DISCUSSION OF PAPER PUBLISHED IN VOLUME LXXXIII

LOSS PREDICTION BY GENERALIZED LEAST SQUARES

LEIGH J. HALLIWELL

DISCUSSION BY MICHAEL D. HAMER

Abstract

The paper by Halliwell [1] and the Discussion of Halliwell's paper by Dr. Schmidt both consider the form of "best" linear unbiased estimators for unknown quantities based on observable values. This paper proposes a general definition of "best" called Uniformly Best (UB) to distinguish it from previous definitions and provides various equivalent forms for the definition. It shows the existence and uniqueness of such UB linear unbiased estimators under fairly general conditions, provides an alternative formulation of the definition of UB for unbiased estimators, and discusses how Dr. Schmidt's proposed optimization problem relates to the proposed UB definition.

ACKNOWLEDGEMENT

I would like to thank Mr. Halliwell and Dr. Schmidt for their valuable comments on earlier drafts of this paper, and Dr. Schmidt for suggesting a shorter version of the proof to Theorem 6.1. Needless to say, the remaining opinions and errors are mine.

1. THE STRUCTURE OF THE VARIABLES

We follow the notation used by Halliwell and Schmidt. An n-dimensional random vector Y is vertically partitioned into a

p-dimensioned vector Y_1 of observable outcomes and an $n-p$ dimensioned vector Y_2 of unobservable outcomes. It is assumed that Y takes the form

$$Y = X\beta_0 + e,$$

where e is an n-dimensional random vector of "error" terms with zero mean and $(n \times n)$ dimensional non-singular variance-covariance matrix S (thus $\mathrm{E}[ee^T] = S$ where e^T represents the transpose of e, and S is positive definite), X is a given $(n \times r)$ "design" matrix, and β_0 is an unknown parameter vector of dimension r.

The matrix X and vector e can also be partitioned so that we may write

$$Y_1 = X_1\beta_0 + e_1 \qquad \text{and} \qquad Y_2 = X_2\beta_0 + e_2,$$

where X_1 is a $(p \times r)$ matrix and X_2 is a $(n-p \times r)$ matrix, and we assume that X_1 is of full rank r.

2. A PROPOSED DEFINITION OF "BEST"—THE OBJECTIVE FUNCTION

Halliwell provides a definition of "best" in Appendix A of [1], where he considers linear unbiased estimators β for the unknown vector β_0. We use this as a basis for proposing a more general definition of a "best" estimator P of a "target" quantity T. We call this definition Uniformly Best to distinguish it from other definitions of "best" used in [1–3].

Firstly, we provide a definition of a non-negative definite matrix:

DEFINITION 2.1 Non-Negative Definite. *An $(n \times n)$ matrix M is non-negative definite if $\alpha^T M \alpha \geq 0$ for any n-dimensional vector α.*

Halliwell provides an extensive review of non-negative definite matrices in Appendix A of [2]. Perhaps the most relevant characteristic for our purposes is that any non-negative definite matrix M can be written in the form $W^T W$ for some matrix W, and conversely that any matrix of the form $W^T W$ is non-negative definite.

We use the concept of non-negative definite in our proposed definition of "best" as follows:

DEFINITION 2.2 Uniformly "Best" (UB) Estimator. *A estimator P^* of a target quantity T is uniformly "best" (UB) if, for any other admissible estimator P, the matrix* $\{\mathrm{Var}(T-P) - \mathrm{Var}(T-P^*)\}$ *is non-negative definite.*

For an n-dimensional random vector z, the upper-case $\mathrm{Var}(z)$ is the $(n \times n)$ dimensional variance-covariance matrix of z where

$$\mathrm{Var}(z) = \mathrm{E}[(z - \mathrm{E}[z])(z - \mathrm{E}[z])^T].$$

Elsewhere, we will use the lower-case $\mathrm{var}(x)$ to denote the variance of a scalar random variable x.

To assist in understanding the nature of a UB estimator, we provide the following "equivalence" result:

LEMMA 2.1 *Suppose we consider estimators P that belong to some given admissible set J. The following statements are equivalent:*

(a) *There exists an estimator P^* in J that is the UB estimator of T.*

(b) *For any admissible P belonging to J, the matrix* $\{\mathrm{Var}(T-P) - \mathrm{Var}(T-P^*)\}$ *is non-negative definite.*

(c) *P^* minimizes* $\alpha^T \mathrm{Var}(T-P)\alpha$ *over all admissible P for any α of appropriate dimension.*

(d) *P^* minimizes* $\mathrm{var}(\alpha^T(T-P))$ *over all admissible P for any α of appropriate dimension.*

Proof (a) and (b) are equivalent from Definition 2.2.

From Definition 2.1 and (b), we have

$$\alpha^T[\mathrm{Var}(T-P) - \mathrm{Var}(T-P^*)]\alpha \geq 0$$

for any suitable α and for any P belonging to J. Then

$$\alpha^T \mathrm{Var}(T-P)\alpha \geq \alpha^T \mathrm{Var}(T-P^*)\alpha$$

for any P belonging to J, and so (c) follows. To show (d), we have

$$\begin{aligned}
\alpha^T \mathrm{Var}(T-P)\alpha &= \alpha^T \mathrm{E}[(T-P-\mathrm{E}[T-P])(T-P-\mathrm{E}[T-P])^T]\alpha \\
&= \mathrm{E}[\alpha^T(T-P-\mathrm{E}[T-P])(T-P-\mathrm{E}[T-P])^T\alpha] \\
&= \mathrm{var}(\alpha^T(T-P)).
\end{aligned}$$

The definition of UB given in (d) provides us with an objective function that we show below is easy to work with, and is perhaps the easiest to conceptualize. $\alpha^T(T-P)$ can be interpreted as the "length" of the projection of the stochastic vector representing the difference between the target T and the estimator P onto any fixed vector α. The UB estimator P^* minimizes the variance of this projection and does so for any given α.

The UB criterion is potentially quite difficult to meet. Expanding out $\mathrm{var}(\alpha^T(T-P))$ we have:

$$\mathrm{var}(\alpha^T(T-P)) = \sum\sum \alpha_i\alpha_j \,\mathrm{cov}(T_i - P_i, T_j - P_j).$$

The UB estimator P^* must minimize this double sum of products for any possible choice of α_i. However, UB estimators do exist for suitable admissible sets and targets, as shown below.

3. CONSTRAINTS ON ADMISSIBLE ESTIMATORS AND TARGETS

The definition of UB does not put any particular constraints on the admissible sets of estimators or on the form of the "target" quantities. However, it may be necessary to do so to ensure the existence of UB estimators.

(a) *Constraints on Admissible Sets Of Estimators*. Following Halliwell and Schmidt, we wish to consider estimators P that are linear in Y_1 and unbiased estimators of their "targets" T, so we define the set J of linear unbiased estimators as follows:

DEFINITION 3.1 The Admissible Set $J = J(Y_1, T)$. *An estimator P belongs to J if it is*

- *linear in Y_1 and hence of the form $P = QY_1$ where Q is a* $(n \times p)$ *matrix;*
- *unbiased, so that* $\mathrm{E}[P] = \mathrm{E}[T]$.

(b) *Constraints on "Targets".* We also need to define the "target" quantity T that is being estimated. For the Gauss–Markov theorem it is β_0, but elsewhere in [1] and in Schmidt's paper Y_2 and Y are also considered. To encompass all these possibilities, we consider a general form

$$T = F_1 Y_1 + F_2 Y_2 + A\beta_0,$$

where F_1, F_2 and A are variables. Since T is a vector of dimension n, F_1 is an $(n \times p)$ matrix, F_2 is an $(n \times n - p)$ matrix, and A is an $(n \times r)$ matrix.

4. EXISTENCE OF A UB LINEAR UNBIASED ESTIMATOR FOR T

The following theorem shows that there are many situations in which a UB solution not only exists but is unique.

THEOREM 4.1 *If $T = F_1 Y_1 + F_2 Y_2 + A\beta_0$ and P belongs to the admissible set J, a unique UB linear unbiased estimator $P^* = Q^* Y_1$ exists, and*

$$P^* = F_1 Y_1 + F_2 y_2(\beta^*) + A\beta^*,$$

where

$$y_2(\beta^*) = X_2\beta^* + S_{21}S_{11}^{-1}(Y_1 - X_1\beta^*) \qquad \textit{and}$$
$$\beta^* = (X_1^T S_{11}^{-1} X_1)^{-1} X_1^T S_{11}^{-1} Y_1.$$

Proof A proof of this theorem is presented in the Appendix.

Note the appearance of the Gauss–Markov estimator β^* and the predictor $y_2(\beta^*)$ discussed by Halliwell and Schmidt.

Theorem 4.1 has several interesting special cases.

CASE 1 The UB estimator for β_0

We set

$$F_1 = F_2 = 0 \qquad \text{and} \qquad A = \begin{bmatrix} I(r) \\ 0 \end{bmatrix}$$

where $I(r)$ is an $(r \times r)$ identity matrix. Then

$$Q^* = \begin{bmatrix} \beta^* \\ 0 \end{bmatrix}$$

as required by the Gauss–Markov Theorem, and the definition of UB is consistent with the Gauss–Markov notion of "best".

CASE 2 The UB estimator for Y_2

We set

$$F_1 = A = 0 \qquad \text{and} \qquad F_2 = \begin{bmatrix} 0 \\ I(n-p) \end{bmatrix},$$

where $I(n-p)$ is an $(n-p \times n-p)$ identity matrix. Then

$$Q^* = \begin{bmatrix} 0 \\ y_2(\beta^*) \end{bmatrix},$$

the form of the "best" predictor suggested by Halliwell.

CASE 3 The UB estimator for Y_1

We set

$$F_2 = A = 0 \qquad \text{and} \qquad F_1 = \begin{bmatrix} I(p) \\ 0 \end{bmatrix}.$$

Then

$$Q^* = \begin{bmatrix} Y_1 \\ 0 \end{bmatrix}.$$

Case 3 seems trivial, for of course the difference between an estimator Y_1 and target Y_1 will have zero variance. However, this result still "fits" our process, because the estimator Y_1 is certainly linear in Y_1 and unbiased.

CASE 4 The UB estimator for Y

We set

$$A = 0, \qquad F_1 = \begin{bmatrix} I(p) \\ 0 \end{bmatrix} \qquad \text{and} \qquad F_2 = \begin{bmatrix} 0 \\ I(n-p) \end{bmatrix}.$$

Then

$$Q^* = \begin{bmatrix} Y_1 \\ y_2(\beta^*) \end{bmatrix}.$$

The UB estimator for Y is thus a linear combination of the UB estimators for Y_1 and Y_2. This last result will be used in Section 6.

5. A FURTHER CHARACTERIZATION OF UB

In his Discussion, Schmidt proposes a related optimization problem in which the objective function to be minimized is $\mathrm{E}[(Y - X\beta)^T S^{-1}(Y - X\beta)]$.

We generalize Schmidt's objective function by replacing S^{-1} with any non-negative definite matrix H, and use this to define another type of estimator, which we will call Generalized Schmidt Best.

DEFINITION 5.1 Generalized Schmidt Best (GSB) Estimator. *An estimator P^* of a target quantity T is GSB if it minimizes*

$$\mathrm{E}[(T - P)^T H(T - P)]$$

over all admissible estimators P for any $(n \times n)$ non-negative definite matrix H.

How does a GSB estimator relate to a UB estimator? Rather surprisingly, the answer is that when the admissible set consists of unbiased estimators, if one exists, then they both exist and are the same.

THEOREM 5.1 *If the admissible estimators P of a general target T are all unbiased, an estimator P^* is UB if and only if it is GSB.*

Proof From our discussion of non-negative matrices, we know we can write $H = WW^T$ for some $(n \times n)$ matrix W. Now let z_i be a vector whose ith component is 1 and whose other components are all zero.

(i) Suppose a UB estimator P^* exists. For any other unbiased estimator P and any $H = WW^T$,

$$\begin{aligned}
&\mathrm{E}[(T-P)^T WW^T (T-P)] \\
&\quad = \mathrm{trace}\{\mathrm{E}[W^T(T-P)(T-P)^T W]\} \\
&\quad = \mathrm{trace}\{W^T \mathrm{Var}(T-P)W\}, \qquad \text{since} \quad \mathrm{E}[T-P]=0 \\
&\quad = \Sigma z_i^T W^T \mathrm{Var}(T-P)Wz_i, \qquad \text{where the sum is over } i \\
&\quad = \Sigma \alpha_i^T \mathrm{Var}(T-P)\alpha_i \qquad \text{for} \quad \alpha_i = Wz_i \\
&\quad \geq \Sigma \alpha_i^T \mathrm{Var}(T-P^*)\alpha_I, \qquad \text{since} \quad P^* \text{ is UB} \\
&\quad = \mathrm{E}[(T-P^*)^T WW^T (T-P^*)].
\end{aligned}$$

Thus P^* is also GSB.

(ii) Suppose a GSB estimator P^* exists but P^* is not UB. This means, for some $\alpha^\#$ and for some admissible P, we must have

$$\alpha^{\#T}\mathrm{Var}(T-P^*)\alpha^\# > \alpha^{\#T}\mathrm{Var}(T-P)\alpha^\#.$$

We can construct the matrix $W^\# = \{\alpha^\#, \alpha^\# \ldots, \alpha^\#)$ so that $\alpha^\# = W^\# z_i$ for any i. Then

$$z_i^T W^{\#T}\mathrm{Var}(T-P^*)W^\# z_i > z_i^T W^{\#T}\mathrm{Var}(T-P)W^\# z_i \qquad \text{for any } z_i.$$

Thus

$$\begin{aligned}&\mathrm{E}[(T-P^*)^T W^\# W^{\#T}(T-P^*)]\\&\qquad > \mathrm{E}[(T-P)^T W^\# W^{\#T}(T-P)],\end{aligned}$$

which contradicts the assumption that P^* is GSB. Thus P^* is also UB.

This proof does not require that the admissible estimators be linear in Y_1, nor does it impose any constraint on the form of the "target" T. But it is likely that a "best" solution will not always exist unless there are further restrictions on the admissible estimator set and the target because the UB and GSB conditions are so strong. When T is linear in Y_1, Y_2 and β_0 and the set J consists of linear unbiased estimators, Theorem 4.1 tells us that a UB estimator does exist, and then, from Theorem 5.1, the GSB estimator will be the same as a UB estimator.

More generally, we can use Theorem 5.1 to state an extended "equivalence" result.

LEMMA 5.1 *If the admissible set only contains unbiased estimators of a general "target" T, the following statements are equivalent (but not necessarily true):*

(a) *There exists a P^* that is the UB estimator of T for all admissible estimators P.*

(b) *For any unbiased P, the matrix* $\{\mathrm{Var}(T-P)-\mathrm{Var}(T-P^*)\}$ *is non-negative definite.*

(c) P^* *minimizes* $\alpha^T\mathrm{Var}(T-P)\alpha$ *over all admissible P for any* α *of appropriate dimension.*

(d) P^* *minimizes* $\mathrm{var}(\alpha^T(T-P))$ *over all admissible P for any* α *of appropriate dimension.*

(e) P^* *minimizes* $\mathrm{E}[(T-P)^T H(T-P)]$ *over all admissible P for any non-negative definite matrix H of appropriate dimension.*

If we expand the objective function in (e), we get

$$\mathrm{E}[(T-P)^T H(T-P)] = \sum\sum h_{ij}\,\mathrm{cov}(T_i - P_i, T_j - P_j),$$

and the UB estimator P^* minimizes this double sum over all possible choices of h_{ij} provided the h_{ij} belong to a non-negative definite matrix. This is more general than (d), which corresponds to the case where $h_{ij} = \alpha_i \alpha_j$. (Note: we can think of any non-negative definite matrix as a possible variance-covariance matrix if we allow the possibility that some of the variances may be zero. In this context, (d) corresponds to the case where all correlations are either $+1$ or -1, and (e) generalizes this to correlations in between.)

6. RELATIONSHIP BETWEEN "BEST" AND SCHMIDT'S OPTIMIZATION PROBLEM

In his Discussion and in [3], Schmidt suggests an optimization problem as a way of justifying the form of the "best" estimators for Y_1 and Y_2. Schmidt shows that his optimization problem can be decomposed into two parts, one involving only Y_1 and the other involving only Y_2. Further, he shows that the solution to the initial optimization problem is achieved by $\beta = \beta^*$, the Gauss–Markov estimator for β_0, and β^* minimizes each of the parts separately. In view of this optimization, Schmidt proposes that the solutions to the separate optimization problems of the parts are "best" estimators for Y_1 and Y_2, respectively.

The objective function for his optimization problem is a special case of the GSB objective function when $H = S^{-1}$ and the target $T = Y$. In addition, however, Schmidt's optimization problem requires that the admissible estimators belong to a set K, where

$$K = \{P : P = X\beta \text{ where } \beta = BY_1 \text{ and } BX_1 = I(r)\}.$$

This constraint means that the estimators in K are linear unbiased estimators of Y, but also the estimators BY_1 are also unbiased estimators of β_0.

Although Schmidt's optimization looks like the GSB objective function and K is a subset of J, the solution to Schmidt's optimization is not in general a UB estimator for Y. This is because K does not include all linear unbiased estimators of Y, and in general (except in the special circumstance detailed below) the UB linear unbiased estimator of Y is not in K.

THEOREM 6.1 *Unless X_1 is square, the UB linear unbiased estimator for Y will not belong to K.*

Proof From Theorem 4.1, the UB estimator of Y among all unbiased linear estimators is

$$P^* = \begin{bmatrix} Y_1 \\ y_2(\beta^*) \end{bmatrix}$$

and it is unique. If P^* belonged to K, we would require $X_1 B^* = I(p)$ as well as $B^* X_1 = I(r)$, where $I(p)$ and $I(r)$ are $(p \times p)$ and $(n \times n)$ identity matrices, respectively. However,

$$\begin{aligned} r = \text{trace}(I(r)) &= \text{trace}(B^* X_1) \\ &= \text{trace}(X_1 B^*), \\ &= \text{trace}(I(p)) = p. \end{aligned}$$

since $\text{trace}(AB) = \text{trace}(BA)$ for any matrices A and B,

This is a contradiction unless $r = p$, in which case X_1 and B are square.

The solution to Schmidt's optimization for a "target" Y is the vector

$$\begin{bmatrix} X_1\beta^* \\ X_2\beta^* \end{bmatrix}$$

which in general is quite different to the UB estimator

$$P^* = \begin{bmatrix} Y_1 \\ y_2(\beta^*) \end{bmatrix}.$$

Nevertheless, Schmidt's analysis does produce the UB estimator for Y_2. To get the "best" estimator for Y_2, Schmidt minimizes $E[(Y_2 - y_2(\beta))^T D^T D(Y_2 - y_2(\beta))]$ for a particular matrix D related to S^{-1}, over possible β belonging to the set $L^* = \{\beta : \beta = BY_1$ and $BX_1 = I(r)\}$. In this case, L^* contains β^*, and the corresponding estimator $y_2(\beta^*)$ belongs to J and is UB. Because of this, we know that $y_2(\beta^*)$ will be a solution to Schmidt's optimization for any matrix D.

The "best" estimator for Y_1 derived by Schmidt's analysis is $X_1\beta^*$, which compares to the UB estimator Y_1. Using the arguments of Theorem 6.1, it can be shown that L^* does not contain a β such that $Y_1 = X_1\beta$ unless X_1 is square.

If the above restrictions on the admissible estimators in Schmidt's optimization are removed, we know from Lemma 5.1 that the resulting solution(s) will be UB. In these circumstances, Schmidt's optimization problem may then be generalized by replacing S^{-1} in the objective function with any non-negative definite matrix of appropriate dimension.

7. SUMMARY

We have proposed a general definition of "best" that we have termed Uniform Best (UB) and that is consistent with the Gauss–Markov Theorem. We have also provided a number of equivalent forms of the UB definition. We have then shown that for a "target" T linear in Y_1 and Y_2 there is always a unique UB linear unbiased estimator of the form QY_1. We have also shown that a generalization of the optimization problem proposed by Schmidt provides yet another characterization of UB. Finally, we have shown that the admissibility conditions imposed by Schmidt on the set of estimators in his optimization problem generally prevent the solution to his problem from being UB, although his "best" and the UB linear unbiased estimators for Y_2 are the same.

REFERENCES

[1] Halliwell, Leigh J., "Loss Prediction By Generalized Least Squares," *PCAS* LXXXIII, 1996, pp. 436–489.

[2] Halliwell, Leigh J., "Conjoint Prediction Of Paid And Incurred Losses," Casualty Actuarial Society *Forum* 1, Summer 1997, pp. 241–380.

[3] Schmidt, Klaus D., "Prediction In The Linear Model: A Direct Approach," *Metrika* 48, 1998, pp. 141–147.

APPENDIX

PROOF OF THEOREM 4.1

Consider two linear unbiased estimators P and P^* for T. Then

$$\mathrm{E}[P] = \mathrm{E}[QY_1] = QX_1\beta_0 = \mathrm{E}[T] = \mathrm{E}[P^*] = Q^*X_1\beta_0.$$

Since this must hold for any β_0, we have $(Q^* - Q)X_1 = 0$.

Then, for any α,

$$\begin{aligned}\mathrm{var}(\alpha^T(T-P)) &= \mathrm{var}(\alpha^T(T-P^*) + \alpha^T(P^*-P))\\ &= \mathrm{var}(\alpha^T(T-P^*)) + \mathrm{var}(\alpha^T(P^*-P))\\ &\quad + 2\,\mathrm{cov}(\alpha^T(T-P^*), \alpha^T(P^*-P)).\end{aligned}$$

Now

$$\begin{aligned}&\mathrm{cov}(\alpha^T(T-P^*), \alpha^T(P^*-P))\\ &\quad = \mathrm{E}[\alpha^T(T-P^*)\alpha^T(P^*-P)]\\ &\quad = \mathrm{E}[\alpha^T(T-P^*)(P^*-P)^T\alpha]\\ &\quad = \mathrm{E}[\alpha^T((F_1-Q^*)Y_1 + F_2Y_2)Y_1^T(Q^*-Q)^T\alpha]\\ &\quad = \alpha^T\{(F_1-Q^*)\mathrm{E}[Y_1Y_1^T] + F_2\mathrm{E}[Y_2Y_1^T]\}(Q^*-Q)^T\alpha\\ &\quad = \alpha^T\{(F_1-Q^*)S_{11} + F_2S_{21}\}(Q^*-Q)^T\alpha.\end{aligned}$$

Suppose $(F_1 - Q^*)S_{11} + F_2S_{21}$ is of the form GX_1^T, so that

$$Q^* = F_1 + F_2S_{21}S_{11}^{-1} - GX_1^TS_{11}^{-1}.$$

Then

$$\begin{aligned}\mathrm{cov}(\alpha^T(T-P^*), \alpha^T(P^*-P)) &= \alpha^TGX_1^T(Q^*-Q)^T\alpha\\ &= 0, \qquad \text{since} \quad (Q^*-Q)X_1 = 0.\end{aligned}$$

So, for any admissible P,

$$\begin{aligned}\mathrm{var}(\alpha^T(T-P)) &= \mathrm{var}(\alpha^T(T-P^*)) + \mathrm{var}(\alpha^T(P^*-P))\\ &\geq \mathrm{var}(\alpha^T(T-P^*)).\end{aligned}$$

Since $P^* = Q^*Y_1$ minimizes $\mathrm{var}(\alpha^T(T-P))$, by Lemma 2.1, it is UB.

We now solve for the form of G. The unbiased property of estimators $P = QY_1$ for T requires that

$$\mathrm{E}[T] = F_1X_1\beta_0 + F_2X_2\beta_0 + A\beta_0 = \mathrm{E}[P] = QX_1\beta_0$$

and, since this holds for any β_0, we have

$$F_1X_1 + F_2X_2 + A = QX_1.$$

Then we have

$$Q^* = F_1X_1 + F_2S_{21}S_{11}^{-1}X_1 - GX_1^TS_{11}^{-1}X_1 = F_1X_1 + F_2X_2 + A,$$

and so

$$G = \{F_2(S_{21}S_{11}^{-1}X_1 - X_2) - A\}(X_1^TS_{11}^{-1}X_1)^{-1}.$$

Substituting this back into the expression for Q^* gives

$$Q^* = F_1 + F_2S_{21}S_{11}^{-1} - \{F_2(S_{21}S_{11}^{-1}X_1 - X_2) - A\}B^*,$$

where

$$B^* = (X_1^TS_{11}^{-1}X_1)^{-1}X_1^TS_{11}^{-1}.$$

Rearranging, we get

$$Q^* = F_1 + F_2\{X_2B^* + S_{21}S_{11}^{-1}(I - X_1B^*)\} + AB^*.$$

Finally, multiplying through by Y_1 gives

$$P^* = F_1Y_1 + F_2y_2(\beta^*) + A\beta^*,$$

where

$$y_2(\beta^*) = X_2\beta^* + S_{21}S_{11}^{-1}(Y_1 - X_1\beta^*) \qquad \text{and}$$
$$\beta^* = B^*Y_1 = (X_1^TS_{11}^{-1}X_1)^{-1}X_1^TS_{11}^{-1}Y_1.$$

So far we have shown the existence of a "best" estimator. Consider another admissible estimator $P^{**} = Q^{**}Y_1$. Because P^* minimizes $\mathrm{var}(\alpha^T(T-P))$, we have from above that

$$\mathrm{var}(\alpha^T(T-P^{**})) = \mathrm{var}(\alpha^T(T-P^*)) + \mathrm{var}(\alpha^T(P^*-P^{**})).$$

If P^{**} also minimizes $\mathrm{var}(\alpha^T(T-P))$, then

$$\mathrm{var}(\alpha^T(T-P^{**})) = \mathrm{var}(\alpha^T(T-P^*)),$$

and so

$$\mathrm{var}(\alpha^T(P^*-P^{**})) = 0 \qquad \text{for any } \alpha.$$

Substituting $(P^*-P^{**}) = (Q^*-Q^{**})Y_1$ into this equation gives

$$\begin{aligned}\mathrm{var}(\alpha^T(P^*-P^{**})) &= \mathrm{var}(\alpha^T(Q^*-Q^{**})Y_1)\\ &= \alpha^T(Q^*-Q^{**})S_{11}(Q^*-Q^{**})^T\alpha = 0.\end{aligned}$$

S_{11}, the variance-covariance matrix of Y_1, is positive definite, so this implies

$$\alpha^T(Q^*-Q^{**}) = 0 \qquad \text{for any } \alpha.$$

Since Q^* and Q^{**} are independent of α, we must have $Q^* = Q^{**}$, and so the "best" estimator $P^* = Q^*Y_1$ is also unique.