Ann. Inst. Statist. Math.
29 (1977), Part A, 35-41

TWO PROBLEMS IN MULTIVARIATE ANALYSIS: BLUS RESIDUALS AND TESTABILITY OF LINEAR HYPOTHESIS*

SOMESH DAS GUPTA

(Received Feb. 28, 1972; revised June 5, 1974)

1. Introduction

In this paper some aspects of a univariate linear model, namely, optimality of BLUS residuals, and testability of a linear hypothesis are considered.

Theil [6], [7] introduced the concept of BLUS residuals and Grossman and Styan [1] proved some optimal properties of these BLUS residuals. In this paper, a simple and direct proof of the results of Grossman and Styan is given *without the restriction* imposed by Theil and Grossman and Styan (see Section 2). This simple method is enough to tackle some other aspects of BLUS residuals considered in many other papers and, moreover, gives proper insight into the problem.

For testing a linear hypothesis in a linear model, Roy and Roy [4], [5] introduced different concepts on testability. A geometric explanation of the different situations is presented here and it is also emphasized that the notions of testability introduced by Roy and Roy is somewhat misleading. Unfortunately, this notion is widely believed and practiced. Again, a recent result of Millikan [2] on estimability is shown without involving unnecessary matrix calculus.

2. On BLUS residuals

Some results on BLUS residuals, which are slightly more general than those obtained by Grossman and Styan [1] and Theil [6], [7], are stated with very simple proofs.

Consider a linear model

$$\boldsymbol{Y}=X\boldsymbol{\beta}+\boldsymbol{u}\ ,$$

where X is a known $n\times q$ matrix of rank q, $\boldsymbol{\beta}$ is unknown, and $\boldsymbol{u}$: $n\times 1$ is a random vector with mean $\mathbf{0}$ and covariance matrix $\sigma^2 I_n$.

* This research was supported by U. S. Army Research Grant DA-ARO-D-31-124-70-G102.

A vector of uncorrelated regression residuals is defined by $\gamma_A = A'Y$, where $A: n \times m$, and

$$\text{(2.1)} \qquad \mathrm{E}\,\gamma_A = 0 \quad \text{and} \quad \mathrm{Cov}\,(\gamma_A) = \sigma^2 I_m .$$

Let $L: n \times e$ $(e = n - q)$ be a matrix such that $L'L = I_e$ and $C(L)$, the vector space spanned by the column vectors of L, is orthogonal to $C(X)$. Note that

$$\mathrm{E}\,(A'Y) \equiv 0 \Longleftrightarrow C(A) \subset C(L) .$$

Hence (2.1) is equivalent to $A = LH$ with $H'H = I_m$. When $m = e$, the above condition is equivalent to $AA' = LL'$. This was obtained by Koerts (see Grossman and Styan [1]) in a lengthy way.

Now, suppose we want to "approximate" $J'u$ by $A'Y$ where

$$\text{(2.2)} \qquad J'J = I_e , \qquad A'A = I_e , \qquad \mathrm{E}\,(A'Y) \equiv 0 .$$

The "best" approximation is done by minimizing

$$\text{(2.3)} \qquad \mathrm{tr}\,[\mathrm{Cov}\,(A'Y - J'u)] .$$

It is clear that if $C(J) = C(L)$ then $J'Y$ is the best approximation since $J'Y = J'u$ in this case.

Note that J can be expressed as

$$\text{(2.4)} \qquad J = J_1 + J_2 ,$$

where $C(J_1) \subset C(L)$, $C(J_2) \perp C(L)$, and

$$\text{(2.5)} \qquad J_1 = LM ,$$

where M is an $e \times e$ matrix. Suppose that the rank of M is s. Then M can be expressed as

$$\text{(2.6)} \qquad M = PDQ' ,$$

where P and Q are $e \times e$ orthogonal matrices,

$$\text{(2.7)} \qquad D = \left[\begin{array}{c|c} D_s & 0 \\ \hline 0 & 0 \end{array}\right] ,$$

and $D_s: s \times s$ is a diagonal matrix with positive diagonals less than or equal to 1. We can also write $A = LT$, where $T: e \times e$ is an orthogonal matrix. Now

$$\text{(2.8)} \qquad \begin{aligned} \Sigma_A &\equiv \mathrm{Cov}\,(A'Y - J'u) \\ &= \mathrm{Cov}\,[(A-J)'u] , \qquad \text{since } A'X = 0 \\ &= \sigma^2 (A-J)'(A-J) \\ &= \sigma^2 (2I_e - A'J_1 - J_1'A) , \end{aligned}$$

using (2.2) and (2.4).

From (2.8) we get

(2.9) $$\Sigma_A=\sigma^2(2I_e-T'M-M'T)\ ,$$

(2.10) $$\operatorname{tr}\Sigma_A=\sigma^2(2e-2\operatorname{tr}(T'M))=2e\sigma^2-2\sigma^2\operatorname{tr}(DG)\ ,$$

where

(2.11) $$G=P'TQ\ .$$

Thus the problem is to maximize $\operatorname{tr}(DG)$, where G is an orthogonal matrix. From the structure of D given in (2.7) it is clear that

(2.12) $$\operatorname{tr}(DG)\leqq\operatorname{tr}(D)$$

and the equality occurs, iff

(2.13) $$G=\left[\begin{array}{c|c} I_s & 0 \\ \hline 0 & G_1 \end{array}\right]$$

where G_1 is an orthogonal matrix. Hence an optimum A is given by

(2.14) $$A^*=LP\left[\begin{array}{c|c} I_s & 0 \\ \hline 0 & G_1 \end{array}\right]Q'\ .$$

Grossman and Styan [1] (1972) and Theil [6], [7] considered the above problem with the following additional assumption on J: For any matrix K: $n\times q$ of rank q such that $K'J=0$, the rank of $K'X$ is q. It can be seen that the above assumption is equivalent to

$$\begin{aligned} L'J &\text{ is nonsingular} \\ &\Longleftrightarrow \operatorname{rank}(J_1)=e \\ &\Longleftrightarrow \operatorname{rank}(M)=e \\ &\Longleftrightarrow s=e\ . \end{aligned}$$

Thus, in this case the optimum A is given by LPQ' which can be expressed as $J_1(J_1'J_1)^{-1/2}$ (using the symmetric square root).

Let A^* correspond to an optimum A in the general case. Next, we shall give a shorter (and direct) proof of an optimality result, more general than that obtained by Grossman and Styan. Their result is (2.19), and in particular,

(2.15) $$Ch_{\max}(\Sigma_A)\geqq Ch_{\max}(\Sigma_{A^*})$$

where A satisfies (2.2) and Σ_A is defined by (2.4), and $Ch_{\max}$ denotes the maximum characteristic root.

Consider an $e\times 1$ vector $\boldsymbol{l}$ with $\boldsymbol{l}'\boldsymbol{l}=1$. Note that

$$l'\Sigma_A l=\sigma^2||Al-Jl||^2 ,$$

where $|| \ ||$ denotes the standard Euclidean norm. For fixed $\boldsymbol{l}$, $\boldsymbol{b}\in C(L)$, $\boldsymbol{b}'\boldsymbol{b}=1$,

$$||\boldsymbol{b}-J\boldsymbol{l}||^2$$

attains its minimum value $2-2||J_1\boldsymbol{l}||$ when $\boldsymbol{b}=J_1\boldsymbol{l}/||J_1\boldsymbol{l}||$, J_1 being defined as in (2.4). This can be seen easily. Thus

$$(2.16)\quad (1/\sigma^2)\boldsymbol{l}'\Sigma_A\boldsymbol{l}\geqq 2-2||J_1\boldsymbol{l}||=2-2(\boldsymbol{l}'M'M\boldsymbol{l})^{1/2}=2-2(\boldsymbol{l}'QD^2Q'\boldsymbol{l})^{1/2} .$$

Taking supremum for both the sides, we get

$$(2.17)\quad (1/\sigma^2)Ch_{\max}(\Sigma_A)\geqq 2-2Ch_{\min}(D)=2Ch_{\max}(I-D)=(1/\sigma^2)Ch_{\max}\Sigma_{A^*}$$

since,

$$(2.18)\qquad \Sigma_{A^*}=\sigma^2(A^*-J)'(A^*-J)=\sigma^2[2I_e-2QDQ'] .$$

Note that, if $s<e$, $Ch_{\max}(\Sigma_{A^*})=2\sigma^2$. This is the result of Grossman and Styan who obtained it in a restricted setup discussed earlier.

Using Courant-Fisher minimax theorem (see Rao [3]), the following can be obtained from (2.16) in a straightforward manner:

$$(2.19)\qquad Ch_i(\Sigma_A)\geqq Ch_i(\Sigma_{A^*}) , \qquad i=1,\cdots,e$$

where Ch_i denotes the ith largest characteristic root. It may be remarked that for any two $p\times p$ p.s.d. matrices Γ_1 and Γ_2,

$$(2.20)\qquad \Gamma_1-\Gamma_2 \text{ is p.s.d.} \Longrightarrow Ch_i(\Gamma_1)\geqq Ch_i(\Gamma_2) , \qquad i=1,\cdots,p$$

although the converse is not necessarily true. It was pointed out by Grossman and Styan that Theil's conjecture "$\Sigma_A-\Sigma_{A^*}$ is p.s.d." is false; however, Theil's conjecture is *almost* true, in the sense described in the above result (2.19).

3. Testability of a linear hypothesis under a linear model

The standard linear model states that the mean μ of a random vector Y: $n\times 1$ is given by $\mu=A'\theta$, where A' is a known $n\times m$ matrix of rank r and θ, $m\times 1$ is unknown. A linear hypothesis is stated as H: $G'\theta=0$, where G' is a known $s\times m$ matrix of rank s (assume, for simplicity). In standard terminology, the hypothesis H is said to be *testable*, iff there exists a matrix B' : $s\times n$ such that $G'\theta\equiv B'\mu$, where $\mu=A'\theta$. In two papers [4], [5] Roy and Roy considered linear hypotheses which may not be testable in the above sense and introduced the concepts of complete testability, partial testability and non-testability of a linear hypothesis.

It is the purpose of this note to indicate that these concepts are somewhat misleading and to clarify the situation through geometric interpretations.

Basically, a linear model states $\mu \in \Omega$, where Ω is a vector subspace of R^n, and a linear hypothesis (under this model) states $\mu \in \omega$, where ω is a vector subspace of Ω. The problem arises when μ is parametrized through θ and the relation between μ and θ is not one-to-one under the model (i.e., when $r<m$). Although in many experiments θ may have some meaningful interpretations, but mathematically θ serves as a coordinate vector of μ with respect to some given spanning vectors (i.e., the columns of A'). However our only interest is on μ and any meaningful hypothesis should be expressed in terms of μ. Here θ serves as an auxiliary parameter and any hypothesis on θ has to be considered only in terms of its equivalent representation in terms of μ.

Let T be the linear transformation of R^m to R^n with the matrix representation A'. Then the model $\mu=A'\theta$, $\theta \in R^m$, can be stated as $\mu \in T(R^m) \equiv W$. The linear hypothesis H: $G'\theta=0$ can also be written as $\theta \in V$, where V is the vector subspace of R^m orthogonal to $C(G)$, the space spanned by the columns of G. Although $\theta \in V \Rightarrow \mu \in T(V)$, but $\mu \in T(V)$ does not imply $\theta \in V$. Thus the hypotheses $\theta \in V$ and $\mu \in T(V)$ may not be identical. This fact is the basis of the papers of Roy and Roy [4], [5].

Note that the hypothesis H: $G'\theta=0$ is testable in the above sense, iff

$$N_T \subset V \Longleftrightarrow V=V+N_T \ , \tag{3.1}$$

where N_T is the null space of T. The cases where $V=V+N_T$ is violated can be classified into two categories according as $V+N_T=R^m$ or $V+N_T$ is a proper subset of R^m. Using a different approach, Roy and Roy suggested the same three categories and called the hypothesis H campletely testable, partially testable or non-testable according as (in our formulation) $V=V+N_T$, $V \subset V+N_T \subset R^m$, or $V \subset V+N_T=R^m$.

The main point of this discussion is that the fact that $\mu \in T(V)$ and $\theta \in V$ are not identical should not bother us. Note that it is not possible to distinguish between θ and θ^* based on our model if $\theta^*-\theta \in N_T$, and so such θ and θ^* should be considered as "equivalent" since both of them are mapped to the same value of μ. Using this idea we should treat two hypotheses H_1: $\theta \in V_1$ and H_2: $\theta \in V_2$ as "equivalent" if

$$T(V_1)=T(V_2) \ . \tag{3.2}$$

The hypotheses $\theta \in V$ and $\mu \in T(V)$ are indistinguishable and should be treated as "equivalent." Note that $\mu \in T(V)$ is identical to $\theta \in V+N_T$, which is testable in the standard sense. The degrees of freedom of the hypothesis H should be defined as

$$h \equiv \dim T(R^m) - \dim T(V) = m - \dim(V + N_T) \leqq s \,. \tag{3.3}$$

It can be seen that $h=s$ iff $N_T \subset V$. Moreover, $h=0$, iff $V+N_T=R^m$ and h can be >0 even when the condition $N_T \subset V$ is violated. When $h=0$, the hypothesis H puts no restriction on the model and such a hypothesis is logically acceptable under the model.

So our viewpoint is that every linear hypothesis is testable when it is translated in terms of μ in the way described above. To illustrate this consider the following simple example: The model states $\mu=\theta_1+\theta_2$. Then consider three hypothesis H_1: $\theta_1=0$, H_2: $\theta_1=\theta_2=0$, and H_3: $\theta_1+\theta_2=0$. Here H_1 is equivalent to the model and H_2 is equivalent to H_3 which is testable in the standard sense.

Note that the condition of linear estimability of $G'\theta$ is given by (3.1). Millikan [2] suggested a computational criterion to check the condition of linear estimability using the following facts:

(i) $\tilde{V} \subset \tilde{N}_T \Longleftrightarrow \dim(P_V \tilde{N}_T) = \dim \tilde{N}_T - \dim \tilde{V}$,

where $\sim$ denotes the orthogonal compliment and P_V denotes the orthogonal projection on V.

(ii) For any matrix B, $\operatorname{tr}(BB^-) = \operatorname{rk}(B)$,

where B^- denotes the (Penrose) generalized inverse of B.

To see (i) more easily than Millikan's proof, note that

$$\dim(P_V \tilde{N}_T) = \dim \tilde{N}_T - \dim(\tilde{N}_T \tilde{V}) \,,$$

since $\tilde{V}$ is the null space of P_V. Now note that $\dim(\tilde{N}_T \tilde{V}) - \dim(\tilde{V})$, iff $\tilde{V} \subset \tilde{N}_T$.

It seems that a more simplified computational criterion is to check whether

$$P_{N_T}(\tilde{V}) = \{0\} \,.$$

This can be checked using only one generalized inverse.

UNIVERSITY OF MINNESOTA

REFERENCES

[1] Grossman, S. I. and Styan, G. P. H. (1970). Uncorrelated regression residuals and singular values, *The Economics Series Tech. Report No. 32*, Stanford University, Stanford California. Revised paper published in 1972 in *J. Amer. Statist. Ass.*, **67**, 672-673.

[2] Millikan, G. A. (1971). New criteria for estimability for linear models, *Ann. Math. Statist.*, **42**, 1588-1594.

[3] Rao, C. R. (1965). *Linear Statistical Inference and its Applications*, John Wiley & Sons, Inc., New York.

[4] Roy, S. N. and Roy, J. (1958). A note on testability in normal model I ANOVA and MANOVA, *Cal. Math. Soc. Golden Jubilee Commenoration Vol., Part III*, Calcutta Math. Soc., Calcutta, 333-339.

[5] Roy, S. N. and Roy, J. (1959). A note on a class of problems in 'normal' multivariate analysis of variance, *Ann. Math. Statist.*, **30**, 577-581.

[6] Theil, Henri (1965). The analysis of disturbances in regression analysis, *J. Amer. Statist. Ass.*, **60**, 1067-1079.

[7] Theil, Henri (1968). A simplification of the BLUS procedure for analyzing regression disturbances, *J. Amer. Statist. Ass.*, **63**, 242-251.