Belgisch Tijdschrift voor Statistiek, Informatica en Operationeel Onderzoek, 4, 1980
Revue Belge de Statistique, d'Informatique et de Recherche Opérationnelle, 4, 1980

MORE ON GOLDBERGER'S PREDICTOR

Ishay Weissman[1)]

1. INTRODUCTION

Let $Y_1 \in R^t$ and $Y_2 \in R^s$ be two random vectors with expectations and co-variance given by

$$(1.1) \qquad E\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta \quad , \quad \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} ,$$

where X_1, X_2 and $\Omega > 0$ are known matrices of orders $t \times k$, $s \times k$ and $(t+s) \times (t+s)$ respectively, $\beta \in R^k$ and rank $X_1 = t$. A linear predictor for Y_2, on the basis of Y_1, was suggested by Goldberger (1962):

$$(1.2) \qquad \hat{Y}_2 = X_2 \hat{\beta} + \Omega_{21} \Omega_{11}^{-1} (Y_1 - X_1 \hat{\beta}) ,$$

where $\hat{\beta} = (X_1' \Omega_{11}^{-1} X_1)^{-1} X_1' \Omega_{11}^{-1} Y_1$ is the Gauss-Markov estimator of β. In a recent paper, Loeff and Leclercq (1976) have shown that $\hat{Y}_2$ is the optimal predictor by each one of the following criteria:

(a) Minimum mean square prediction error (MMSPE) among linear unbiased predictors;

(b) Maximum likelihood prediction under normality;

(c) Minimum generalized variance of the prediction error among linear unbiased predictors.

---

1) Department of Statistics, Iowa State University. (On leave from the Technion, Haifa, Israel).

All of their results are proved by differentiation and then solving the resulting system of equations. The purpose of this note is to derive criteria (a) - (c) and

(d) MMSPE among linear predictors with bounded error

by algebraic-geometric arguments only, i.e. by using orthogonal projections.

2. NOTATIONS AND PRELIMINARIES

Let $A = \{\mu = X_1\beta : \beta \in R^k\}$ be the regression sub-space, then the orthogonal projection on A with respect to the inner-produce $(x, \Omega_{11}^{-1} z) \equiv x'\Omega_{11}^{-1} z$ $(x, z \in R^t)$ is

$$P = X_1(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1} .$$

The Gauss-Markov estimator of $\mu = EY_1$ based on the observation Y_1 is $\hat{\mu} = PY_1$ (e.g., Kruskal, 1961). Suppose we are told that $EY_2 = T\mu$ for some linear transformation $T : R^t \to R^s$, whenever $EY_1 = \mu$. We shall show in Section 3 that

$$T\hat{\mu} + \Omega_{21}\Omega_{11}^{-1}(Y_1 - \hat{\mu}) = (TP + \Omega_{21}\Omega_{11}^{-1}Q)Y_1 \equiv C_o Y_1 \tag{2.1}$$

is the "best" predictor of Y_2 by each one of the criteria (a) - (d) (here $Q = I - P$).

Remarks. The definition of T is not unique. If $EY_2 = T\mu = T^*\mu$ for all $\mu \in A$, then in particular $T\hat{\mu} = T^*\hat{\mu}$. Hence (2.1) is unaffectea when T is replaced by T^*. One choice of T is $T = X_2(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1}$. Note that $T = TP$, $TQ = 0$ and that (1.2) and (2.1) are identical.

3. THE OPTIMALITY OF C_oY_1

In this section we prove that C_oY_1 is the optimal predictor of Y_2 by each one of the criteria (a) - (d).

Proof of (a). The most general linear predictor ot Y_2 is $CY_1 + b$, where $C : R^t \to R^s$

is a linear transformation and $b \epsilon R^s$. The unbiasedness implies $C\mu + b = T\mu$ for all $\mu \epsilon A$. Hence $b = 0$ and $C = T$ on A. Let $\mathcal{C}_T = \{C : Cx = Tx \text{ for all } x \epsilon A\}$, then we have to prove that

$$(3.1) \qquad \min_{C \epsilon \mathcal{C}_T} E\|CY_1 - Y_2\|^2 = E\|C_oY_1 - Y_2\|^2 .$$

Assume first that $\Omega_{21} = 0$, then for $C \epsilon \mathcal{C}_T$

$$(3.2) \qquad E\|CY_1 - Y_2\|^2 = E\|C(Y_1 - \mu) - (Y_2 - T\mu)\|^2$$

$$= E\|C(Y_1 - \mu)\|^2 + E\|Y_2 - T\mu\|^2 .$$

The last term is unaffected by the choice of C. For the other term, we have

$$(3.3) \qquad E\|C(Y_1 - \mu)\|^2 = E\|CP(Y_1 - \mu)\|^2 + E\|CQY_1\|^2 + 2E(CQY_1, CP(Y_1 - \mu))$$

$$= E\|TP(Y_1 - \mu)\|^2 + E\|CQY_1\|^2 + 2E(QY_1, T'TP(Y_1 - \mu)),$$

since $C'C = T'T$ on A. It is clear now that the $C \epsilon \mathcal{C}_T$ which will minimize (3.3) is the one for which $CQ = 0$, i.e. $C = TP = C_o$, where C_o is defined in (1.3).

This proves (a) for the case $\Omega_{21} = 0$. For the general case we write

$$(3.4) \qquad CY_1 - Y_2 = (C - \Omega_{21}\Omega_{11}^{-1})Y_1 - (Y_2 - \Omega_{21}\Omega_{11}^{-1}Y_1) \equiv C^* Y_1 - Y_2^*$$

as a difference between two uncorrelated random vectors. As we have just shown, the optimal predictor C^*Y_1 of Y_2^* is one with $C^* = (T - \Omega_{21}\Omega_{11}^{-1})P$, and hence the optimal C for predicting Y_2 is $(T - \Omega_{21}\Omega_{11}^{-1})P + \Omega_{21}\Omega_{11}^{-1} = C_o$.

<u>Proof of (b).</u> Under normality, the likelihood function is of the form

$$L(Y_1, Y_2, \mu) = K \exp\left\{-\tfrac{1}{2}\left(\begin{bmatrix} Y_1 - \mu \\ Y_2 - T\mu \end{bmatrix}, \Omega^{-1}\begin{bmatrix} Y_1 - \mu \\ Y_2 - T\mu \end{bmatrix}\right)\right\} \equiv K \exp\{-\tfrac{1}{2}U\},$$

where $K > 0$ does not depend on Y_1, Y_2 and μ. In order to maximize L we have to minimize U. In the case of $\Omega_{21} = 0$, we have

$$(3.5) \qquad U = (Y_1 - \mu, \Omega_{11}^{-1}(Y_1 - \mu)) + (Y_2 - T\mu, \Omega_{22}^{-1}(Y_2 - T\mu))$$

$$= (P(Y_1 - \mu), \Omega_{11}^{-1} P(Y_1 - \mu)) + (QY_1, \Omega_{11}^{-1} QY_1) + (Y_2 - T\mu, \Omega_{22}^{-1}(Y_2 - T\mu))$$

$$\geq (QY_1, \Omega_{11}^{-1} QY_1)$$

with equality for $\mu = \hat{\mu} = PY_1$ and $Y_2 = T\hat{\mu} = C_o Y_1$. Hence $C_o Y_1$ minimizes U.

In the general case, we make use of the one-to-one transformation $\{\mu, Y_2\} \rightarrow \{\mu, Y_2^*\}$, where Y_2^* is defined in (3.4). Now, Y_2^* and Y_1 are uncorrelated, and hence the optimal values for μ and Y_2^* are $\hat{\mu} = PY_1$ and $\hat{Y}_2^* = (T - \Omega_{21}\Omega_{11}^{-1})PY_1$ respectively, resulting in $\hat{Y}_2 = C_o Y_1$.

<u>Proof of (c).</u> The covariance operator (or matrix) of $CY_1 - Y_2$ for $C \in \mathcal{C}_T$ (and $\Omega_{12} = 0$) is

$$G(C) = C\Omega_{11}C' + \Omega_{22} = TP\Omega_{11}P'T' + CQ\Omega_{11}Q'C' + \Omega_{22},$$

and we have to show that det G(C) is minimized over $\mathcal{C}_T$ by $C_o = TP$. Since $CQ\Omega_{11}Q\,C' \geq 0$, it is clear that

$$\det G(C) \geq \det\{TP\Omega_{11}P'T' + \Sigma_{22}\} = \det G(C_o)$$

(see Rao (1965) p. 56, 9(ii)). When $\Omega_{12} \neq 0$, we use (3.4) and the argument thereafter.

<u>Proof of (d).</u> Criteria (a) and (c) restricted the choice of the optimal C to the class $\mathcal{C}_T$. Suppose now that we predict Y_2 by $CY_1 + b$. Then, assuming $\Omega_{12} = 0$, we have

(3.6) $E\|CY_1 + b - Y_2\|^2 = E\|C(Y_1 - \mu)\|^2 + E\|Y_2 - T\mu\|^2 + \|(C - T)\mu + b\|^2.$

Under criteria (d), (3.6) is bounded for all $\mu \epsilon A$. This can only happen if $C = T$ on A, i.e., $C \epsilon \mathcal{C}_T$. The best choice for b is then $b = 0$. If indeed $b = 0$ and $C \epsilon \mathcal{C}_T$ then (3.6) is equal to (3.2) and the proof follows from the proof of (a). The same is true in the general case where $\Omega_{12} \neq 0$.

Concluding Remarks. The fact that bounded error (with $b = 0$) implies unbiasedness is also true in estimation (Kruskal 1961).

When Ω is known up to an unknown positive constant σ^2, $\Omega = \sigma^2 \Omega_0$, say, then C_o depends only on Ω_o and not on σ^2. When $\Omega_{12} = 0$, the problem of predicting Y_2 is equivalent to the problem of estimating $EY_2 = TEY_1$ on the basis of Y_1.

Loeff and Leclercq (1976) minimize $E\|B(CY_1 - Y_2)\|^2$ for arbitrary regular matrix B rather than (3.1). This is not a more general problem but rather a reparametrization of the same problem: to see this, define $\tilde{C} = BC$, $\tilde{Y}_2 = BY_2$, and the problem is to find the best $\tilde{C}$ for predicting $\tilde{Y}_2$.

REFERENCES

Goldberger, A.S. (1962). Best linear unbiased prediction in the generalized regression model. J. Amer. Statist. Assoc., 57, 369-375.

Kruskal, W.H. (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.

Rao, C.R. (1965). Linear Statistical Inference and Its Applications. Wiley, New York.

Van der Loeff, S.S. and Leclercq, L. (1976). A note on Goldberger's best linear unbiased predictor in the generalized regression model. La Rev. Belge de Statist. l'Inform. et de Rech Oper., 3, 37-42.