^{-n} f(R_1(\frac{z_1 - b - \theta}{\sigma}) + R_2(\frac{z_2}{\sigma}))$$

$$= \sigma^{-n} f(R_1(\frac{z_1 - (\theta+b)}{\sigma}) + R_2(\frac{z_2}{\sigma})) = h(z;\ \theta^*,\ \sigma)$$

with $\theta^* = \theta + b \in \mathbb{R}^p$ and $\sigma > 0$.
Hence, the problem remains unchanged after the transformations $z = g_1(w)$, $g_1 \in G_1$.
In the second place, consider G_2. If $z = g_2(w)$ for $c \neq 0$, the inverse transformation is

$$w = g_2^{-1}(z) = \frac{z}{c},$$

and the absolute value of the Jacobian equals $|c|^{-n}$.
The p.d.f. of z becomes

$$|c|^{-n} h(g_2^{-1}(z);\ \theta,\ \sigma) = (|c|\sigma)^{-n} f(R_1(\frac{z_1/c - \theta}{\sigma}) + R_2(\frac{z_2}{c\sigma}))$$

$$= (|c|\sigma)^{-n} f(R_1(\frac{z_1 - c\theta}{|c|\sigma}) + R_2(\frac{z_2}{|c|\sigma})) = h(z;\ \theta^*,\ \sigma^*).$$

where $\theta^* = c\,\theta \in \mathbb{R}^p$ and $\sigma^* = |c|\sigma > 0$.
Here use has been made of $f(-x) = f(x)$. The above argument shows that the transformations $z = g_2(w)$, $g_2 \in G_2$ leave the problem unchanged.

Out of all tests $\phi(w)$ we now restrict attention to the invariant tests, i.e., tests which satisfy

(A.9) $\quad \phi(g(w)) = \phi(w)$ for all $g \in G$,

where G is the group generated by G_1 and G_2.

Since a function is invariant with respect to G if and only if it is a function of a maximal invariant (see Lehmann [10], Ch. 6, pp. 215-218), the totality of invariant tests can be characterized by a maximal invariant statistic with respect to G.

A statistic t(w) is maximal invariant with respect to G if and only if

(i) $t(g(w)) = t(w)$ for all $g \in G$

(ii) $t(w^*) = t(w)$ implies that $w_* = g(w)$ for some $g \in G$.

As is shown by Lehmann [10], Ch. 6, p. 218, Theorem 2, a maximal invariant can be derived in steps corresponding to the subgroups G_1 and G_2 of G.

In our case it is easily verified that the function $t_1(w) = w_2$ is maximal invariant with respect to G_1. In the space of w_2 the group G_2 induces the following group G_2^* of transformations g_2^*:

$$G_2^*:\ g_2^*(w_2) = cw_2 \text{ for all } c \in \mathbb{R}^1,\ c \neq 0.$$

If the elements of the (n-p)-vector w_2 are denoted by V_i, i = 1, 2, ..., n-p, it is again easily verified that the function

$$t_2^*(w_2) = (\frac{V_1}{V_{n-p}}, \frac{V_2}{V_{n-p}}, \ldots, \frac{V_{n-p-1}}{V_{n-p}})$$

is maximal invariant with respect to G_2^*, where it is assumed that $n-p > 1$. When $n-p \leq 1$, there are no maximal invariants and the only invariant functions are the constant functions.

Since this stepwise procedure yields a maximal invariant with respect to G, it follows that

$$\text{(A.10)} \qquad T = t(w) = (T_1, T_2, \ldots, T_{n-p-1})$$

$$= (\frac{V_1}{V_{n-p}}, \frac{V_2}{V_{n-p}}, \ldots, \frac{V_{n-p-1}}{V_{n-p}})$$

is a maximal invariant statistic with respect to G.

As we saw above a test $\phi(w)$ is invariant under G if and only if it is a function of T. This means that the class of invariant tests is precisely equal to the class of tests based on T as defined in (A.10) (i.e., we only consider tests or critical functions $\psi(T)$).
In other words, invariance considerations reduce the sample w to a maximal invariant statistic T (a reduction form $\mathbb{R}^n$ to $\mathbb{R}^{n-p-1}$).
Now, as is typically the case, invariance considerations not only reduce the sample space but also the parameter space. In general, a group of transformation in the sample space induces a group of transformations in the parameter space and it can be shown that the probability distribution of a maximal invariant statistic only depends on the parameters through a corresponding maximal invariant function in the parameter space (see Lehmann [10], Ch. 6, p. 220, Theorem 3).
The parameter space in our problem consists of all points (θ, σ, f) with $\theta \in \mathbb{R}^1$, $\sigma > 0$ and $f \in \{f_0, f_1\}$.
It is not difficult to see that the group G of transformations g induces a group $\bar{G}$ of transformations $\bar{g}$ in the parameter space which is generated by the following subgroups (corresponding to G_1 and G_2, respectively):

$$\bar{G}_1: \bar{g}_1(\theta, \sigma, f) = (\theta+b, \sigma, f) \text{ for all } b \in \mathbb{R}^p$$

$$\bar{G}_2: \bar{g}_2(\theta, \sigma, f) = (c\theta, |c|\sigma, f) \text{ for all } c \in \mathbb{R}^1, c \neq 0.$$

Again it is easily verified that the function

$$\text{(A.11)} \qquad v(\theta, \sigma, f) = f$$

is maximal invariant with respect to the induced group $\bar{G}$.
This shows that the probability distribution of the maximal invariant statistic T only depends on f and no longer on θ and σ. In other words, invariance reduces the sample space to the space of T and the parameter space to $\{f_0, f_1\}$. That is, the original problem is reduced by invariance to a problem of testing two simple hypotheses on the basis of the observation T.
If the p.d.f. of T is denoted by $k_i(t_1, t_2, \ldots, t_{n-p-1})$ when $f = f_i$, $i = 0, 1$, it follows that k_i $(i = 0, 1)$ is completely specified and that invariance reduces the original problem to:

$$H_0\colon T \sim k_0(t_1, t_2, \dots, t_{n-p-1})$$

against

$$H_1\colon T \sim k_1(t_1, t_2, \dots, t_{n-p-1}).$$

By the well-known lemma of Neyman and Pearson the most powerful (MP) test for the latter problem rejects when

$$\text{(A.12)} \qquad \frac{k_1(T_1, T_2, \dots, T_{n-p-1})}{k_0(T_1, T_2, \dots, T_{n-p-1})} > c,$$

and it follows that this test (which does not depend on θ and σ) is UMP invariant for the original problem. It remains to show that the critical region (A.12) of the UMP invariant test can be written in the form (A.4). To this extend we first derive the p.d.f. of the random vector $w_2 = (V_1\ V_2\ \dots\ V_{n-p})'$ for arbitrary $f \in \mathcal{F}$.
Let the function $\ell(x)$ be defined as

$$\text{(A.13)} \qquad \ell(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(R_1 z + R_2 x)dz_1 \dots dz_p,$$

where $x = (x_1\ x_2\ \dots\ x_{n-p})'$ and $z = (z_1\ z_2\ \dots\ z_p)'$.
Then it is easily seen that $\ell(x)$ is a p.d.f. and $\ell(-x) = \ell(x)$. By integrating on the elements of w_1 and using the transformation $z = (w_1 - \theta)/\sigma$ it follows from (A.8) that the p.d.f. of w_2 becomes:

$$\text{(A.14)} \qquad h_2(w_2; \sigma) = \sigma^{-(n-p)}\ell(\frac{w_2}{\sigma}) = \sigma^{-(n-p)}\ell(\frac{v_1}{\sigma}, \frac{v_2}{\sigma}, \dots, \frac{v_{n-p}}{\sigma}),$$

where $w_2 = (v_1\ v_2\ \dots\ v_{n-p})'$.
Note that the p.d.f. of w_2 no longer depends on θ. Next we shall derive the p.d.f. of T by using the transformation

$$t_i = \frac{v_i}{v_{n-p}}, \quad i = 1, 2, \dots, n-p-1$$

$$t_{n-p} = v_{n-p}.$$

The inverse transformation is

$$v_i = t_i t_{n-p},\ i = 1, 2, \ldots, n-p-1$$

$$v_{n-p} = t_{n-p},$$

with absolute value of the Jacobian $|t_{n-p}|^{n-p-1}$. Then it follows from (A.14) that the p.d.f. of $T_1, T_2, \ldots, T_{n-p-1}, T_{n-p}$, where $T_{n-p} = V_{n-p}$, has the form:

$$|t_{n-p}|^{n-p-1}\sigma^{-(n-p)}\ell\left(\frac{t_1 t_{n-p}}{\sigma}, \frac{t_2 t_{n-p}}{\sigma}, \ldots, \frac{t_{n-p-1} t_{n-p}}{\sigma}, \frac{t_{n-p}}{\sigma}\right).$$

After integration on the variable t_{n-p}, making use of the substitution $\eta = t_{n-p}/\sigma$ and the fact that $\ell(-x) = \ell(x)$, we obtain the p.d.f. of the maximal invariant statistic T for arbitrary $f \in \mathcal{F}$:

$$\text{(A.15)} \qquad k(t_1, t_2, \ldots, t_{n-p-1}) = 2\int_0^\infty \eta^{n-p-1}\ell(\eta t_1, \ldots, \eta t_{n-p-1}, \eta)d\eta.$$

Note that this p.d.f. only depends on f (through ℓ) and not on θ and σ. If ℓ_i $(i = 0, 1)$ is defined by (A.13) with f replaced by f_i $(i = 0, 1)$, it is seen from (A.12) that the UMP invariant test takes the form

$$\text{(A.16)} \qquad \psi(T) = \begin{cases} 1 \text{ if } \dfrac{\int_0^\infty \eta^{n-p-1}\ell_1(\eta T_1, \ldots, \eta T_{n-p-1}, \eta)d\eta}{\int_0^\infty \eta^{n-p-1}\ell_0(\eta T_1, \ldots, \eta T_{n-p-1}, \eta)d\eta} > c \\ \gamma \text{ if } \quad '' \quad = c \\ 0 \text{ if } \quad '' \quad < c. \end{cases}$$

By using $T = t(w)$ (see (A.10)) and defining $\phi(w) = \psi(t(w))$, it follows that $\phi(w)$ is the UMP invariant test in terms of w, i.e.,:

$$\text{(A.17)} \qquad \phi(w) = \begin{cases} 1 \text{ if } \dfrac{\int_0^\infty \eta^{n-p-1}\ell_1\left(\frac{\eta}{V_{n-p}} w_2\right)d\eta}{\int_0^\infty \eta^{n-p-1}\ell_0\left(\frac{\eta}{V_{n-p}} w_2\right)d\eta} > c \\ \gamma \text{ if } \quad '' \quad = c \\ 0 \text{ if } \quad '' \quad < c \end{cases}$$

where $w_2' = (V_1\ V_2\ \dots\ V_{n-p})$.

When $n-p \leq 1$, the test $\phi(w) \equiv \alpha$ is UMP invariant (a purely randomized test which rejects with probability α regardless of the observations). The above analysis is carried out under the assumption $p > 0$. If $p = 0$, we can always take $R = I$, that is $w = y$. In that case we have $G = G_2$ and $\ell_i = f_i$ $(i = 0, 1)$, and it follows that the UMP invariant test is of the form (A.16) or (A.17) with p replaced by 0, provided that $n > 1$. In the case $n = 1$, again the test $\phi(w) \equiv \alpha$ is UMP invariant. Finally we shall write the test (A.17) in terms of the original observations y. Since $w = R'y$, it follows that the test $\phi^*(y) = \phi(R'y)$ is the UMP invariant test in terms of y. In order to express $\phi^*(y)$ in terms of the p.d.f.'s f_0 and f_1 we proceed as follows.
From (A.13) and $f_i(-x) = f_i(x)$ it is seen that

$$\int_0^\infty \eta^{n-p-1} \ell_i\left(\frac{\eta}{V_{n-p}} w_2\right) d\eta =$$

$$\int_0^\infty \int_{-\infty}^\infty \dots \int_{-\infty}^\infty \eta^{n-p-1} f_i\left(R_1 z + \frac{\eta}{|V_{n-p}|} R_2 w_2\right) dz_1 \dots dz_p d\eta.$$

Consider the new integration variables $\delta \in \mathbb{R}^p$ and $\sigma > 0$ which are related to z and η through

$$\text{(A.18)} \qquad z = \frac{1}{\sigma}(-R_1' C\delta + w_1), \qquad \eta = \frac{|V_{n-p}|}{\sigma}.$$

Since by construction $C = R_1 A$ for some nonsingular $p \times p$ matrix A, it follows that the Jacobian of (A.18) is equal to

$$|\det(A)|\ |V_{n-p}|\ \sigma^{-(p+2)}.$$

By making using of $R_1 R_1' C = C$ and $R_1 w_1 + R_2 w_2 = Rw = y$ we get:

$$\text{(A.19)} \qquad \int_0^\infty \eta^{n-p-1} \ell_i\left(\frac{\eta}{V_{n-p}} w_2\right) d\eta =$$

$$|\det(A)|\ |V_{n-p}|^{n-p} \int_0^\infty \int_{-\infty}^\infty \dots \int_{-\infty}^\infty \sigma^{-(n+1)} f_i\left(\frac{y-C\delta}{\sigma}\right) d\delta_1 \dots d\delta_p d\sigma,$$

for $i = 0, 1$.

Substitution of (A.19) into (A.17) yields

(A.20)

$$\phi^*(y) = 1 \text{ if } \frac{\int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \sigma^{-(n+1)} f_1(\frac{y-C\delta}{\sigma}) d\delta_1 \cdots d\delta_p d\sigma}{\int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \sigma^{-(n+1)} f_0(\frac{y-C\delta}{\sigma}) d\delta_1 \cdots d\delta_p d\sigma} > c$$

$$= \gamma \text{ if } \quad " \quad = c$$

$$= 0 \text{ if } \quad " \quad < c.$$

The proof of (A.4) is completed by observing that there exists a 1-1 correspondence between the group G in w-space and the group G' in y-space as defined in (A.2). This fact can easily be verified with the aid of the 1-1 transformation $w = R'y$ with inverse $y = Rw$. In the case $p = 0$ it can be shown in a similar way that the test (A.5) is UMP invariant.

When $n-p \leq 1$, the test $\phi^*(y) \equiv \alpha$ is UMP invariant. In order to obtain the UMP invariant test with level α, we have to choose γ and the critical value c in such a way that the size of the test is equal to a preassigned significance level α. Usually we can take $\gamma = 0$.

In applications it is sometimes more easy to use the form (A.17) instead of (A.4) or (A.5).

An interesting application of the above procedure to the problem of testing for serial correlation in least-squares regression is given by Durbin and Watson [4].

Appendix B

The probability distribution of the test statistic

In this appendix we shall derive the probability distribution of the statistic

$$\text{(B.1)} \qquad S = \frac{x'V_1^{-1}x}{x'V_0^{-1}x},$$

where V_0 and V_1 are given symmetric, positive-definite n×n matrices and where

$$\text{(B.2)} \qquad x \sim n(0, V_0).$$

Since V_0 is symmetric and positive definite, there exists a nonsingular n×n matrix Γ_0 such that $V_0 = \Gamma_0\Gamma_0'$.

Let z be defined by $z = \Gamma_0^{-1}x$, then we have

$$\text{(B.3)} \qquad z \sim n(0, I)$$

and

$$\text{(B.4)} \qquad S = \frac{z'\Gamma_0'V_1^{-1}\Gamma_0 z}{z'z}.$$

Let $\lambda_i > 0$, $i = 1, 2, \ldots, n$ be the eigenvalues of $\Gamma_0'V_1^{-1}\Gamma_0$, Λ the diagonal matrix with main-diagonal elements λ_i and H the corresponding orthogonal matrix of eigenvectors, then S can be rewritten as follows:

$$\text{(B.5)} \qquad S = \frac{\xi'\Lambda\xi}{\xi'\xi},$$

where $\xi = H'z \sim n(0, I)$.
Hence, if $\xi' = (\xi_1\ \xi_2\ \ldots\ \xi_n)$, we have

$$\text{(B.6)} \qquad S = \frac{\sum_{i=1}^{n} \lambda_i \xi_i^2}{\sum_{i=1}^{n} \xi_i^2},$$

where $\xi_1^2, \xi_2^2, \ldots, \xi_n^2$ are mutually stochastically independent and $\xi_i^2 \sim \chi^2(1)$ for $i = 1, 2, \ldots, n$.

As is typically the case, not all eigenvalues λ_i are different. Suppose that $\Gamma_0' V_1^{-1} \Gamma_0$ has M different eigenvalues τ_j with multiplicities m_j, $j = 1, 2, \ldots, M$, where $\sum_{j=1}^{M} m_j = n$.

If we define

$$\eta_j = \sum_{i=m_{j-1}+1}^{m_{j-1}+m_j} \xi_i^2, \; j = 1, 2, \ldots, M \text{ and } m_0 = 0,$$

it is seen from (B.6) that

$$\text{(B.7)} \qquad S = \frac{\sum_{j=1}^{M} \tau_j \eta_j}{\sum_{j=1}^{M} \eta_j},$$

where $\eta_1, \eta_2, \ldots, \eta_M$ are mutually stochastically independent and $\eta_j \sim \chi^2(m_j)$.

In order to find the distribution function

$$\text{(B.8)} \qquad F(s) = P(S \leq s)$$

of S, we introduce the auxiliarly random variable

$$\text{(B.9)} \qquad Q(s) = \sum_{j=1}^{M} (\tau_j - s)\eta_j.$$

Since the event

$$\frac{\sum_{j=1}^{M} \tau_j \eta_j}{\sum_{j=1}^{M} \eta_j} \leq s$$

is equivalent to

$$\sum_{j=1}^{M} (\tau_j - s)\eta_j \leq 0,$$

it follows from (B.8) and (B.9) that

(B.10) $$F(s) = P(Q(s) \leq 0).$$

The characteristic function $\phi(t;\, s) = E(e^{itQ(s)})$ (where i denotes the imaginary unit) of Q(s) can easily be found from (B.9) by making use of the fact that the η_j's are mutually independent and $\eta_j \sim \chi^2(m_j)$.
We get:

(B.11) $$\phi(t;\, s) = \prod_{j=1}^{M} (1 - 2i(\tau_j - s)t)^{-\frac{m_j}{2}}.$$

For random variables of the type Q(s) the inversion formula for characteristic functions can be written in the following form, provided that M > 1:

(B.12) $$P(Q(s) \leq x) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{1}{t}\, |\phi(t;\, s)| \sin\{\arg(\phi(t;\, s)) - tx\}dt,$$

where $x \in \mathbb{R}^1$ and $|\phi(t;\, s)|$ and $\arg(\phi(t;\, s))$ denote the modulus and argument of the complex-valued function $\phi(t;\, s)$, respectively.
From (B.11) it can be deduced that:

(B.13) $$|\phi(t;\, s)| = \prod_{j=1}^{M} (1 + 4(\tau_j - s)^2 t^2)^{-\frac{m_j}{4}}$$

and

(B.14) $$\arg(\phi(t;\, s)) = \sum_{j=1}^{M} \frac{m_j}{2} \operatorname{arctg}[2(\tau_j - s)t].$$

The above results yield the following expression for F(s):

(B.15) $$F(s) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin \varepsilon(u;\, s)}{u\, \gamma(u;s)} du,$$

where

(B.16) $$\gamma(u;\ s) = \prod_{j=1}^{M} (1 + (\tau_j - s)^2 u^2)^{\frac{m_j}{4}}$$

and

(B.17) $$\varepsilon(u;\ s) = \sum_{j=1}^{M} \frac{m_j}{2} \operatorname{arctg}[(\tau_j - s)u].$$

The value of the integrand in (B.15) at u = 0 is defined by

(B.18) $$\left.\frac{\sin \varepsilon(u;\ s)}{u\ \gamma(u;\ s)}\right|_{u=0} = \lim_{u \downarrow 0} \frac{\sin \varepsilon(u;\ s)}{u\ \gamma(u;\ s)}$$

$$= \frac{1}{2} \sum_{j=1}^{M} (\tau_j - s)m_j = \frac{1}{2} \sum_{j=1}^{M} \tau_j m_j - \frac{1}{2} ns.$$

Note that $u\ \gamma(u;\ s)$ is a monotone increasing function of u.
The formula (B.15) shows that we have determined F(s) up to an integral. That is, for any value of s we can compute F(s) through numerical integration. This method of computing F(s) is known as Imhof's method and there exists several computer programs for the numerical integration of (B.15), see for instance Imhof [6] and Koerts and Abrahamse [8].
The process of calculating (or approximating) F(s) consists of two parts.

(i) The improper integral $\int_0^\infty \frac{\sin \varepsilon(u;\ s)}{u\ \gamma(u;\ s)} du$ is approximated by the proper integral $\int_0^U \frac{\sin \varepsilon(u;\ s)}{u\ \gamma(u;\ s)} du$.

(ii) The integral $\int_0^U \frac{\sin \varepsilon(u;\ s)}{u\ \gamma(u;\ s)} du$ is approximated by using (the compound) Simpson's rule.

The errors arising from (i) and (ii) can be made arbitrarily small.
As far as the truncation error from (i) is concerned this can be seen from the fact that for arbitrary A we have:

$$\left| \int_U^\infty \frac{\sin \varepsilon(u;\ s)}{u\ \gamma(u;\ s)} du \right| \leq A$$

for

$$U = [\frac{n}{2} \cdot A \prod_{j=1}^{M} |\tau_j - s|^{\frac{m_j}{2}}]^{-\frac{2}{n}}, \quad s \neq \tau_j.$$

When $s = \tau_i$ for some i, we get $U = [\frac{n}{2} \cdot A \prod_{\substack{j=1 \\ j \neq i}}^{M} |\tau_j - s|^{\frac{m_j}{2}}]^{-\frac{2}{n}}$.

Hence, F(s) can be approximated to any desired degree of accuracy.
In the above derivation the eigenvalues τ_j and corresponding multiplicities m_j are supposed to be given.
For the computation of the eigenvalues of the matrix $\Gamma_0' V_1^{-1} \Gamma_0$ it is not necessary to find a matrix Γ_0 which satisfies $V_0 = \Gamma_0 \Gamma_0'$. This follows from the fact that the matrices $\Gamma_0' V_1^{-1} \Gamma_0$ and $V_1^{-1} V_0$ have the same eigenvalues, as can easily be verified.
Finally consider the probability distribution of $S = \frac{x' V_1^{-1} x}{x' V_0^{-1} x}$ under the assumption that $x \sim n(0, V_1)$.
If $F_*(s) = P(S \leq s)$ when $x \sim n(0, V_1)$, a similar analysis shows that:

$$\text{(B.19)} \qquad F_*(s) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin \varepsilon_*(u; \frac{1}{s})}{u\, \gamma_*(u; \frac{1}{s})} du,$$

where

$$\gamma_*(u; s) = \prod_{j=1}^{M} (1 + (\frac{1}{\tau_j} - s)^2 u^2)^{\frac{m_j}{4}}$$

(B.20)

$$\varepsilon_*(u; s) = \sum_{j=1}^{M} \frac{m_j}{2} \operatorname{arctg}[(\frac{1}{\tau_j} - s)u].$$

Here use has been made of $P(S \leq s) = 1 - P(\frac{1}{S} \leq \frac{1}{s})$ and the fact that if τ_j is an eigenvalue of $V_1^{-1} V_0$ with multiplicity m_j, then $\frac{1}{\tau_j}$ is an eigenvalue of $(V_1^{-1} V_0)^{-1} = V_0^{-1} V_1$ with multiplicity m_j.

In the above case we have

$$\left. \frac{\sin \varepsilon_*(u; s)}{u\, \gamma_*(u; s)} \right|_{u=0} = \frac{1}{2} \sum_{j=1}^{M} \frac{m_j}{\tau_j} - \frac{1}{2} ns.$$

Appendix C

The eigenvalues of P

In this appendix X and Z denote given matrices. X is a $n\times k$ matrix with rank k and $X'X = I_{(k)}$. Z is a $n\times \ell$ matrix with rank ℓ and $Z'Z = I_{(\ell)}$. Throughout this appendix, the linear subspace $\mathbb{R}^n$ spanned by the columnvectors of a $n\times m$ matrix A is denoted by $M(A)$.
Let $p = \dim(M(X) \cap M(Z))$ and let the columnvectors of the $n\times p$ matrix C be an orthonormal basis for $M(X) \cap M(Z)$, provided that $p > 0$. Then the $n\times n$ matrix P is defined as

$$\text{(C.1)} \qquad P = I_{(n)} + XX' - \tfrac{1}{2}ZZ' - \tfrac{1}{2}ZZ'XX' - CC'.$$

When $p = 0$, i.e., $M(X) \cap M(Z) = \{0\}$, we define P as

$$\text{(C.2)} \qquad P = I_{(n)} + XX' - \tfrac{1}{2}ZZ' - \tfrac{1}{2}ZZ'XX'.$$

Now we shall show that the eigenvalues (and corresponding multiplicities) of P can be deduced from the eigenvalues of the $k\times k$ matrix $X'ZZ'X$ (or equivalently, from the eigenvalues of the $\ell\times\ell$ matrix $Z'XX'Z$). To this extend we first consider the eigenvalues of the matrices $X'ZZ'X$ and $Z'XX'Z$. It is not difficult to verify that:

(i) If λ is an eigenvalue of $X'ZZ'X$ then $0 \leq \lambda \leq 1$. The same result holds for $Z'XX'Z$.
(ii) $X'ZZ'X$ and $Z'XX'Z$ have the same nonzero eigenvalues.
(iii) If $p = \dim(M(X) \cap M(Z)) > 0$, then $X'ZZ'X$ (and $Z'XX'Z$) has an eigenvalue 1 with multiplicity p and vice versa.

Let ρ_j be an eigenvalue of $X'ZZ'X$ (and $Z'XX'Z$) with $0 < \rho_j < 1$ and multiplicity r_j, $j = 1, 2, \ldots, R$.
When r is defined by $r = p + \sum_{j=1}^{R} r_j$, it can be shown that (see Bouman [2]):

(iv) $\text{rank}(X'Z) = \text{rank}(Z'X) = r$
(v) $X'ZZ'X$ has an eigenvalue 0 with multiplicity $k-r$
$Z'XX'Z$ has an eigenvalue 0 with multiplicity $\ell-r$

(vi) $\dim(M(X) \cap M(Z)^{\perp}) = k-r$, $\dim(M(X)^{\perp} \cap M(Z)) = \ell-r$ and $\dim(M(X)^{\perp} \cap M(Z)^{\perp}) = n+p-k-\ell$, where $M(A)^{\perp}$ denotes the orthogonal complement (with respect to $\mathbb{R}^n$) of $M(A)$.

Further we define the linear subspace V by

$$V = (M(X) \cap M(Z)) \oplus (M(X) \cap M(Z)^{\perp}) \oplus (M(X)^{\perp} \cap M(Z)) \oplus (M(X)^{\perp} \cap M(Z)^{\perp}), \tag{C.3}$$

where the symbol $\oplus$ denotes the direct sum of two linear subspaces. It follows that

$$\begin{aligned} &\dim(V) = n-2(r-p) \\ &\dim(V^{\perp}) = 2(r-p) = 2 \sum_{j=1}^{R} r_j. \end{aligned} \tag{C.4}$$

Let h_j be an eigenvector of $X'ZZ'X$ corresponding to the eigenvalue ρ_j, i.e.,

$$X'ZZ'Xh_j = \rho_j h_j, \; j = 1, 2, \ldots, R. \tag{C.5}$$

Note that for each j there are r_j linear independent vectors h_j. If we define w_j and q_j by

$$\begin{aligned} &w_j = Xh_j \\ &q_j = w_j + a_j ZZ'w_j, \; a_j \in \mathbb{R}^1, \end{aligned} \tag{C.6}$$

it is easily verified that

$$q_j \in V^{\perp}, \text{ for any } a_j \in \mathbb{R}^1 \tag{C.7}$$

and

$$\begin{aligned} &C'q_j = 0 \\ &XX'q_j = (1 + \rho_j a_j)w_j \\ &ZZ'q_j = (1 + a_j)ZZ'w_j, \end{aligned} \tag{C.8}$$

for any $a_j \in \mathbb{R}^1$.
From the above results we can determine the eigenvalues of the matrix P.
First suppose that $p > 0$, then we get:

a) Let $q \in M(X) \cap M(Z)$, then

$$Pq = q + XX'q - \tfrac{1}{2}ZZ'q - \tfrac{1}{2}ZZ'XX'q - CC'q =$$

$$q + q - \tfrac{1}{2}q - \tfrac{1}{2}q - q = 0$$

It follows that 0 is an eigenvalue of P with multiplicity p.

b) Let $q \in M(X) \cap M(Z)^{\perp}$, then

$$Pq = q + XX'q - \tfrac{1}{2}ZZ'q - \tfrac{1}{2}ZZ'XX'q - CC'q =$$

$$q + q - 0 - 0 - 0 = 2q$$

That is, 2 is an eigenvalue of P with multiplicity k-r.

c) Let $q \in M(X)^{\perp} \cap M(Z)$, then

$$Pq = q + XX'q - \tfrac{1}{2}ZZ'q - \tfrac{1}{2}ZZ'XX'q - CC'q =$$

$$q + 0 - \tfrac{1}{2}q - 0 - 0 = \tfrac{1}{2}q$$

Therefore, $\frac{1}{2}$ is an eigenvalue of P with multiplicity ℓ-r.

d) Let $q \in M(X)^{\perp} \cap M(Z)^{\perp}$, then

$$Pq = q + XX'q - \tfrac{1}{2}ZZ'q - \tfrac{1}{2}ZZ'XX'q - CC'q =$$

$$q + 0 - 0 - 0 - 0 = q$$

It follows that 1 is an eigenvalue of P with multiplicity $n+p-k-\ell$.

e) Let q_j be as defined in (C.6), then $q_j \in V^{\perp}$ and we shall show that there exist values of a_j such that q_j is an eigenvector of P. Moreover, we shall determine the corresponding eigenvalue λ_j. In order to obtain the desired result we solve the equation $Pq_j = \lambda_j q_j$ for a_j and λ_j. First we compute the left-hand side of this equation. From (C.8) we get:

$$\begin{aligned} Pq_j &= q_j + XX'q_j - \tfrac{1}{2}ZZ'q_j - \tfrac{1}{2}ZZ'XX'q_j - CC'q_j \\ &= q_j + (1 + \rho_j a_j)w_j - \tfrac{1}{2}(1 + a_j)ZZ'w_j - \tfrac{1}{2}(1 + \rho_j a_j)ZZ'w_j \\ &= (2 + \rho_j a_j)w_j + \tfrac{1}{2}((1 - \rho_j)a_j - 2)ZZ'w_j. \end{aligned}$$

The right-hand side $\lambda_j q_j$ becomes:

$$\lambda_j q_j = \lambda_j w_j + \lambda_j a_j ZZ'w_j$$

This gives:

$$(2 + \rho_j a_j)w_j + \tfrac{1}{2}((1 - \rho_j)a_j - 2)ZZ'w_j = \lambda_j w_j + \lambda_j a_j ZZ'w_j.$$

Hence, we are looking for values of a_j and λ_j which satisfy

$$\text{(C.9)} \quad \begin{aligned} 2 + \rho_j a_j &= \lambda_j \\ \tfrac{1}{2}((1 - \rho_j)a_j - 2) &= \lambda_j a_j. \end{aligned}$$

Solving (C.9) for a_j we get:

$$\tfrac{1}{2}((1 - \rho_j)a_j - 2) = (2 + \rho_j a_j)a_j,$$

which yields:

$$\text{(C.10)} \quad 2\rho_j a_j^2 + (3 + \rho_j)a_j + 2 = 0.$$

The roots of this equation are

$$a_{1j} = \frac{-(3 + \rho_j) - \sqrt{(1 - \rho_j)(9 - \rho_j)}}{4\rho_j}$$

(C.11)

$$a_{2j} = \frac{-(3 + \rho_j) + \sqrt{(1 - \rho_j)(9 - \rho_j)}}{4\rho_j} .$$

Substitution of (C.11) into $\lambda_j = 2 + \rho_j a_j$ gives

$$\lambda_{1j} = \frac{1}{4}(5 - \rho_j) - \frac{1}{4}\sqrt{(1 - \rho_j)(9 - \rho_j)}$$

(C.12)

$$\lambda_{2j} = \frac{1}{4}(5 - \rho_j) + \frac{1}{4}\sqrt{(1 - \rho_j)(9 - \rho_j)}.$$

Note that $\lambda_{1j}\lambda_{2j} = 1$ and let $\frac{1}{2} < \lambda_{1j} < 1$.
It follows that λ_{1j} is an eigenvalue of P with multiplicity r_j and corresponding eigenvectors $q_{1j} = w_j + a_{1j}ZZ'w_j$ for $j = 1, 2, \ldots, R$.
Similarly, λ_{2j} is an eigenvalue of P with multiplicity r_j and corresponding eigenvector $q_j = w_j + a_{2j}ZZ'w_j$, $j = 1, 2, \ldots, R$.

Note that the sum of the multiplicities is equal to:

$$p + (k-r) + (\ell-r) + (n+p-k-\ell) + 2 \sum_{j=1}^{R} r_j = n.$$

In the second place, consider the case $p = 0$, then P is as defined in (C.2). It is easily verified that P has an eigenvalue 2 with multiplicity $(k-r)$, an eigenvalue $\frac{1}{2}$ with multiplicity $\ell-r$, an eigenvalue 1 with multiplicity $n-k-\ell$, an eigenvalue λ_{1j} with multiplicity r_j, $j = 1, 2, \ldots, R$ and an eigenvalue λ_{2j} with multiplicity r_j, $j = 1, 2, \ldots, R$, where λ_{1j} and λ_{2j} are as given in (C.12) and where $r = \sum_{j=1}^{R} r_j$.

The above results make it clear that the eigenvalues and multiplicities of the $n\times n$ matrix P can easily be determined from the eigenvalues (1 and ρ_j) and multiplicities (p and r_j) of the $k\times k$ matrix $X'ZZ'X$ (or equivalently, the $\ell\times\ell$ matrix $Z'XX'Z$), where it should be emphasized that in the applications k and ℓ are often considerably smaller than n.

Appendix D

The distance between two t distributions

In this appendix we shall compute the distance from a $t_v(0, I)$ distribution to a $t_v(\theta, V)$ distribution.
Let $f_0(x)$ be the p.d.f. of a $t_v(0, I)$ distribution and $f_1(x)$ the p.d.f. of a $t_v(\theta, V)$ distribution, i.e.,

$$\text{(D.1)} \qquad f_0(x) = \frac{\Gamma(\frac{n+v}{2})v^{\frac{v}{2}}}{\pi^{n/2}\Gamma(\frac{v}{2})}(v + x'x)^{-\frac{n+v}{2}}$$

and

$$\text{(D.2)} \qquad f_1(x) = \frac{\Gamma(\frac{n+v}{2})v^{\frac{v}{2}}}{\pi^{n/2}\Gamma(\frac{v}{2})\sqrt{\det(V)}}(v + (x-\theta)'V^{-1}(x-\theta))^{-\frac{n+v}{2}},$$

where $x \in \mathbb{R}^n$, $v > 0$, $n = 1, 2, 3, \ldots$, $\theta \in \mathbb{R}^n$ and V is a symmetric, positive-definite matrix of the order n×n.
As a measure of distance we consider

$$\text{(D.3)} \qquad d = E_0\left(\frac{\left[\frac{f_0(X)}{f_1(X)}\right]^{\frac{2}{n+v}} - 1}{\frac{2}{n+v}}\right), \quad n = 1, 2, 3, \ldots, \; v > 0,$$

where X is a n-dimensional random vector and E_0 denotes the expectation taken with respect to the p.d.f. f_0.
We first compute the ratio $f_0(x)/f_1(x)$. From (D.1) and (D.2) we get

$$\text{(D.4)} \qquad \frac{f_0(x)}{f_1(x)} = [\det(V)]^{\frac{1}{2}}\left(\frac{v + (x-\theta)'V^{-1}(x-\theta)}{v + x'x}\right)^{\frac{n+v}{2}}.$$

It follows from (D.3) that

$$\text{(D.5)} \qquad d = \tfrac{1}{2}(n+v)[\det(V)]^{\frac{1}{n+v}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{v + (x-\theta)'V^{-1}(x-\theta)}{v + x'x}\right) f_0(x)dx_1 \ldots dx_n - \frac{n+v}{2}.$$

The ratio $f_0(x)/(v + x'x)$ takes the form

$$\frac{\Gamma(\frac{n+v}{2})v^{\frac{v}{2}}}{\pi^{n/2}\,\Gamma(\frac{v}{2})}(v + x'x)^{-\frac{n+v+2}{2}}$$

and this can be rewrittten in the following way

$$\text{(D.6)}\qquad \frac{f_0(x)}{v + x'x} = \frac{\Gamma(\frac{n+v}{2})v^{\frac{v}{2}}}{\pi^{n/2}\,\Gamma(\frac{v}{2})}\left(\frac{v}{v+2}\right)^{-\frac{n+v+2}{2}}\left[(v+2) + \frac{x'x}{(\frac{v}{v+2})}\right]^{-\frac{n+v+2}{2}} = \frac{1}{n+v}\cdot f_*(x),$$

where $f_*(x)$ is the p.d.f. of a $t_{v+2}(0, \frac{v}{v+2} I)$ distribution, i.e.,

$$\text{(D.7)}\qquad f_*(x) = \frac{\Gamma(\frac{n+v+2}{2})(v+2)^{\frac{v+2}{2}}\left(\frac{v}{v+2}\right)^{-\frac{n}{2}}}{\pi^{n/2}\,\Gamma(\frac{v+2}{2})}\left[(v+2) + \frac{x'x}{(\frac{v}{v+2})}\right]^{-\frac{n+v+2}{2}}.$$

Substitution of (D.6) into (D.5) yields

$$\text{(D.8)}\qquad d = \tfrac{1}{2}[\det(V)]^{\frac{1}{n+v}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(v + (x-\theta)'V^{-1}(x-\theta))f_*(x)dx_1\ldots dx_n - \frac{n+v}{2}$$

$$= \tfrac{1}{2}[\det(V)]^{\frac{1}{n+v}}E_*(v + (X-\theta)'V^{-1}(X-\theta)) - \frac{n+v}{2},$$

where E_* denotes the expectation with respect to the p.d.f. f_* as given in (D.7).

We have

$$\text{(D.9)}\qquad E_*(v + (X-\theta)'V^{-1}(X-\theta)) = v + E_*((X-\theta)'V^{-1}(X-\theta))$$

$$= v + \mathrm{tr}(V^{-1}E_*[(X-\theta)(X-\theta)'])$$

$$= v + \mathrm{tr}(V^{-1}[E_*(XX') - E_*(X)\theta' - \theta(E_*(X))' + \theta\theta']).$$

Since it follows from (D.7) that

(D.10)
$$E_*(X) = 0, \quad v+2 > 1$$
$$E_*(XX') = \frac{v+2}{v}\,\frac{v}{v+2}\, I, \quad v+2 > 2,$$

we get

(D.11) $$E_*(v + (X-\theta)'V^{-1}(X-\theta)) = v + tr(V^{-1}) + \theta'V^{-1}\theta,$$

for all $v > 0$ and $n = 1, 2, 3, \ldots$.

Substitution of (D.11) into (D.8) gives

(D.12) $$d = \tfrac{1}{2}[\det(V)]^{\frac{1}{n+v}}(v + tr(V^{-1}) + \theta'V^{-1}\theta) - \frac{n+v}{2},$$

$v > 0$, $n = 1, 2, 3, \ldots$.
Note that in the special case $V = I$ we get $d = \frac{1}{2}\,\theta'\theta$.

Suppose that we want to compute the distance d from a $t_v(\theta_0, V_0)$ distribution to a $t_v(\theta_1, V_1)$ distribution, where d is as defined in (D.3) with $f_i(x)$ the p.d.f. of a $t_v(\theta_i, V_i)$ distribution, $i = 0, 1$. Consider the random vector X and the transformation

(D.13) $$Z = \Gamma_0^{-1}(X - \theta_0),$$

where Γ_0 is a nonsingular n×n matrix which satisfies

(D.14) $$\Gamma_0\Gamma_0' = V_0.$$

Then it is easily seen that $Z \sim t_v(0, I)$ if $X \sim t_v(\theta_0, V_0)$ and $Z \sim t_v(\theta, V)$ if $X \sim t_v(\theta_1, V_1)$, where

(D.15)
$$\theta = \Gamma_0^{-1}(\theta_1 - \theta_0)$$
$$V = \Gamma_0^{-1}V_1(\Gamma_0')^{-1}.$$

It follows that d can be found from (D.12) through the substitution of (D.15).

Since

$$\det(V) = \det(\Gamma_0^{-1} V_1 (\Gamma_0')^{-1}) = \det(V_0^{-1} V_1) = \frac{1}{\det(V_0 V_1^{-1})}$$

(D.16) $$\mathrm{tr}(V^{-1}) = \mathrm{tr}(\Gamma_0' V_1^{-1} \Gamma_0) = \mathrm{tr}(\Gamma_0 \Gamma_0' V_1^{-1}) = \mathrm{tr}(V_0 V_1^{-1})$$

$$\theta' V^{-1} \theta = (\theta_1 - \theta_0)'(\Gamma_0')^{-1} \Gamma_0' V_1^{-1} \Gamma_0 \Gamma_0^{-1} (\theta_1 - \theta_0) = (\theta_1 - \theta_0)' V_1^{-1} (\theta_1 - \theta_0),$$

we get

(D.17) $$d = \tfrac{1}{2}[\det(V_0 V_1^{-1})]^{-\frac{1}{n+v}}(v + \mathrm{tr}(V_0 V_1^{-1}) + (\theta_1 - \theta_0)' V_1^{-1} (\theta_1 - \theta_0)) - \frac{n+v}{2}.$$

Note that in the special case $V_0 = V_1 = V_*$ we have $d = \frac{1}{2}(\theta_1 - \theta_0)' V_*^{-1} (\theta_1 - \theta_0)$.

Appendix E

A large sample approximation to the probability distribution of the test statistic

In this appendix we shall derive a large sample approximation to the distribution of the test statistic S.

From the results of Section 5 and Appendix B it follows that under H_0' the test statistic S can be written as

$$(E.1) \qquad S = \frac{\sum_{i=1}^{n-p} \lambda_i \xi_i}{\sum_{i=1}^{n-p} \xi_i},$$

where $\frac{1}{2} \leq \lambda_1 \leq \lambda_2 \ldots \leq \lambda_{n-p} \leq 2$ are the nonzero eigenvalues of the matrix P as defined in (5.32) of Section 5, and where $\xi_1, \xi_2, \ldots, \xi_{n-p}$ are mutually independent random variables with $\xi_i \sim \chi^2(1)$, $i = 1, 2, \ldots, n-p$.

There are m different λ_i's denoted by $\frac{1}{2} \leq \tau_1 < \tau_2 < \ldots \tau_M \leq 2$ with corresponding multiplicities $m_1, m_2, \ldots, m_M$, as shown in Table 1 of Section 5, where $\sum_{j=1}^{M} m_j = n-p$. Note that we always have $P(\tau_1 \leq S \leq \tau_M) = 1$.

We only consider the case $M > 1$, since for the trivial case $M = 1$ we get $P(S = \tau_1) = 1$.

Moreover we know that one of the τ_j's is equal to 1, say $\tau_{j*} = 1$, with corresponding multiplicity $m_{j*} = n+p-k-\ell$. This implies that

$$\frac{m_{j*}}{n-p} \to 1 \text{ if } n\to\infty$$

and also

$$\frac{m_j}{n-p} \to 0 \text{ if } n\to\infty,\ j \neq j_*.$$

Let $F_0(s) = P(S \leq s;\ H_0')$, then we get

$$F_0(s) = P(S \leq s;\ H_0') = P\Big(\frac{\Sigma\, \lambda_i \xi_i}{\Sigma\, \xi_i} \leq s\Big) = P(\Sigma(\lambda_i - s)\xi_i \leq 0).$$

This yields

$$(E.2)\qquad F_0(s) = P(Q_s \leq 0),$$

where

$$(E.3)\qquad Q_s = \sum_{i=1}^{n-p} (\lambda_i - s)\xi_i .$$

We first prove the following result:

$$(E.4)\qquad \frac{Q_s - (n-p)(a_n - s)}{\sqrt{2(n-p)[(a_n-s)^2 + b_n]}} \overset{F}{\to} n(0,\ 1) \text{ if } s \neq 1 \text{ and } n\to\infty,$$

where $\overset{F}{\to}$ denotes convergence in distribution and where a_n and b_n are defined by

$$(E.5)\qquad \begin{aligned} a_n &= \frac{1}{n-p} \sum_{j=1}^{n-p} m_j \tau_j \\ b_n &= \frac{1}{n-p} \sum_{j=1}^{n-p} m_j \tau_j^2 - a_n^2 . \end{aligned}$$

Proof: We first observe that $a_n \to 1$ and $b_n \to 0$ if $n\to\infty$. This easily follows from $\tau_{j*} = 1$ and $\frac{1}{2} \leq \tau_j \leq 2$ for all n.
Also note that $\tau_1 < a_n < \tau_M$ and $b_n > 0$ since $M > 1$. Let $s \neq 1$ and define $X_i = (\lambda_i - s)\xi_i$. Then $X_1, X_2, \ldots, X_{n-p}$ are mutually independent random variables and

$$(i)\qquad \begin{aligned} \mu_i &= E(X_i) = \lambda_i - s \\ \sigma_i^2 &= Var(X_i) = 2(\lambda_i - s)^2 . \end{aligned}$$

We also have

$$E(|X_i - \mu_i|^3) = E(|(\lambda_i - s)\xi_i - (\lambda_i - s)|^3) = |\lambda_i - s|^3 E(|\xi_i - 1|^3).$$

Since $E(|\xi_i-1|^3) \leq 28$ and $|\lambda_i-s|^3 \leq (2+|s|)^3$, it follows that

(ii) $\quad E(|X_i-\mu_i|^3) \leq 28(2+|s|)^3 < \infty.$

Moreover, we get

$$\sum_{i=1}^{n-p} \mu_i = \sum_{i=1}^{n-p} (\lambda_i - s) = \sum_{i=1}^{n-p} \lambda_i - (n-p)s$$

$$= \sum_{j=1}^{M} m_j \tau_j - (n-p)s = (n-p)(a_n - s)$$

and

$$\sum_{i=1}^{n-p} \sigma_i^2 = 2 \sum_{i=1}^{n-p} (\lambda_i - s)^2 = 2[\sum_{i=1}^{n-p} \lambda_i^2 - 2s \sum_{i=1}^{n-p} \lambda_i + (n-p)s^2]$$

$$= 2[\sum_{j=1}^{M} m_j \tau_j^2 - 2s \sum_{j=1}^{M} m_j \tau_j + (n-p)s^2]$$

$$= 2(n-p)[b_n + a_n^2 - 2sa_n + s^2]$$

$$= 2(n-p)[(a_n - s)^2 + b_n].$$

This gives:

$$\frac{\sum_{i=1}^{n-p} E(|X_i-\mu_i|^3)}{(\sum_{i=1}^{n-p} \sigma_i^2)^{3/2}} \leq \frac{(n-p)[28(2+|s|)^3]}{(\sum_{i=1}^{n-p} \sigma_i^2)^{3/2}} = \frac{1}{\sqrt{n-p}} \cdot \frac{7\sqrt{2}(2+|s|)^3}{[(a_n-s)^2+b_n]^{3/2}}.$$

Since $s \neq 1$ it follows from $a_n \to 1$ and $b_n \to 0$ that

$$\lim_{n\to\infty} \frac{7\sqrt{2}(2+|s|)^3}{[(a_n-s)^2+b_n]^{3/2}} = 7\sqrt{2}\left(\frac{2+|s|}{|1-s|}\right)^3 < \infty.$$

Together with $\lim_{n\to\infty} \frac{1}{\sqrt{n-p}} = 0$ and the above inequality this implies that

(iii) $$\lim_{n\to\infty} \frac{\sum_{i=1}^{n-p} E(|X_i-\mu_i|^3)}{(\sum_{i=1}^{n-p} \sigma_i^2)^{3/2}} = 0.$$

According to Liapounov's theorem, see Cramer [3], pp. 215-218, the mutual independence of the X_i's and (i), (ii), (iii) imply that

$$\frac{\sum_{i=1}^{n-p} X_i - \sum_{i=1}^{n-p} \mu_i}{\sqrt{\sum_{i=1}^{n-p} \sigma_i^2}} = \frac{Q_s - (n-p)(a_n-s)}{\sqrt{2(n-p)[(a_n-s)^2 + b_n]}} \overset{F}{\to} n(0, 1)$$

if $s \neq 1$ and $n\to\infty$, which completes the proof of (E.4).
In the case $s = 1$ the random variable Q_1 becomes a sum of $k+\ell-2p$ terms since λ_i-1 vanishes when $\lambda_i = 1$ and this root has multiplicity $n+p-k-\ell$. Hence, the theorem of Liapounov cannot be applied and (E.4) does not hold for $s = 1$.
From (E.2) and (E.4) we obtain the following large sample approximation of $F_0(s)$ when $s \neq 1$:

$$F_0(s) = P(Q_s \leq 0) = P\left(\frac{Q_s - (n-p)(a_n-s)}{\sqrt{2(n-p)[(a_n-s)^2 + b_n]}} \leq \frac{\sqrt{n-p}\,(s-a_n)}{\sqrt{2[(s-a_n)^2 + b_n]}}\right)$$

and this yields for $s \neq 1$

(E.6) $$F_0(s) \approx N\left(\frac{\sqrt{n-p}\,(s-a_n)}{\sqrt{2[(s-a_n)^2 + b_n]}}\right) \text{ for large } n,$$

where $N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$ and where a_n and b_n are as given in (E.5).
In order to compute a_n and b_n it is often more easy to use

$$a_n = \frac{1}{n-p} \operatorname{tr}(P)$$

$$b_n = \frac{1}{n-p} \operatorname{tr}(P^2) - a_n^2.$$

From the definition of P we get:

$$(E.7)\quad \begin{aligned} a_n &= \frac{1}{n-p}\,(n+k - \tfrac{1}{2}\ell - p - \tfrac{1}{2}tr(AB)) \\ b_n &= \frac{1}{n-p}\,(n+3k - \tfrac{3}{4}\ell - p - \tfrac{5}{2}tr(AB) + \tfrac{1}{4}tr[(AB)^2]) - a_n^2, \end{aligned}$$

where $A = (X'X)^{-1}X'Z$ en $B = (Z'Z)^{-1}Z'X$.

The formulae (E.7) can also be obtained from (E.5) and Table 1 of Section 5.

In applications we often need the value of c which satisfies $F_0(c) = \alpha$ for some α between 0 and $\frac{1}{2}$. We shall now derive a large sample approximation of c based on (E.6).

From $F_0(c) = \alpha$ and (E.6) it is seen that we are looking for an approximation of c, say x, which satisfies

$$N\left(\frac{\sqrt{n-p}\,(x-a_n)}{\sqrt{2[(x-a_n)^2 + b_n]}}\right) = \alpha,\ 0 < \alpha \leq \tfrac{1}{2}.$$

If t_α is the (100α)th percentile of the n(0, 1) distribution the above expression can also be written as

$$(E.8)\qquad \frac{\sqrt{n-p}\,(x-a_n)}{\sqrt{2[(x-a_n)^2 + b_n]}} = t_\alpha,$$

where $t_\alpha \leq 0$ since $0 < \alpha \leq \frac{1}{2}$.

In other words, we are looking for a value of x which satisfies (E.8) and $x \leq a_n$.

By taking the square at both sides of (E.8) we obtain

$$(n-p-2t_\alpha^2)(x-a_n)^2 - 2t_\alpha^2 b_n = 0.$$

If we assume that $n-p-2t_\alpha^2 > 0$ we get

$$(x-a_n)^2 - q_n(\alpha)b_n = 0,$$

or equivalently,

(E.9) $$x^2 - 2a_n x + a_n^2 - q_n(\alpha)b_n = 0,$$

where

(E.10) $$q_n(\alpha) = \frac{2t_\alpha^2}{n-p-2t_\alpha^2} \geq 0.$$

Note that for $\alpha \leq \frac{1}{2}$ the condition $n-p-2t_\alpha^2 > 0$ is equivalent to $\alpha > N(-\sqrt{\frac{n-p}{2}})$.

The discriminant D of the right-hand side of equation (E.9) is equal to

$$D = 4a_n^2 - 4[a_n^2 - q_n(\alpha)b_n] = 4q_n(\alpha)b_n.$$

Since $D \geq 0$ it follows that (E.9) has two real roots:

(E.11)
$$x_1 = a_n - \sqrt{q_n(\alpha)b_n}$$
$$x_2 = a_n + \sqrt{q_n(\alpha)b_n}.$$

Note that $x_1 = x_2 = a_n$ if and only if $\alpha = \frac{1}{2}$.

It is easily seen that only the root x_1 satisfies (E.8) with $t_\alpha \leq 0$.

Hence, the large sample approximation of the level α critical value c (satisfying $F_0(c) = \alpha$) with $0 < \alpha \leq \frac{1}{2}$ becomes:

(E.12) $$c \approx c_A = a_n - \sqrt{q_n(\alpha)b_n}.$$

The approximation c_A makes no sense if $c_A \leq \tau_1$ since we know in advance that $c > \tau_1$, which follows from $P(\tau_1 < S < \tau_M) = 1$.

When $\alpha \leq \frac{1}{2}$ (i.e., $t_\alpha \leq 0$) it is easily verified from (E.12) that $c_A > \tau_1$ if and only if

(E.13) $$\alpha > N(d_1),$$

where

(E.14) $$d_1 = \frac{\sqrt{n-p}(\tau_1-a_n)}{\sqrt{2[(\tau_1-a_n)^2 + b_n]}}.$$

Note that $d_1 < 0$ and also that $\alpha > N(d_1)$ implies $n-p-2t_\alpha^2 > 0$. In other words, if we choose a value of α with $N(-\sqrt{\frac{n-p}{2}}) < \alpha \leq N(d_1)$ we get an approximation $c_A \leq \tau_1$.

However, unless $\tau_1 = 1$, we have $d_1 \to -\infty$ if $n\to\infty$ and therefore $N(d_1) \to 0$ if $n\to\infty$.

That is, if $\tau_1 < 1$ the condition (E.13) is no restriction on α for large n. In order to see this, suppose that $\tau_1 < 1$, then there exists a constant g such that $\tau_1 \leq g < 1$ for all n. This implies that for sufficiently large n we have $\tau_1 \leq g < a_n$ ($a_n \to 1$ if $n \to \infty$).

Hence, for large n we get $\tau_1-a_n \leq g-a_n < 0$ and therefore $(\tau_1-a_n)^2 \geq (g-a_n)^2 > 0$, which implies that

$$\frac{b_n}{(\tau_1-a_n)^2} \leq \frac{b_n}{(g-a_n)^2}.$$

Since $\lim\limits_{n\to\infty} \frac{b_n}{(g-a_n)^2} = \frac{0}{(g-1)^2} = 0$, it follows that $\lim\limits_{n\to\infty} \frac{b_n}{(\tau_1-a_n)^2} = 0$, which yields

$$\text{(E.15)} \qquad \lim_{n\to\infty} \frac{1}{\sqrt{1 + \frac{b_n}{(\tau_1-a_n)^2}}} = 1.$$

Now d_1 can be written as

$$d_1 = -\sqrt{\frac{n-p}{2}} \cdot \frac{1}{\sqrt{1 + \frac{b_n}{(\tau_1-a_n)^2}}}$$

and it is easily seen from (E.15) that

$$\text{(E.16)} \qquad d_1 \to -\infty \text{ if } n\to\infty.$$

Finally we consider a large sample approximation to the distribution of S under H_1'.

As is shown in Section 5 and Appendix B, under H_1' we have

(E.17) $$S = \frac{\sum_{i=1}^{n-p} \xi_i}{\sum_{i=1}^{n-p} \frac{1}{\lambda_i} \xi_i},$$

with the λ_i's and ξ_i's as before.
From

$$P(S \leq s;\ H_1') = P(\frac{1}{S} \geq \frac{1}{s};\ H_1') = 1 - P(\frac{1}{S} \leq \frac{1}{s};\ H_1')$$

$$= 1 - P\Big(\frac{\Sigma \frac{1}{\lambda_i} \xi_i}{\Sigma \xi_i} \leq \frac{1}{s}\Big) = 1 - P\Big(\Sigma(\frac{1}{\lambda_i} - \frac{1}{s})\xi_i \leq 0\Big),$$

it follows in a similar way that for large n and $s \neq 1$ we can approximate $P(S \leq s;\ H_1')$ by

(E.18) $$P(S \leq s;\ H_1') \approx 1 - N\Big(\frac{\sqrt{n-p}(\frac{1}{s} - a_n^*)}{\sqrt{2[(\frac{1}{s} - a_n^*)^2 + b_n^*]}}\Big)$$

$$= N\Big(\frac{\sqrt{n-p}\ (a_n^* - \frac{1}{s})}{\sqrt{2[(a_n^* - \frac{1}{s})^2 + b_n^*]}}\Big),$$

where $a_n^* = \frac{1}{n-p} \sum_{j=1}^{M} \frac{m_j}{\tau_j}$ and $b_n^* = \frac{1}{n-p} \sum_{j=1}^{M} \frac{m_j}{\tau_j^2} - (a_n^*)^2$.

LIST OF REPORTS 1985

8500 "Publications of the Econometric Institute Second Half 1984: List of Reprints 378-400, Abstracts of Reports".

8501/O **R.M. Karp, J.K. Lenstra, C.J.H. McDiarmid and A.H.G. Rinnooy Kan,** "Probabilistic analysis of combinatorial algorithms: an annotated bibliography", 26 pages.

8502/E **M.E. Homan,** "Verschillen in consumptie tussen één en tweekostwinnerhuishoudens: een eerste analyse", 13 pages.

8503/O **A.W.J. Kolen,** "The round-trip p-center and covering problem on a tree", 17 pages.

8504/O **H.C.P. Berbee, C.G.E. Boender, A.H.G. Rinnooy Kan, C.L. Scheffer, R.L. Smith and J. Telgen,** "Hit-and-run algorithms for the identification of nonredundant linear inequalities", 32 pages.

8505/S **H. Brozius and L. de Haan,** "On limiting laws for the convex hull of a sample", 10 pages.

8506/O **J.B.G. Frenk and A.H.G. Rinnooy Kan,** "On the rate of convergence, to optimality of the LPT rule - postscript", 5 pages.

8507/E **P. Kooiman, H.K. van Dijk and A.R. Thurik,** "Likelihood diagnostics and Bayesian analysis of a micro-economic disequilibrium model for retail services", 35 pages.

8508/O **C.G.E. Boender, A.H.G. Rinnooy Kan and J.R. de Wit,** "A Bayesian procedure for the (s,Q) inventory problem", 26 pages.

8509/E **B.M.S. van Praag, S. Dubnoff and N.L. van der Sar,** "From judgments to norms: measuring the social meaning of income, age and educaton", 35 pages.

8510/S **B. Bode, J. Koerts and A.R. Thurik,** "On shopkeepers' pricing behaviour", 19 pages.

8511/M **H. Bart, I. Gohberg and M.A. Kaashoek,** "Exponentially dichotomous operators and inverse Fourier transforms", 70 pages.

8512/M **J. Brinkhuis,** "Normal integral bases and embedding of fields", 13 pages.

8513/E **P.M.C. de Boer,** "On the relationship between Revankar's and Lu-Fletcher's production function", 5 pages.

8514/S **L. de Haan,** "Extremes in higher dimensions: the model and some statistics", 15 pages.

8515 Publications of the Econometric Intitute First Half 1985: List of Reprints 401-414, Abstracts of Reports.

8516/A **A.R. Thurik and A. Kleijweg,** "Cyclical effects in retail labour productivity", 20 pages.

8517/C **B.M.S. van Praag and J. van Weeren,** "The impact of past experiences and anticipated future on individual income judgements", 24 pages.

8518/A **A.R. Thurik and J. Koerts,** "Behaviour of retail entrepreneurs". 16 pages.

8519/B **M. Hazewinkel, J.F. Kaashoek and B. Leynse,** "Pattern formation for a one dimensional evolution equation based on Thom's River Basin Model". 24 pages.

8520/A **H.K. van Dijk, T. Kloek and C.G.E. Boender,** "Posterior moments computed by mixed integration", 25 pages.

8521/B **J. Brinkhuis,** "Testing concavity and quasi-concavity is easy", 12 pages.

8522/A **R. Harkema,** "Minimum sample size requirements for maximum likelihood estimation of some demand models", 25 pages.

8523/B **A.C.F. Vorst,** "The general linear group of discrete Hodge algebras", 10 pages.

8524/A **A.M.H. Gerards and A.W.J. Kolen,** "Polyhedral combinatorics in combinatorial optimization", 25 pages.

8525/A **B.J. Lageweg, J.K. Lenstra, A.H.G. Rinnooy Kan and L. Stougie,** "Stochastic integer programming by dynamic programming", 19 pages.

8526/A **J.B. Orlin,** "A dual version of Tardo's algorithm for linear programming", 9 pages.

8527/A **B.M.S. van Praag,** "Household cost functions and equivalence scales", 24 pages (report 8424/E revised)

8528/A **S. Schim van der Loeff,** "Limited information maximum likelihood estimation of a subsystem of nonlinear equations", .. pages.

8529/A **B.S. van der Laan and A.S. Louter,** "On the number and the amount of damage of a passenger car traffic accidents in the Netherlands", ..pages.

8530/A **B.S. van der Laan and A.S Louter,** "A statistical model for the costs of passenger car traffic accidents", .. pages.

8531/A **B.M.S. van Praag, J. de Leeuw and T. Kloek,** "The population-sample decomposition approach to multivariate estimation methods", 34 pages.

8532/A **M.E. Homan, B.M.S. van Praag and A.J.M. Hagenaars,** "Household cost functions and the value of home production in one-and two earner families", .. pages.

8533/B **R.J. Stroeker,** "An inequality for YFF's analogue of the Brocard angel of a plane triangle", 18 pages.

8534/A **J. Bouman**, "Testing nonnested linear hypothesis: A Bayesian approach based on imcompletely specified prior distributions", 118 pages.

Until report 8515 this series was devided in E(conometrics), S(tatistics), M(athematics) and O(perations Research). From report 8516 on it will be

A. Economics, Econometrics and Operations Research

B. Mathematics

C. Miscellaneous