The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 70–75

WHEN ARE GAUSS-MARKOV AND LEAST SQUARES ESTIMATORS IDENTICAL? A COORDINATE-FREE APPROACH[1]

BY WILLIAM KRUSKAL

University of Chicago

1. Introduction. In the framework of the general linear hypothesis, it is well known that the Gauss-Markov (minimum variance linear unbiased) and least squares estimators may be the same even when the underlying covariance structure is not a multiple of the identity. The main purpose of this note is to develop a (known) condition for this phenomenon simply in terms of the coordinate-free approach to the subject.

For general background and bibliography see Zyskind (1967) and Watson (1967). Much of the present paper is effectively a simpler form of part of the material in the Zyskind and Watson papers. For discussion of the coordinate-free approach see Kruskal (1961). Two very simple examples may be useful at the outset.

EXAMPLE 1. *One-way layout with unequal variances in known ratios.* The observations correspond to uncorrelated random variables Y_{ij} ($i = 1, \cdots, I$; $j = 1, \cdots, J_i$), $EY_{ij} = \beta_i$, and $\operatorname{Var} Y_{ij} = \lambda_i^2\sigma^2$. Here the β_i are unrestricted, unknown numbers; σ^2 is an unknown positive number; and the λ_i^2 are known positive numbers. It is readily seen that the Gauss-Markov and least square estimators of β_i are both $\bar{Y}_{i\cdot} = J_i^{-1} \sum_j Y_{ij}$. The corresponding conventional estimators of σ^2 are, however, not the same, unless all the λ_i^2 are one. These conventional estimators are $[\sum (J_i - 1)]^{-1}$ times

$$\textstyle\sum_{ij} (Y_{ij} - \bar{Y}_{i\cdot})^2 \qquad \text{(for Least Squares)},$$

$$\textstyle\sum_{ij} (Y_{ij} - \bar{Y}_{i\cdot})^2/\lambda_i^2 \qquad \text{(for Gauss-Markov)}.$$

(Of course if the λ_i^2 are equal, the two expressions are proportional, and if all $\lambda_i^2 = 1$, the expressions are identical. To obtain unbiased estimators of σ^2, the first expression must be divided by $\sum [(J_i - 1)\lambda_i^2]$ and the second by $\sum(J_i - 1)$.)

EXAMPLE 2. *Single sample with permutation-invariant covariance structure.* The observations correspond to random variables Y_i ($i = 1, \cdots, n$), $EY_i = \beta$, $\operatorname{Var} Y_i = \sigma^2$, and $\operatorname{Cov}(Y_i, Y_{i'}) = \sigma^2\gamma$ for $i \neq i'$. Here β is an unrestricted unknown number, σ^2 is an unknown positive number, and γ is a known number

Received 25 July 1967.

[1] This research was carried out in the Department of Statistics, University of Chicago, under partial sponsorship of the Statistics Branch and the Army Research Office, Office of Naval Research, and Air Force Office of Scientific Research by Contract No. Nonr 2121(23), NR 342-043 and in part by Research Grant No. NSF GP 6562 from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation. Reproduction in whole or in part is permitted for any purpose of the United States Government.

satisfying (for positive definiteness) the condition $1 > \gamma > -1/(n-1)$. It is readily seen here also that the Gauss-Markov and least squares estimators of β both are $\bar{Y}.$, the sample average. The conventional corresponding estimators of σ^2 are $(n-1)^{-1}$ times

$$\sum (Y_i - \bar{Y}.)^2 \qquad \text{(for Least Squares)},$$

$$(1-\gamma)^{-1} \sum (Y_i - \bar{Y}.)^2 \qquad \text{(for Gauss-Markov)}.$$

While these are proportional, they are the same if and only if $\gamma = 0$.

2. Identity of estimators if and only if regression manifold is invariant. In general, let Y be a random vector in an n-dimensional inner product space. Without essential loss of generality, the space may be taken as ordinary n-dimensional coordinate space with the conventional inner product, but it is simpler to proceed without specific coordinates. The inner product of vectors x, z is denoted by (x, z). Let μ be the expectation of Y and Σ its covariance operator. This is equivalent to saying that $E(x, Y) = (x, \mu)$ and that $\text{Cov}\,[(x, Y), (z, Y)] = (x, \Sigma z)$ for all vectors x, z in the space. Of course the moments are assumed to exist.

Now suppose that μ is assumed to lie in a given linear manifold, Ω, and that Σ is taken as known up to an unknown positive scalar constant, $\Sigma = \sigma^2 V$, say, where σ^2 is the unknown scalar and V is a given symmetric positive definite linear transformation. (Positive definiteness will later be relaxed.)

The least squares estimator of μ, μ^*, is given by the orthogonality condition that μ^* is the unique member of Ω satisfying

$$(Y - \mu^*, z) = 0 \qquad \text{for all} \qquad z \in \Omega.$$

The Gauss-Markov estimator, $\hat{\mu}$, is given by the corresponding orthogonality condition in terms of the new inner product

$$((x, z)) = (x, V^{-1}z).$$

The condition is that $\hat{\mu}$ is the unique vector in Ω satisfying

$$((Y - \hat{\mu}, z)) = 0 \qquad \text{for all} \qquad z \in \Omega.$$

In this setting, we have

THEOREM 1. *The two estimators, μ^* and $\hat{\mu}$, are the same if and only if Ω is invariant under V.*

PROOF. To say that Ω is invariant under V is to say that $Vx \in \Omega$ for all $x \in \Omega$. By non-singularity of V, this is equivalent to $V\Omega = \Omega$, and this in turn is equivalent to $V^{-1}\Omega = \Omega$ and to $V^{-1}x \in \Omega$ for all $x \in \Omega$. Now write the two defining conditions as

$$(Y - \mu^*, z) = (Y - \hat{\mu}, V^{-1}z) = 0 \qquad \text{for all} \qquad z \in \Omega.$$

If Ω is invariant under V (or V^{-1}), then the two defining conditions are the same, so that, by uniqueness, $\mu^* = \hat{\mu}$. On the other hand, if $\mu^* = \hat{\mu}$, then $Y - \mu^*$ must always be orthogonal (in terms of the original inner product) to both Ω and $V^{-1}\Omega$. Unless $\Omega = V^{-1}\Omega$ this leads to a contradiction. This completes the proof; a second proof will be given later.

As an illustration, consider Example 2, for which Ω is the equiangular line, spanned by the column vector $e = (1, 1, \cdots, 1)'$. The matrix corresponding to V has i, i' component $\delta_{ii'}(1 - \gamma) + \gamma$, where $\delta_{ii'}$ is the Kronecker delta. Multiplying this matrix into e gives $(1 + (n - 1)\gamma)e \, \varepsilon \, \Omega$.

Note that $\hat{\mu} = \mu^*$ means that, for all estimable linear functionals of μ, the Gauss-Markov and least squares estimators are the same. In other words, $\hat{\mu} = \mu^*$ if and only if $(x, \hat{\mu}) = (x, \mu^*)$ for all $x \, \varepsilon \, \Omega$.

Note also that to say $V\Omega = \Omega$ is to say that Ω is spanned by dim Ω characteristic vectors of V. For if $V\Omega = \Omega$, consider V as a linear transform on Ω alone. It must have dim Ω mutually orthogonal non-zero characteristic vectors in Ω, since it is symmetric positive definite. Yet these remain characteristic vectors for V as a transformation in the whole space by the assumed invariance. On the other hand, if Ω is spanned by dim Ω characteristic vectors (necessarily non-zero) of Ω, then clearly $V\Omega = \Omega$.

3. Another proof in terms of an intrinsic characterization of $\hat{\mu}$. The following characterization of $\hat{\mu}$ has intuitive appeal and is often useful.

THEOREM 2. *Assume that V is non-singular. A linear transform AY of Y is the Gauss-Markov estimator $\hat{\mu}$ of μ, if and only if these three conditions hold,*

1. $AY \, \varepsilon \, \Omega$,
2. $Ax = x$ *for all* $x \, \varepsilon \, \Omega$,
3. AY *and* $Y - AY$ *are uncorrelated.*

PROOF. Condition 3 means that the covariance between any linear functional of AY and any linear functional of $Y - AY$ is zero. Let $\langle x, z \rangle$ be an inner product for which the covariance operation of Y is the identity. Such an inner product must always exist by non-singularity of Y. If $AY = \hat{\mu}$, then AY is the orthogonal projection of Y onto Ω with respect to $\langle \, , \rangle$ and the three conditions are immediately verified.

If the conditions are satisfied by A, then, by condition 3, for all vectors x, z

$$\begin{aligned} 0 &= \text{Cov}\,[\langle x, AY \rangle, \langle z, (I - A)Y \rangle] \\ &= \text{Cov}\,[\langle A'x, Y \rangle, \langle (I - A)'z, Y \rangle] \\ &= \langle x, A(I - A)'z \rangle \end{aligned}$$

so that $A = AA'$, whence $A' = AA' = A$ and $A = A^2$. Hence A is symmetric and idempotent, and thus an orthogonal projection. By condition 1, A is into Ω; by condition 2, A is onto Ω. Hence A gives orthogonal projection onto Ω, and $AY = \hat{\mu}$.

In terms of Theorem 2, a second proof of Theorem 1 may be given. One wants to show that $V\Omega \subset \Omega$ if and only if (1) $\mu^* \, \varepsilon \, \Omega$, (2) $E\mu^* = \mu$, and (3) μ^* and $Y - \mu^*$ are uncorrelated. But (1) and (2) always hold for μ^*. For convenience, denote by P_Ω orthogonal projection onto Ω with respect to (,), and set $Q_\Omega = I - P_\Omega$; thus $\mu^* = P_\Omega Y$ and $Y - \mu^* = Q_\Omega Y$. It is readily computed that condition (3) is equivalent to $Q_\Omega V P_\Omega = 0$. Now if $V\Omega \subset \Omega$, $VP_\Omega x \varepsilon \Omega$ for all x and hence $Q_\Omega(VP_\Omega x) = 0$ since $Q_\Omega\Omega = 0$. In the other direction, if $Q_\Omega V P_\Omega x = 0$ for all x, then $(Q_\Omega z, VP_\Omega x) = 0$ for all x and z. This says that $VP_\Omega x$ is orthogonal to the orthogonal complement of Ω, or that $VP_\Omega x$ is in Ω.

4. Equality of residual sums of squares. If V were the identity, the conventional estimator of μ would be $P_\Omega Y = \mu^* = \hat{\mu}$, and the conventional estimator of σ^2 would be $(n\text{-dim}\,\Omega)^{-1}$ times $||Y - P_\Omega Y||^2 = ||Q_\Omega Y||^2$, where $||x||^2 = (x, x)$. For arbitrary non-singular V, the conventional estimator of μ is $\hat{\mu}$ and the conventional estimator of σ^2 is $(n\text{-dim}\,\Omega)^{-1}$ times $|||Y - \hat{\mu}|||^2$, where $|||x|||^2 = ((x, x)) = (x, V^{-1}x)$. The following question arises: suppose that $\mu^* = \hat{\mu}$ (i.e., that $V\Omega = \Omega$); under what circumstances are the two residual sums of squares also equal? That is, given $\hat{\mu} = \mu^*$, when do we also have

$$||Y - \mu^*||^2 = |||Y - \hat{\mu}|||^2?$$

The answer is given by

THEOREM 3. *If $\mu^* = \hat{\mu}$, then $||Y - \mu^*||^2 = |||Y - \hat{\mu}|||^2$ if and only if V is the identity operator on the orthogonal complement $\Omega^\perp$ of Ω.*

PROOF. Orthogonal complementation here is with respect to (,). Write the desired necessary and sufficient condition as

$$(Q_\Omega y, Q_\Omega y) = (Q_\Omega y, V^{-1}Q_\Omega y), \qquad \text{all } y,$$

or as

$$(x, x) = (x, V^{-1}x), \qquad \text{all } x \,\varepsilon\, \Omega^\perp.$$

By the standard identity $4(x, z) = ||x + z||^2 - ||x - z||^2$, this is equivalent to

$$(x, z) = ((x, z)) = (x, V^{-1}z), \qquad \text{all } x, z \,\varepsilon\, \Omega^\perp,$$

or to $V^{-1}z = z$ for all $z \,\varepsilon\, \Omega^\perp$. That is, V^{-1} (and V) is the identity operator on $\Omega^\perp$. This completes the proof.

By virtually the same argument, if $\mu^* = \hat{\mu}$, then $||Y - \mu^*||^2 = c\,|||Y - \hat{\mu}|||^2$ for a scalar c if and only if V as an operator on $\Omega^\perp$ is c times the identity.

Consider Example 1. A generic vector in $\Omega^\perp$ has i, j coordinates $y_{ij} - \bar{y}_{i\cdot}$ for some set of y_{ij}. The $i, j; i', j'$ element of V, regarded as a matrix, is zero unless $i = i', j = j'$; then it is $\lambda_i^2 : \delta_{ii'}\delta_{jj'}\lambda_i^2$. Thus the i, j element of the product of V and the generic vector in $\Omega^\perp$ is

$$\textstyle\sum_{i',j'} \delta_{ii'}\lambda_i^2(y_{i'j'} - y_{i\cdot}).$$

Hence V is the identity on $\Omega^\perp$ just when all $\lambda_i^2 = 1$, and V is proportional to the identity on $\Omega^\perp$ just when all λ_i^2 are equal.

The condition that V be the identity on $\Omega^\perp$ may of course be restated thus: the characteristic roots corresponding to characteristic vectors of V in $\Omega^\perp$ must all be unity. (There must be at least n-dim Ω such characteristic vectors because of the assumed invariance of Ω.)

5. The case in which V is singular. Suppose now that V is singular, but still, of course, symmetric and positive semi-definite. Let R and N be, respectively, the range and the null space of V. Because of symmetry of V, R and N are orthogonal and their direct sum is the entire space. With probability one, Y lies in $R + P_N\mu$.

Thus $P_N\mu$ is estimated without error by P_NY. Actually part of the component of μ in R may in general also be estimated without error, but we do not require this detail for present purposes.

It is simpler to work with linear manifolds than with flats (translated manifolds). Accordingly, let ν be any vector in Ω that satisfies $P_N\nu = P_N\mu = P_NY$. Such a ν must exist, since EY itself satisfies the condition, and in practice one may pick a convenient ν; which ν is immaterial so long as it satisfies the condition. Then let $Z = Y - \nu$. Clearly $Z \in R$ with probability one, and EZ ranges over $\Omega \cap R$; further, Z as a random vector in R alone is nonsingular, with covariance transformation σ^2V, regarded as a (nonsingular) transformation from R onto R. Of course $P_NZ = 0$. The Gauss-Markov estimator of EZ is $\hat{\mu}(Z) \in \Omega \cap R$, a well-defined quantity, and the Gauss-Markov estimator of EY is $\hat{\mu}(Y) = \hat{\mu}(Z) + \nu$. By Theorem 1, $\hat{\mu}(Z) = P_{\Omega\cap R}Z$ if and only if $V(\Omega \cap R) = \Omega \cap R$. We now sketch a proof that Theorem 1 holds even if V is singular.

Suppose first that $\hat{\mu}(Y) = P_\Omega(Y)$. Then $\hat{\mu}(Z) = P_\Omega Y - \nu = P_\Omega Z$. Projecting onto N, $P_N\hat{\mu}(Z) = 0 = P_NP_\Omega Z$. Hence $P_\Omega Z \in \Omega \cap R$. By making the orthogonal decomposition $\Omega = (\Omega \cap R) + (\Omega - (\Omega \cap R))$, where the notation in the second summand means the orthogonal complement of $\Omega \cap R$ relative to Ω, we find that $\Omega - (\Omega \cap R) \subset N$, or that $\Omega = (\Omega \cap R) + P_N\Omega$ with summands orthogonal. It follows that $P_\Omega Z = P_{\Omega\cap R}Z$, whence, by Theorem 1, $V(\Omega \cap R) = \Omega \cap R$. For $\mu \in \Omega$, $\mu = \mu_1 + \mu_2$ with $\mu_1 \in \Omega \cap R$ and $\mu_2 \in N$; hence $V\mu = V\mu_1 \in \Omega \cap R \subset \Omega$. Thus $V\Omega \subset \Omega$ and we have invariance.

In the other direction, suppose that $V\Omega \subset \Omega$. Then $V(\Omega \cap R) \subset \Omega$ and of course $V(\Omega \cap R) \subset R$; hence $V(\Omega \cap R) \subset \Omega \cap R$, and in fact $V(\Omega \cap R) = \Omega \cap R$ by nonsingularity of V in R. By Theorem 1, $\hat{\mu}(Z) = P_{\Omega\cap R}Z$, and it remains to show that $P_{\Omega\cap R}Z = P_\Omega Z$. To this end make another orthogonal decomposition so that for $\mu \in \Omega$,

$$\mu = P_{\Omega\cap R}\mu + P_{R-(\Omega\cap R)}\mu + P_N\mu.$$

Since $V(\Omega \cap R) = \Omega \cap R$ and V is nonsingular in R, the second summand on the right cannot, if non-zero, be in $\Omega \cap R$, but clearly it is in R and (since $VP_{R-(\Omega\cap R)}\mu = V\mu - VP_{\Omega\cap R}\mu \in \Omega$) it is also in Ω. Hence the second summand is zero and $\Omega = (\Omega \cap R) + P_N\Omega$. Thus $P_{\Omega\cap R}Z = P_\Omega Z$. The proof is complete and we have

THEOREM 3. *Theorem 1 holds even if V is singular.*

REFERENCES

KRUSKAL, WILLIAM (1961). The coordinate-free approach to Gauss-Markov estimation, and its application to missing and extra observations. *Fourth Berkeley Symp. Math. Statist. Prob.* **1** 435–451.

WATSON, GEOFFREY S. (1967). Linear least squares regression. *Ann. Math. Statist.* **38** 1679–1699.

ZYSKIND, GEORGE (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38** 1092–1109.