JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 68B, No. 4, October–December 1964

Weak Generalized Inverses and Minimum Variance Linear Unbiased Estimation

A. J. Goldman and M. Zelen* [1]

(August 25, 1964)

This paper presents a unified account of the theory of least squares and its adaptations to statistical models more complicated than the classical one. First comes a development of the properties of weak generalized matrix inverses, a useful variant of the more familiar pseudo-inverse. These properties are employed in a proof of the usual Gauss theorem, and in analyzing the case in which known linear restraints are obeyed by the parameters. Another situation treated is that of a singular variance-covariance matrix for the observations. Applications include the case of equi-correlated variables (including estimation despite ignorance of the correlation), linear "restraints" subject to random error, and stepwise linear estimation.

1. Introduction and Summary

The aim of this paper is to present a unified account of the theory of least squares, and in particular to describe the necessary modifications when the customary statistical model is complicated in certain ways required for greater realism. The paper contains (probably) new results, (probably) new proofs of known results, and an (almost certainly) new overall treatment of the subject. Our hesitancy to make stronger claims arises because many of the theorems associated with least squares are part of the "folk-lore" of the field, and because the relevant literature is growing rapidly and much of it is "disguised" in the context of other branches of mathematics or science. The most closely related paper of which we are aware of is that of Rao [1962]; our work was done independently of his. (The relevance of the very recent paper of Chipman and Rao [1964], which contains other references of interest, is detailed at the end of section 5.2.) Valuable summaries of various aspects of the theory of least squares can be found in Deming [1943], Plackett [1949, 1960], Rao [1946], and Scheffe [1959].

The foundation of least-squares estimation theory is the well-known Gauss [2] theorem which can be proved in a number of ways, e.g., by linear vector space techniques as in Scheffe (*op. cit*) or by the method of Lagrange multipliers as in Plackett [1960]. We shall present a proof suggested by the properties of generalized inverses of matrices, an idea motivated quite naturally by the possible singularity of the coefficient matrix in the usual normal equations. It will be shown that any one of a wider class of matrices, which we call *weak generalized inverses*, can serve equally well. The properties of weak generalized inverses appear interesting in their own right; they are developed in section 2, are applied to the derivation of the Gauss theorem in section 3, and are involved implicitly or explicitly throughout the rest of the paper as well.

One complication of the customary statistical model which often arises in practice is the imposition of known linear restraints on the parameters. In section 4 the Gauss theorem is extended to this case. For a careful analysis it is important to distinguish clearly between artificial constraints (imposed to obtain unique solutions) and "real" ones, and among the latter class to exploit the distinction between those constrained functions which were estimable before the restraints were imposed and those which are estimable only by virtue of the restraints.

Another frequent complication, the possibility of a singular variance-covariance matrix for the observations, is discussed in section 5. It is shown how this deviation from the "standard model" can be replaced by the adjunction of linear restraints, and *vice versa*. Models involving *both* kinds of complications are treated. Applications of the general theory are made to the case of equicorrelated variables (including the possibility of estimation in some cases despite ignorance of the correlation), and to the case of linear "restraints" subject to random error. The topic of *stepwise linear estimation*, which has aroused considerable interest recently, is examined in section 5.5(cf. Freund, Vail, and Clunies-Ross [1961], Goldberger and Jockems [1961]).

The style of the paper represents a compromise between (1) the desire to have it serve as a useful

---

*Part of this author's work was supported by the Mathematics Research Center, U.S. Army, Madison, Wis. under Contract DA–11–022–Ord–2059. This is a revision and extension of Mathematics Research Center Technical Report 314, May 1962.

[1] Present address: National Cancer Institute, National Institutes of Health, Bethesda, Md., 20014.

[2] The literature refers to Gauss' fundamental work as the Markov or the Gauss-Markov theorem. Since Markov's contribution consisted essentially of bringing attention to Gauss' work, it does not appear necessary to hyphenate the theorem with the name of Markov.

statistical reference as well as a vehicle of research communication, and (2) the need to avoid a length and prolixity which surely would induce "battle fatigue" in readers and authors alike. On the one hand, additional information and "sidelights" appear throughout as corollaries and informal remarks. Also, the more familiar matrix techniques have been used in preference to vector space concepts, at the cost of some awkwardness at points where the "linear geometry" approach is the really natural one. Proofs have been written out in fairly full detail (except for matrix-algebraic steps). It is hoped that these policies make the paper more valuable and accessible to a wider range of readers. On the other hand, it has been necessary to presuppose a rather mature grasp of matrix theory and manipulations. A serious expository gap (which we hope some colleague will fill) is the omission of any discussion of computational methods for the calculations required in utilizing the theory, and also the absence of concrete and nontrivial numerical examples. Inclusion of such material, though desirable for completeness, would have interrupted the logical pattern of the theoretical development.

It is a pleasure to acknowledge the many fruitful, often heated, but always stimulating discussions with J. M. Cameron (NBS Statistical Engineering Laboratory) which have continued over many years. Without his constant interest, this paper would never have been written. Colleagues at the Mathematics Research Center, whose helpful comments have influenced the present version of the material, include H. Reinhardt and J. C. Boot. We also acknowledge with thanks a constructive reading of our paper by T. N. E. Greville.

2. Weak Generalized Inverses

In this section we define weak generalized inverses and develop some of their properties. Let X be a $p \times n$ matrix. As a special case of what follows, we shall show that there exists an $n \times p$ matrix X^+ with the properties [3]

$$
\begin{aligned}
&\text{(a)} \quad XX^+X = X \\
&\text{(b)} \quad X^+XX^+ = X^+ \\
&\text{(c)} \quad (X^+X)' = X^+X \\
&\text{(d)} \quad (XX^+)' = XX^+.
\end{aligned}
\tag{2.1}
$$

The matrix X^+ is unique (this will not be proved in the present paper) and is called the generalized inverse of X. Further details on this topic can be found in the excellent review paper by Greville [1959]. Sometimes X^+ is called a pseudo-inverse or a Moore-Penrose inverse, the latter association referring to Moore [1935] who originally discovered its properties, and to Penrose [1953] who later rediscovered and developed them further.

[3] A superscript prime will always denote (vector or matrix) transposition; the original definitions involved the complex-conjugate transpose, but we deal only with real matrices.

Our approach to this material is based on the following lemmas whose proofs (although simple) are given for completeness.

Lemma 1. *Let* A *be a* p×p *symmetric matrix of rank* q (q < p), *and* K *a* p×r *matrix of rank* r = p − q. *Then there exists a* p×r *matrix* H *with the properties*

$$
\begin{aligned}
&\text{(a)} \quad H'A = 0 \\
&\text{(b)} \quad \det(H'K) \neq 0
\end{aligned}
\tag{2.2}
$$

if and only if the square symmetric matrix

$$M = \begin{bmatrix} A & K \\ K' & 0 \end{bmatrix}$$

is nonsingular. In this case any H *of rank* r *obeying* (2.2a) *can be used as the* H *in* (2.2b). *Furthermore,* M^{-1} *has the form*

$$M^{-1} = \begin{bmatrix} C & H(K'H)^{-1} \\ (H'K)^{-1}H' & 0 \end{bmatrix}.$$

Proof. First assume M nonsingular, and let

$$M^{-1} = \begin{bmatrix} C & C_1 \\ C_1' & C_2 \end{bmatrix}$$

where C is symmetric and $p \times p$, C_2 is symmetric and $r \times r$, and C_1 is $p \times r$. The multiplication [4] $MM^{-1} = I$ implies

$$AC + KC_1' = I.$$

Now choose any $p \times r$ matrix H of rank r obeying (2.2a). Such matrices certainly exist. Premultiply the last equation by H' to obtain $H'KC_1' = H'$; since H' is of rank r, $H'K$ must have rank $\geqslant r$ (and thus exactly r since K has rank r), so that (2.2b) holds. To prove the converse, let H be any $p \times r$ matrix obeying (2.2). Then by (2.2b) H must have rank $r = p - q$, and from this and (2.2a) it follows that any $p \times p$ matrix B with $H'B = 0$ has the form $B = AD$ for some $p \times p$ matrix D. Now specialize to $B = I - K(H'K)^{-1}H'$ and use the resulting matrix D to define [5]

$$C = [I - H(K'H)^{-1}K']D.$$

If this matrix, together with $C_1 = H(K'H)^{-1}$ and $C_2 = 0$, is substituted in the M^{-1} formula given above, then it is easily verified that $MM^{-1} = I$ so that M is nonsingular and the proof is complete.

[4] The symbol I will always denote an identity matrix of appropriate dimension.

[5] We remark that the matrix C can be written explicitly as $C = [I - H(K'H)^{-1}K'][A + KH']^{-1}$. One verification employs the properties (2.3b) and (3.12) of the "true C," whose existence is shown in Lemma 2, to check that using the indicated formula in the upper left block in M^{-1} does in fact lead to $MM^{-1} = I$. Another formula not requiring knowledge of H, and verifiable using (3.12b) and its consequence $(A + KK')^{-1}K = H(K'H)^{-1}$, is $C = (A + KK')^{-1} - (A + KK')^{-1}KK'(A + KK')^{-1}$.

LEMMA 2. *Let* A, K, M *be as in Lemma* 1 *and assume* M *nonsingular. Then there is a unique* $p \times p$ *symmetric matrix* C *associated to* K, *with the property that for at least one* H *obeying* (2.2),

$$\begin{aligned} &\text{(a)} \quad K'C = 0 \\ &\text{(b)} \quad AC = I - K(H'K)^{-1}H'. \end{aligned} \tag{2.3}$$

Furthermore C *obeys* (2.3b) *for every* H *satisfying* (2.2), *and has the additional properties.*

$$\begin{aligned} &\text{(a)} \quad C = CAC, \qquad A = ACA \\ &\text{(b)} \quad C \textit{ is of rank } q. \end{aligned} \tag{2.4}$$

PROOF. For any H obeying $H'A=0$ and $\det(H'K) \neq 0$, it is easily verified that any symmetric $p \times p$ matrix C satisfying (2.3) must be a block of M^{-1} placed as in the formula for M^{-1} given in Lemma 1. Furthermore, such a C *does* satisfy (2.3). Since M depends only on K (i.e., not on the choice of H), the same is true of M^{-1} and therefore of C. Since $CK=0$, premultiplication of (2.3b) by C yields $CAC=C$, which implies that the rank of C is at most that of A. Since $H'A=0$, postmultiplication of (2.3b) by A yields $ACA=A$, which implies that the rank of A is at most that of C. Thus (2.4) is proved.[5a]

It is interesting to observe, from eqs (2.2) through (2.4), that the relationship between the pairs (A, H) and (C, K) is symmetric. Also, property (2.4a) shows that C enjoys properties (2.1a) and (2.1b) of A^+. Since A and C are symmetric, (2.1c) and (2.1d) read

$$AC = CA. \tag{2.5}$$

This will certainly hold (by (2.3b)) if $K=H$, an allowable choice of K in accordance with (2.2) since $\det(H'H) \neq 0$ if $H(p \times r)$ is of rank r. Equation (2.5) will not hold in general,[6] but we shall not require it and so can permit ourselves the freedom of choosing K different from H. The case $q=p$ (i.e., A nonsingular) can be included by appropriate formal conventions concerning "vacuous blocks" in the block matrix M and its inverse; this will be assumed done wherever appropriate, the result (by (2.4a)) being of course $C=A^{-1}$.

The next lemma and its use in the following theorem are not strictly necessary for our purposes, but are included to round out the theory.

LEMMA 3. *Let* A *be a symmetric* $p \times p$ *matrix. Then every symmetric* $p \times p$ *matrix* C *related to* A *by* (2.4a) *arises from some* K *as above.*

PROOF. Let q and r be as above, and let H be any $p \times r$ matrix of rank r such that $H'A=0$. Let $K(p \times r)$ consist of r columns of $I-AC$ in the same positions as r independent columns of H'. Since $H'(I-AC)=H'$, it follows that $H'K$ is nonsingular and thus that K has rank r. Also since $C(I-AC)=0$, it follows that $CK=0$ and therefore $K'C=0$. To verify (2.3b), first observe that (2.4a) implies (2.4b), so that the equation $C(I-AC)=0$ proves $I-AC$ to have rank not exceeding $p-q=r$. Thus the columns of $I-AC$ not in K are linear combinations of the columns of K, i.e., we can write

$$I - AC = KE$$

for some $r \times p$ matrix E. Then

$$H' = H'(I-AC) = H'KE$$

so that

$$(H'K)^{-1}H' = E$$

and therefore

$$I - AC = K(H'K)^{-1}H',$$

completing the proof.

Now let X be a $p \times n$ matrix. A *weak generalized inverse* of X is an $n \times p$ matrix X^- with the first three of properties (2.1), i.e.,

$$\begin{aligned} &\text{(a)} \quad XX^-X = X \\ &\text{(b)} \quad X^-XX^- = X^- \\ &\text{(c)} \quad (X^-X)' = X^-X. \end{aligned} \tag{2.6}$$

The following theorem, which characterizes the class of all weak generalized inverses of X, in particular establishes the existence of at least one such inverse.

THEOREM. *Let* X *be a* $p \times n$ *matrix. The* $n \times p$ *matrix* X^- *is a weak generalized inverse of* X, *if and only if*

$$X^- = X'C$$

for some C *associated to* $A = XX'$ *as in Lemma* 2.

PROOF. First suppose $X^- = X'C$ with C associated to A as above. Property (2.6a) reads $ACX=X$, and follows from (2.3b) upon noting that $H'AH=0$ implies[7] $H'X=0$. Property (2.6b) reads $X'CAC=X'C$ and follows from (2.4a), while (2.6c) asserts the symmetry of $X'CX$ and is a consequence of the symmetry of C. To prove the converse, assume X^- is any $n \times p$ matrix obeying (2.6). By (2.6b) and (2.6c),

$$X^- = (X^-X)X^- = X'[(X^-)'X^-].$$

By Lemma 3, it suffices to prove that $C = (X^-)'X^-$ obeys (2.4a). This follows from

$$\begin{aligned} CAC &= (X^-)'X^-X(X^-X)'X^- = (X^-)'X^-XX^-XX^- \\ &= (X^-)'X^-XX^- = (X^-)'X^- = C, \\ ACA &= X(X^-X)'X^-XX' = XX^-XX^-XX' \\ &= XX^-XX' = XX' = A. \end{aligned}$$

[5a] From the last formula in footnote 5, we see that C is determined by K only *via* KK'; e.g. C is unchanged if K is replaced by some $(p \times r)\ KL$ with $LL' = I$.

[6] For a specific example in which eq (2.5) fails, take the rows of A to be (1, 0) and (0, 0), $H' = (0, 1)$, $K' = (1, 1)$; the rows of C are $(1, -1)$ and $(-1, 1)$.

[7] The last equation implies that the sum of the squares of the entries in each row of $H'X$ vanishes.

We note in passing that with $X^-=X'C$, X^- obeys (2.1d) if and only if eq (2.5) holds. Thus (2.1d) holds if the choice $K=H$ is made (this completes the proof that X has a generalized inverse X^+), but as mentioned earlier we shall not have to impose this requirement.

In what follows the notations H, K, C, X^- will have the same significance as in this section, A will stand for XX', and the notation X^+ will be reserved for the generalized inverse. Note that X^- and C need not be uniquely determined by X (although X^+ is), but depend on the choice of K. The relations

$$ACX=X \tag{2.7}$$

$$H'X=0 \tag{2.8}$$

obtained in connection with the last proof are recorded here for subsequent reference.

3. Fundamental Gauss Theorem for Linear Estimation

The methods of least squares have been in use now for over 150 years. Gauss [1873] in 1821 (collected works 1873) is now credited with placing the method on a sound theoretical basis without any assumptions that the random variables follow a normal distribution. Gauss's contribution was for a time neglected until Markov [1912] "rediscovered" the work of Gauss. It should be noted that Legendre [1806] in 1806 was the first to publish the method of least squares, although apparently Gauss had known about it some years previous. For a more detailed historical introduction consult Merriman [1877], Plackett [1949], and Eisenhart [1964].

In this section we apply the properties of the weak generalized inverse to obtain a proof of the Gauss theorem. The relevance of the generalized inverse to the theory of least squares has been noted by Bjerhammar [1951], Greville [1960], and Penrose [1956]. The fundamental result used in these papers is that for an over-determined system of linear equations

$$X'b=y,$$

$$y'=(y_1, y_2, \ldots, y_n),\ b'=(b_1, b_2, \ldots, b_p),$$

the selection of b which minimizes the sum of squares of residuals, $(y-X'b)'(y-X'b)$, is given by

$$b=(X')^+y$$

where $(X')^+$ is the generalized inverse of X'. It is easily verified using (2.1) that $(X')^+=(X^+)'$, so that by (2.1a) and (2.1c)

$$Ab=XX'(X^+)'y=X(X^+X)'y=(XX^+X)y=Xy$$

i.e., b must be a solution of the usual normal equations $Ab=Xy$ of least-squares theory. More recently Rao [1962] has used property (2.4a) to demonstrate some of the well-known results associated with minimum variance linear unbiased estimation.

Before stating the theorem we review the central idea of an estimable parameter, cf. Bose [1944]. Let $Y'=(Y_1, \ldots, Y_n)$ be a vector of random variables having a distribution which depends on a parameter θ. A function $g(Y)$ of the random vector Y is called an *unbiased estimate of* θ if $E[g(Y)]=\theta$, for all values of θ, where this last phrase may reflect limitations on the possible values of θ imposed by the problem at hand. The parameter θ is called *estimable* if it has at least one unbiased estimate of some form prescribed by the context. In this paper we deal only with *linear* estimates

$$g(Y)=d'Y+c$$

where $d'=(d_1, \ldots, d_n)$ is a $1\times n$ vector and c is a scalar. A *best* (unbiased linear) estimate of θ is one which has minimum variance among the class of unbiased linear estimates of θ.

THEOREM 1: (Gauss). *Let* X *be a* $p\times n$ *matrix* ($p\leq n$) *of known constants having rank* q, β *a* $p\times 1$ *vector of unknown parameters, and* Y *an* $n\times 1$ *vector of random variables such that* [8]

$$\begin{aligned} E(Y)&=X'\beta \\ \mathrm{var}(Y)&=\sigma^2 I. \end{aligned} \tag{3.1}$$

The minimum variance unbiased linear estimate of any estimable linear function $\theta=l'\beta$ *of* β *(where* l *is a* $p\times 1$ *vector) is*

$$\hat{\theta}=l'(X^-)'Y=l'CXY.$$

For all such θ, $\hat{\theta}$ *can be obtained as* $l'\hat{\beta}$ *where* $\hat{\beta}$ *(independent of* l*) is any vector minimizing the quadratic form* $(Y-X'\beta)'(Y-X'\beta)$, *or equivalently is any solution of the normal equations*

$$A\hat{\beta}=XY \qquad (A=XX') \tag{3.2}$$

whose general solution can be written

$$\hat{\beta}=CXY+(I-CA)z \tag{3.3}$$

with z *an arbitrary* $p\times 1$ *vector.*

PROOF: Let d be an $n\times 1$ vector. Then we remark that the unbiased linear estimates of $\theta=l'\beta$ are precisely the linear forms $d'Y$ with d obeying

$$Xd=l, \tag{3.4}$$

so that θ is estimable if and only if (3.4) has a solution d. Indeed, the function $d'Y+c$ is an unbiased estimate of θ if and only if, for all values of θ,

$$\begin{aligned} l'\beta=\theta&=E(d'Y+c) \\ &=d'E(Y)+c=d'X'\beta+c. \end{aligned}$$

[8] If $Y'=(Y_1, \ldots, Y_n)$, then $E(Y)$ is the vector with $E(Y_k)$ as kth component, and var(Y) is the $n\times n$ matrix with cov(Y_i, Y_j) as (i, j)th entry.

Whether $l=0$ (so that 0 is the only value of θ) or $l\neq 0$ (so that θ assumes all real values), this will be true if and only if $c=0$ and d obeys the system (3.4).

The key idea is to seek a linear change of (random) variable from Y to $\hat{\beta}=B'Y$, where B is an $n\times p$ matrix so chosen that for each estimable $\theta=l'\beta$, at least one unbiased linear estimate of θ can be written in the form $l'\hat{\beta}$. That is, for at least one d_l obeying (3.4) the identity

$$l'B'Y=l'\hat{\beta}=d_l'Y$$

is to hold, or equivalently[9] $d_l=Bl$. For any pair of vectors l and d related by (3.4) we would have

$$Xd=l=Xd_l=XBl=XBXd,$$

and since every d is related to some l by (3.4) (just define l by (3.4)) the equality between the end terms of the last display is an identity in d. This shows that B must be chosen to obey

$$X=XBX. \tag{3.5}$$

Conversely if (3.5) holds then for each d and l related by (3.4) we can set $d_l=BXd$ so that

$$Bl=BXd=d_l \text{ and } Xd_l=XBXd=Xd=l$$

as desired. Therefore (3.5) is exactly the desired relationship, and its resemblance to (2.6a) suggests our setting $B=X^-$ so that

$$d_l=X^-l=X'Cl.$$

The variance of $\hat{\theta}=d'Y$ is $\text{var}(\hat{\theta})=(d'd)\sigma^2$, which is to be minimized by a proper choice of d subject to (3.4). Define an unknown $n\times 1$ vector δ by $d=d_l+\delta$, so that

$$d'd=d_l'd_l+d_l'\delta+\delta'd_l+\delta'\delta.$$

However, since d_l satisfies (3.4) we have

$$d_l'\delta=l'CX(d-X'Cl)=l'Cl-l'CACl=0,$$

so that

$$\text{var}(\hat{\theta})=(d_l'd_l+\delta'\delta)\sigma^2,$$

which is minimized if and only if $\delta=0$, i.e., $d=d_l$. (Incidentally this shows d_l independent of the choice $B=X^-$.) Thus the unique "best estimate" is

$$\hat{\theta}=d_l'Y=l'CXY \tag{3.6}$$

and its variance is

$$\text{var}\,(\hat{\theta})=(d_l'd_l)\sigma^2=(l'CXX'Cl)\sigma^2=l'Cl\sigma^2. \tag{3.7}$$

We have shown that $l'\hat{\beta}$ is a best estimate of θ if and only if

$$l'\hat{\beta}=l'CXY,$$

which, since $l=Xd_l=ACl$, is equivalent to

$$l'C(A\hat{\beta}-XY)=0. \tag{3.8}$$

This shows that any solution $\hat{\beta}$ of the normal equation $A\hat{\beta}=XY$ yields a best estimate $l'\hat{\beta}$ of θ. Conversely[10] if (3.8) is to hold for *all* estimable $\theta=l'\beta$ (i.e. for all l such that (3.4) has a solution d), then since every d is related to some l by (3.4) we have

$$d'X'CA\hat{\beta}=d'X'CXY$$

as an identity in d, so that $X'CA\hat{\beta}=X'CXY$ and premultiplication by X (together with (2.4a) and (2.7)) shows that $\hat{\beta}$ must be a solution of the normal equations.

Since CXY is a solution of the normal equations, the general solution can be written

$$\hat{\beta}=CXY+\eta$$

where η is an arbitrary $p\times 1$ vector such that $A\eta=0$. For any $p\times 1$ vector z,

$$\eta=(I-CA)z$$

satisfies this condition by (2.4a), while conversely any η obeying $A\eta=0$ has the form $(I-CA)z$ with $z=\eta$.

It only remains to show that the solutions $\hat{\beta}$ of the normal equations are precisely the vectors β which minimize the quadratic form

$$Q=(Y-X'\beta)'(Y-X'\beta).$$

For this purpose set $\beta=\hat{\beta}+\delta$ and observe $X(Y-X'\hat{\beta})=0$, so that

$$\begin{aligned}Q&=(Y-X'\hat{\beta})'(Y-X'\hat{\beta})+(X'\delta)'(X'\delta)\\&\geqslant (Y-X'\hat{\beta})'(Y-X'\hat{\beta})\end{aligned}$$

where equality holds if and only if $X'\delta=0$ and thus[10a] if and only if $A\delta=0$, i.e., if and only if β (as well as $\hat{\beta}$) satisfies the normal equations.

The preceding analysis essentially contains the description of the class of estimable functions. We rephrase this in the following corollary.[11]

COROLLARY 1.1: *The parametric function* $\theta=l'\beta$ *is estimable if and only if*

$$(I-AC)l=0 \tag{3.9}$$

or equivalently

$$H'l=0.$$

PROOF. We know that θ is estimable if and only if there exists a vector $d_l=X'Cl$ with $Xd_l=l$. Substi-

[9] We assume for this motivation that the distribution of Y is not concentrated on some lower dimensional subset of n-dimensional space.

[10] This converse, which makes the role of the normal equations precise, was not explicitly stated as part of the theorem.

[10a] Clearly $X'\delta=0$ implies $A\delta=XX'\delta=0$; conversely $XX'\delta=0$ implies $(X'\delta)'(X'\delta)=0$ and thus $X'\delta=0$.

[11] We again remind the reader that H and K are assumed chosen as in section 2, i.e., obeying (2.2).

tuting (2.3b) yields

$$l - Xd_l = (I - AC)l = 0 = K(H'K)^{-1}H'l$$

which implies

$$H'l = 0$$

as desired. Conversely if $(I - AC)l = 0$ then

$$l = ACl = Xd_l \quad (d_l = X'Cl)$$

and if $H'l = 0$ then by (2.3b), $(I - AC)l = 0$ as well.

We observe in particular that the components of $\hat{\beta}$ are best estimates of the corresponding components of β if and only if these components are in fact estimable; by (3.4) this requires that every unit $p \times 1$ vector, and thus every $p \times 1$ vector, be a linear combination of the columns of X. In other words "$\hat{\beta}$ is an estimate of β" makes sense only in the special case $q = p$, when $C = A^{-1}$.

The next corollary pertains to finding a solution of the normal equations by adjoining "dummy" quantities to obtain a system of full rank.

COROLLARY 1.2: *Let* λ *and* m *be* $r \times 1$ *vectors of constants where* m *is arbitrary. Then the unique solution of the system of* $(p+r)$ *simultaneous linear equations*

$$\begin{bmatrix} A & K \\ K' & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \lambda \end{bmatrix} = \begin{bmatrix} XY \\ m \end{bmatrix} \tag{3.10}$$

yields a solution of the normal equations; the same holds for the unique solution of the system

$$(A + HK')\hat{\beta} = XY + Hm. \tag{3.11}$$

PROOF. The system (3.10) is of full rank since its coefficient matrix is the M of Lemma 1 in section 2. Therefore the solution can be written

$$\begin{bmatrix} \hat{\beta} \\ \lambda \end{bmatrix} = \begin{bmatrix} C & H(K'H)^{-1} \\ (H'K)^{-1}H' & 0 \end{bmatrix} \begin{bmatrix} XY \\ m \end{bmatrix}$$
$$= \begin{bmatrix} CXY + H(K'H)^{-1}m \\ (H'K)^{-1}H'XY \end{bmatrix}. \tag{3.11a}$$

However since $H'X = 0$, the vector λ is identically zero. Since K has rank r, m can be written as $K'z$ where z is a $p \times 1$ vector; then the solution vector $\hat{\beta}$ is $\hat{\beta} = CXY + (I - CA)z$ which is the general solution of the normal equations. It also follows that the $\hat{\beta}$ of (3.10)'s solution satisfies (3.11), since $K'\hat{\beta} = m$. It only remains to prove that (3.11)'s solution is unique, i.e., that $A + HK'$ is nonsingular. This is true since it can be directly verified using (2.3b), (2.3a), and (2.8) that

$$(A + HK')^{-1} = C + [H(K'H)^{-1} - CH]\,(H'H)^{-1}H'. \tag{3.12}$$

For situations in which a suitable K is known but a suitable H is not at hand, it may be desirable to replace (3.11) by an analogous system not involving H. Such a system is given by

$$(A + KK')\hat{\beta} = XY + Km, \tag{3.12a}$$

which is satisfied by the $\hat{\beta}$ of (3.10)'s solution, and which has only one solution since

$$(A + KK')^{-1} = C + H(K'H)^{-1}(H'K)^{-1}H' \tag{3.12b}$$

as can be directly verified using (2.3). Lacking H, one might still want to know C in order to check estimability by (3.9).

From the criterion (3.9) and the fact that K is of rank r, it follows that the elements of $K'\beta$ are an independent set of nonestimable functions with the additional property that $[A, K]$ has independent rows. The analysis of (3.10), together with the Gauss theorem, shows that the values of these nonestimable linear functions can be prescribed in any way (i.e., $K'\beta = m$) without affecting the best estimates of the estimable functions; $\hat{\beta}$ depends on m but $\hat{\theta} = l'\hat{\beta}$ (where $\theta = l'\beta$ is estimable) does not. The results of prescribing (in a self-consistent way) the values of an arbitrary set of linear forms in β are treated in section 4.

It is natural to inquire as to the significance of $l'\hat{\beta}$ when $\theta = l'\beta$ is *not* necessarily estimable. One form of the answer is given in the next corollary.

COROLLARY 1.3: *Let* $\hat{\beta} = CXY$. *Then for any* $\theta = l'\beta$, *there is a unique* [11a] *estimable function* $\theta_1 = l_1'\beta$, *namely* $\theta_1 = l'CA\beta$, *such that* $l\hat{\beta}$ *is the best estimate of* θ_1.

PROOF. First assume $l_1 = ACl$ and $\theta_1 = l_1'\beta$, so that θ_1 is estimable by Corollary 1.1. Then the Gauss theorem implies that $l'\hat{\beta}$ is the best estimate of θ_1, since by (2.4)

$$l'\hat{\beta} = l'CXY = l'CACXY$$
$$= (ACl)'CXY = l_1'\hat{\beta}.$$

To prove the uniqueness of θ_1, consider *any* estimable $\theta_1 = l_1'\beta$ such that $l'\hat{\beta}$ is the best estimate of θ_1. Then by Corollary 1.1 $H'l_1 = 0$, so that $l_1 = A\eta$ for some $p \times 1$ vector η. Also we must have

$$l'CXY = l'\hat{\beta} = l_1'\hat{\beta} = \eta'ACXY = \eta'XY,$$

so that $X'Cl = X'\eta$ and therefore

$$l_1 = XX'\eta = XX'Cl = ACl$$

as asserted.

For completeness we include some known facts about the vector of residuals

$$\delta = Y - X'\hat{\beta}$$

(where $\hat{\beta}$ is any solution of the normal equations [12]) and

[11a] The uniqueness assertion requires the assumption mentioned in footnote 9. Note that a definite choice of C is assumed.

[12] Note that δ is independent of the choice of $\hat{\beta}$.

the usefulness in estimating σ^2 of its squared length (the residual sum of squares)

$$S^2=\delta'\delta=(Y-X'\hat{\beta})'(Y-X'\hat{\beta})=Y'Y-\hat{\beta}'A\hat{\beta}.$$

COROLLARY 1.4: *The residual vector is uncorrelated with the estimate of any estimable function; in fact* [12a]

$$\mathrm{Cov}(\hat{\beta},\delta)=0.$$

Furthermore we have

$$\mathrm{E}(S^2)=(n-q)\sigma^2.$$

PROOF. The expected value of the residual vector is

$$E(\delta)=E(Y)-E(X'\hat{\beta})=0.$$

Therefore

$$\begin{aligned}\mathrm{Cov}(\hat{\beta},\delta)&=E\{\hat{\beta}\,\delta'\}-E\{E(\hat{\beta}')\delta'\}\\&=E\{[CXY][Y-X'\hat{\beta}]'\}=CXE(YY')(I-X'CX).\end{aligned}$$

Since

$$E(YY')=\sigma^2I+(X'\beta)(X'\beta)'$$

we have

$$\begin{aligned}\mathrm{Cov}(\hat{\beta},\delta)&=CX[\sigma^2I+X'\beta\beta'X][I-X'CX]\\&=C[\sigma^2I+A\beta\beta'][X-ACX]=0\end{aligned}$$

because $X=ACX$ from (2.7).

To prove the second assertion we use the general formula

$$E(Y'BY)=E(Y')BE(Y)+\mathrm{trace}\,(B\,\mathrm{var}\,(Y))$$

for the mean of a quadratic form, with

$$B=(I-X'CX)^2=I-X'CX,$$

to obtain

$$\begin{aligned}E(S^2)&=E(\delta'\delta)\\&=\beta'X(I-X'CX)X'\beta+\sigma^2\,\mathrm{trace}\,(I-X'CX)\\&=\sigma^2\,\mathrm{trace}\,(I-X'CX)\\&=\sigma^2[n-\mathrm{trace}\,(X'CX)].\end{aligned}$$

By the general formula trace $(M_1M_2')=$ trace $(M_2'M_1)$ where M_1 and M_2 are rectangular matrices of the same dimensions, we have

$$\begin{aligned}\mathrm{trace}\,(X'CX)&=\mathrm{trace}\,(XX'C)=\mathrm{trace}\,(AC)\\&=\mathrm{trace}\,(I-K(H'K)^{-1}H')\\&=p-\mathrm{trace}\,[(H'K)(H'K)^{-1}]\\&=p-r=q,\end{aligned}$$

completing the proof.

A final comment deals with the maximum possible number of linearly independent nonestimable parametric functions $\theta=l'\beta$. If $q=p$ (i.e., A is nonsingular) then this number is *zero*; every linear function of β is estimable since β itself is estimable (see the remarks after Corollary 1.1). If $q<p$, however, then the number is p rather than r (as is occasionally suggested). This can be seen by partitioning $A=[A_1, A_2]$, where A_1 consists of q independent columns of A. Then the $p\times p$ matrix $[H, A_1]$ is nonsingular, since $H'H$ and $A_1'A_1$ are nonsingular and

$$\begin{bmatrix}(H'H)^{-1}H'\\(A_1'A_1)^{-1}A_1'\end{bmatrix}[H, A_1]=I.$$

Therefore, if h is a column of H and $\alpha_1, \alpha_2, \ldots \alpha_q$ are the columns of A_1, then

$$N=[H, \alpha_1+h, \alpha_2+h, \ldots, \alpha_q+h]$$

has the same determinant as $[H, A_1]$ and so is also nonsingular. The p columns of N are therefore the vectors "l" of coefficients of p independent parametric functions, which are all nonestimable since the nonsingularity of $H'H$ implies that no column of

$$H'N=[H'H, H'h, H'h, \ldots, H'h]$$

is the zero vector.

4. Gauss Theorem With Given Restraints

Often experimental situations arise in which the parameters (components of β) are connected by known linear relations. It is not generally realized that some of the linear forms whose values are prescribed by these given restraints may be estimable with respect to the equations of condition $E(Y)=X'\beta$ where as before we assume X is $p\times n$ $(p\leq n)$ and of rank q. In this section we discuss the appropriate extension of the Gauss theorem when these equations of condition are supplemented by known linear constraints. It will be shown that several applications of the "simple" Gauss theorem of section 3 suffice to reduce such problems to purely matrix-theoretic questions.

We introduce the term *pre-estimable* to be used in this section for those parametric functions (linear in β) which are estimable with respect to $E(Y)=X'\beta$. The term estimable will refer to the parametric functions which are estimable with all the given information including the restraints. Clearly every pre-estimable parameter is also estimable, but the converse need not hold; for example a nonpre-estimable function whose value is specified by one of the given constraints is obviously estimable.

We will find it convenient to assume that the constraints have been brought into an "irreducible form" in a sense made precise in this and the next few

[12a] By definition, $\mathrm{Cov}(\hat{\beta},\delta)$ is the matrix $E\{[\hat{\beta}-E(\hat{\beta})][\delta-E(\delta)]'\}$.

paragraphs. Suppose the initially given restraints are

$$L'\beta=\bar{m}$$

where L' is $k\times p$ with rank k, and $\bar{m}$ is $k\times 1$. The matrix L can be partioned into $L=(L_1, L_2)$ where L_i is $p\times s_i$ $(k=s_1+s_2)$ with rank s_i such that $L_1'\beta=\bar{m}_1$ is nonpre-estimable and $L_2'\beta=\bar{m}_2$ is pre-estimable; i.e., $H'L_1$ has no zero column and $H'L_2=0$. Since H has rank $r=p-q$, we know that $s_2\leq q$. Furthermore from the remarks at the end of section 3, the maximum number of linearly independent nonpre-estimable restraints [13] is p; hence $s_1\leq p$.

Let the rank of the $r\times s_1$ matrix $H'L_1$ be ν. There will then exist a $s_1\times(s_1-\nu)$ matrix G with rank $s_1-\nu$ such that $H'L_1G=0$. Also there will exist a $s_1\times\nu$ matrix F with rank ν such that $F'G=0$. We can for example take F' to consist of ν linearly independent rows of $H'L_1$. Then the square matrix of order s_1, $S=(F, G)$, has rank s_1. Hence we can premultiply the nonpre-estimable restraints $L_1'\beta$ by S' to obtain [14]

$$\begin{bmatrix} F' \\ G' \end{bmatrix} L_1'\beta=\begin{bmatrix} F'L_1'\beta \\ G'L_1'\beta \end{bmatrix}=\begin{bmatrix} F'\bar{m}_1 \\ G'\bar{m}_1 \end{bmatrix}.$$

Since $H'L_1G=0$, the $(s_1-\nu)$ restraints $G'L_1'\beta$ are pre-estimable. It is clear that the restraints $F'L_1'\beta$ are nonpre-estimable, as $H'L_1F$ has no zero column by virtue of $F'G=0$. Thus the original k restraints $L'\beta=\bar{m}$ may be regarded as being transformed into two sets

$$K_1'\beta=m_1, \qquad m_1=F'\bar{m}_1$$

$$K_2'\beta=m_2, \qquad m_2=\begin{bmatrix} G'\bar{m}_1 \\ \bar{m}_2 \end{bmatrix}$$

where K_i is $p\times k_i$ with rank k_i such that

$$K_1'=F'L_1', \; k_1=\nu$$

$$K_2'=\begin{bmatrix} G'L_1' \\ L_2' \end{bmatrix}, \; k_2=(s_1-\nu)+s_2$$

with $H'K_2=0$. Furthermore the rank of the $r\times k_1$ matrix $H'K_1=H'L_1F$ $(k_1<p)$ is [14a] k_1 and hence the rank of K_1 is also k_1.

When the given s_1 nonpre-estimable restraints $L_1'\beta$ are such that the rank ν of $H'L_1$ is s_1 (equal to the number of nonpre-estimable restraints) then these restraints will be termed *irreducible* restraints. Alternatively if the rank ν of $H'L_1$ is $<s_1$ (smaller than the number of nonpre-estimable restraints) the restraints $L_1'\beta$ will be called *reducible* restraints since it is then possible (as was just shown) to obtain pre-estimable restraints from them. Unless otherwise indicated the given restraints in this section will be denoted by

$$\begin{bmatrix} K_1'\beta \\ K_2'\beta \end{bmatrix}=\begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

where K_i is $p\times k_i$ and has rank k_i. Furthermore the restraints $K_1'\beta$ are a set of k_1 irreducible nonpre-estimable restraints and $K_2'\beta$ denotes a set of k_2 pre-estimable restraints; i.e., $H'K_2=0$. Since k_1 and r are the ranks of $H'K_1$ and H respectively, we must have $k_1\leq r$.

THEOREM 2. *Let* X, β, *and* Y *be as before, satisfying*

$$\mathrm{E(Y)=X'}\beta, \ \mathrm{var\,Y}=\sigma^2\mathrm{I}.$$

Also let there be given known linear restraints among the parameters of the form

$$\mathrm{K}_1'\beta=\mathrm{m}_1, \ \mathrm{K}_2'\beta=\mathrm{m}_2$$

where $\mathrm{K_i}$ *is* $\mathrm{p}\times\mathrm{k_i}$ *with rank* $\mathrm{k_i}$ *and* $\mathrm{m_i}$ *is* $\mathrm{k_i}\times 1$. *The* k_1 *restraints* $\mathrm{K}_1'\beta=\mathrm{m}_1$ *are irreducible and nonpre-estimable whereas the* k_2 *restraints* $\mathrm{K}_2'\beta=\mathrm{m}_2$ *are pre-estimable. With* H *as before, let* $\mathrm{H}=[\mathrm{H}_0, \mathrm{H}_1]$ *correspond to a partition such that* H_0 *is* $\mathrm{p}\times(\mathrm{r}-\mathrm{k}_1)$ *and* H_1 *is* $\mathrm{p}\times\mathrm{k}_1$ *where* $\det \mathrm{H}_1'\mathrm{K}_1\neq 0$. *Then the minimum variance linear unbiased estimate of the estimable function* $\theta=l'\beta$ *is*

$$\hat{\theta}=l'\{\mathrm{CXY}+\mathrm{H}_1(\mathrm{K}_1'\mathrm{H}_1)^{-1}\mathrm{m}_1 +\mathrm{CK}_2(\mathrm{K}_2'\mathrm{CK}_2)^{-1}(\mathrm{m}_2-\mathrm{K}_2'\mathrm{CXY})\}$$

where the matrix C *is obtained from Lemma 2 with* K *taken to be* $\mathrm{K}=[\mathrm{K}_0, \mathrm{K}_1]$ *and* $\mathrm{K}_0(\mathrm{p}\times(r-\mathrm{k}_1))$ *chosen such that* $\det \mathrm{H'K}\neq 0$.

PROOF. [14b] A partition $H=[H_0, H_1]$ with the desired properties can be formed by taking H_1' to consist of k_1 rows of H' in the same positions as k_1 linearly independent rows of $H'K_1$. We first show that the unbiased linear estimates of $\theta=l'\beta$ are precisely the linear forms

$$g(Y)=d'Y+d_1'm_1+d_2'm_2, \tag{4.1}$$

for which

$$Xd+K_1d_1+K_2d_2=l, \tag{4.2}$$

where d is an $n\times 1$ vector and d_i is a $k_i\times 1$ vector. Thus θ is estimable if and only if (4.2) has a solution $[d', d_1', d_2']$. The proof is based on the observation that

$$Z'=[Y', m_1', m_2']$$

defines an $(n+k_1+k_2)\times 1$ random vector Z (recall

[13] We will use the term "restraint" to refer to a constrained linear form as well as to the constraint equation itself.

[14] Since S is nonsingular, $L_1'\beta=\bar{m}_1$ is logically equivalent to $S'L_1'\beta=S'\bar{m}_1$.

[14a] For some nonsingular $\nu\times\nu$ matrix U, we have $F=\bar{F}U$ where $\bar{F}(\nu\times s_1)$ consists of $\nu=k_1$ independent rows of $H'L_1$. Also $L_1'H=\bar{F}P$, where P is a $k_1\times r$ matrix of rank k_1. Then $K_1'H=F'L_1'H=U'\bar{F}'\bar{F}P$; since $U'\bar{F}'\bar{F}$ is nonsingular, $K_1'H$ is also of rank k_1.

[14b] We point out in advance that rearranging the columns of H and/or K does not alter the properties required of H and K (i.e., (2.2)).

that a constant is a special case of a random variable), and that

$$E(Y)=X'\beta, \qquad K_1'\beta=m_1, \qquad K_2'\beta=m_2$$

are equivalent to $E(Z)=[X, K_1, K_2]'\beta$. Thus the assertion is proved by the proof of (3.4), with Z replacing Y, $[X, K_1, K_2]$ replacing X, and $[d', d_1', d_2']$ replacing d'.

The nonsingularity of $H_1'K_1$ can be used to solve (4.2) for d_1 after premultiplying it by H_1'. The result is

$$d_1=(H_1'K_1)^{-1}H_1'l,$$

so that the unbiased linear estimates of $\theta=l'\beta$ are precisely the linear forms

$$g(Y)=d'Y+d_2'm_2+l'H_1(K_1'H_1)^{-1}m_1 \tag{4.3}$$

for which

$$Xd+K_2d_2=[I-K_1(H_1'K_1)^{-1}H_1']l=l^*. \tag{4.4}$$

Thus θ is estimable if and only if (4.4) has a solution $[d', d_2']$.

Since $\operatorname{var}[g(Y)]=\operatorname{var}(d'Y)=\sigma^2d'd$, as before (in section 3) the objective is to minimize $d'd$, but now the side condition on d is the existence of a d_2 related to d by (4.4). Initially regard d_2 as fixed; then the minimization of $d'd$ subject to $Xd=l^*-K_2d_2$ is identical with the problem of finding a best estimate for $(l^*-K_2d_2)'\beta$ subject to (3.1). By the Gauss theorem of section 3, the unique solution (as a function of d_2) is

$$d=X'C(l^*-K_2d_2)=X'C\{[I-K_1(H_1'K_1)^{-1}H_1']l-K_2d_2\}. \tag{4.5}$$

In (4.5) the matrix C is obtained from Lemma 2 for *some* appropriate K; choosing[15] K as in the statement of the theorem yields $CK_1=0$ (since C is symmetric and (2.3a) holds); so that (4.5) simplifies to

$$d=X'Cl-X'CK_2d_2. \tag{4.6}$$

This simplification was the purpose for choosing the indicated form $K=[K_0, K_1]$.

Let $Y^*=X'Cl$ and $X^*=K_2'CX$. Then the quadratic form $d'd$ to be minimized becomes, by (4.6),

$$Q^*=(Y^*-(X^*)'d_2)'(Y^*-(X^*)'d_2),$$

and the condition on d_2 is that it be related to some d by (4.4). This condition is, however, automatically satisfied for *any* d_2, which can be seen as follows. First, the estimability of θ implies that l^* can be written in at least one way in the form (4.4), say

$$l^*=X\delta+K_2\delta_2.$$

[15] To show that such a choice is possible in at least one way, select any $p\times(r-k_1)$ matrix K_0 of rank $r-k_1$ such that $H_1'K_0=0$ and $AK_0=0$. If $H_0=K_0$ held, then (2.2a) would be satisfied and

$$H'K=[H_0, H_1]'[K_0, K_1]$$

would have nonsingular square blocks $H_0'H_0$ and $H_1'K_1$ on its main diagonal, implying the desired relation $\det(H'K)\neq 0$. Since by Lemma 1 this relation (for fixed K) is independent of the particular choice of H, it persists even if $H_0\neq K_0$.

Second, the pre-estimability of $K_2'\beta$ (i.e.; the fact $H'K_2=0$) implies that $K_2=XB$ for some $n\times k_2$ matrix B. Combining these observations gives (for any d_2)

$$l^*-K_2d_2=X(\delta+B\delta_2-Bd_2)=Xd \qquad (d=\delta+B\delta_2-Bd_2)$$

as desired.

The choices of d_2 (now unrestrained) which minimize Q^* are known by the Gauss theorem to be precisely the vectors

$$d_2=C^*X^*Y^*+(I-C^*A^*)z, \tag{4.7}$$

where z is an arbitrary $k_2\times 1$ vector and C^* is related to

$$A^*=X^*(X^*)'=K_2'CXX'CK_2=K_2'CK_2 \tag{4.8}$$

as C is to A. We shall however show below that $I-A^*C^*=0$, so that A^* is nonsingular and the solution becomes uniquely

$$d_2=(A^*)^{-1}X^*Y^*=(K_2'CK_2)^{-1}K_2'Cl. \tag{4.9}$$

Substitution of (4.6) and (4.9) into (4.3) gives the best estimate $\hat{\theta}$ as asserted in the statement of the theorem.

Since K_2 has linearly independent columns, we can prove $I=A^*C^*$ by showing that

$$(I-A^*C^*)K_2'=0.$$

For this purpose write $K_2=XB$ as above, and use eq (2.6a) to obtain

$$K_2'=B'(XX^-X)'=B'X'(X^-)'X'=K_2'(X'C)'X'=X^*X'.$$

Then the version $A^*C^*X^*=X^*$ of (2.7) gives

$$(I-A^*C^*)K_2'=(I-A^*C^*)X^*X'=0$$

as desired, completing the proof of the theorem. We shall frequently use the consequence

$$ACK_2=K_2 \tag{4.10}$$

of $K_2=XB$ and $ACX=X$.

COROLLARY 2.1: *The parametric function* $\theta=l'\beta$ *is estimable if and only if*

$$\mathrm{H_0}'[\mathrm{I}-\mathrm{K_1}(\mathrm{H_1}'\mathrm{K_1})^{-1}\mathrm{H_1}']l=0. \tag{4.11}$$

Furthermore if $\mathrm{H_0}$ *is chosen such that* $\mathrm{H_0}'\mathrm{K_1}=0$, *then the condition reduces to*

$$\mathrm{H_0}'l=0 \tag{4.12}$$

PROOF. The necessary and sufficient condition for θ to be estimable was shown to be (4.4), i.e.,

$$Xd+K_2d_2=[I-K_1(H_1'K_1)^{-1}H_1']l \tag{4.13}$$

holds for some d and d_2. Since $H_0'X=0$ and $H_0'K_2=0$, eq (4.11) holds. Conversely if (4.11) holds, then

$$H'[I-K_1(H_1'K_1)^{-1}H_1']l=0$$

744-268 O-65—2

which implies that (4.13) holds for some d and d_2, which in turn means that $\theta = l'\beta$ is estimable.

The only restrictions on $H=[H_0,\ H_1]$ and $K=[K_0,\ K_1]$ are that $\det\ H'K \neq 0$, $H'A=0$, $\det H_1K_1 \neq 0$ and that both H and K have rank $r=p-q$. The matrix H_0 can be chosen in any way subject to satisfying the above conditions. H_0 can always be taken to satisfy $H_0'K_1=0$ by taking an initial $\tilde{H}_0$ for which the above conditions hold and letting

$$H_0=[I-H_1(K_1'H_1)^{-1}K_1']\tilde{H}_0.$$

It is easy to verify that $H_0'A=0$, and further that $H_0'K_1=0$. By Lemma 1, if the matrix $H=[H_0, H_1]$ has rank r, then $\det\ (H'K) \neq 0$. To prove that H has the required rank r, we write the formula for H_0 as

$$H_0=\tilde{H}_0-H_1P \qquad (P=(K_1'H_1)^{-1}K_1'\tilde{H}_0)$$

and observe that

$$[H_0, H_1]=[H_0, H_1]\begin{bmatrix} I & 0 \\ -P & I \end{bmatrix}$$

where the first factor on the right-hand side has rank r while the second is $r\times r$ and nonsingular.

Note that the previous construction did not depend on K_0. We now show, in addition, that K_0 can be so chosen that $H_1'K_0=0$. Simply replace an initial $\tilde{K}_0$ by

$$K_0=[I-K_1(H_1'K_1)^{-1}H_1']\tilde{K}_0;$$

then $H_1'K_0=0$, and $H_0'K_0$ has the required rank $r-k_1$ since it coincides with $H_0'\tilde{K}_0$.

For simplicity we shall assume in what follows that H_0' and K_0 are chosen so that both $H_0'K_1=0$ and $H_1'K_0=0$. Thus the estimability condition is given by $H_0'l=0$, and the frequently occurring inverse $(H'K)^{-1}$ takes the simple form

$$(H'K)^{-1}=\mathrm{diag}\ [(H_0'K_0)^{-1}, (H_1'K_1)^{-1}]. \tag{4.14}$$

Then the general form of the vector $\hat{\beta}$, such that the best estimate of *every* estimable $l'\beta$ is $l'\hat{\beta}$, is given by

$$\begin{aligned}\hat{\beta} = CXY+H_0(K_0'H_0)^{-1}m_0+H_1(K_1'H_1)^{-1}m_1 \\ +CK_2(K_2'CK_2)^{-1}(m_2-K_2'CXY),\end{aligned} \tag{4.15}$$

where m_0 is an arbitrary $(r-k_1)\times 1$ vector.

The next corollary formulates some systems of equations involving "dummy" variables (μ_0, μ_1, λ) and artificial restraints ($K_0'\beta=m_0$) which can be used to solve for $\hat{\beta}$ of (4.15).

COROLLARY 2.2: *Let* μ_0 *and* m_0 *be* $(\mathrm{r}-\mathrm{k}_1)\times 1$ *vectors,* μ_1 *a* $\mathrm{k}_1\times 1$ *vector and* λ *a* $\mathrm{k}_2\times 1$ *vector. Then every solution* $\hat{\beta}$ *of the system*

$$\begin{bmatrix} \mathrm{A} & \mathrm{K}_2 \\ \mathrm{K}_1' & 0 \\ \mathrm{K}_2' & 0 \end{bmatrix}\begin{bmatrix} \hat{\beta} \\ \lambda \end{bmatrix}=\begin{bmatrix} \mathrm{XY} \\ \mathrm{m}_1 \\ \mathrm{m}_2 \end{bmatrix} \tag{4.16}$$

can be used in the best estimate $\theta=l\hat{\beta}$ *of any estimable function* $\theta=l'\beta$. *The same holds for the unique solutions of each of the systems*

$$\begin{bmatrix} \mathrm{A} & \mathrm{K}_0 & \mathrm{K}_1 & \mathrm{K}_2 \\ \mathrm{K}_0' & 0 & 0 & 0 \\ \mathrm{K}_1' & 0 & 0 & 0 \\ \mathrm{K}_2' & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \hat{\beta} \\ \mu_0 \\ \mu_1 \\ \lambda \end{bmatrix}=\begin{bmatrix} \mathrm{XY} \\ \mathrm{m}_0 \\ \mathrm{m}_1 \\ \mathrm{m}_2 \end{bmatrix}, \tag{4.17}$$

$$\begin{bmatrix} \mathrm{A} & \mathrm{K}_2 \\ \mathrm{K}_0' & 0 \\ \mathrm{K}_1' & 0 \\ \mathrm{K}_2' & 0 \end{bmatrix}\begin{bmatrix} \hat{\beta} \\ \lambda \end{bmatrix}=\begin{bmatrix} \mathrm{XY} \\ \mathrm{m}_0 \\ \mathrm{m}_1 \\ \mathrm{m}_2 \end{bmatrix} \tag{4.18}$$

$$\begin{bmatrix} \mathrm{A}+\mathrm{H}_0\mathrm{K}_0'+\mathrm{H}_1\mathrm{K}_1' & \mathrm{K}_2 \\ \mathrm{K}_2' & 0 \end{bmatrix}\begin{bmatrix} \hat{\beta} \\ \lambda \end{bmatrix}=\begin{bmatrix} \mathrm{XY}+\mathrm{H}_0\mathrm{m}_0+\mathrm{H}_1\mathrm{m}_1 \\ \mathrm{m}_2 \end{bmatrix}, \tag{4.19}$$

as well as the vector

$$\hat{\beta}=\hat{\beta}_0+\mathrm{CK}_2(\mathrm{K}_2'\mathrm{CK}_2)^{-1}(\mathrm{m}_2-\mathrm{K}_2'\mathrm{CXY}) \tag{4.20}$$

where $\hat{\beta}_0$ *is obtained from the unique solution of*

$$\begin{bmatrix} \mathrm{A} & \mathrm{K}_0 & \mathrm{K}_1 \\ \mathrm{K}_0' & 0 & 0 \\ \mathrm{K}_1' & 0 & 0 \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \mu_0 \\ \mu_1 \end{bmatrix}=\begin{bmatrix} \mathrm{XY} \\ \mathrm{m}_0 \\ \mathrm{m}_1 \end{bmatrix}. \tag{4.21}$$

PROOF. System (4.16) does not have a unique solution for $k_1<r$, but for any solution $[\hat{\beta}', \lambda']$ we can *define* a vector m_0 by $K_0'\hat{\beta}=m_0$ and observe that $[\hat{\beta}', \lambda']$ satisfies (4.18). Thus the discussion of (4.16) reduces to that of (4.18).

Since the $\hat{\beta}$ of (4.15) clearly obeys $K_i'\hat{\beta}=m_i$, and also (premultiply $\hat{\beta}$ by CA) satisfies

$$C\{A\hat{\beta}+K_2\lambda-XY\}=0$$

where

$$\lambda=(K_2'CK_2)^{-1}(K_2'CXY-m_2), \tag{4.22}$$

we find that $\hat{\beta}$ and λ obey (4.18). Thus the solution of (4.18), once it is proved unique, must have the form (4.15).

The first subsystem

$$A\hat{\beta}+K_0\mu_0+K_1\mu_1+K_2\mu_2=XY$$

of (4.17), when premultiplied by $(H_0'K_0)^{-1}H_0'$, yields $\mu_0=0$; then premultiplication of

$$A\hat{\beta}+K_1\mu_1+K_2\mu_2=XY$$

by $(H_1'K_1)^{-1}H_1'$ yields $\mu_1=0$. Thus every solution of

(4.17) is a solution of (4.18), so (4.17) does not require further discussion. Note that the dummy variables μ_0 and μ_1 are zero vectors in the solution.

It is trivial to check that any solution of (4.18) is also a solution of (4.19). Thus the results for (4.16) through (4.19) will be proved once we show that (4.19) has a unique solution. The utility of (4.19) is that it is a smaller system than those preceding it. We write the first subsystem of (4.19) in the form

$$(A+HK')\hat{\beta} = XY+Hm-K_2\lambda$$

where $m'=(m_0', m_1')$. By Corollary 1.2, $A+HK'$ is nonsingular so that $\hat{\beta}$ is directly determined in terms of λ by

$$\hat{\beta} = (A+HK')^{-1}(XY+Hm)-(A+HK')^{-1}K_2\lambda.$$

Corollary 1.2 shows that the first term on the right is just $\hat{\beta}_0$, while since $H'K_2=0$ the formula (3.12) for $(A+HK')^{-1}$ shows that we have

$$\hat{\beta} = \hat{\beta}_0 - CK_2\lambda. \tag{4.23}$$

After premultiplying by $(K_2'CK_2)^{-1}K_2'$, noting $K_2'\hat{\beta} = m_2$, we obtain a unique λ.

To treat (4.20) we first use (4.14) to obtain

$$H(K'H)^{-1} = [H_0(K_0'H_0)^{-1}, H_1(K_1'H_1)^{-1}]$$

and then apply corollary 1.1 (see (3.11a)) to the system (4.21) to show that its unique solution has

$$\hat{\beta}_0 = CXY + H_0(K_0'H_0)m_0 + H_1(K_1'H_1)^{-1}m_1. \tag{4.24}$$

Thus $\hat{\beta}$ given by (4.20) coincides with (4.15).

Corollary 2.3. *With the particular choice* $\hat{\beta} = \mathrm{CXY} + \mathrm{H_1(K_1'H_1)^{-1}m_1} + \mathrm{CK_2(K_2'CK_2)^{-1}(m_2 - K_2'CXY)}$ *for any* $\theta = l'\beta$ *there is a unique*[15a] *estimable* $\theta_1 = l_1'\beta_1$ *given by* $l_1 = [\mathrm{AC} + \mathrm{K_1(H_1'K_1)^{-1}H_1'}]l$ *such that for all possible* $\mathrm{m_1}, \mathrm{m_2}$ *and* $\beta, l'\hat{\beta}$ *is the best estimate of* θ_1.

Proof. First assume $l_1 = [AC + K_1(H_1'K_1)^{-1}H_1']l$ and $\theta_1 = l_1'\beta$. Then $H_0'l_1 = 0$, so θ_1 is estimable, and we have

$$l_1'\hat{\beta} = l'[CA + H_1(K_1'H_1)^{-1}K_1']\hat{\beta} = l'\hat{\beta}$$

by direct calculation (using $CAC=C, AH_1=0, K_1'C=0$) so that $l'\hat{\beta}$ is the best estimate of θ_1. To prove uniqueness, consider any estimable $\theta_1 = l_1'\beta$ such that $l'\hat{\beta}$ is the best estimate of θ_1 for all m_1 and m_2. Note with the aid of (2.2a), that det $(H_1'K_1) \neq 0$ implies that $[A, K_1]$ has rank $q+k_1$. Since $H_0'[A, K_1]=0$ and H_0 has rank $p-(q+k_1)$, it follows from $H_0'l_1=0$ that $l_1 = Ad + K_1d_1$ for some vectors d and d_1. Using $ACX=X$ and $ACK_2=K_2$ we obtain

$$l_1'\hat{\beta} = d'[(I-K_2(K_2'CK_2)^{-1}K_2'C)XY + K_2(K_2'CK_2)^{-1}m_2] + d_1'm_1.$$

Setting $m_i = K_i'\beta$ $(i=1, 2)$ and equating the coefficients of Y and β in $l_1'\beta$ and $l'\hat{\beta}$, we obtain

$$X'(I-CK_2(K_2'CK_2)^{-1}K_2')(d-Cl)=0, \tag{4.25}$$

$$K_2(K_2'CK_2)^{-1}K_2'(d-Cl) = K_1[(H_1'K_1)^{-1}H_1'l - d_1] \tag{4.26}$$

Multiplication of the second equation by H_1' leads to $H_1'l = H_1'K_1d_1$ and thus to

$$d_1 = (H_1'K_1)^{-1}H_1'l$$

as desired. Substitution of this into (4.26) yields a result which when substituted into (4.25) gives

$$X'(d-Cl)=0$$

implying that $Ad=ACl$ as desired.

We turn now to the residual vector $\delta = Y - X'\hat{\beta}$ and the residual sum of squares $S^2 = \delta'\delta$.

Corollary 2.4. *The residual sum of squares can be written as*

$$S^2 = \delta'\delta = (Y - X'\hat{\beta}_0)'(Y - X'\hat{\beta}_0) + \lambda'(K_2'CK_2)\lambda \tag{4.27}$$

and has the expected value

$$E(S^2) = (n-q+k_2)\sigma^2,$$

where $\hat{\beta}_0$ *is the estimate ignoring the preestimable restraint* $\mathrm{K_2'\beta = m_2}$ *and*

$$\lambda = (K_2'CK_2)^{-1}(K_2'CXY - m_2).$$

Proof. The residual sum of squares can be written

$$S^2 = \delta'\delta = (Y - X'\hat{\beta}_0 + X'CK_2\lambda)'(Y - X'\hat{\beta}_0 + X'CK_2\lambda)$$

$$= (Y-X'\hat{\beta}_0)'(Y-X'\hat{\beta}_0) + \lambda'(K_2'CK_2)\lambda + 2(Y-X'\hat{\beta}_0)'X'CK_2\lambda.$$

However we have

$$(Y-X'\hat{\beta}_0)'X'CK_2\lambda = (Y'X'C - \hat{\beta}_0'AC)K_2\lambda$$

$$= (Y'X'C - Y'X'CAC)K_2\lambda = 0$$

and thus

$$S^2 = (Y-X'\hat{\beta}_0)'(Y-X'\hat{\beta}_0) + \lambda'(K_2'CK_2)\lambda.$$

From Corollary 1.4 of section 3 we have

$$E\{(Y-X'\hat{\beta}_0)'(Y-X'\hat{\beta}_0)\} = (n-q)\sigma^2.$$

Furthermore

$$E(\lambda) = (K_2'CK_2)^{-1}(K_2'\beta - m_2) = 0, \text{ var } \lambda = (K_2'CK_2)^{-1}\sigma^2.$$

Making use of the formula for finding the expectation

[15a] The uniqueness assertion requires the assumption mentioned in footnote 9.

of a quadratic form [16] gives

$$E(\lambda' K_2' C K_2 \lambda) = \text{trace}\,\{(K_2' C K_2)(\text{var}\,\lambda)\}$$

$$= \text{trace}\,\{(K_2' C K_2)(K_2' C K_2)^{-1}\sigma^2\} = k_2\sigma^2,$$

and thus the result is proved. Note that the formula for S^2 is composed of two parts, the second of which measures the deviation between the values of the pre-estimable restraints $K_2'\beta$ as estimated from the data (Y), and the given values m_2 of these restraints.

COROLLARY 2.5. *The residual vector* $\delta = \mathrm{Y} - \mathrm{X}'\hat{\beta}$ *is uncorrelated with any estimable function; in fact*

$$\text{Cov}\,(\delta, \hat{\beta}) = 0. \tag{4.28}$$

PROOF. We write $\delta_0 = Y - X'\hat{\beta}_0$ as the residual vector if the restraints $K_2'\beta = m_2$ had been ignored. By (4.20) $\delta = \delta_0 + X'CK_2\lambda$ and $\hat{\beta} = \hat{\beta}_0 - CK_2\lambda$ and we can write

$$\text{Cov}\,(\delta, \hat{\beta}) = \text{cov}\,(\delta_0, \hat{\beta}_0) - \text{cov}\,(\delta_0, \lambda)K_2'C$$

$$+ X'CK_2\,\text{cov}\,(\lambda, \hat{\beta}_0) - [E(\lambda\lambda')]K_2'C. \tag{4.29}$$

From Corollary 1.4 of section 3 we have $\text{cov}(\delta_0, \hat{\beta}_0) = 0$. For the second term in (4.29) we calculate

$$\text{cov}(\delta_0, \lambda) = \text{cov}(\delta_0, K_2'\hat{\beta}_0 - m_2)(K_2'CK_2)^{-1} \tag{4.30}$$

$$= \text{cov}(\delta_0, \hat{\beta}_0)K_2(K_2'CK_2)^{-1} = 0.$$

For the third term we calculate

$$\text{cov}(\lambda, \hat{\beta}_0) - [E(\lambda\lambda')]K_2'C \tag{4.31}$$

$$= (K_2'CK_2)^{-1}\,\text{cov}(K_2'\hat{\beta}_0 - m_2, \hat{\beta}_0) - (\text{var}\,\lambda)K_2'C$$

$$= (K_2'CK_2)^{-1}\{K_2'\,\text{var}(\hat{\beta}_0) - K_2'C\sigma^2\}$$

$$= (K_2'CK_2)^{-1}\{K_2'\,\text{var}(CXY) - K_2'C\sigma^2\} = 0.$$

Substituting in (4.29) we obtain the desired result (4.28).

It is possible to develop the extension of the Gauss theorem in a manner which leans more heavily on properties of the weak generalized inverse. However, the final form of the solution is not useful for practical applications. One possible advantage of this alternative approach is that there is no need to make a distinction between pre-estimable and nonpre-estimable functions. These results are contained in the following theorem.

THEOREM 3. *Let* X, β *and* Y *be as before, satisfying*

$$\mathrm{E(Y)} = \mathrm{X}'\beta,\ \mathrm{var(Y)} = \sigma^2\mathrm{I}$$

and also the restraints

$$\mathrm{L}'\beta = \mathrm{m} \tag{4.32}$$

where L *is a* $\mathrm{p}\times\mathrm{k}$ *matrix of known constants and* m *is a* $\mathrm{k}\times 1$ *vector of known constants. The minimum variance unbiased linear estimate of any estimable linear function* $\theta = l'\beta$ *is* $\hat{\theta} = l'\hat{\beta}$, *where* $\hat{\beta}$ *(independent of* l*) is given by*

$$\hat{\beta} = (\mathrm{I} - \mathrm{LL}^-)'\tilde{\mathrm{C}}(\mathrm{I} - \mathrm{LL}^-)\mathrm{XY}$$

$$+ [\mathrm{I} - (\mathrm{I} - \mathrm{LL}^-)'\tilde{\mathrm{C}}(\mathrm{I} - \mathrm{LL}^-)\mathrm{A}](\mathrm{L}^-)'\mathrm{m} \tag{4.33}$$

and $\tilde{\mathrm{C}}$ *is related to* $\tilde{\mathrm{A}} = (\mathrm{I} - \mathrm{LL}^-)\mathrm{A}(\mathrm{I} - \mathrm{LL}^-)'$ *as* C *is to* A.

PROOF. We first show that the unbiased linear estimates of $\theta = l'\beta$ are precisely the linear forms

$$g(Y) = d'Y + \rho' m \tag{4.34}$$

for which

$$Xd + L\rho = l \tag{4.35}$$

where d is an $n\times 1$ vector and ρ is a $k\times 1$ vector. Thus θ is estimable if and only if (4.35) has a solution (d, ρ). The proof is based on the observation that

$$Z' = [Y', m'] \tag{4.36}$$

defines an $(n+k)\times 1$ random vector Z (recall that a constant is a special case of a random variable), and that (4.32) and $E(Y) = X'\beta$ are equivalent to $E(Z) = [X, L]'\beta$. Thus the assertion is proved by the proof of eq (3.4), with Z replacing Y, $[X, L]$ replacing X, and $[d', \rho']$ replacing d'.

The variance of $g(Y)$ given by (4.34) is $(d'd)\sigma^2$, so that finding a best estimate of θ is equivalent to minimizing $d'd$ by a proper choice of d, subject to the condition that there exist a ρ related to d by (4.35). The choice of such a ρ is immaterial (as long as one exists) since ρ appears in (4.34) only in the combination

$$\rho' m = (L\rho)'\beta$$

which by eq (4.35) is determined by d and l. If d is such that some ρ obeys (4.35), then by eq (2.6a)

$$l - Xd = L\rho = LL^-L\rho = LL^-(l - Xd)$$

with L^- any weak generalized inverse of L, so that

$$(I - LL^-)(l - Xd) = 0 \tag{4.37}$$

and a particular solution of (4.35) is $\rho^* = L^-(l - Xd)$. Conversely if (4.37) is satisfied then ρ^* provides a solution of eq (4.35) and we can take

$$g(Y) = d'(Y - X'(L^-)'m) + l'(L^-)'m. \tag{4.38}$$

It has been shown that finding a best estimate of θ is equivalent to minimizing $d'd$ subject to condition (4.37), which can be rewritten as $\tilde{X}d = \tilde{l}$ with

$$\tilde{X} = (I - LL^-)X,\ \tilde{l} = (I - LL^-)l.$$

[16] If the column vector Z is such that $E(Z) = 0$, var $Z = \sigma^2\Sigma$, then $E(Z'AZ) = \sigma^2$ tr $A\Sigma$.

This is analogous to the problem (treated in the proof of the Gauss theorem) of minimizing $d'd$ subject to $Xd=l$, and so the unique solution is

$$\hat{d}=\tilde{X}^-\tilde{l}=X'(I-LL^-)'\tilde{C}(I-LL^-)l, \tag{4.39}$$

from which eq (4.33) follows by substitution into (4.38).

Still another approach to the material of the section can be based on the random variable Z defined by eq (4.36). Namely, as regards the first and second moments with which least-squares theory is exclusively concerned, the model specified by $E(Y)=X'\beta$ and $L'\beta=m$ is equivalent to the model

$$E(Z)=[X, L]'\beta,\ \mathrm{var}(Z)=\sigma^2\begin{bmatrix} I_n & 0 \\ 0 & 0_k \end{bmatrix} \tag{4.40}$$

where I_n is the $n\times n$ identity matrix and 0_k is the $k\times k$ zero matrix. Thus a model with linear restraints is equivalent to a "restraintless" model which however involves a singular variance-covariance matrix. Least-squares estimation in such models is discussed in the next section.

5. Gauss Theorem With Arbitrary Variance-Covariance Matrix

The results of the previous sections were derived assuming that the vector of random variables $Y'=(Y_1, Y_2, \ldots, Y_n)$ were uncorrelated and had common variance; i.e., var $Y=\sigma^2 I$. This section considers some ramifications when var $Y=\sigma^2 V$ where V is a known $n\times n$ matrix with rank m $(m\leq n)$. The case when $m=n$ has been investigated by Aitken [1937]. His result is generalized to include the possibility of a singular variance-covariance matrix.

5.1. Preliminaries

Before discussing the extension of Aitken's results it will be convenient to record the implications of having a singular variance-covariance matrix. When V is singular with rank m $(m<n)$, then there will exist a $n\times s$ $(s=n-m)$ matrix F with rank s such that $F'V=0$. However, this also implies that the s components of $F'Y$ have var $F'Y=(F'VF)\sigma^2=0$ which is equivalent to $F'Y$ being equal to a constant.[17] Since $E(Y)=X'\beta$, we have as the value of this constant

$$F'Y=F'E(Y)=F'X'\beta. \tag{5.1}$$

Then the distribution of $Y'=(Y_1, Y_2, \ldots, Y_n)$ is singular and can be reduced to a distribution in m random variables. In most applications when (5.1) holds we generally have $F'X'=0$. However, it is quite possible that $F'X'\neq 0$. In order to discuss this more general problem, we write $F=(F_1, F_2)$ in partitioned form where F_i is $n\times s_i$ with rank s_i $(i=1, 2)$ and $s_1+s_2=s$. Furthermore we have

$$F_1'X'=0 \tag{5.2}$$

$$\text{rank } F_2'X'=s_2 \qquad (s_2<p).$$

[17] The qualifying phrase "with probability one" should be added but we omit such distinctions.

Note that (5.2) combined with (5.1) results in

$$F_1'Y=0 \tag{5.3}$$

$$F_2'X'\beta=F_2'Y.$$

That is, there are s_1 independent linear relations among $(Y_1, Y_2, \ldots, Y_n)$ and s_2 restraints among the β which are preestimable by virtue of $H'(XF_2)=0$.

Another preliminary aspect of the problem is the existence of an $n\times n$ orthogonal matrix P such that

$$P'VP=\begin{bmatrix} 0 & 0 \\ 0 & \Lambda \end{bmatrix} \tag{5.4}$$

where Λ is the $m\times m$ diagonal matrix whose elements are the m nonzero characteristic roots of the symmetric positive semidefinite matrix V. Let G be a $n\times m$ matrix such that the columns of G are the m (normalized) characteristic vectors of V; i.e.,

$$VG=G\Lambda, \qquad G'G=I$$

Then the orthogonal matrix P in (5.4) can be taken to be

$$P=[F,G] \tag{5.5}$$

where F is the $n\times s$ matrix mentioned previously, chosen (as is possible) so that $F'F=I$ and $G'F=0$. By virtue of this partition we have

$$V=G\Lambda G'. \tag{5.6}$$

We also note that V^+ is given by

$$V^+=G\Lambda^{-1}G'.$$

The necessary four properties (2.1a–d) follow from

$$V^+V=G\Lambda^{-1}G'G\Lambda G'=GG'$$

as $G'G=I$.

A frequently occurring case is when $V^2=cV$ where c is a scalar. Then it can readily be verified that the generalized inverse of V is $V^+=c^{-2}V$.

Also there will be need for writing the matrix V^+ as

$$V^+=TT',\ T=G\Lambda^{-1/2} \tag{5.7}$$

where $\Lambda^{-1/2}$ denotes the matrix obtained from Λ by replacing the diagonal terms by the reciprocals of their positive square roots. Note also that $T'VT=I$.

5.2. Arbitrary Variance-Covariance Matrix

In this subsection we give some of the main results associated with an arbitrary variance-covariance matrix. The notation used will correspond to that of the preceding sections.

THEOREM 4. *Consider the vector of random variables* Y *having* E(Y)$=$X$'\beta$, var (Y)$=\sigma^2$V *where* V *is an* n$\times$n *symmetric positive semidefinite matrix with rank* m (m$\leq$n). *Then the minimum variance linear unbiased estimate of* $\theta=l'\beta$ *coincides with its best* *estimate found from the model*

$$E(\tilde{Y})=\tilde{X}'\beta, \text{ var } \tilde{Y}=\sigma^2 I. \tag{5.8}$$

$$\tilde{K}'\beta=\tilde{m}$$

where $\tilde{X}=XT$, $\tilde{Y}=T'Y$, $\tilde{K}=XF_2$ *and* $\tilde{m}=F_2'Y$. *Thus if* F_2 *is null,* $\hat{\beta}$ *can be chosen as any solution of the normal equations*

$$(XV^+X')\hat{\beta}=XV^+Y. \tag{5.9}$$

PROOF. As in the proof of the Gauss theorem of section 3, obtaining a best estimate $\Delta'Y+c$ of θ is equivalent to choosing an $n\times 1$ vector Δ, subject to $X\Delta=l$, so as to minimize

$$\text{var}(\Delta'Y)=(\Delta'V\Delta)\sigma^2.$$

On the other hand, from the beginning of the proof of Theorem 3 we see that finding a best estimate in the model (5.8) is equivalent to choosing a pair $[d', \rho']$, where d is an $m\times 1$ vector and ρ an $s_2\times 1$ vector, so as to minimize $d'd$ subject to

$$\tilde{X}d+\tilde{K}\rho=l,$$

i.e.,

$$X(Td+F_2\rho)=l.$$

Thus the theorem will be proved if we show how to associate to each vector Δ a pair $[d', \rho']$, and to each pair $[d', \rho']$ a vector Δ, such that in each case

$$X\Delta=X(Td+F_2\rho),\ \Delta'V\Delta=d'd. \tag{*}$$

Given $[d', \rho']$ we simply set $\Delta=Td+F_2\rho$; the second relation in (*) then follows from $F'V=0$ and $T'VT=I$. Given Δ, we employ the orthogonal matrix

$$P=[F, G]=[F_1, F_2, G]$$

to define an $n\times 1$ vector (and thus define d and ρ) by

$$[\rho_1', \rho', d'\Lambda^{-1/2}]'=P^{-1}\Delta;$$

then we have

$$\Delta=P[\rho_1', \rho', d'\Lambda^{-1/2}]'=F_1\rho_1+F_2\rho+Td$$

The first requirement of (*) is satisfied because $XF_1=0$, and the second for the same reasons as above.

Finally, that the normal equations corresponding to the first line of (5.8) are given by (5.9) follows from substitution for the tilde quantities, together with $TT'=V^+$.

COROLLARY 4.1 (Aitken). *If* V *is non-singular then the best estimate of any estimable* $\theta=l'\beta$ *is given by* $\hat{\theta}=l'\hat{\beta}$ *where* $\hat{\beta}$ *is any solution of the normal equations*

$$(XV^{-1}X')\hat{\beta}=XV^{-1}Y. \tag{5.10}$$

PROOF. When V is nonsingular, the matrix F is null and $V^+=V^{-1}$, so the result follows from (5.9).

COROLLARY 4.2. *Let* $\tilde{X}=XT$ *have rank* q. *Then the minimum variance linear unbiased estimate of an estimable function is* $\hat{\theta}=l'\hat{\beta}$ *where*

$$\hat{\beta}=\tilde{C}XV^+Y + \tilde{C}\tilde{K}(\tilde{K}'\tilde{C}\tilde{K})^{-1}(\tilde{m}-\tilde{K}'\tilde{C}XV^+Y)+H(K'H)^{-1}m_0.$$

The matrix $\tilde{C}$ *is related to* $\tilde{A}=\tilde{X}\tilde{X}'=XV^+X'$ *and* $\tilde{K}$ *by* Lemmas 1 *and* 2; $\tilde{K}$ *is a* $p\times r$ $(r=p-q)$ *matrix of rank* r *such that* det $H'\tilde{K}\neq 0$, *and* m_0 *is an arbitrary* $r\times 1$ *vector.*

PROOF. Since $\tilde{X}=XT$ has rank q, H has the same relation to $\tilde{X}$ as to X. Because $H'\tilde{K}=0$, the restraints $\tilde{K}'\beta=\tilde{m}$ are pre-estimable in the model (5.8). The result follows from (4.15) upon noting that here H_1 and K_1 are null, while $H_0=H$ and $K_0=\tilde{K}$.

COROLLARY 4.3. *If* $\tilde{X}$ *has rank* q, *then the quantity*

$$S^2=(\tilde{Y}-\tilde{X}'\tilde{\beta}_0)'(\tilde{Y}-\tilde{X}'\tilde{\beta}_0)+\tilde{\lambda}'(\tilde{K}'\tilde{C}\tilde{K})\tilde{\lambda},$$

in which

$$\tilde{\beta}_0=\tilde{C}\tilde{X}\tilde{Y}=\tilde{C}XV^+Y$$

$$\tilde{\lambda}=(\tilde{K}'\tilde{C}\tilde{K})^{-1}(\tilde{K}'\tilde{C}\tilde{X}\tilde{Y}-\tilde{m}),$$

has expectation

$$E(S^2)=(m-q+s_2)\sigma^2.$$

PROOF. This corollary is an application of corollary 2.4 of section 4 to the model (5.8). Note that $\tilde{X}$ is $p\times m$, which is why m appears in place of n in the formula for $E(S^2)$.

The first two sentences of the proof of corollary 4.2 show that if $\tilde{X}=XT$ has rank q, then the class of pre-estimable functions is not reduced in passing to the model (5.8). However if $\tilde{X}$ has rank $\tilde{q}(\tilde{q}<q)$, then in passing to (5.8) H is replaced by a $p\times\tilde{r}$ matrix $\tilde{H}$ of rank $\tilde{r}=p-\tilde{q}$ such that $\tilde{H}'X=0$. Such a matrix can be obtained as $\tilde{H}=[H_0, H]$ where H_0 is an appropriate $p\times(q-\tilde{q})$ matrix.

It is desirable to have a system of equations for $\hat{\beta}$ (in Theorem 4) when F_2 is *not* null. Such systems can be obtained (and other information derived) by applying the material of section 4 to the model (5.8). In doing so it should be kept in mind that the restraints $\tilde{K}'\beta=\tilde{m}$ must be separated into those which are pre-estimable (this is the sole class when $\tilde{q}=q$, as already noted), and those which are not; the latter must be examined for irreducibility (see the paragraph preceding Theorem 2) and "reduced" if necessary.

It is natural, as a next step, to consider a model which involves *both* the complications of linear restraints on the β and an arbitrary variance-covariance matrix $V\sigma^2$. This requires no new extension of the theory, since the only addition is that of the restraints $K_i'\beta = m_i$ $(i=1, 2)$ where $K_2'\beta$ represent pre-estimable restraints. It is quite possible that some of the restraints $K'\beta$ coincide with the restraints $\tilde{K}'\beta$ in which case the duplicate restraints in $K_2'\beta$ (or $\tilde{K}_2'\beta$) may be dropped. Aside from this duplication one will then have the situation

$$E(Y)=X'\beta, \text{ var } Y=\sigma^2 V \ (V \text{ has rank } m \ (m \leqslant n))$$

$$K_i'\beta = m_i \ (i=1, 2)$$

which is identical for purposes of estimating $\theta = l'\beta$ with

$$\begin{aligned} &E(\tilde{Y})=\tilde{X}'\beta, \text{ var } \tilde{Y}=\sigma^2 I \\ &K_i'\beta = m_i, \ \tilde{K}'\beta = \tilde{m} \end{aligned} \tag{5.11}$$

where the tilde (˜) quantities are defined as in Theorem 4. A formal proof is obtained by applying Theorem 4 to

$$E(Y^*)=(X^*)'\beta, \text{ var } Y^* = V^*\sigma^2$$

in which

$$(Y^*)'=[Y', m_1', m_2'], X^*=[X, K_1, K_2],$$

$$V^* = \begin{bmatrix} V & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Suppose for example that X is of full rank (i.e., $q=p$) and that V is nonsingular (i.e., $m=n$), and consider the model [17a]

$$E(Y)=X'\beta, \text{ var}(Y)=\sigma^2 V \tag{5.11a}$$

$$K'\beta = m,$$

where K is a $p\times k$ matrix of rank k and m is a $k\times 1$ vector, both consisting of known constants. By the prescription given in the last paragraph, and from the fact that $m=n$ implies that $\tilde{K}$ in Theorem 4 is null, we see that an appropriate $\hat{\beta}$ will be one for the model

$$E(\tilde{Y})=\tilde{X}'\beta, \text{ var } (\tilde{Y})=\sigma^2 I \tag{5.11b}$$

$$K'\beta = m \tag{5.11c}$$

where $\tilde{X}=XT$, $\tilde{Y}=T'Y$, and $T=G\Lambda^{-1/2}$. G is of rank $m=n$ and so T is $n\times n$ nonsingular, implying (since $q=p$) that $\tilde{X}$ is of rank p. Hence the constraints $K'\beta$ are all pre-estimable with respect to (5.11b), i.e. $K=K_2$ in our previous notation. Applying (4.20), we see that we can take

$$\hat{\beta} = \tilde{\beta}_0 + \tilde{C}K(K'\tilde{C}K)^{-1}(m - K'\tilde{C}\tilde{X}\tilde{Y}) \tag{5.11d}$$

[17a] In the rest of this subsection, use of the symbol m both for the rank of V (here $m=n$), and for the $k\times 1$ vector in (5.11a), should cause no confusion.

where $\tilde{C}$ is related to $\tilde{A}=\tilde{X}\tilde{X}'=XV^{-1}X'$ as C is to A, and where $\tilde{\beta}_0$ is the unique solution of $\tilde{A}\tilde{\beta}_0=\tilde{X}\tilde{Y}$. Since $\tilde{A}$ is $p\times p$ nonsingular, we have

$$\tilde{C}=\tilde{A}^{-1}=(XV^{-1}X')^{-1},$$

$$\tilde{\beta}_0=\tilde{C}\tilde{X}\tilde{Y}=(XV^{-1}X')^{-1}XV^{-1}Y,$$

so that substitution in (5.11d) yields

$$\begin{aligned} \hat{\beta} =& (XV^{-1}X')^{-1}\{XV^{-1}Y \\ &+ K[K'(XV^{-1}X')^{-1}K]^{-1}(m - K'(XV^{-1}X')^{-1}XV^{-1}Y)\}, \end{aligned}$$

in agreement with the result obtained for this special case by Chipman and Rao [1964].

5.3. Simplification of the Normal Equations

In the model

$$E(Y)=X'\beta, \qquad \text{var } Y=\sigma^2 V,$$

we will have $XF=0$ (so that the normal equations are given by (5.9)) if and only if $X=XGG'$. For, if this condition holds then

$$XF=(XG)(G'F)=0,$$

while if $XF=0$ then $X=MG'$ for some $p\times m$ matrix M (i.e., the rows of X are linear combinations of the orthonormalized characteristic vectors of V), and postmultiplication by G yields $M=XG$.

In particular, this will be the case if

$$XV^+=BX, \qquad X=BXV \tag{5.12}$$

for some nonsingular $p\times p$ matrix B. For, the first condition in (5.12) yields $X=MG'$ with $M=B^{-1}XG\Lambda^{-1}$, while the second yields it with $M=BXG\Lambda$. The two conditions of (5.12) are logically equivalent, for the first implies

$$X=XGG'=(XG\Lambda^{-1}G')(G\Lambda G')=(XV^+)V=BXV$$

while the second implies

$$XV^+=BXVV^+=BXGG'=BX.$$

If (5.12) holds then the normal equations are

$$(XV^+X')\hat{\beta}=XV^+Y,$$

and become

$$(BXX')\hat{\beta}=BXY,$$

which are equivalent to the usual normal equations $A\hat{\beta}=XY$ obtained when $V=I$. This result seems to have been first noted by T. W. Anderson [1948], and Muller and Watson [1959] have discussed it in the context of randomization theory.

For the rest of this section we assume $XF=0$ (i.e., $X=XGG'$), and ask when a simplification of the normal equations something like the one described above is possible. Note that $X=XGG'$ implies $q \leqslant m$. If $q=m$ we can partition

$$X'=[X_1', X_2'], \quad \beta'=[\beta_1', \beta_2']$$

where X_1 is $q \times n$ of rank q, X_2 is $r \times n$, β_1 is $q \times 1$ and β_2 is $r \times 1$. The normal equations become

$$(X_1V^+X_1')\hat{\beta}_1+(X_1V^+X_2')\hat{\beta}_2=X_1V^+Y,$$

$$(X_2V^+X_1')\hat{\beta}_1+(X_2V^+X_2')\hat{\beta}_2=X_2V^+Y.$$

Since $X=XGG'$ implies $X_1=X_1GG'$, X_1G has rank $q=m$ and hence

$$X_1V^+=X_1G\Lambda^{-1}G'=B_1X_1, \qquad B_1=(X_1G)\Lambda^{-1}(X_1G)^{-1}$$

where B_1 is $q \times q$ nonsingular. Premultiplication of the first normal equations by B_1^{-1} (after substitution of B_1X_1 for X_1V^+) yields

$$(X_1X_1')\hat{\beta}_1+(X_1X_2')\hat{\beta}_2=X_1Y \tag{5.13}$$

$$(X_2X_1'B_1')\hat{\beta}_1+(X_2V^+X_2')\hat{\beta}_2=X_2V^+Y.$$

Thus at least the first subsystem of the normal equations has been somewhat simplified.

If in particular $p=q=m$ then we have (5.12) and the resulting full simplification. If $p>q=m$ but $X_1X_2'=0$, then the normal equations reduce to

$$(X_1X_1')\hat{\beta}_1=X_1Y \tag{5.14}$$

$$(X_2V^+X_2')\hat{\beta}_2=X_2V^+Y,$$

and the solutions for $\hat{\beta}_1$ are the same as if var $Y=\sigma^2I$. Without assuming $X_1'X_2=0$, we can observe that $X_2'=X_1'N$ for some $q \times r$ matrix N and that $A_1=X_1X_1'$ is $q \times q$ nonsingular; thus the first subsystem of (5.13) can be solved for $\hat{\beta}_1$ as

$$\hat{\beta}_1=A_1^{-1}X_1Y-N\hat{\beta}_2,$$

and the second subsystem becomes

$$N'(B_1-A_1B_1'A_1^{-1})A_1N\hat{\beta}_2=N'(B_1-A_1B_1'A_1^{-1})X_1Y$$

which is to be solved for $\hat{\beta}_2$. If in addition X_2 has rank r (which requires $r \leqslant g$), then N does too and one can first find the unique $\tilde{\beta}_1$ such that $A_1\tilde{\beta}_1=X_1Y$, and then satisfy the second subsystem by solving $N\hat{\beta}_2=\tilde{\beta}_1$, i.e. $\hat{\beta}_2=(N'N)^{-1}N'\tilde{\beta}_1$.

If $q<m$ the situation is more complicated. This is illustrated by the following example (due to K. Goldberg, NBS) in which $p=q=1$ and $n=m=2$. Take $X=[1, 0]$ and

$$G=G'=2^{-1/2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}, \qquad \Lambda=\begin{bmatrix}1 & 0\\ 0 & \frac{1}{2}\end{bmatrix}$$

Then $XV^+=XG\Lambda^{-1}G'=[3/2, -1/2]$ but BX has the form $[t, 0]$ for all 1×1 matrices B; hence (5.12) or its analog $X_1V^+=B_1X_1$ cannot hold.

Even when $q<m$, some simplification is possible if there is a partition $G=[G_1, G_2]$, with G_1 ($n \times q$, such that $X_1G_2=0$. For then, if

$$\Lambda=\operatorname{diag}(\Lambda_1, \Lambda_2)$$

denotes the decomposition of Λ corresponding to the partition of G, we have

$$X_1=X_1GG'=X_1(G_1G_1'+G_2G_2')=X_1G_1G_1',$$

$$X_1V^+=X_1(G_1\Lambda_1^{-1}G_1'+G_2\Lambda_2^{-1}G_2')=X_1G_1\Lambda_1^{-1}G_1'$$

and can mimic the procedure for $q=m$ (up to and including (5.14)) using G_1 and Λ_1 instead of G and Λ. At present it is not clear what other cases admit analogous simplification if $q<m$. One such situation arises if we change the dimensions of the partition of X' so that X_i is $p_i \times n$ ($p_1+p_2=p$), β_i is $p_i \times 1$, and X_1 has rank p_1 (implying $p_1 \leqslant q$). If there is a partition $G=[G_1, G_2]$, with G_1 ($p_1 \times n$), such that $X_1G_2=0$, then the preceding analysis still applies.

5.4. Equicorrelated Variables

In many experimental situations the covariances between the observations are not zero, but to a reasonable degree of approximation may be regarded as being equal; i.e., cov $(Y_i, Y_j)=\rho\sigma^2$ $(i \neq j)$. Therefore we can write var $Y=V\sigma^2$ where

$$V=(1-\rho)I+\rho J \tag{5.15}$$

and J is an $n \times n$ matrix with all elements unity.

The matrix V has the two distinct characteristic roots $[1+(n-1)\rho]$ and $(1-\rho)$ with multiplicities one and $(n-1)$ respectively. However since $V\sigma^2$ is a variance-covariance matrix, it is positive semidefinite and the roots are nonnegative. Consequently

$$1+(n-1)\rho \geqslant 0, \qquad 1-\rho \geqslant 0$$

and we obtain the bounds $-(n-1)^{-1} \leqslant \rho \leqslant 1$. When either $\rho=-(n-1)^{-1}$ or $\rho=1$, a characteristic root will be zero and V will be singular. If the Y_i are in creasing linear functions of one another, ρ will be equal to unity. The case $\rho=-(n-1)^{-1}$ implies that $\sum_{i=1}^{n} Y_i=\text{constant}$, $V^+=(n-1)n^{-1}\{I-J/n\}$, and that the sum of the elements in any row or column of V is zero.

When $\rho \neq 1$ or $\rho \neq (n-1)^{-1}$, V has an inverse which is given by

$$V^{-1}=(1-\rho)^{-1}\{I-\rho[1+(n-1)\rho]^{-1}J\}.$$

Therefore using (5.10) the normal equations can be written

$$\{[A-\rho[1+(n-1)\rho]^{-1}XJX'\}\hat{\beta}=XY-\rho[1+(n-1)\rho]^{-1}XJY. \tag{5.16}$$

where $A = XX'$. The conditions for estimability of a parametric function only involve first moments and hence are not dependent on ρ. Therefore the function $\theta = l'\beta$ is estimable if and only if $H'l = 0$ where $H'X = 0$. The solution of (5.16) involves knowledge of ρ. However we wish to determine the parametric functions which can be estimated *without* knowledge of ρ.

Let $\mathbf{1}$ denote an $n \times 1$ vector of ones, so that $J = \mathbf{1}\,\mathbf{1}'$. Since $\mathbf{1}$ and therefore $n^{-1/2}\mathbf{1}$ is a characteristic vector of V corresponding to the characteristic root $[1 + (n-1)\rho]$, we can take the matrix G of subsection 5.1 as $G = [n^{-1/2}\mathbf{1}, M]$. Here the $n-1$ columns of M are characteristic vectors of V corresponding to the characteristic root $(1-\rho)$, and

$$M'M = I_{n-1}, \qquad \mathbf{1}'M = 0, \qquad VM = (1-\rho)M.$$

Since G is square, $G'G = I$ implies $GG' = I$ and therefore

$$MM' = I - n^{-1}J.$$

Also, in the notation of subsection 5.1,

$$T = G\Lambda^{-1/2} = [n^{-1/2}(1 + (n-1)\rho)^{-1/2}\mathbf{1}, (1-\rho)^{-1/2}M],$$

so in applying Theorem 4

$$\tilde{X} = XT = [n^{-1/2}(1 + (n-1)\rho)^{-1/2}X\mathbf{1}, (1-\rho)^{-1/2}XM],$$

$$\tilde{Y} = T'Y = \begin{bmatrix} n^{-1/2}(1 + (n-1)\rho)^{-1/2}\mathbf{1}'Y \\ (1-\rho)^{-1/2}M'Y \end{bmatrix}.$$

The equation $E(\tilde{Y}) = \tilde{X}'\beta$ of (5.8) therefore becomes equivalent to

$$E(\mathbf{1}'Y) = \mathbf{1}'X'\beta, \tag{5.17}$$

$$E(\bar{Y}) = \bar{X}'\beta \qquad (\bar{X} = XM, \bar{Y} = M'Y) \tag{5.18}$$

and it is readily verified that

$$\text{var}\,(\bar{Y}) = (1-\rho)\sigma^2 I, \qquad \text{cov}\,(\mathbf{1}'Y, \bar{Y}) = 0. \tag{5.19}$$

The unbiased estimates of any estimable $\theta = l'\beta$ have the form

$$g(Y) = d'\bar{Y} + e(\mathbf{1}'Y)$$

where d is an $(n-1) \times 1$ vector, e is a scalar, and

$$\bar{X}d + X\mathbf{1}e = l. \tag{5.20}$$

Also,

$$\text{var}\,[g(Y)] = (1-\rho)\sigma^2 d'd + e^2 n[1 + (n-1)\rho]\sigma^2$$
$$= (1-\rho)\sigma^2[d'd + e^2 n] + e^2 n^2 \rho\sigma^2.$$

First suppose the rank of $\bar{X} = XM$ is less than the rank q of X. This rank must be $q-1$. Then there exists an $m \times 1$ vector h such that $h'X = \mathbf{1}$, and (5.20) yields

$$h'l = h'XMd + h'X\mathbf{1}e = ne,$$

so that

$$e = e_0 = n^{-1}(l'h)$$

in every estimate $g(Y)$.[18] Thus the minimization of var $[g(Y)]$ subject to (5.20) is achieved by choosing d to minimize $d'd$ subject to

$$Xd = l - X\mathbf{1}e_0 = l - Ahe_0.$$

This, however, coincides with the problem of finding a best estimate for $(l - Ahe_0)'\beta$ in the model specified by (5.18) and (5.19); the Gauss theorem yields the solution as

$$d'\bar{Y} = (l - Ahe_0)'\tilde{\beta}$$

where $\tilde{\beta}$ is any solution of the normal equations obtained using (5.18) and (5.19). Since $MM' = I - n^{-1}J$, we find that these normal equations are

$$X[I - n^{-1}J]X'\tilde{\beta} = X[I - n^{-1}J]Y. \tag{5.21}$$

Thus the best estimate $g(Y)$ of $\theta = l'\beta$ is

$$\hat{\theta} = (l - Ahe_0)'\tilde{\beta} + \mathbf{1}'Ye_0$$
$$= l'\tilde{\beta} - l'h(n^{-1}h'A\tilde{\beta}) + l'h(n^{-1}\mathbf{1}'Y)$$
$$= l'\hat{\beta}$$

where $\hat{\beta} = \tilde{\beta} + n^{-1}h(\mathbf{1}'Y - \mathbf{1}'X'\tilde{\beta})$. Use of (5.21) and $h'X = \mathbf{1}'$ leads to $A\hat{\beta} = XY$. Conversely if $\hat{\beta}$ is any solution of $A\hat{\beta} = XY$, then choosing $\tilde{\beta} = \hat{\beta}$ yields a solution of (5.21), and also

$$\hat{\theta} = (l - Ahe_0)'\hat{\beta} + \mathbf{1}'Ye_0 = l'\hat{\beta}.$$

Now assume $\bar{X} = XM$ has the same rank q as X. To minimize var $[g(Y)]$, first treat e as fixed; as in the previous case we are led to the choice

$$d'\bar{Y} = (l - X\mathbf{1}e)'\tilde{\beta}$$

where $\tilde{\beta}$ is any solution of (5.21). The rank hypothesis implies that the same H, and thus the same K and C, work for $\bar{X}\bar{X}'$ as for A, and so we may choose $\tilde{\beta} = C\bar{X}\bar{Y}$. Now

$$g(Y) = l'\tilde{\beta} + e\mathbf{1}'(Y - X'\tilde{\beta})$$

$$\text{var}\,[g(Y)] = \text{var}\,(l'\tilde{\beta}) + e^2\,\text{var}\,[\mathbf{1}'(Y - X'\tilde{\beta})]$$
$$+ 2e\,\text{cov}\,[l'\tilde{\beta}, \mathbf{1}'(Y - X'\tilde{\beta})].$$

The range of e in the remaining minimization problem is that of all real numbers. To prove this, note that by the rank hypothesis $l = \bar{X}d_0$ for some $(n-1) \times 1$

[18] A simple sufficient condition for the existence of such a vector h is that the columns of X all sum to some nonzero constant k. That is, $\mathbf{1}'X = k\mathbf{1}'$ and thus we may take $h = k^{-1}\mathbf{1}$.

vector d_0 (i.e., each estimable form is also estimable with respect to (5.18)), and also $X=\bar{X}N$ for some matrix N, so that for any real number e

$$l-X\mathbf{1}e=\bar{X}d \qquad (d=d_0-N\mathbf{1}e)$$

as required by (5.20). The solution of the minimization problem is therefore

$$e=-\operatorname{cov}[l'\tilde{\beta},\ \mathbf{1}'(Y-X'\tilde{\beta})]/\operatorname{var}[\mathbf{1}'(Y-X'\tilde{\beta})]$$

$$=(1-\rho)l'CX\mathbf{1}/\{n[1+(n-1)\rho]+(1-\rho)\mathbf{1}'X'CX\mathbf{1}\}.$$

This is independent of ρ if and only if the numerator vanishes, i.e.,

$$l'CX\mathbf{1}=0, \tag{5.22}$$

and in that event the best estimate reduces to

$$l'\tilde{\beta}=l'CX(I-n^{-1}J)Y=l'CXY=l'\hat{\beta}$$

where $\hat{\beta}$ is a solution of $A\hat{\beta}=XY$. Note that $l'CX\mathbf{1}=0$ will hold for *all* estimable functions if and only if $X\mathbf{1}=0$.

Before assembling these results (with a few more substitutions) into a formal theorem, we remark that XM has the same rank as

$$XMM'X'=X(I-n^{-1}J)X'=X(I-n^{-1}J)^2X',$$

and thus the same rank as the matrix

$$X^*=X[I-n^{-1}J]$$

obtained from X by simply taking deviations from the mean, i.e.,

$$x^*_{i\alpha}=x_{i\alpha}-\bar{x}_i; \qquad \bar{x}_i=n^{-1}\sum_{\alpha=1}^{n}x_{i\alpha}.$$

Thus when X^* and X have the same rank, a solution of (5.21) can be obtained as

$$\tilde{\beta}=CX[I-n^{-1}J]Y=\hat{\beta}-n^{-1}CXJY.$$

THEOREM 5. *Let*

$$\mathrm{E(Y)}=\mathrm{X}'\beta,\ \operatorname{var}(\mathrm{Y})=\sigma^2[(1-\rho)\mathrm{I}+\rho\mathrm{J}],$$

$-(\mathrm{n}-1)^{-1}<\rho<1$. *If the rank* q^* *of* $\mathrm{X}^*=\mathrm{X}[\mathrm{I}-\mathrm{n}^{-1}\mathrm{J}]$ *is* $\mathrm{q}-1$ (*i.e., there exists an* $\mathrm{n}\times 1$ *vector* h *such that* $\mathrm{h}'\mathrm{X}=\mathbf{1}'$), *the normal equations are* $\mathrm{A}\hat{\beta}=\mathrm{XY}$ *and do not depend on* ρ. *When* $\mathrm{q}^*=\mathrm{q}$, *the only estimable functions* $\theta=l'\beta$ *with best estimate independent of* ρ *are those with*

$$l'CX\mathbf{1}=0, \tag{5.22}$$

and for these the best estimate is $l'\hat{\beta}$ *with* $\mathrm{A}\hat{\beta}=\mathrm{XY}$. *If* (5.22) *does not hold, the best estimate of* θ *is*

$$\hat{\theta}=l'[\hat{\beta}-\mathrm{n}^{-1}\mathrm{CXJY}+\{\mathrm{n}[\mathrm{n}\rho(1-\rho)^{-1}+1]+\mathbf{1}'\mathrm{X}'\mathrm{CX}\mathbf{1}\}^{-1}$$

$$(\mathrm{Y}-\mathrm{X}'\hat{\beta}+\mathrm{n}^{-1}\mathrm{X}'\mathrm{CXJY})]$$

with $\hat{\beta}$ *as above.*

COROLLARY 5.1. *Let the deviations* Y^* *and* X^* *be defined by* $\mathrm{Y}^*=\mathrm{M}\bar{\mathrm{Y}}=\mathrm{MM}'\mathrm{Y}$, $\mathrm{X}^*=\bar{\mathrm{X}}\mathrm{M}'=\mathrm{XMM}'$. *Then the quantity*

$$\mathrm{S}^2=(\mathrm{Y}^*-\mathrm{X}^{*\prime}\hat{\beta})'(\mathrm{Y}^*-\mathrm{X}^{*\prime}\hat{\beta})$$

has expectation

$$\mathrm{E}(\mathrm{S}^2)=(\mathrm{n}-\mathrm{q})(1-\rho)\sigma^2$$

if a vector h *exists for which* $\mathrm{h}'\mathrm{X}=\mathbf{1}'$; *otherwise the expectation of* S^2 *is*

$$\mathrm{E}(\mathrm{S}^2)=(\mathrm{n}-\mathrm{q}-1)(1-\rho)\sigma^2.$$

PROOF. The expectation of $(\bar{Y}-\bar{X}'\hat{\beta})'(\bar{Y}-\bar{X}'\hat{\beta})$ is $(n-q)(1-\rho)\sigma^2$ if $\bar{X}$ has rank $q-1$ (i.e., a vector h exists for which $h'X=1$). When X has rank q, the expectation is $(n-q-1)(1-\rho)\sigma^2$. These results immediately follow by applying corollary 1.4 of the Gauss theorem. Since

$$\bar{Y}-\bar{X}'\hat{\beta}=M'(Y-X'\hat{\beta})$$

and $(MM')^2=MM'=I-n^{-1}J$, we have

$$\begin{aligned}(\bar{Y}-\bar{X}'\hat{\beta})'(\bar{Y}-\bar{X}'\hat{\beta})&=(Y-X'\hat{\beta})'MM'(Y-X'\hat{\beta})\\&=(Y^*-X^{*\prime}\hat{\beta})'(Y^*-X^{*\prime}\hat{\beta}).\end{aligned}$$

The problem arises as to what to do if there does not exist an h for which $h'X=\mathbf{1}'$ (i.e., if XM has the same rank q as X), ρ is unknown, and we wish to estimate an estimable function for which $l'CX\mathbf{1}\neq 0$. Estimates of $\theta=l'\beta$ can be obtained *if* we are willing to consider the alternate estimation problem where

$$E(\bar{Y})=\bar{X}'\beta,\ \operatorname{var}\bar{Y}=(1-\rho)\sigma^2I \tag{5.23}$$

subject to the restraint

$$\mathbf{1}'X'\beta=\mathbf{1}'Y,$$

which must be pre-estimable.

Application of (4.16) results in the normal equations

$$X(I-n^{-1}J)X'\hat{\beta}+X\mathbf{1}\lambda=X(I-n^{-1}J)Y,$$

$$\mathbf{1}'X'\hat{\beta}=\mathbf{1}'Y,$$

which reduce to

$$A\hat{\beta}+X\mathbf{1}\lambda=XY,$$

$$\mathbf{1}'X'\hat{\beta}=\mathbf{1}'Y.$$

After premultiplication by the nonzero $1\times p$ vector $\mathbf{1}'X'$, the first equation can be solved for λ to obtain

$$\lambda=(\mathbf{1}'X'X\mathbf{1})^{-1}\mathbf{1}'X'(XY-A\hat{\beta}),$$

and then the normal equations for $\hat{\beta}$ alone are obtained as

$$[I-(\mathbf{1}'X'X\mathbf{1})^{-1}XJX']A\hat{\beta}=[I-(\mathbf{1}X'X\mathbf{1})^{-1}XJX']XY$$

$$\mathbf{1}'X'\hat{\beta}=\mathbf{1}'Y.$$

Alternatively, we can apply Theorem 2 to the model given by (5.23) and the restraint $\mathbf{1}'X'\beta=\mathbf{1}'Y$. Since $\bar{X}$ has the same rank as X, H has the same relation to $\bar{X}$ as to X. Thus, by Lemmas 1 and 2, the matrices K and C are the same for $\bar{X}$ as for X. Here H_1 and K_1 are null, while H_0 and K_0 correspond to H and K, respectively. The result of applying Theorem 2 is given next as another theorem; note that the estimate $\hat{\theta}=l'\hat{\beta}$ coincides with that given in Theorem 5 when $l'CX\mathbf{1}=0$.

THEOREM 6. *For the model*

$$\mathrm{E(Y)=X'}\beta,\qquad \mathrm{var\ Y}=\sigma^2[(1-\rho)\mathrm{I}+\rho\mathrm{J}],$$

where $-(\mathrm{n}-1)^{-1}<\rho<1$, *the parameter* ρ *is unknown, and* $\bar{\mathrm{X}}=\mathrm{XM}$ *has the same rank as* X, *the best estimate conditional on* $\mathbf{1}'X'\beta=\mathbf{1}'Y$ *of the estimable function* $\theta=l'\beta$ *is given by*

$$\hat{\theta}=l'\hat{\beta}=l'\mathrm{CX}\{(\mathrm{I}-\mathrm{n}^{-1}\mathrm{J})$$

$$+(\mathbf{1}'\mathrm{X}'\mathrm{CX}\mathbf{1})^{-1}\mathrm{J}[\mathrm{I}-\mathrm{X}'\mathrm{CX}(\mathrm{I}-\mathrm{n}^{-1}\mathrm{J})]\}\mathrm{Y}.\quad(5.24)$$

COROLLARY 6.1. *The quantity*

$$\mathrm{S}^2=(\mathrm{Y}^*-(\mathrm{X}^*)'\hat{\beta})'(\mathrm{Y}^*-(\mathrm{X}^*)'\hat{\beta}),$$

where $\mathrm{X}^*=\mathrm{X}(\mathrm{I}-\mathrm{n}^{-1}\mathrm{J})$ *and* $\mathrm{Y}^*=(\mathrm{I}-\mathrm{n}^{-1}\mathrm{J})\mathrm{Y}$, *has the conditional expectation*

$$\mathrm{E}(\mathrm{S}^2|\mathbf{1}'\mathrm{Y})=(1-\rho)\sigma^2(\mathrm{n}-\mathrm{q}).$$

PROOF. We first observe that

$$S^2=(\bar{Y}-\bar{X}'\hat{\beta})'(\bar{Y}-\bar{X}'\hat{\beta}),$$

so that $E(S^2|\mathbf{1}'Y)$ can be found by applying Corollary 2.4 to the model consisting of (5.23) and the restraint $\mathbf{1}'X'\beta=\mathbf{1}'Y$. Here $k_2=1$, n is replaced by $n-1$ since $\bar{X}$ is $p\times(n-1)$, and σ^2 is replaced by $(1-\rho)\sigma^2$. This proof also shows, by (4.27), that S^2 can be written as

$$S^2=(Y^*-(X^*)'\hat{\beta}_0)'(Y^*-(X^*)\hat{\beta}_0)+\lambda^2(\mathbf{1}'X'CX\mathbf{1})$$

where $\hat{\beta}_0=CX(I-n^{-1}J)Y$ and

$$\lambda=(\mathbf{1}'X'CX\mathbf{1})^{-1}[\mathbf{1}'X'CX(I-n^{-1}J)Y-\mathbf{1}'Y].$$

5.5. Two Stage Least Squares

An application of Theorem 4 arises in two stage least squares estimation which has recently been discussed by Freund, Vail, and Clunies-Ross (1961) and Goldberger and Jockems (1961). We shall consider some further generalizations and discuss the matter more fully. Consider the model

$$E(Y)=X_1'\beta_1+X_2'\beta_2,\ \mathrm{var}\ Y=\sigma^2I\qquad(5.25)$$

where X_i are $p_i\times n$ matrices and β_i are $p_i\times 1$ vectors for $i=1,2$. Instead of considering the full model, in the first stage we ignore β_2 variables and take $E(Y)=X_1'\beta_1$. Then the normal equations will yield the solution

$$\tilde{\beta}_1=C_1X_1Y\qquad(5.26)$$

where C_1 is related to X_1X_1' as C is to A.

Define the residual vector

$$\delta=Y-X_1'\tilde{\beta}_1=(I-X_1'C_1X_1)Y$$

and the idempotent matrix

$$V=I-X_1'C_1X_1.$$

Then we have

$$E(\delta)=VX_2'\beta_2,$$

$$\mathrm{var}\,(\delta)=V\sigma^2,$$

and these equations serve as the model for the *second* stage. Now apply Theorem 4 to this model: $V=V^+$ since V is idempotent, the analogs of X and $F'=F_1'$ are X_2V and X_1 respectively with $X_1(X_2V)'=0$ since $X_1V=0$, and so the result is the equation

$$(X_2VX_2')\hat{\beta}_2=X_2V\delta=X_2VY\qquad(5.27)$$

with solution

$$\hat{\beta}_2=C_2X_2V\delta=C_2X_2VY,\qquad(5.28)$$

where C_2 is related to X_2VX_2' as is C to A.

Suppose $\theta=l_1'\beta_1+l_2'\beta_2$ is estimable in the full model. Then (see (3.4)) there exists an $n\times 1$ vector d such that $X_id=l_i(i=1,2)$, and so $\theta_1=l_1'\beta_1$ is estimable in the first-stage model. Its best estimate in that model is

$$\tilde{\theta}_1=l_1'\tilde{\beta}_1,$$

and in the full model

$$E(\tilde{\theta}_1)=l_1'C_1X_1(X_1'\beta_1+X_2'\beta_2)=\theta_1+l_1'C_1X_1X_2'\beta_2.\quad(5.29)$$

The procedure to be described involves adding a term to $\tilde{\theta}_1$ to obtain an unbiased estimate $\hat{\theta}_1$ of θ_1. Clearly this will be possible only if θ_1 is in fact estimable in the full model. We therefore are led to determine what condition on the partition $[X_1', X_2']$ will ensure that $\theta_1=l_1'\beta_1$ is estimable in the full model whenever $\theta=l'\beta$ is. First suppose the partition has this property. Since the rows of $X'\beta$ are estimable in the full model, the same must hold for the rows of $X_1'\beta_1$ and thus for the rows of

$$X_2'\beta_2=X'\beta-X_1'\beta_1.$$

By (3.4) there is an $n\times n$ matrix B such that $X_1B=0$ and $X_2B=X_2$. $X_1B=0$ implies that $B=VB_1$ for some $p_1\times n$ matrix B_1, and so $X_2VB_1=X_2$. The last equation shows that the rows of $X_2'\beta_2$ are estimable in the second-stage model, or equivalently (by Corollary 1.1)

$$X_2=X_2VX_2'C_2X_2. \tag{5.30}$$

Conversely, suppose (5.30) holds and that

$$\theta=l_1'\beta_1+l_2'\beta_2$$

is any parametric function estimable in the full model. By (3.4), there exists an $n\times 1$ vector d such that

$$X_id=l_i\ (i=1,2).$$

Then

$$X_1(I-VX_2'C_2X_2)d=l_1,\ X_2(I-VX_2'C_2X_2)d=0$$

so that by (3.4) $\theta_1=l_1'\beta_1$ is estimable in the full model. Hence (5.30) is exactly the required condition on $[X_1', X_2']$, and is assumed in what follows.

An unbiased estimate of θ_1 in the full model can now be given as

$$\hat{\theta}_1=\tilde{\theta}_1-l_1'C_1X_1X_2'C_2X_2VY=l_1'\hat{\beta}_1.$$

Since θ and θ_1 are estimable in the full model, the same is true of

$$\theta_2=l_2'\beta_2=\theta-\theta_1,$$

so that

$$X_1d_2=0,\ X_2d_2=l_2$$

for some $n\times 1$ vector d_2. From this and (5.30) it can be verified that

$$\hat{\theta}_2=l_2'\hat{\beta}_2$$

is an estimate (therefore the best estimate) of θ_2 in the second-stage model, and also an unbiased estimate of θ_2 in the full model.

It has been shown that an unbiased estimate of the estimable function

$$\theta=l_1'\beta_1+l_2'\beta_2$$

is given by

$$\hat{\theta}=\hat{\theta}_1+\hat{\theta}_2=l_1'\hat{\beta}_1+l_2'\hat{\beta}_2,$$

where

$$\hat{\beta}_1=C_1[I+X_1X_2'C_2X_2X_1'C_1]X_1Y-C_1X_1X_2'C_2X_2Y,$$
$$\hat{\beta}_2=C_2X_2[I-X_1'C_1X_1]Y. \tag{5.31}$$

The solutions (5.31) can be shown by substitution to satisfy the normal equations

$$\begin{bmatrix} X_1X_1' & X_1X_2' \\ X_2X_1' & X_2X_2' \end{bmatrix}\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}=\begin{bmatrix} X_1Y \\ X_2Y \end{bmatrix}$$

of the full model, and so $\hat{\theta}$ is the minimum variance linear unbiased estimate of θ.

For the same reason, $\hat{\theta}_i=l_i'\hat{\beta}_i$ is the best estimate of $\theta_i=l_i'\beta_i$ in the full model. In terms of this model alone, the following result has been proved: If the portions of every estimable function which respectively involve the β_1 and β_2 variables are separately estimable, then the best estimate of each such function is simply the sum of the best estimates of its portions. In this sense the condition (5.30) can be regarded as a generalization of orthogonality ($X_2X_1'=0$); in the orthogonal case the normal eqs (5.27) of the second-stage model are simply

$$(X_2X_2')\hat{\beta}_2=X_2Y$$

in direct analogy to those of the first-stage model. Note also that (5.30) automatically holds if $q=p$ (i.e., if A is nonsingular), since then every parametric function $l'\beta$, in particular $l_1'\beta_1$, is estimable.

5.6. Restraints Subject to Uncertainty

Occasionally situations arise in which the given restraints $K'\beta=m$ are themselves subject to variation. Such may be the case when the value of $K'\beta$ is not known but prior information is available which can be summarized as a value of a random vector $\hat{m}$ with $E(\hat{m})=K'\beta$ and with precision described by var $(\hat{m})=V_m\sigma^2$. A circumstance where this may occur is when data are available from another source which is believed to be without bias or systematic error.

Let $E(Y)=X'\beta$, var $Y=\sigma^2I$ and let the k "given" restraints consist of unbiased estimates $\hat{m}_i$ $(i=1,2)$ of $K_i'\beta$, where K_i is $p\times k_i$ of rank k_i, and $\hat{m}'=(\hat{m}_1', \hat{m}_2')$ obeys var $(\hat{m})=V_m\sigma^2$. Further it is assumed that the restraints $K_1'\beta$ are nonpre-estimable functions and $K_2'\beta$ are pre-estimable functions with respect to the observational equations $E(Y)=X'\beta$. It is desired to perform estimation subject to the additional conditions $K_i'\hat{\beta}=\hat{m}_i$, i.e., to fit the new data so that the quantities $K'\beta$ are exactly equal to $\hat{m}$. We may assume without loss of generality that $K_2'K_2=I$ and that the restraints $K_1'\beta$ are irreducible.

It will be convenient to introduce the expression *undisturbed* to refer to those estimable functions $\theta=l'\beta$ whose best estimate $\hat{\theta}=l'\hat{\beta}$ is not altered by the requirement that $\hat{\beta}$ be chosen to satisfy $K'\hat{\beta}=\hat{m}$. Not all estimable functions are undisturbed in general; for example we have no freedom in choosing $\hat{\theta}$ when θ is a linear combination of the rows of $K_2'\beta$. The subclass of the estimable functions, consisting of those which are undisturbed, is a matter of *choice* and its selection would presumably depend on the problem at hand, but it should not contain any nonzero linear combinations of the rows of $K_2'\beta$. (If for example there is skepticism concerning the prior information, then this subclass would chosen to include, so far as possible, those functions for which a minimum variance estimate is of particular importance.) The class of undisturbed functions may be chosen, of the maximum possible dimension, as the class of all linear combinations of the rows of $L'\beta$ where L is a $p\times(q-k_2)$ matrix of rank $q-k_2$ such that $H'L=0$ and $[K_2, L]$ has rank q. We may assume $L'L=I$ without loss of generality.

Because $H'[K_2, L]=0$, and $[K_2, L]$ has the same rank q as X, there exist a $k_2\times n$ matrix P_2 and a $(q-k_2)\times n$ matrix P such that

$$X=[K_2, L]\begin{bmatrix} P_2 \\ P \end{bmatrix}=K_2P_2+LP.$$

These matrices can be found explicitly, in terms of the inverse N^{-1} of the $q\times q$ nonsingular matrix

$$N=\begin{bmatrix} K_2' \\ L' \end{bmatrix}[K_2, L]=\begin{bmatrix} I & K_2'L \\ L'K_2 & I \end{bmatrix},$$

as

$$\begin{bmatrix} P_2 \\ P \end{bmatrix}=N^{-1}\begin{bmatrix} K_2' \\ L' \end{bmatrix}X.$$

Since $E(\hat{m}_2)=K_2'\beta$, we find that $E(Y)=X'\beta$ is equivalent to

$$E(\bar{Y})=\bar{X}'\beta \tag{5.32}$$

where $\bar{Y}=Y-P_2'\hat{m}_2$ and $\bar{X}=LP=X-K_2P_2$. Similarly, under the assumption $\operatorname{cov}(Y, \hat{m}_2)=0$ which is implicit in our situation, it follows that $\operatorname{var}(Y)=\sigma^2I$ is equivalent to

$$\operatorname{var}(\bar{Y})=\bar{V}\sigma^2, \tag{5.33}$$

where $\bar{V}=I+P_2'V_2P_2$ and V_2 is defined by

$$\operatorname{var}(\hat{m}_2)=V_2\sigma^2.$$

Thus the original model $E(Y)=X'\beta$, $\operatorname{var} Y=\sigma^2I$, ignoring the restraint $K'\beta=\hat{m}$, is equivalent to the one given by (5.32) and (5.33).

From the fact that equality holds throughout the sequence

$$q=\operatorname{rank}(X)=\operatorname{rank}(K_2P_2+LP)$$

$$\leqslant \operatorname{rank}(K_2P_2)+\operatorname{rank}(LP)$$

$$\leqslant \operatorname{rank}(K_2)+\operatorname{rank}(L)=k_2+(q-k_2)=q$$

of inequalities, it follows in particular that $\bar{X}=LP$ has the same rank $q-k_2$ as L. Therefore the class of functions estimable with respect to (5.32) consists of all linear combinations of the rows of $L'\beta$.

We next prove that an analog of $\underline{H}$ for (5.32) is given by $\bar{H}=[H, (I-LL')K_2]$. Since $\bar{H}'\bar{X}=0$, it suffices to show that $\bar{H}$ has rank at least $p-(q-k_2)=r+k_2$; since

$$H'(I-LL')K_2=0$$

and H has rank r, it suffices to show that $(I-LL')K_2$ has at least rank k_2. This however follows from the consequence

$$\begin{bmatrix} I \\ -L'K_2 \end{bmatrix}=N^{-1}\begin{bmatrix} K_2' \\ L' \end{bmatrix}(I-LL')K_2$$

of the identity

$$(I-LL')K_2=[K_2, L]\begin{bmatrix} I \\ -L'K_2 \end{bmatrix}.$$

From the irreducibility of the restraints $K_1'\beta$ in the original model, we can deduce that the restraints $K'\beta$, where $K=[K_1, K_2]$, are nonestimable and irreducible with respect to (5.32). Namely,

$$\bar{H}'K=\begin{bmatrix} H'K_1 & 0 \\ K_2'(I-LL')K_1 & K_2'(I-LL')K_2 \end{bmatrix}$$

can be shown to have rank k_1+k_2. For this purpose, observe that $H'K_1$ has rank k_1 so that the same holds for the first block column in $\bar{H}'K$. Also, since $(I-LL')K_2$ has rank k_2, the same is true of

$$\{(I-LL')K_2\}'\{(I-LL')K_2\}=K_2'(I-LL')K_2$$

and thus of the second block column. The presence of the zero block then ensures the result.

Theorem 4 can be applied to the model consisting of (5.32) and (5.33), to obtain a new model analogous to (5.8), and the restraints $K'\beta$ will remain nonestimable in this new model. Thus the conditions $K'\hat{\beta}=\hat{m}$ can simply be adjoined to the normal equations of the new model without affecting the best estimates $\hat{\theta}$ of the functions estimable in this new model . . . i.e. the linear combinations of the rows of $L'\beta$. Thus, as desired, these linear forms have their best estimates "undisturbed" by requiring $K'\beta=\hat{m}$. (Here K plays the role of K_0 in Theorem 2.) If in particular $\bar{V}$ is nonsingular, then by Corollary 4.1 the normal equations become

$$(\bar{X}\bar{V}^{-1}\bar{X}')\hat{\beta}=\bar{X}\bar{V}^{-1}\bar{Y},$$

$$K'\hat{\beta}=\hat{m}.$$

It may also be appropriate to adjoin artificial non-pre-estimable restraints to secure a unique solution for β.

The previous material also permits us to arrive at unbiased estimates, consistent with $K'\hat{\beta}=\hat{m}$, of functions $\theta=l'\beta$ which are estimable in the original model but are *not* linear combinations of the rows of $L'\beta$. From

$$l=ACl=(K_2P_2+LP)X'Cl$$

it follows that

$$\theta=l'CXP_2'K_2'\beta+l'CXP'L'\beta,$$

so that an unbiased estimate is

$$\hat{\theta}=l'CXP_2'\hat{m}_2+l'CXP'L'\hat{\beta}=l'CX(P_2'K_2'+P'L')\hat{\beta}=l'\hat{\beta}$$

with $\hat{\beta}$ as in the last paragraph. Note that although the second summand $(l'CXP'L'\hat{\beta})$ in $\hat{\theta}$ is the best estimate of the second summand of θ, $\hat{\theta}$ as a whole does not coincide with the best estimate of θ in the original model since $\hat{\beta}$ comes from a set of normal equations other than $A\hat{\beta}=XY$. Thus θ *has* been disturbed.

The previous material takes an especially simple form when $L'K_2=0$, i.e., when the estimable functions whose minimum-variance estimation is to be emphasized $(L'\beta)$ are orthogonal to those whose estimates are prescribed $(K_2'\beta)$. Here premultiplication of

$$X=K_2P_2+LP$$

by K_2' and L', respectively, shows that $P_2=K_2'X$ and $P=L'X$. Thus $\overline{H}=[H, K_2]$, and the model (5.32) and (5.33) becomes

$$E(\overline{Y})=\overline{X}'\beta,\ \mathrm{var}(\overline{Y})=\overline{V}\sigma^2$$

with

$$\overline{Y}=Y-X'K_2\hat{m}_2,\ \overline{X}=LL'X,$$

$$\overline{V}=I+X'K_2V_2K_2'X.$$

For a simple but artificial example, suppose

$$X=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},\ K=K_2=\begin{bmatrix}1\\ 0\end{bmatrix},\ L=\begin{bmatrix}0\\ 1\end{bmatrix}$$

i.e., estimation of the second component β_2 of β is of principal importance. Suppose also that

$$\mathrm{var}(\hat{m}_2)=V_2\sigma^2=\tau^2\sigma^2,$$

so that τ indicates the relative precision of the prior information relative to the new measurements under discussion. The previous paragraph applies, and we are led to the model

$$E(\overline{Y})=\overline{X}'\beta,\ \mathrm{var}(\overline{Y})=\overline{V}\sigma^2$$

with

$$\overline{Y}=\begin{bmatrix}y_1-\hat{m}_2\\ y_2\end{bmatrix},\ \overline{X}=\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix},$$

$$\overline{V}=\begin{bmatrix}1+\tau^2 & 0\\ 0 & 1\end{bmatrix}.$$

This can be rewritten

$$E(\tilde{Y})=\tilde{X}'\beta,\ \mathrm{var}(\tilde{Y})=\sigma^2 I$$

with

$$\tilde{X}=\overline{X},\ \tilde{Y}'=((y_1-\hat{m}_2)(1+\tau^2)^{-1/2},\ y_2)'.$$

The normal equations of the new model are $(\tilde{X}\tilde{X}')\hat{\beta}=\tilde{X}\tilde{Y}$, i.e.,

$$0\hat{\beta}_1+0\hat{\beta}_2=0(y_1-\hat{m}_2)(1+\tau^2)^{-1/2}+0y_2,$$

$$0\hat{\beta}_1+\hat{\beta}_2=0(y_1-\hat{m}_2)(1+\tau^2)^{-1/2}+y_2,$$

to which we adjoin $K'\hat{\beta}=\hat{m}_2$, i.e., $\hat{\beta}_1=\hat{m}_2$. Thus the result is

$$\hat{\beta}'=(\hat{m}_2,y_2)',$$

whereas without the requirement $K'\hat{\beta}=\hat{m}_2$ we would have

$$\hat{\beta}'=(y_1,y_2)'.$$

The estimate assigned to $\theta=\beta_1+\beta_2$, which is not a linear combination of the rows of $L'\beta$, is

$$\theta=\hat{\beta}_1+\hat{\beta}_2=\hat{m}_2+y_2$$

and has variance $\tau^2\sigma^2+\sigma^2$, whereas the best estimate of θ in the original model is y_1+y_2 with variance $2\sigma^2$. Thus the requirement $K'\hat{\beta}=\hat{m}_2$ decreases or increases the variance of the estimate of θ according as $\tau<1$ or $\tau>1$, i.e., according as the prior measurement of m_2 was more or less precise than the new measurement of y_1.

6. References

A. C. Aitken (1937), On least squares and linear combination of observations, Proc. Roy. Soc. Edinburgh, Sect. A, **55**, 42–47.

T. W. Anderson (1948), On the theory of testing serial correlation, Skandinavisk Aktvarietidskrift **XXXI**, 88–116.

A. Bjerhammar (1951), Application of calculus of matrices to the method of least squares, with special references to geodetic calculations. Trans. Roy. Inst. Tech. Stockholm **49**, 1–86.

R. C. Bose (1944), The fundamental theorem of linear estimation, Proc. 31st Indian Sci. Cong., 2–3 (abstract).

J. S. Chipman and M. M. Rao (1964), The treatment of linear restrictions in regression analysis, Econometrica **32**, 198–209.

W. E. Deming (1943), Stastical Adjustment of Data (John Wiley & Sons, Inc., New York, N.Y.).

C. Eisenhart (1964), The meaning of "least" in least squares, J. Washington Acad. Sci. **54**, 24–33.

R. J. Freund, R. W. Vail and C. W. Clunies-Ross (1961), Residual analysis, J. Amer. Stat. Assoc. **56**, 98–104.

C. F. Gauss (1873), Werke, Vol. 4, Theoria combinatonis observationum erroribus minimus obnoxiae, Gottingen, 1–93. (English translation by Hale F. Trotter, Technical Report 5 Statistical Techniques Research Group, Princeton University.)

A. S. Goldberger and D. B. Jockems (1961), Note on stepwise least squares, J. Amer. Stat. Assoc. **56**, 105–110.

T. N. E. Greville (1959), The pseudoinverse of a rectangular or singular matrix and its application to the solution of systems of linear equations, SIAM Review **1**, 38–43.

T. N. E. Greville (1960), Some applications of the pseudoinverse of a matrix, SIAM Review **2**, 15–22.

A. M. Legendre (1806), Nouvelles methodes pour la determination des orbites des cometes, 2d edition, Paris.

A. A. Markov (1912), Wahrscheinlichkeitsrechnung, H. Lieberman, tr. 2d ed., Leipzig.

M. Merriman (1877–1882), A list of writings relating to the method of least squares with historical and critical notes, Trans. Conn. Acad. Arts and Sciences, **IV**, 151–231.

E. H. Moore (1935), General Analysis I, Mem. Am. Phil. Soc. **I**, p. 197

M. E. Muller and G. S. Watson (1959), Randomization and linear least squares estimation. Statistical Techniques Research Group Technical Report No. 32, Princeton University.

R. Penrose (1955), A generalized inverse for matrices, Proc. Camb. Phil. Soc. **51**, 406–413.

R. Penrose (1956), On best approximate solutions of linear matrix equations, Proc. Camb. Phil. Soc. **52**, 17–19.

R. L. Plackett (1949), A historical note on the method of least squares, Biometrika **36**, 458–460.

R. L. Plackett (1960), Regression Analysis, Oxford University Press.

C. R. Rao, (1946), On the linear combination of observations and the general theory of least squares, Sankhya **7**, 237–256.

C. R. Rao (1962), A note on a generalized inverse of a matrix with applications to problems in mathematical statistics, J. Roy. Stat. Soc. (B) **24**, 152–158.

H. Scheffé (1959), The Analysis of Variance (John Wiley & Sons, Inc., New York, N.Y.).

(Paper 68B4–128)