RT-MAE 9627

SIMPLE FORMS OF THE BEST LINEAR UNBIASED PREDICTOR IN THE LINEAR REGRESSION MODEL

by

Silvia N. Elian

Palavras-Chave: **Prediction; Generalized Least Squares Estimator; Covariance Matrix.**
(Key words)

Classificação AMS: 62J05.
(AMS Classification)

- Novembro de 1996 -

Simple Forms of the Best Linear Unbiased Predictor in the Linear Regression Model

Silvia N. Elian
University of São Paulo
São Paulo, Brazil

Abstract

We investigate conditions to be satisfied by the best linear unbiased predictor of future observations in the general linear regression model in order to have a simpler form than the one that would be obtained in the uncorrelated case.

Key words: Prediction; Generalized Least Squares Estimator; Covariance Matrix.

1. INTRODUCTION

In the general linear model

$$\underset{\sim}{y} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\epsilon}\,, \tag{1}$$

where $\mathbf{X}$ is a $n \times p$ matrix of rank p, $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$, $\text{Var}(\underset{\sim}{\epsilon}) = \Sigma$, and Σ is a known positive definite matrix, the generalized least squares estimator $\underset{\sim}{\hat{\beta}}_G = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\underset{\sim}{y}$ is the best linear unbiased estimator (BLUE) of $\underset{\sim}{\beta}$.

In the derivation of this estimator, we premultiply the assumed model (1) by a $n \times n$ non-singular matrix $\mathbf{P}^{-1}$ such that $\Sigma = \mathbf{PP}'$, obtaining the transformed model

$$\underset{\sim}{z} = \mathbf{Q}\underset{\sim}{\beta} + \underset{\sim}{\epsilon}^*, \tag{2}$$

where $\underset{\sim}{z} = \mathbf{P}^{-1}\underset{\sim}{y}$, $\mathbf{Q} = \mathbf{P}^{-1}\mathbf{X}$ and $\underset{\sim}{\epsilon}^* = \mathbf{P}^{-1}\underset{\sim}{\epsilon}$.

After this transformation, the covariance matrix of $\underset{\sim}{\epsilon}^*$ in (2) is $\text{Var}(\underset{\sim}{\epsilon}^*) = \mathbf{P}^{-1}\Sigma\mathbf{P}'^{-1} = \mathbf{I}$ and then, in this model, the BLUE of $\underset{\sim}{\beta}$ is $(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\underset{\sim}{z}$, which reduces to $\underset{\sim}{\hat{\beta}}_G$. Since $\underset{\sim}{\beta}$ is the same in both models (1) and (2), the result follows.

At this point, most of the regression text books finishes the subject and nothing about prediction under the model (1) is mentioned.

In the model (2), it is well known that the predictor of m future observations $\underset{\sim}{z}_f$, if $\mathbf{Q}_f$ is the regressor matrix, is given by $\hat{\underset{\sim}{z}}_f = \mathbf{Q}_f(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\underset{\sim}{z}$. This is also the best linear unbiased predictor (BLUP) of $\underset{\sim}{z}_f$ and, since after the variable transformation, we usually turn to the original model, one could conclude that $\hat{\underset{\sim}{y}}_f = \mathbf{X}_f\underset{\sim}{\hat{\beta}}_G$ would be the BLUP of $\underset{\sim}{y}_f$ with matrix of regressors $\mathbf{X}_f$. Furthermore, since is supposed that the relation $\underset{\sim}{y} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$ still holds for $\underset{\sim}{y}_f$ given $\mathbf{X}_f$, this conclusion seems natural. Unfortunately, this is wrong.

As shown by Goldberger (1962), the BLUP of $\underset{\sim}{y}_f$ is

$$\hat{\underset{\sim}{y}}_f^* = \mathbf{X}_f \hat{\underset{\sim}{\beta}}_G + \mathbf{W}\Sigma^{-1}(\underset{\sim}{y} - \mathbf{X}\hat{\underset{\sim}{\beta}}_G), \tag{3}$$

where $\mathbf{W}$ is the supposed known $m \times n$ covariance matrix of $\underset{\sim}{y}_f$ with $\underset{\sim}{y}$.

Johnston (1984) discussed the prediction problem under the simple regression linear model with intercept, assuming the first-order autoregressive process (AR(1)) for the disturbance terms of $\underset{\sim}{\epsilon}$ in (1). It has been shown that, in this case, the BLUP of one future observation y_{n+1} is not $\hat{y}_{n+1} = a + bx_{n+1}$, where $\hat{\underset{\sim}{\beta}}_G = [a\ b]'$, because the relation $\mathbf{P}^{-1}\underset{\sim}{y} = \mathbf{P}^{-1}\mathbf{X}\underset{\sim}{\beta} + \mathbf{P}^{-1}\underset{\sim}{\epsilon}$ implies that

$$y_j - \rho y_{j-1} = \alpha(1-\rho) + \beta(x_j - \rho x_{j-1}) + \epsilon_j, \quad j = 2, 3, \ldots,$$

or equivalently,

$$y_j = \alpha + \beta x_j + \rho(y_{j-1} - \alpha - \beta x_{j-1}) + \epsilon_j\ .$$

The last equation would suggest that

$$\hat{y}_{n+1} = \hat{\alpha} + \hat{\beta}x_{n+1} + \rho(y_n - \hat{\alpha} - \hat{\beta}x_n),$$

which will be a predictor of y_{n+1} if ρ is known. In addition, if we take $[\hat{\alpha}\ \hat{\beta}]' = \hat{\underset{\sim}{\beta}}_G$, $\hat{y}_{n+1}$ is the BLUP obtained by (3), because in this case

$$\begin{aligned} \mathbf{W} &= \sigma^2[\rho^n\ \rho^{n-1} \cdots \rho^2\ \rho] \qquad \text{and} \\ \mathbf{W}\Sigma^{-1} &= \rho[0\ 0 \cdots 0\ 1] \qquad \text{(see appendix).} \end{aligned}$$

The prediction of y_{n+1} under this model is also discussed in Neter, Kutner, Nachtsheim and Wasserman (1996) and Montgomery and Peck (1992).

Returning to the general linear model (1), if we can assume that the future observations are uncorrelated with the sample observations so that $\mathbf{W} = \mathbf{0}$, it's easy to see that the BLUP of $\underset{\sim}{y}_f$ reduces to $\mathbf{X}_f\hat{\underset{\sim}{\beta}}_G$. Also, it should be expected that, except in this very special case, the BLUP predictor of $\underset{\sim}{y}_f$ had always the form shown in (3). This is not true.

In the next section we present necessary and sufficient conditions for $\mathbf{X}_f\hat{\underset{\sim}{\beta}}_G$ to be the BLUP of $\underset{\sim}{y}_f$ under the model (1).

2. THE SIMPLIFYING CONDITIONS

Several authors have discussed conditions for which $\hat{\underset{\sim}{\beta}}_G$ coincides with the ordinary least squares estimators $\hat{\underset{\sim}{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underset{\sim}{y}$. Graybill (1976) stated that $\hat{\underset{\sim}{\beta}}_G = \hat{\underset{\sim}{\beta}}$ if and only if there exists a nonsingular matrix $\mathbf{F}$ such that $\Sigma\mathbf{X} = \mathbf{X}\mathbf{F}$. Milliken and Albohali (1984) present another necessary and sufficient condition. They show that $\hat{\underset{\sim}{\beta}}_G = \hat{\underset{\sim}{\beta}}$ if and only if $\mathbf{X}'\Sigma^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbf{0}$.

Our problem is different but, as we will see later, some similarities with these situations can be found. The first result is presented in the next theorem.

Theorem 1. Under the model (1), with $\underset{\sim}{y}_f = \mathbf{X}_f\underset{\sim}{\beta} + \underset{\sim}{\epsilon}_f$, $E(\underset{\sim}{\epsilon}_f) = \underset{\sim}{0}$, $\mathrm{Var}(\underset{\sim}{\epsilon}_f) = \Sigma_f$ and $\mathrm{Cov}(\underset{\sim}{y}_f, \underset{\sim}{y}) = \mathbf{W}$, $\mathbf{X}_f\hat{\underset{\sim}{\beta}}_G$ is the BLUP of $\underset{\sim}{y}_f$ if and only if

$$\mathbf{W}\Sigma^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}) = \mathbf{0}\ . \tag{4}$$

Proof. The predictor $X_f \tilde{\beta}_G$ will be the BLUP of $\tilde{y}_f$ if and only if

$$X_f \tilde{\beta}_G = W\Sigma^{-1}\tilde{y} + (X_f - W\Sigma^{-1}X)\tilde{\beta}_G ,$$

which is equivalent to

$$W\Sigma^{-1}\tilde{y} = W\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\tilde{y} ,$$

or

$$W\Sigma^{-1}(I - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1})\tilde{y} = 0 \quad \text{for all } \tilde{y}.$$

Since last equality holds if and only if $W\Sigma^{-1}(I - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}) = 0$, the result is proved. □

The condition (4) is very similar with Milliken and Albohali's condition. It requires knowing Σ^{-1} and W. We will see in the examples that it is necessary to know only the form of these matrices and not the numerical values of its elements.

Another necessary and sufficient condition is given in the following theorem.

Theorem 2. Under the same conditions of Theorem 1, $X_f \tilde{\beta}_G$ is the BLUP of $\tilde{y}_f$ if and only if there exist a $p \times m$ matrix D such that

$$W' = XD . \tag{5}$$

Proof. We must show that

$$X_f \tilde{\beta}_G = W\Sigma^{-1}\tilde{y} + (X_f - W\Sigma^{-1}X)\tilde{\beta}_G$$

if and only if (5) holds.

These two quantities are equal for all $\tilde{y}$ if and only if

$$W\Sigma^{-1} = W\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1},$$

or

$$W = W\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'$$

which is equivalent to

$$W' = XD, \text{ where } D = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}W' .$$ □

The result of Theorem 2 gives us a condition that, as Graybill's condition, depends on the existence of the matrix D.

It's easy to see that, if $W = 0$, both conditions (4) and (5) hold. The following examples ilustrate that these conditions can be satisfied for other nontrivial forms of W.

Example 1. Consider the linear regression model with $p - 1$ regressor variables and an intercept term and let $\Sigma = \sigma^2[(1-\rho)I + \rho J_n]$, where J_n is a $n \times n$ matrix of ones.

Suppose that $\mathrm{Cov}(y_k, y_i) = \gamma\sigma^2$, for $k = n+1, n+2, \ldots$ and $i = 1, 2, \ldots, n$, and consider the prediction of m future observations $\underset{\sim}{y}_f = [y_{n+1}, y_{n+2}, \ldots, y_{n+m}]'$.

Since $\mathbf{W} = \gamma\sigma^2\mathbf{J}_{m,n}$, where $\mathbf{J}_{m,n}$ is a $m \times n$ matrix of ones, it is easy to see that $\mathbf{W}' = \mathbf{XD}$ for

$$\mathbf{D} = \begin{bmatrix} \gamma & \gamma & \cdots & \gamma \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{p\times m} \sigma^2 .$$

Thus, by Theorem 2, $\hat{\underset{\sim}{y}}_f = \mathbf{X}_f \underset{\sim}{\beta}_G$. We can also note (Graybill, 1976, Theorem 6.8.1) that in this case, $\underset{\sim}{\beta}_G = \underset{\sim}{\hat{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underset{\sim}{y}$. These facts imply that prediction of new observations can be made without knowing ρ or γ. These parameters are generally unknown and then, we got very suitable simplifications.

Example 2. In the location model, where $\mathbf{X} = \mathbf{J}_{n,1}$, condition (5) holds if and only if $\mathbf{W}' = \mathrm{Cov}(\underset{\sim}{y}, \underset{\sim}{y}_f) = \mathbf{J}_{n,1}[d_{n+1}\ d_{n+2}\ \cdots\ d_{n+m}]$. This is equivalent to, for each $k = n+1, n+2, \ldots, n+m$, $\mathrm{Cov}(y_i, y_k) = \mathrm{d}_k$ for all $i = 1, 2, \ldots, n$. Milliken and Albohali emphasize that in this model, $\underset{\sim}{\beta}_G = \underset{\sim}{\hat{\beta}}$ only if $\Sigma = \sigma^2[\rho\mathbf{I} + (1-\rho)\mathbf{J}_n]$. If both conditions are satisfied, the BLUP of $\underset{\sim}{y}_f$ is $\mathbf{X}_f\underset{\sim}{\hat{\beta}} = \mathbf{J}_{m,1}\bar{y}$, where $\bar{y}$ is the sample mean of the Y variable. So, in this example, we also can get the BLUP of future observations without knowing the parameters of Σ and $\mathbf{W}$.

3. CHARACTERIZATION OF W

Condition (5) gives us a characterization of $\mathbf{W}$, because it is equivalent to $\mathrm{Cov}(y_i, y_k) = \underset{\sim}{x}_i'\underset{\sim}{\delta}_k$, $i = 1, 2, \ldots, n$, $k = n+1, n+2, \ldots, n+m$, where $\underset{\sim}{x}_i'$ is the i-th row of $\mathbf{X}$ and $\underset{\sim}{\delta}_k$ is a p-dimensional column vector that would correspond to the k-th column of $\mathbf{D}$.

Example 3. Milliken and Albohaly show that $\underset{\sim}{\beta}_G = \underset{\sim}{\hat{\beta}}$ in the linear regression model with an intercept term and $\Sigma = \mathbf{I} + a\underset{\sim}{x}(\underset{\sim}{x}'\underset{\sim}{x})^{-1}\underset{\sim}{x}'$, where $\underset{\sim}{x}$ is a column of $\mathbf{X}$ and a is such that Σ is positive definite.

It's easy to see that in this case,

$$\mathrm{Cov}(y_i, y_k) = \frac{a x_{ji} x_{jk}}{\sum_{l=1}^{n} x_{jl}^2} ,$$

where $i, k = 1, 2, \ldots, n$ and X_j is any regressor variable, $j = 1, 2, \ldots, p$.

Suppose that this kind of covariance still holds for the future observations, that is

$$\mathrm{Cov}(y_i, y_k) = a\frac{x_{ji} x_{jk}}{\sum_{l=1}^{n} x_{jl}^2} ,$$

for $i = 1, 2, \ldots, n$ and $k = n+1, n+2, \ldots, n+m$.

This expression can be written as $\underset{\sim}{x}_i'\underset{\sim}{\delta}_k$ for

$$\underset{\sim}{\delta}_k = \frac{a}{\sum_{l=1}^{n} x_{jl}^2}[0 \cdots 0\ x_{jk}\ 0 \cdots 0]'$$

and then, the BLUP of $\underset{\sim}{y}_f$ is $X_f \underset{\sim}{\hat{\beta}}_G = X_f \underset{\sim}{\hat{\beta}}$.

4. CONCLUDING REMARKS

In the examples, $\Sigma X = XF$ and also (5) held. This will not always happen. These conditions work with different parameters and it is not possible to get some relation between them. However, it seems that in some cases, the type of covariance between y_i and y_k, $i \neq k$, implies the validity of both. When this holds, we can get the BLUP of $\underset{\sim}{y}_f$ without knowing any parameter of Σ or W. Otherwise, if these parameters are unknown, usually they are substituted by its estimates. Then, if Σ and W have different parameters, validity of just one of conditions implies in less parameters to be estimated.

APPENDIX

If ϵ_i, $i = 1, 2, \ldots$ follows a first-order autoregressive process, then

$$\Sigma = \text{Var}(\underset{\sim}{\epsilon}) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & & \rho^{n-2} \\ \rho^2 & \rho & 1 & & \rho^{n-3} \\ \vdots & & & & \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & & 1 \end{bmatrix}$$

and

$$\Sigma^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & & -\rho & 1 \end{bmatrix}.$$

REFERENCES

Graybill, F.A. (1976), *Theory and Application of the Linear Model*, North Scituate, Mass.: Duxbury Press.

Goldberger, A.S. (1962), "Best Linear Unbiased Prediction in the Generalized Linear Regression Model", *Journal of the American Statistical Association*, 57, 369–375.

Johnston, J. (1984), *Econometric Methods* (3rd ed.), McGraw-Hill, Inc.

Milliken, G.A. and Albohali, M. (1984), "On Necessary and Sufficient Conditions for Ordinary Least Squares Estimators to Be Best Linear Unbiased Estimators", *The American Statistician*, 38, 298–299.

Montgomery, D.C. and Peck, E. (1992), *Introduction to Linear Regression Analysis* (2nd ed.), John Wiley & Sons, Inc.

Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996), *Applied Linear Statistical Models* (4th ed.), Richard D. Irwin, Inc.

9611 - **GIMENEZ, P.; BOLFARINE, H.** Unbiased Score Functions in Error-In-Variables Models. São Paulo, IME-USP, 1996. 23p. (RT-MAE-9611)

9612 - **BUENO, V.C.** Comparing component redundancy allocation in K-out-of-n system. IME-USP, 1996. 10p. (RT-MAE-9612)

9613 - **GALEA, M.; PAULA, G.A.; BOLFARINE, H.** Local influence in elliptical linear regression models. IME-USP, 1996. 14p. (RT-MAE-9613)

9614 - **LIMA, C.R.; BOLFARINE, H.; SANDOVAL, M.C.** Linear calibration in multiplicative measurement error models. IME-USP, 1996. 12p. (RT-MAE-9614)

9615 - **AVERBACH, M.; CUTAIT, R.; WECHSLER, S.; CORRÊA, P.; BORGES, J.L.A.** Aplicação da inferência bayesiana no estudo das probabilidades de diagnóstico, por colonoscopia, das afecções colorretais em portadores da síndrome da imunodeficiência adquirida com diarréia. IME-USP, 1996. 10p. (RT-MAE-9615)

9616 - **YOSHIDA, O.S.; LEITE, J.G.; BOLFARINE, H.** Inferência bayesiana do número de espécies de uma população. IME-USP, 1996. 33p. (RT-MAE-9616)

9617 - **CARMONA, S.C.; TANAKA, N.I.** Exponential estimates for "Not Very Large Deviations" and ware front propagation for a class of reaction-diffusion equations. IME-USP, 1996. 30p. (RT-MAE-9617)

9618 - **IRONY, T.Z.; PEREIRA, C.A.B.; TIWARI, R.C.** On The Comparison Between Two Correlated Proportions in 2 x 2 Tables. IME-USP, 1996. 15p. (RT-MAE-9618)

9619 - **ANDRÉ, C.D.S.; ELIAN, S.N.; NARULA, S.C.; AUBIN, E.C.** Stepwise procedures for selecting variables in the minimum sum of absolute errors regression. IME-USP, 1996.9p. (RT-MAE-9619)

9620 - **ANDRÉ-C.D.S.; NARULA, S.C.; PERES, C.A.P.; VENTURA.G.A.** Asymptotic properties of the MSAE estimators in the dose-response model. IME-USP, 1996. 9p. (RT-MAE-9620)

9621 - **FERRARI, S.L.P.; CRIBARI-NETO, F.** On bootstrap and analytical bias corrections. IME-USP, 1996. 9p. (RT-MAE-9621)

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

9601 - **MENTZ, R.P.; MORETTIN, P.A. and TOLOI, C.M.C.** Bias correction for estimators of the residual variance in the ARMA (1,1) Model. São Paulo, IME-USP, 1996. 21p. (RT-MAE-9601)

9602 - **SINGER, J.M.; PERES, C.A.; HARLE, C.E.** Performance of Wald's test for the Hardy-Weiberg equilibrium with fixed sample sizes. São Paulo, IME-USP, 1996.15p. (RT-MAE-9602).

9603 - **SINGER, J.M. and E. SUYAMA, E.** Dispersion structure, Hierarchical models, Random effects models, Repeated measures. São Paulo, IME-USP, 1996. 21p. (RT-MAE-9603).

9604 - **LIMA, A.C.P. and SEN, P.K.** A Matrix-Valued Counting Process with First-Order Interactive Intensities. São Paulo, IME-USP, 1996, 25p. (RT-MAE-9604)

9605 - **BOTTER, D.A. and SINGER, J.M.** Experimentos com Intercâmbio de Dois Tratamentos e Dois Períodos: Estratégias para Análise e Aspectos Computacionais, São Paulo, IME-USP, 1996. 18p. (RT-MAE-9605)

9606- **MORETTIN, P.A.** From Fourier to Wavelet Analysis of Time Series. São Paulo, IME-USP, 1996. 11p. (RT-MAE-9606)

9607 - **SINGER, J.M.; HO, L.L.** Regression Models for Bivariate Counts. São Paulo, IME-USP, 1996. 20p. (RT-MAE-9607)

9608 - **CORDEIRO, G.M. and FERRARI, S.L.P.** A method of mements for finding Bartlett-type corrections. São Paulo, IME-USP, 1996. 11p. (RT-MAE-9608)

9609 - **VILCA-LABRA, F.; ARELLANO-VALLE, R.B.; BOLFARINE, H.** Elliptical functional models. São Paulo, IME-USP, 1996. 20p. (RT-MAE-9609)

9610 - **BELITSKY, V.; FERRARI, P.A.; KONNO, N.** A Refinement of Harris-FKG Inequality for Oriented Percolation. São Paulo, IME-USP, 1996. 13p. (RT-MAE-9610)

9622 - NARULA, S.C.; SALDIVA, P.H.N.; ANDRÉ, C.D.S.; ELIAN, S.N.; AUBIN, E.C.Q. MSAE regression: a more robust alternative to the least squares regression. IME-USP, 1996. 10p. (RT-MAE-9622)

9623 - BRANCO, M.; BOLFARINE, H; IGLESIAS, P. Bayesian calibration under a student-*t* model. IME-USP, 1996. 14p. (RT-MAE-9623)

9624 - GASCO, L.; BOLFARINE, H.; SANDOVAL, M.C. Regression estimators under multiplicative measurement error superpopulation models. IME-USP, 1996. 11p. (RT-MAE-9624)

9625 - BOTTER, D.A.; CORDEIRO, G.M. Improved estimators for generalized linear models with dispersion covariates. IME-USP, 1996. 12p. (RT-MAE-9625)

9626 - SUYAMA, E.; SINGER, J.M. Identification of random effects models for longitudinal data. IME-USP, 1996. 22p. (RT-MAE-9626)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05315-970 - São Paulo, Brasil