Transformation and Reparametrization
for Best Linear Unbiased Estimation of a
Linear Model with Arbitrary Known Variance

by

R. Schall¹

and

T.T. Dunne²

Technical Report 86-16, July 1986.

---

¹Messerschmitt-Bölkow-Blohm GmbH, D-800 München and University of Cape Town.

²University of Cape Town Rondebosch 7700, RSA and Rice University.

TRANSFORMATION AND REPARAMETRIZATION FOR BEST LINEAR
UNBIASED ESTIMATION OF A LINEAR MODEL WITH ARBITRARY
KNOWN VARIANCE

R. Schall

T.T. Dunne

Messerschmitt-Bölkow-Blohm GmbH
D - 8000 München
and University of Cape Town

University of Cape Town
Rondebosch 7700
RSA

Key Words and Phrases: best linear unbiased estimation; testing of hypothesis; arbitrary variance; reparametrizations; transformations; singular value decomposition; QR-decomposition

ABSTRACT

Two approaches to the problem of best linear unbiased estimation and testing of a linear hypothesis in a linear model under known arbitrary variance are presented which lead directly to efficient and numerically stable algorithms for the computation of estimates and test statistics.

Further, it is shown that every linear model with design matrix X and covariance structure $\sigma^2 V$, possibly not of full rank, can be reduced by a transformation and subsequent reparametrization of the model to a linear model with a full rank design matrix and covariance structure $\sigma^2 I$ which is statistically equivalent to the original model.

1. INTRODUCTION

The problem of best linear unbiased estimation (BLUE) and the related problem of testing a linear hypothesis in a linear model under arbitrary variance are quite old, and have been solved by several authors, e.g. Goldman and Zelen (1964), Zyskind and Martin (1969) and Rao (1971).

Those approaches emphasize the theoretical solution of the problem rather than the development of efficient and numerically stable algorithms for the computation of estimates and test statistics.

In the following we present two approaches to the problem of BLU-estimation and testing of a linear hypothesis in a linear model under arbitrary variance. We will call those approaches the T - ("by transformation") and R - ("by reparametrization") method respectively.

Many approaches to this problem are already available. The possible merit of the methods presented here might lie in the fact that they directly lead to efficient, numerically stable and easily programmable algorithms for the computation of BLU-estimates and test-statistics.

Concerning the problem of testing a linear hypothesis, they allow for the check of consistency of a linear hypothesis, check for testability of the hypothesis and for the computation of degrees of freedom of the associated statistics.

The methods are related to the work of Goldman and Zelen, Zyskind and Martin and Rao.

The approaches are unified here in the sense that they include the well known Gauß-Markov-model (with design matrix X of full rank and covariance structure $\sigma^2 I$) as a special case, as well as best linear constrained unbiased estimation (BLICUE).

Insight into the nature of a linear model under arbitrary variance is provided and it is shown that every linear model under arbitrary variance can be reduced through a transformation and subsequent reparametrization to a model with full rank design matrix

and covariance structure $E(ee') = \sigma^2 I$, which is statistically equivalent to the original model.

2. THE MODEL

We consider the well known linear model (LM)

$$y = X\beta + e; E(ee') = \text{cov}(e) = \sigma^2 V; \quad (2.1)$$

where y the vector of observations is $n \times 1$, X is the known design matrix of order $n \times p$ and rank $r \leq p$, β is the $p \times 1$ vector of parameters and e is $n \times 1$ with zero expectation and covariance structure $\sigma^2 V$, V nonnegative definite.

In the following we rely heavily on the result of Mitra and Rao (1968) that BLUE's are invariant under a nonsingular transformation of the model (2.1).

Further, a reparametrized model is equivalent in every way to the original, in respect of estimation (Pringle and Rayner, 1971).

Thus the BLUE $X\hat{\beta}$ for $X\beta$ in model (2.1) is the same as in

$$Ty = TX\beta + Te; E(Tee'T') = \sigma^2 TVT'; \quad (2.2)$$

and

$$y = X^*\beta^* + e = XUW\beta + e; E(ee') = \sigma^2 V; \quad (2.3)$$

the transformed and reparametrized models respectively, where T is any nonsingular $n \times n$ - matrix and $X^* = XU$, $\beta^* = W\beta$ any rank preserving reparametrization of (2.1).

3. BEST LINEAR UNBIASED ESTIMATION

Let $R(X)$ and $C(X)$ denote respectively the row and column spaces of a matrix X .

3.1 The T - method

The BLUE $X\hat{\beta}$ for $X\beta$ in the model (2.1) can be obtained by two subsequent transformations of the model:

Algorithm 3.1

Step 1: Choose a transformation T_1 such that

$$T_1 V T_1' = \begin{bmatrix} I_k & : & 0 \\ 0 & : & 0 \end{bmatrix} = \begin{bmatrix} I & : & 0 \\ 0 & : & 0 \end{bmatrix}, \quad k = \text{rank}(V)$$

and set $y^{(1)} = \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \end{bmatrix} := T_1 y$

$$X^{(1)} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \end{bmatrix} := T_1 X \quad \text{conformably partitioned,}$$

to obtain the model

$$\begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \end{bmatrix} \beta + e^{(1)}; \quad \text{cov}(e^{(1)}) = \sigma^2 \begin{bmatrix} I & : & 0 \\ 0 & : & 0 \end{bmatrix}. \quad (3.1)$$

Step 2: Choose a transformation $T_2 = \begin{bmatrix} I & : & -A \\ 0 & : & I \end{bmatrix}$

and set $y^{(2)} = \begin{bmatrix} y_1^{(2)} \\ y_2^{(2)} \end{bmatrix} := T_2 \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \end{bmatrix} = \begin{bmatrix} y_1^{(1)} - A y_2^{(1)} \\ y_2^{(1)} \end{bmatrix}$

$$X^{(2)} = \begin{bmatrix} X_1^{(2)} \\ X_2^{(2)} \end{bmatrix} := T_2 \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} - A X_2^{(1)} \\ X_2^{(1)} \end{bmatrix}$$

such that $R(X_1^{(2)}) \cap R(X_2^{(2)}) = \{0\}$. With

$$T_2 \begin{bmatrix} I & : & 0 \\ 0 & : & 0 \end{bmatrix} T_2' = \begin{bmatrix} I & : & 0 \\ 0 & : & 0 \end{bmatrix}$$

we obtain the model

$$\begin{bmatrix} y_1^{(2)} \\ y_2^{(2)} \end{bmatrix} = \begin{bmatrix} X_1^{(2)} \\ X_2^{(2)} \end{bmatrix} \beta + e^{(2)}; \quad \text{cov}(e^{(2)}) = \sigma^2 \begin{bmatrix} I & : & 0 \\ 0 & : & 0 \end{bmatrix}. \quad (3.2)$$

Step 3: The BLUE $X^{(2)} \hat{\beta}$ for $X^{(2)} \beta$ in model (3.2) is the straightforward least-squares-solution, i.e.

$$X^{(2)} \hat{\beta} = X^{(2)} (X^{(2)'} X^{(2)})^{-1} X^{(2)'} y \quad (3.3)$$

and thus the BLUE $X \hat{\beta}$ for $X \beta$ in the original model (2.1) is

$$X \hat{\beta} = T_1^{-1} T_2^{-1} X^{(2)} \hat{\beta} = T^{-1} X^{(2)} \hat{\beta}, \text{ where } T := T_2 T_1. \quad (3.4)$$

Notes on the steps of the algorithm:

1. T_1 can be obtained by standard methods and we name but one in particular: let $V = P \begin{bmatrix} \Delta & : & 0 \\ 0 & : & 0 \end{bmatrix} P'$ be the (complete)

singular value decomposition (SVD) of V , i.e. P is an orthogonal matrix of the same order as V and Δ is a $k \times k$ diagonal matrix containing the positive eigenvalues of V .

$$\text{Then } T_1 = \begin{bmatrix} \Delta^{-1/2} & : & 0 \\ 0 & : & I \end{bmatrix} P' \quad (\text{say}).$$

Program libraries like NAG and IMSL provide the SVD of a matrix, using the algorithm of Golub and Reinsch (1970) which is efficient and numerically stable.

2. Without loss of generality A can be taken as

$$A = X_1^{(1)} X_2^{(1)'} (X_2^{(1)} X_2^{(1)'})^{-1}.$$

$$\begin{aligned} \text{Then } X_1^{(2)} &= X_1^{(1)} - A X_2^{(1)} = X_1^{(1)} (I_p - X_2^{(1)} (X_2^{(1)} X_2^{(1)'})^{-1} X_2^{(1)'}) =: \\ &:= X_1^{(1)} P_{R^\perp(X_2^{(2)})}, \text{ where } P_{R^\perp(X_2^{(2)})} \text{ is the orthogonal} \\ &\text{projector on } R^\perp(X_2^{(2)}), \text{ the orthogonal complement of } R(X_2^{(2)}). \end{aligned}$$

Thus $R(X_1^{(2)}) \perp R(X_2^{(2)})$ which implies of course

$$R(X_1^{(2)}) \cap R(X_2^{(2)}) = \{0\}.$$

A is most effectively computed using the QR-factorization of $X_2^{(1)}$, (see appendix A). If $X_2^{(1)} = Q_1 [R_1 : D] M$

then $A = X_1^{(1)} Q_1 [(R_1^{-1})' : 0] M$ (say) and $X_1^{(2)} = X_1^{(1)} (I_p - Q_1 Q_1')$.

$$3. \text{ If } T_1 = \begin{bmatrix} \Delta^{-1/2} & 0 \\ 0 & I \end{bmatrix} P' \text{ then } T_1^{-1} = P \begin{bmatrix} \Delta^{1/2} & 0 \\ 0 & I \end{bmatrix}$$

$$\text{and if } T_2 = \begin{bmatrix} I & -A \\ 0 & I \end{bmatrix} \text{ then } T_2^{-1} = \begin{bmatrix} I & A \\ 0 & I \end{bmatrix},$$

thus no further computations are necessary to obtain the inverses T_1^{-1} , T_2^{-1} . $X^{(2)} \hat{\beta}$ in (3.3) is computed using the QR-factorization of $X^{(2)}$ (see appendix A).

Linking this approach to known theory we note that Goldman and Zelen (1964) obtained the transformed model (3.1) and in effect proceeded by minimizing

$$(y_1^{(1)} - X_1^{(1)} \beta)' (y_1^{(1)} - X_1^{(1)} \beta) \text{ subject to}$$

$$y_2^{(1)} = X_2^{(1)} \beta,$$

i.e. by solving constrained least squares, whereas we proceed in step 2 of the algorithm by disconnecting the constraints from the "statistical" part of the data.

It is here, in step 2 of the algorithm, that BLICU-estimation fits in. If we have the constraints $c = C\beta$, we augment $y_2^{(1)}$ by c and $X_2^{(1)}$ by C and proceed with $y_2^{(c)} = \begin{bmatrix} y_2^{(1)} \\ c \end{bmatrix}$ and $X_2^{(c)} = \begin{bmatrix} X_2^{(1)} \\ C \end{bmatrix}$.

Further, we note that in (3.4) $X \hat{\beta}$ can be written as

$$\begin{aligned} X \hat{\beta} &= T^{-1} X^{(2)} \hat{\beta} \\ &= X (X' T' T X)^{-1} X' T' T y \\ &= X (X' V^* X)^{-1} X' V^* y, \quad V^* = T' T. \end{aligned}$$

V^* is easily shown to be a g-inverse of V and it can be shown that $VV^*X = XB$ for some B and that $\text{rank}(X'V^*X) = \text{rank}(X)$. Thus V^* is a g-inverse of V as given by Zyskind and Martin (1969) and Rao (1971). (Algorithm 3.1 can then be used for the computation of such a g-inverse. For a proof of the relationship $VV^*X = XB$ see appendix B.)

That (3.3) gives the BLUE $X^{(2)} \hat{\beta}$ for $X^{(2)} \beta$ in model (3.2) is obvious if $R(X_1^{(2)}) \perp R(X_2^{(2)})$. But $R(X_1^{(2)}) \cap R(X_2^{(2)}) = \{0\}$ is sufficient indeed for (3.3) to be true, which follows from a generalization by Dunne (1982, Th. 2.10) of a result by Chipman (1964, p. 1083).

3.2 The R-method

The BLUE $X \hat{\beta}$ for $X \beta$ in the model (2.1) can be obtained by a transformation and subsequent reparametrization of the model:

Algorithm 3.2

Step 1: Same as step 1 in algorithm 3.1.

Step 2: Compute the QR-factorization of $X_2^{(1)'} , i.e.$
 $X_2^{(1)} = M[R' : 0'] Q'$ and reparametrize the model (3.1) as
 $X^* = X^{(1)} Q ; \beta^* = Q' \beta$ to obtain the model

$$\begin{bmatrix} y_1^{(1)} \\ y_2^{(3)} \end{bmatrix} = \begin{bmatrix} X_{11}^* : X_{12}^* \\ R' : 0' \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} + e^{(1)}; \text{cov}(e^{(1)}) = \sigma^2 \begin{bmatrix} I : 0 \\ 0 : 0 \end{bmatrix} \quad (3.5)$$

where $y_2^{(3)} := M y_2^{(1)}$, and β_1^* is $r_1 \times 1$, $r_1 = \text{rank}(X_2^{(1)})$.

Step 3: Compute $\beta_1^* = \hat{\beta}_1^*$ from

$$y_2^{(3)} = \begin{bmatrix} R_1' \\ D' \end{bmatrix} \beta_1^* \quad (D, R_1 \text{ as in the appendix A}).$$

Step 4: The BLUE $X_{12}^* \hat{\beta}_2^*$ for $X_{12}^* \beta_2^*$ is

$$X_{12}^* \hat{\beta}_2^* = X_{12}^* (X_{12}^{*'} X_{12}^*)^{-1} X_{12}^{*'} (y_1^{(1)} - X_{11}^* \hat{\beta}_1^*)$$

and the BLUE $X\hat{\beta}$ for $X\beta$ is

$$X\hat{\beta} = T_1^{-1} \begin{bmatrix} X_{11}^* & : & X_{12}^* \\ R' & : & 0' \end{bmatrix} \hat{\beta}^*$$

Notes on some steps of the algorithm:

$$3. \text{ If } y_2^{(3)} = \begin{bmatrix} y_{21}^{(3)} \\ y_{22}^{(3)} \end{bmatrix} \text{ is conformably partitioned with } \begin{bmatrix} R'_1 \\ D' \end{bmatrix}$$

then of course $\beta_1^* = \hat{\beta}_1^*$ is computed from

$$y_{21}^{(3)} = R'_1 \beta_1^*$$

which is particularly easy and stable since R'_1 is triangular.

$$4. X_{12}^* \hat{\beta}_2^* \text{ is again computed using the QR-factorization of } X_{12}^*.$$

This approach can also be linked to the work of Zyskind and Martin (1969) and Rao (1971): Since in model (3.5) we have

$$C\left(\begin{bmatrix} X_{12}^* \\ 0 \end{bmatrix}\right) \subset C(V) \text{ and } C\left(\begin{bmatrix} X_{11}^* \\ R' \end{bmatrix}\right) \cap C(V) = \{0\}, V = \begin{bmatrix} I : 0 \\ 0 : 0 \end{bmatrix}$$

we have that $V^* = (V + \begin{bmatrix} X_{11}^* \\ R' \end{bmatrix} [X_{11}^{*'} : R])^-$ is a g-inverse of V as given

by the above authors in model (3.5), i.e. $X\hat{\beta} = X(X'V^+X)^-X'V^+y$ is the BLUE for $X\beta$ in model (2.1), where $V^+ = T_1' V^* T_1$. Thus algorithm 3.2 is also a method to compute such a g-inverse.

4. REDUCTION OF A LINEAR MODEL AND TESTING OF A LINEAR HYPOTHESIS

With steps 1 and 2 of algorithm 3.2 we have reduced the linear model (2.1) to the form (3.5). Since $\beta_1^* = b_1 := (R'_1)^{-1} y_{21}^{(3)}$ with probability 1, the model

$$(y_1^{(1)} - X_{11}^* b_1) = X_{12}^* \beta_2^* + e_1^{(1)}; \text{ cov } (e_1^{(1)}) = \sigma^2 I \quad (4.1)$$

is statistically equivalent to (3.5), under normal theory.

By a further reparametrization of (4.1) we can obtain a model with full rank design matrix (if X_{12}^* is not of full rank in the first place).

If we wish to test the linear hypothesis $H_0 : H\beta = h$ we can proceed as follows:

Algorithm 4.1:

Step 1: Reparametrize H_0 as $H^* = HQ$, $\beta^* = Q'\beta$ where Q follows from Step 2 of algorithm 3.2. Now, conformably partitioned, H_0 can be written as

$$H_0 : \begin{bmatrix} H_1^* & H_2^* \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} = h.$$

Step 2: Since $H_1^* \beta_1^*$ is constant and known (β_1^* from step 3 of algorithm 3.2), H_0 can be written as

$$H_0 : H_2^* \beta_2^* = (h - H_1^* \beta_1^*).$$

Step 3: (i) check for consistency:

$$H_0 \text{ consistent} \iff (h - H_1^* \beta_1^*) \in C(H_2^*).$$

(ii) check for testability:

$$H_0 \text{ testable if and only if } R(H_2^*) \subset R(X_{12}^*).$$

(iii) degrees of freedom for hypothesis: df = rank (H_1^*).

(iv) test statistic:

Will be provided by any standard regression program: simply test $H_0 : H_2^* \beta_2^* = (h - H_1^* \beta_1^*)$ in the model (4.1) with covariance structure $\sigma^2 I$.

APPENDIX A

Let X be a matrix of order $n \times p$ with rank $r \leq \min(n, p)$. Then by applying r Householder-transformations X can be decomposed as

$$X = Q \begin{bmatrix} R_1 & : & D \\ O_1 & : & O_2 \end{bmatrix} M = Q \begin{bmatrix} R \\ 0 \end{bmatrix} M$$

where Q orthogonal of order $n \times n$
 R_1 upper triangular of order $r \times r$
 D some matrix of order $r \times (p - r)$
 O_1 a null-matrix of order $(n - r) \times r$
 O_2 a null-matrix of order $(n - r) \times (p - r)$
 $R = [R_1 : D]$, $O = [O_1 : O_2]$
 M a permutation matrix of order $p \times p$, permuting the columns of $Q'X$ to ensure that R_1 is nonsingular. We can see to it that M is symmetric.

Some or all of the matrices D , O_1 , O_2 may vanish:

D vanishes if $r = p$, i.e. X is of full column rank
 O_1 vanishes if $r = n$, i.e. X is of full row rank
 O_2 vanishes if D or O_1 vanishes.

M may be taken as $M = I_p$ if the first r columns of X are linearly independent.

If $Q = [Q_1 : Q_2]$ is conformably partitioned, i.e. Q_1 is of order $n \times r$ and Q_2 is of order $n \times (n - r)$, then $X = Q_1 [R_1 : D] M$.

Important are the following identities:

$$(i) \quad X(X'X)^{-} X' = Q_1 Q_1'$$

and

$$(ii) \quad (X'X)^{-} X' = M' \begin{bmatrix} R_1^{-1} \\ 0 \end{bmatrix} Q_1' = M' \begin{bmatrix} R_1^{-1} Q_1' \\ 0 \end{bmatrix} \quad (\text{say}).$$

APPENDIX B

That $VV^*X = XB$ for some B , where $V^* = T'T$ can be verified by using a theorem by Zyskind (1967):

"A linear function $l'y$ of y is BLUE for its expectation $l'XB$ in the $LM(y, XB, \sigma^2 V)$ if and only if $Vl \in C(X)$."

$$\begin{aligned}\text{Then } X\hat{\beta} &= X(X'V^*X)^{-}X'V^*y \\ &= X(X'T'TX)^{-}X'T'Ty \text{ BLUE for } XB\end{aligned}$$

$$\begin{aligned}\Rightarrow C(VT'TX(X'T'TX)^{-}X') &\subset C(X) \\ \Rightarrow C(VT'TX(X'T'TX)^{-}X'T'TX) &\subset C(X) \\ \Rightarrow C(VT'A(A'A)^{-}A'A) &\subset C(X), \quad A:=TX \\ \Rightarrow C(VT'A) &\subset C(X), \\ &\text{since } A(A'A)^{-}A'A = A \text{ for any matrix } A \text{ and} \\ &\text{any g-inverse } (A'A)^{-} \text{ of } A'A \text{ (see e.g. Rao, 1973).} \\ \Rightarrow C(VT'TX) &\subset C(X) \\ \Rightarrow C(VV^*X) &\subset C(X) \\ \Rightarrow VV^*X &= XB, \text{ for some } B.\end{aligned}$$

BIBLIOGRAPHY

- Chipman, J.S. (1964). On least squares with insufficient observations. Journal of the American Statistical Association, 59, 1078 - 1111.
- Dunne, T.T. (1982). Contributions to the theory of generalized inverses, the linear model and outliers. Ph D thesis, University of Cape Town.
- Goldman, A.J. and Zelen, M. (1964). Weak generalized inverses and minimum variance linear unbiased estimation. Journal of Research of the National Bureau of Standards, 68 B, 151 - 172.
- Golub, G.H. and Reinsch, C. (1970). Singular value decomposition and least squares solutions. Numerische Mathematik, 14, 403 - 420.
- Mitra, S.K. and Rao, C.R. (1968). Some results in estimation and tests of linear hypothesis under the Gauß-Markov-model. Sankhya, Ser. A, 30, 281 - 290.

.Pringle, R.M. and Rayner, A.A. (1971). Generalized inverse matrices with applications to statistics. Griffin, London.

Rao, C.R. (1971). Unified theory of linear estimation. Sankhya, Ser. A, 33, 371 - 394 .

Rao, C.R. (1973). Linear statistical inference and its applications. Wiley, New York.

Zyskind, G. (1967). On canonical forms, nonnegative covariance matrices and best and simple least squares linear estimators in linear models. Annals of Mathematical Statistics, 38, 1092 - 1109.

Zyskind, G. and Martin, F.B. (1969). A general Gauß-Markoff theorem for linear models with arbitrary non-negative covariance structure. SIAM Journal of Applied Mathematics, 17, 1190 - 1202.