Chapter 10
General Gauss–Markov Models

A general Gauss–Markov model is a model

$$Y = X\beta + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma^2 V,$$

where V is a known matrix. Linear models can be divided into four categories depending on the assumptions made about V:

(a) V is an identity matrix,
(b) V is nonsingular,
(c) V is possibly singular but $C(X) \subset C(V)$,
(d) V is possibly singular.

The categories are increasingly general. Any results for category (d) apply to all other categories. Any results for category (c) apply to categories (a) and (b). Any results for (b) apply to (a).

The majority of Chapters 1 through 9 have dealt with category (a). In Sections 2.7 and 3.8, models in category (b) were discussed. In this chapter, categories (c) and (d) are discussed. Section 1 is devoted to finding BLUEs for models in categories (c) and (d). Theorem 10.1.2 and the discussion following it give the main results for category (c). The approach is similar in spirit to Section 2.7. The model is transformed into an equivalent model that fits into category (a), and BLUEs are found for the equivalent model. Although similar in spirit to Section 2.7, the details are considerably more complicated because of the possibility that V is singular. Having found BLUEs for category (c), the extension to category (d) is very simple. The extension follows from Theorem 10.1.3. Finally, Section 1 contains some results on the uniqueness of BLUEs for category (d).

Section 2 contains a discussion of the geometry of estimation for category (d). In particular, it points out the need for a consistency criterion and the crucial role of projection operators in linear unbiased estimation. Section 3 examines the problem of testing a model against a reduced model for category (d). Section 4 discusses the extension of least squares estimation to category (d) in light of the consistency

R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer Texts in Statistics, DOI 10.1007/978-1-4419-9816-3_10,

requirement of Section 2. Section 4 also contains the very important result that least squares estimates are BLUEs if and only if $C(VX) \subset C(X)$.

Section 5 considers estimable parameters that can be known with certainty when $C(X) \not\subset C(V)$ and a relatively simple way to estimate estimable parameters that are not known with certainty. Some of the nastier parts in Sections 1 through 4 are those that provide sufficient generality to allow $C(X) \not\subset C(V)$. The simpler approach of Section 5 seems to obviate the need for much of that. Groß (2004) surveyed important results in linear models with possibly singular covariance matrices.

10.1 BLUEs with an Arbitrary Covariance Matrix

Consider the model

$$Y = X\beta + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma^2 V, \tag{1}$$

where V is a known matrix. We want to find the best linear unbiased estimate of $\mathrm{E}(Y)$.

Definition 10.1.1. Let Λ' be an $r \times p$ matrix with $\Lambda' = P'X$ for some P. An estimate B_0Y is called a *best linear unbiased estimate* (*BLUE*) of $\Lambda'\beta$ if

(a) $\mathrm{E}(B_0Y) = \Lambda'\beta$ for any β, and
(b) if BY is any other linear unbiased estimate of $\Lambda'\beta$, then for any $r \times 1$ vector ξ

$$\mathrm{Var}(\xi' B_0 Y) \le \mathrm{Var}(\xi' BY).$$

Exercise 10.1 Show that A_0Y is a BLUE of $X\beta$ if and only if, for every estimable function $\lambda'\beta$ such that $\rho'X = \lambda'$, $\rho'A_0Y$ is a BLUE of $\lambda'\beta$.

Exercise 10.2 Show that if $\Lambda' = P'X$ and if A_0Y is a BLUE of $X\beta$, then $P'A_0Y$ is a BLUE of $\Lambda'\beta$.

In the case of a general covariance matrix V, it is a good idea to reconsider what the linear model (1) is really saying. The obvious thing it is saying is that $\mathrm{E}(Y) \in C(X)$. From Lemma 1.3.5, the model also says that $e \in C(V)$ or, in other notation, $Y \in C(X,V)$. If V is a nonsingular matrix, $C(V) = \mathbf{R}^n$; so the conditions on e and Y are really meaningless. When V is a singular matrix, the conditions on e and Y are extremely important.

For any matrix A, let M_A denote the perpendicular projection matrix onto $C(A)$. M without a subscript is M_X. Any property that holds with probability 1 will be said to hold almost surely (a.s.). For example, Lemma 1.3.5 indicates that $e \in C(V)$ a.s. and, adding $X\beta$ to $Y - X\beta$, the lemma gives $Y \in C(X,V)$ a.s.

If R and S are any two random vectors with $R = S$ a.s., then $\mathrm{E}(R) = \mathrm{E}(S)$ and $\mathrm{Cov}(R) = \mathrm{Cov}(S)$. In particular, if $R = S$ a.s. and R is a BLUE, then S is also a BLUE.

Results on estimation will be established by comparing the estimation problem in model (1) to the estimation problem in two other, more easily analyzed, models.

Before proceeding to the first theorem on estimation, recall the eigenvector decomposition of the symmetric, nonnegative definite matrix V. One can pick matrices E and D so that $VE = ED$. Here $D = \mathrm{Diag}(d_i)$, where the d_is are the, say m, positive eigenvalues of V (with the correct multiplicities). E is a matrix of orthonormal columns with the ith column an eigenvector corresponding to d_i. Define $D^{1/2} \equiv \mathrm{Diag}\left(\sqrt{d_i}\right)$. Write

$$Q \equiv ED^{1/2} \quad \text{and} \quad Q^- \equiv D^{-1/2}E'.$$

Useful facts are

1. $C(V) = C(E) = C(Q)$
2. $M_V = EE' = QQ^-$
3. $I_m = Q^-Q$
4. $V = QQ'$
5. $QQ^-Q = Q$
6. $Q^-VQ^{-\prime} = I_m$
7. $Q^{-\prime}Q^- = V^-$.

Consider the linear model

$$Q^-Y = Q^-X\beta + Q^-e, \quad \mathrm{E}(Q^-e) = 0, \quad \mathrm{Cov}(Q^-e) = \sigma^2 I_m. \tag{2}$$

Models (1) and (2) are equivalent when $C(X) \subset C(V)$. Clearly, (2) can be obtained from (1) by multiplying on the left by Q^-. Moreover, with $C(X) \subset C(V)$, each of Y, $C(X)$, and e are contained in $C(V)$ a.s.; so multiplying (2) on the left by Q gives $M_VY = M_VX\beta + M_Ve$, which is model (1) a.s. Note that $Q^-Y \in \mathbf{R}^m$ and that the Gauss–Markov theorem can be used to get estimates in model (2). Moreover, if $C(X) \subset C(V)$, then $X = M_VX = QQ^-X$; so $X\beta$ is estimable in model (2).

Theorem 10.1.2. If $C(X) \subset C(V)$, then A_0Y is a BLUE of $X\beta$ in model (1) if and only if $(A_0Q)Q^-Y$ is a BLUE of $X\beta$ in model (2).

PROOF. If this theorem is to make any sense at all, we need to first show that $\mathrm{E}(A_0Y) = X\beta$ iff $\mathrm{E}(A_0QQ^-Y) = X\beta$. Recall that $Y \in C(X,V)$ a.s., so in this special case where $C(X) \subset C(V)$, we have $Y \in C(V)$ a.s. Thus, for the purposes of finding expected values and covariances, we can assume that $Y = M_VY$. Let B be an arbitrary $n \times n$ matrix. Then $\mathrm{E}(BY) = \mathrm{E}(BM_VY) = \mathrm{E}(BQQ^-Y)$, so $\mathrm{E}(A_0Y) = X\beta$ iff $\mathrm{E}(A_0QQ^-Y) = X\beta$. It is also handy to know the following fact:

$$\mathrm{Var}(\rho'BY) = \mathrm{Var}(\rho'BM_VY) = \mathrm{Var}(\rho'BQQ^-Y).$$

Now suppose that A_0Y is a BLUE for $X\beta$ in model (1). We show that A_0QQ^-Y is a BLUE of $X\beta$ in model (2). Let BQ^-Y be another unbiased estimate of $X\beta$ in model (2). Then $X\beta = \mathrm{E}(BQ^-Y)$, so BQ^-Y is an unbiased estimate of $X\beta$ in model (1) and, since A_0Y is a BLUE in model (1), $\mathrm{Var}(\rho' A_0QQ^-Y) = \mathrm{Var}(\rho' A_0Y) \le \mathrm{Var}(\rho' BQ^-Y)$. Thus A_0QQ^-Y is a BLUE of $X\beta$ in model (2).

Conversely, suppose that A_0QQ^-Y is a BLUE of $X\beta$ in model (2). Let BY be an unbiased estimate for $X\beta$ in model (1). Then BQQ^-Y is unbiased for $X\beta$ in model (2) and $\mathrm{Var}(\rho' A_0Y) = \mathrm{Var}(\rho' A_0QQ^-Y) \le \mathrm{Var}(\rho' BQQ^-Y) = \mathrm{Var}(\rho' BY)$; so $\rho' A_0Y$ is a BLUE of $X\beta$ in model (1). □

Note that with $C(X) \subset C(V)$, $A_0Y = A_0M_VY = A_0QQ^-Y$ a.s., so Theorem 10.1.2 is really saying that A_0Y is a BLUE in model (1) if and only if A_0Y is a BLUE in model (2). The virtue of Theorem 10.1.2 is that we can actually find a BLUE for $X\beta$ in model (2). From Exercises 10.1 and 10.2, a BLUE of $X\beta = QQ^-X\beta$ from model (2) is

$$\begin{aligned} X\hat{\beta} &= QM_{Q^-X}Q^-Y \\ &= Q\left[Q^-X(X'Q^{-\prime}Q^-X)^-X'Q^{-\prime}\right]Q^-Y \\ &= M_VX(X'V^-X)^-X'V^-Y \\ &= X(X'V^-X)^-X'V^-Y. \end{aligned} \tag{3}$$

It is useful to observe that we can get a BLUE from any choice of V^- and $(X'V^-X)^-$. First, notice that $X'V^-X$ does not depend on V^-. Since $C(X) \subset C(V)$, we can write $X = VC$ for some matrix C. Then $X'V^-X = C'VV^-VC = C'VC$. Second, $Q^-X(X'Q^{-\prime}Q^-X)^-X'Q^{-\prime}$ does not depend on the choice of $(X'Q^{-\prime}Q^-X)^-$. Therefore, $X(X'V^-X)^-X'V^-$ does not depend on the choice of $(X'V^-X)^-$. Moreover, for any $Y \in C(V)$, $X(X'V^-X)^-X'V^-$ does not depend on the choice of V^-. To see this, write $Y = Vb$. Then $X'V^-Y = (C'V)V^-(Vb) = C'Vb$. Since $Y \in C(V)$ a.s., $X(X'V^-X)^-X'V^-Y$ is a BLUE of $X\beta$ for any choices of V^- and $(X'V^-X)^-$.

To obtain the general estimation result for arbitrary singular V, consider the linear model

$$Y_1 = X\beta + e_1, \quad \mathrm{E}(e_1) = 0, \quad \mathrm{Cov}(e_1) = \sigma^2\left(V + XUX'\right), \tag{4}$$

where U is any symmetric nonnegative definite matrix.

Theorem 10.1.3. A_0Y is a BLUE for $X\beta$ in model (1) if and only if A_0Y_1 is a BLUE for $X\beta$ in model (4).

PROOF. Clearly, for any matrix B, BY is unbiased if and only if BY_1 is unbiased and both are equivalent to the condition $BX = X$. In the remainder of the proof, B will be an arbitrary matrix with $BX = X$ so that BY and BY_1 are arbitrary linear unbiased estimates of $X\beta$ in models (1) and (4), respectively. The key fact in the proof is that

$$\mathrm{Var}(\rho' BY_1) = \sigma^2 \rho' B\left(V + XUX'\right)B'\rho$$

$$
\begin{aligned}
&= \sigma^2 \rho' BVB' \rho + \rho' BXUX'B' \rho \\
&= \mathrm{Var}(\rho' BY) + \sigma^2 \rho' XUX' \rho .
\end{aligned}
$$

Now suppose that $A_0 Y$ is a BLUE for $X\beta$ in model (1). Then

$$\mathrm{Var}(\rho' A_0 Y) \le \mathrm{Var}(\rho' BY).$$

Adding $\sigma^2 \rho' XUX' \rho$ to both sides we get

$$\mathrm{Var}(\rho' A_0 Y_1) \le \mathrm{Var}(\rho' BY_1),$$

and since BY_1 is an arbitrary linear unbiased estimate, $A_0 Y_1$ is a BLUE.

Conversely, suppose $A_0 Y_1$ is a BLUE for $X\beta$ in model (4). Then

$$\mathrm{Var}(\rho' A_0 Y_1) \le \mathrm{Var}(\rho' BY_1),$$

or

$$\mathrm{Var}(\rho' A_0 Y) + \sigma^2 \rho' XUX' \rho \le \mathrm{Var}(\rho' BY) + \sigma^2 \rho' XUX' \rho .$$

Subtracting $\sigma^2 \rho' XUX' \rho$ from both sides we get

$$\mathrm{Var}(\rho' A_0 Y) \le \mathrm{Var}(\rho' BY),$$

so $A_0 Y$ is a BLUE. □

As with Theorem 10.1.2, this result is useful because a BLUE for $X\beta$ can actually be found in one of the models. Let $T = V + XUX'$. If U is chosen so that $C(X) \subset C(T)$, then Theorem 10.1.2 applies to model (4) and a BLUE for $X\beta$ is $X(X'T^- X)^- X'T^- Y$. Exercise 10.3 establishes that such matrices U exist. Since this is an application of Theorem 10.1.2, $X(X'T^- X)^- X'T^- Y$ does not depend on the choice of $(X'T^- X)^-$ and, for $Y \in C(T)$, it does not depend on the choice of T^-. Proposition 10.1.4 below shows that $C(X) \subset C(T)$ implies $C(T) = C(X,V)$; so $Y \in C(T)$ a.s., and any choice of T^- gives a BLUE.

Exercise 10.3 The BLUE of $X\beta$ can be obtained by taking $T = V + XX'$. Prove this by showing that

(a) $C(X) \subset C(T)$ if and only if $TT^- X = X$, and
(b) if $T = V + XX'$, then $TT^- X = X$.

Hint: Searle and Pukelsheim (1987) base a proof of (b) on

$$
\begin{aligned}
0 &= (I - TT^-) T (I - TT^-)' \\
&= (I - TT^-) V (I - TT^-)' + (I - TT^-) XX' (I - TT^-)'
\end{aligned}
$$

and the fact that the last term is the sum of two nonnegative definite matrices.

Proposition 10.1.4. If $C(X) \subset C(T)$, then $C(V) \subset C(T)$ and $C(X,V) = C(T)$.

PROOF. We know that $C(X) \subset C(T)$, so $X = TG$ for some matrix G. If $v \in C(V)$, then $v = Vb_1$, and $v = Vb_1 + XUX'b_1 - XUX'b_1 = Tb_1 - TG(UX'b_1)$; so $v \in C(T)$. Thus, $C(X,V) \subset C(T)$. But clearly, by the definition of T, $C(T) \subset C(X,V)$; so we have $C(X,V) = C(T)$. □

To complete our characterization of best linear unbiased estimates, we will show that for practical purposes BLUEs of $X\beta$ are unique.

Theorem 10.1.5. Let AY and BY be BLUEs of $X\beta$ in model (1), then $\Pr(AY = BY) = 1$.

PROOF. It is enough to show that $\Pr[(A-B)Y = 0] = 1$. We need only observe that $\mathrm{E}[(A-B)Y] = 0$ and show that for any ρ, $\mathrm{Var}[\rho'(A-B)Y] = 0$.

Remembering that $\mathrm{Var}(\rho'AY) = \mathrm{Var}(\rho'BY)$, we first consider the unbiased estimate of $X\beta$, $\frac{1}{2}(A+B)Y$:

$$\mathrm{Var}(\rho'AY) \leq \mathrm{Var}\left(\rho'\frac{1}{2}(A+B)Y\right)$$

but

$$\mathrm{Var}\left(\rho'\frac{1}{2}(A+B)Y\right) = \frac{1}{4}\left[\mathrm{Var}(\rho'AY) + \mathrm{Var}(\rho'BY) + 2\mathrm{Cov}(\rho'AY, \rho'BY)\right],$$

so

$$\mathrm{Var}(\rho'AY) \leq \frac{1}{2}\mathrm{Var}(\rho'AY) + \frac{1}{2}\mathrm{Cov}(\rho'AY, \rho'BY).$$

Simplifying, we find that

$$\mathrm{Var}(\rho'AY) \leq \mathrm{Cov}(\rho'AY, \rho'BY).$$

Now look at $\mathrm{Var}(\rho'(A-B)Y)$,

$$0 \leq \mathrm{Var}\big(\rho'(A-B)Y\big) = \mathrm{Var}(\rho'AY) + \mathrm{Var}(\rho'BY) - 2\mathrm{Cov}(\rho'AY, \rho'BY) \leq 0. \quad \square$$

Finally, the most exact characterization of a BLUE is the following:

Theorem 10.1.6. If AY and BY are BLUEs for $X\beta$ in model (1), then $AY = BY$ for any $Y \in C(X,V)$.

PROOF. It is enough to show that $AY = BY$, first when $Y \in C(X)$ and then when $Y \in C(V)$. When $Y \in C(X)$, by unbiasedness $AY = BY$.

We want to show that $AY = BY$ for all $Y \in C(V)$. Let $\mathscr{M} = \{Y \in C(V) | AY = BY\}$. It is easily seen that $\mathscr{M}$ is a vector space, and clearly $\mathscr{M} \subset C(V)$. If $C(V) \subset \mathscr{M}$, then $\mathscr{M} = C(V)$, and we are done.

From unbiasedness, $AY = X\beta + Ae$ and $BY = X\beta + Be$, so $AY = BY$ if and only if $Ae = Be$. From Theorem 10.1.5, $\Pr(Ae = Be) = 1$. We also know that $\Pr(e \in C(V)) = 1$. Therefore, $\Pr(e \in \mathscr{M}) = 1$.

Computing the covariance of e we find that

$$\sigma^2 V = \int ee' dP = \int_{e \in \mathscr{M}} ee' dP;$$

so, by Exercise 10.4, $C(V) \subset \mathscr{M}$. □

Exercise 10.4 Ferguson (1967, Section 2.7, page 74) proves the following:

Lemma (3) If S is a convex subset of $\mathbf{R}^n$, and Z is an n-dimensional random vector for which $\Pr(Z \in S) = 1$ and for which $\mathrm{E}(Z)$ exists and is finite, then $\mathrm{E}(Z) \in S$.

Use this lemma to show that

$$C(V) = C\left[\int_{e \in \mathscr{M}} ee' dP\right] \subset \mathscr{M}.$$

After establishing Theorem 10.1.5, we only need to find any one BLUE, because for any observations that have a chance of happening, all BLUEs give the same estimates. Fortunately, we have already shown how to obtain a BLUE, so this section is finished. There is only one problem. Best linear unbiased estimates might be pure garbage. We pursue this issue in the next section.

10.2 Geometric Aspects of Estimation

The linear model

$$Y = X\beta + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma^2 V, \tag{1}$$

says two important things:

(a) $\mathrm{E}(Y) = X\beta \in C(X)$, and
(b) $\Pr(e \in C(V)) = 1$.

Note that (b) also says something about $\mathrm{E}(Y)$:

(b′) $X\beta = Y - e \in Y + C(V)$ a.s.

Intuitively, any reasonable estimate of $\mathrm{E}(Y)$, say $X\hat{\beta}$, should satisfy the following definition for consistency.

Definition 10.2.1. An estimate $\widehat{X\beta}$ of $X\beta$ is called a *consistent* estimate if

(i) $\widehat{X\beta} \in C(X)$ for any Y, and
(ii) $\widehat{X\beta} \in Y + C(V)$ for any $Y \in C(X,V)$.

$\widehat{X\beta}$ is called *almost surely consistent* if conditions (i) and (ii) hold almost surely.

Note that this concept of consistency is distinct from the usual large sample idea of consistency. The idea is that a consistent estimate, in our sense, is consistent with respect to conditions (a) and (b′).

EXAMPLE 10.2.2. Consider the linear model determined by

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad V = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

If this model is graphed with coordinates (x,y,z), then $C(X)$ is the x,y plane and $C(V)$ is the plane determined by the y-axis and the line $[x = z, y = 0]$. (See Figure 10.1.)

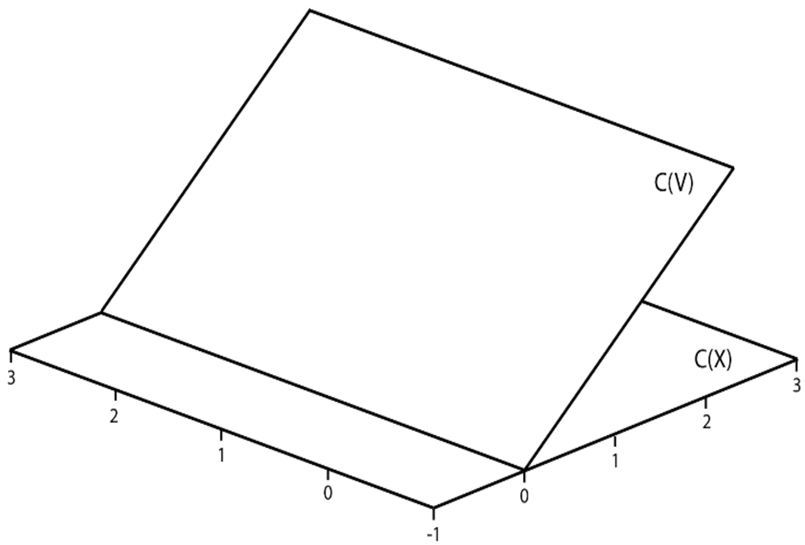

Fig. 10.1 Estimation space and singular covariance space for Example 10.2.2.

Suppose that $Y = (7,6,2)'$. Then (see Figure 10.2) $\mathrm{E}(Y)$ is in $C(X)$ (the x,y plane) and also in $Y + C(V)$ (which is the plane $C(V)$ in Figure 10.1, translated over until it contains Y). The intersection of $C(X)$ and $Y + C(V)$ is the line $[x = 5, z = 0]$, so any consistent estimate of $X\beta$ will be in the line $[x = 5, z = 0]$. To see this, note that $C(X)$ consists of vectors with the form $(a,b,0)'$, and $Y + C(V)$ consists of vectors like $(7,6,2)' + (c,d,c)'$. The intersection is those vectors with $c = -2$, so they are of the form $(5,6+d,0)'$ or $(5,b,0)'$.

The problem with BLUEs of $X\beta$ is that there is no apparent reason why a BLUE should be consistent. The class of linear unbiased estimates (LUEs) is very broad. It consists of all estimates AY with $AX = X$. There are many linear unbiased estimates that are not consistent. For example, Y itself is a LUE and it satisfies condition (ii) of consistency; however, one would certainly not want to use it as an estimate.

EXAMPLE 10.2.2 CONTINUED. MY is a LUE and satisfies condition (i) of consistency; however, with $Y = (7,6,2)'$, $MY = (7,6,0)'$, but $(7,6,0)'$ is not in $C(X) \cap [Y + C(V)] = [x = 5, z = 0]$. (See Figure 10.2.)

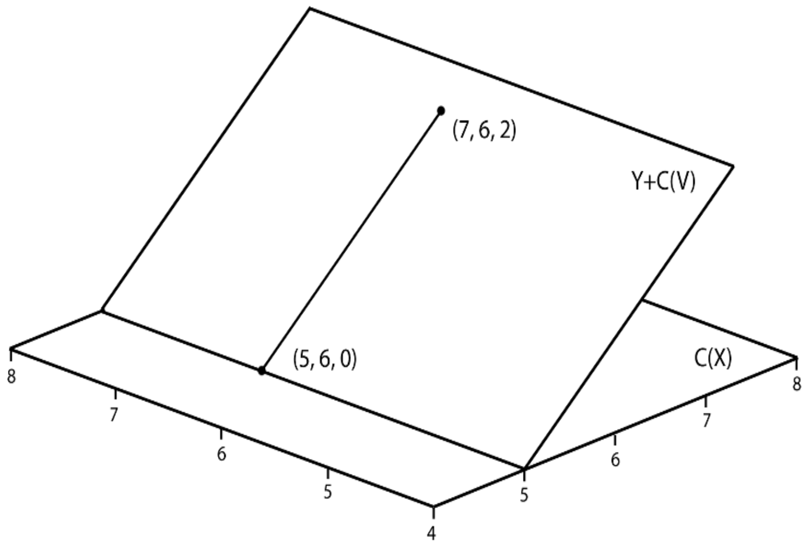

Fig. 10.2 Consistent estimation for Example 10.2.2.

Before showing that BLUEs of $X\beta$ are almost surely consistent, several *observations* will be made. The object is to explicate the case $C(X) \subset C(V)$, give alternative conditions that can be used in place of conditions (i) and (ii) of consistency, and display the importance of projections in linear estimation.

(a) In practice, when a covariance matrix other than $\sigma^2 I$ is appropriate, most of the time the condition $C(X) \subset C(V)$ will be satisfied. When $C(X) \subset C(V)$, then $Y \in C(X,V) = C(V)$ a.s., so the condition $\widehat{X\beta} \in Y + C(V)$ a.s. merely indicates that $\widehat{X\beta} \in C(V)$ a.s. In fact, any estimate of $X\beta$ satisfying condition (i) of consistency also satisfies condition (ii). (Note: The last claim is not restricted to linear estimates or even unbiased estimates.)

(b) An estimate AY satisfies condition (i) of consistency if and only if $A = XB$ for some matrix B.

(c) An estimate AY satisfies condition (ii) of consistency if and only if $(I-A)Y \in C(V)$ for any $Y \in C(X,V)$.

(d) If AY is a LUE of $X\beta$, then AY satisfies condition (ii) of consistency if and only if $AY \in C(V)$ for any $Y \in C(V)$, i.e., iff $C(AV) \subset C(V)$.

PROOF. $Y \in C(X,V)$ iff $Y = x + v$ for $x \in C(X)$ and $v \in C(V)$. $(I-A)Y = (x-x) + (v-Av)$. $v - Av \in C(V)$ iff $Av \in C(V)$. $\square$

(e) AY is a LUE satisfying condition (i) if and only if A is a projection matrix (not necessarily perpendicular) onto $C(X)$.

PROOF. From unbiasedness, $AX = X$, and from (b), $A = XB$. Hence $AA = AXB = XB = A$, and so A is idempotent and a projection onto $C(A)$. Since $AX = X$, we have $C(X) \subset C(A)$; and since $A = XB$, we have $C(A) \subset C(X)$. Therefore, $C(A) = C(X)$. $\square$

Exercise 10.5 Prove observations (a), (b), and (c).

Notice that in general all consistent linear unbiased estimates (CLUEs) are projections onto $C(X)$ and that if $C(X) \subset C(V)$, all projections onto $C(X)$ are CLUEs. This goes far to show the importance of projection matrices in estimation. In particular, the BLUEs that we actually found in the previous section satisfy condition (i) by observation (b). Thus, by observation (e) they are projections onto $C(X)$.

Before proving the main result, we need the following proposition:

Proposition 10.2.3. For $T = V + XUX'$ with U nonnegative definite and $A = X(X'T^-X)^-X'T^-$ with $C(X) \subset C(T)$, $AV = VA'$.

PROOF. Consider the symmetric matrix $X(X'T^-X)^-X'$, where the generalized inverses are taken as in Corollary B.41. By the choice of T, write $X = TG$.

$$X' = G'T = G'TT^-T = X'T^-T,$$
$$X(X'T^-X)^-X' = X(X'T^-X)^-X'T^-T = A(V + XUX').$$

Since A is a projection operator onto $C(X)$, $AXUX' = XUX'$; thus,

$$X(X'T^-X)^-X' - XUX' = AV.$$

The lefthand side is symmetric. $\square$

Proposition 10.2.4. There exists a BLUE of $X\beta$ in model (1) that is a CLUE.

PROOF. $AY = X(X'T^-X)^-X'T^-Y$ is a BLUE satisfying condition (i). In Proposition 10.2.3, we proved that $AV = VA'$. By observation (d), the proposition holds. □

Theorem 10.2.5. If AY is a BLUE of $X\beta$ in model (1), then AY is almost surely consistent.

PROOF. By Proposition 10.2.4, a consistent BLUE exists. By Theorem 10.1.5, any BLUE must almost surely satisfy consistency. □

10.3 Hypothesis Testing

We consider the problem of testing the model

$$Y = X\beta + e, \quad e \sim N\big(0, \sigma^2 V\big) \tag{1}$$

against the reduced model

$$Y = X_0\gamma + e, \quad e \sim N\big(0, \sigma^2 V\big), \tag{2}$$

where $C(X_0) \subset C(X)$. In particular, we use the approach of looking at reductions in the sum of squares for error. First we need to define what we mean by the sum of squares for error.

In the remainder of this section, let $A = X(X'T^-X)^-X'T^-$, where T^- and $(X'T^-X)^-$ are chosen (for convenience) as in Corollary B.41 and as usual $T = V + XUX'$ for some nonnegative definite matrix U such that $C(X) \subset C(T)$. AY is a BLUE of $X\beta$ in model (1). We will define the sum of squares for error (SSE) in model (1) as

$$SSE = Y'(I-A)'T^-(I-A)Y. \tag{3}$$

The idea of using a quadratic form in the residuals to estimate the error is reasonable, and any quadratic form in the residuals, when normalized, gives an unbiased estimate of σ^2. The terminology SSE literally comes from the use of the quadratic form $Y'(I-A)'(I-A)Y$. In particular, if $V = I$, then

$$Y'(I-A)'(I-A)Y/\sigma^2 = Y'(I-A)'[\sigma^2 I]^{-1}(I-A)Y$$

has a χ^2 distribution and is independent of AY. For an arbitrary covariance matrix, an analogous procedure would be to use $Y'(I-A)'V^-(I-A)Y$ to get an estimate of σ^2 and develop tests. To simplify computations, we have chosen to define SSE as in

(3). However, we will show that for $Y \in C(X,V)$, $SSE = Y'(I-A)'V^-(I-A)Y$ for any choice of V^-.

The sum of squares error in model (2) (SSE_0) is defined similarly. Let $T_0 = V + X_0U_0X_0'$ for a nonnegative definite matrix U_0 for which $C(X_0) \subset C(T_0)$. Let $A_0 = X_0(X_0'T_0^-X_0)^-X_0'T_0^-$, where again for convenience take T_0^- and $(X_0'T_0^-X_0)^-$ as in Corollary B.41.

$$SSE_0 = Y'(I-A_0)'T_0^-(I-A_0)Y.$$

The test will be, for some constant K, based on

$$K\left(SSE_0 - SSE\right)\big/ SSE,$$

which will be shown to have an F distribution. We will use Theorem 1.3.6 and Theorem 1.3.9 to obtain distribution results.

Theorem 10.3.1. If $Y \in C(T)$, then

(a) $Y'(I-A)'T^-(I-A)Y = Y'(I-A)'V^-(I-A)Y$ for any V^-,
(b) $Y'(I-A)'T^-(I-A)Y = 0$ if and only if $(I-A)Y = 0$.

PROOF. The proof is given after Proposition 10.3.6. □

These results also hold when A is replaced by A_0. Denote $C \equiv (I-A)'T^-(I-A)$ and $C_0 \equiv (I-A_0)'T_0^-(I-A_0)$. Thus $SSE = Y'CY$ and $SSE_0 = Y'C_0Y$. The distribution theory is:

Theorem 10.3.2.

(a) $Y'CY/\sigma^2 \sim \chi^2(\mathrm{tr}(CV),0)$.
If $X\beta \in C(X_0,V)$, then
(b) $Y'(C_0-C)Y/\sigma^2 \sim \chi^2(\mathrm{tr}(C_0V-CV), \beta'X'C_0X\beta)$
and
(c) $Y'CY$ and $Y'(C_0-C)Y$ are independent.

PROOF. The proof consists of checking the conditions in Theorems 1.3.6 and 1.3.9. There are many conditions to be checked. This is done in Lemmas 10.3.7 through 10.3.9 at the end of the section. □

The last result before stating the test establishes behavior of the distributions under the two models. Model (2) is true if and only if $X\beta \in C(X_0)$.

Theorem 10.3.3.

(a) If $X\beta \in C(X_0)$, then $\Pr(Y \in C(X_0,V)) = 1$ and $\beta'X'C_0X\beta = 0$.

(b) If $X\beta \notin C(X_0)$, then either $X\beta \in C(X_0,V)$ and $\beta'X'C_0X\beta > 0$ or $X\beta \notin C(X_0,V)$ and $\Pr(Y \notin C(X_0,V)) = 1$

PROOF.

(a) The first part is clear, the second is because $C_0X_0 = 0$.
(b) If $X\beta \notin C(X_0)$, then either $X\beta \in C(X_0,V)$ or $X\beta \notin C(X_0,V)$. If $X\beta \in C(X_0,V)$, then by Theorem 10.3.1b, $\beta'X'C_0X\beta = 0$ if and only if $(I-A_0)X\beta = 0$ or $X\beta = A_0X\beta$. Since $X\beta \notin C(X_0)$, $\beta'X'C_0X\beta > 0$. If $X\beta \notin C(X_0,V)$, suppose $e \in C(V)$ and $Y \in C(X_0,V)$, then $X\beta = Y - e \in C(X_0,V)$, a contradiction. Therefore either $e \notin C(V)$ or $Y \notin C(X_0,V)$. Since $\Pr(e \in C(V)) = 1$, we must have $\Pr(Y \notin C(X_0,V)) = 1$. □

The test at the α level is to reject H_0 that model (2) is adequate if $Y \notin C(X_0,V)$ or if

$$\frac{(SSE_0 - SSE)/\mathrm{tr}[(C_0 - C)V]}{SSE/\mathrm{tr}(CV)} > F(1-\alpha, \mathrm{tr}[(C_0 - C)V], \mathrm{tr}(CV)).$$

This is an α level test because, under H_0, $\Pr(Y \notin C(X_0,V)) = 0$. The power of the test is at least as great as that of a noncentral F test and is always greater than α because if $\beta'X'C_0X\beta = 0$ under the alternative, Theorem 10.3.3 ensures that the test will reject with probability 1.

In the next section we consider extensions of least squares and conditions under which such extended least squares estimates are best estimates.

Proofs of Theorems 10.3.1 and 10.3.2.

Before proceeding with the proofs of the theorems, we need some background results.

Proposition 10.3.4. $A'T^-A = A'T^- = T^-A.$

PROOF.

$$\begin{aligned} A'T^- &= T^-X(X'T^-X)^-X'T^- = T^-A, \\ A'T^-A &= T^-X(X'T^-X)^-X'T^-X(X'T^-X)^-X'T^-; \end{aligned}$$

but, as discussed earlier, A does not depend on the choice of $(X'T^-X)^-$ and $(X'T^-X)^-X'T^-X(X'T^-X)^-$ is a generalized inverse of $X'T^-X$, so $A'T^-A = T^-A$. □

Corollary 10.3.5. $(I-A)'T^-(I-A) = (I-A)'T^- = T^-(I-A).$

In the remainder of this discussion we will let $X = TG$ and $V = TBT$. (The latter comes from Proposition 10.1.4 and symmetry.)

Proposition 10.3.6.

(a) $VT^-(I-A) = TT^-(I-A)$,
(b) $VT^-(I-A)V = (I-A)V$.

PROOF. From Corollary 10.3.5 and unbiasedness, i.e., AX=X,

$$TT^-(I-A) = T(I-A)'T^- = V(I-A)'T^- = VT^-(I-A).$$

With $V = TBT$ and, from Proposition 10.2.3, AV symmetric, we have

$$\begin{aligned} VT^-(I-A)V &= TT^-(I-A)V = TT^-V(I-A)' \\ &= TT^-TBT(I-A)' = TBT(I-A)' = V(I-A)' = (I-A)V. \quad \square \end{aligned}$$

PROOF OF THEOREM 10.3.1. Recalling Proposition 10.1.4, if $Y \in C(T)$, write $Y = Xb_1 + Vb_2$.

(a) Using Proposition 10.2.3,

$$\begin{aligned} Y'(I-A)'T^-(I-A)Y &= (Xb_1+Vb_2)'(I-A)'T^-(I-A)(Xb_1+Vb_2) \\ &= b_2'V(I-A)'T^-(I-A)Vb_2 \\ &= b_2'(I-A)VT^-(I-A)Vb_2 \\ &= b_2'(I-A)(I-A)Vb_2 \\ &= b_2'(I-A)V(I-A)'b_2 \\ &= b_2'(I-A)VV^-V(I-A)'b_2 \\ &= b_2'V(I-A)'V^-(I-A)Vb_2 \\ &= Y'(I-A)'V^-(I-A)Y. \end{aligned}$$

(b) From the proof of (a),

$$Y'(I-A)'T^-(I-A)Y = b_2'(I-A)V(I-A)'b_2.$$

Recall that we can write $V = EDE'$ with $E'E = I$, $D = \text{Diag}(d_i)$, and $d_i > 0$ for all i.

$$\begin{aligned} Y'(I-A)'T^-(I-A)Y = 0 \quad &\text{iff} \quad b_2'(I-A)V(I-A)'b_2 = 0 \\ &\text{iff} \quad E'(I-A)'b_2 = 0 \\ &\text{iff} \quad (I-A)'b_2 \perp C(E) \\ &\text{iff} \quad (I-A)'b_2 \perp C(V) \end{aligned}$$

$$\begin{aligned} &\text{iff} \quad V(I-A)'b_2 = 0 \\ &\text{iff} \quad (I-A)Vb_2 = 0 \\ &\text{iff} \quad (I-A)Y = 0. \end{aligned}$$ □

The following lemmas constitute the proof of Theorem 10.3.2.

Lemma 10.3.7.

(a) $CVC = C$ and $C_0VC_0 = C_0$,
(b) $CVC_0V = CV$,
(c) $VCVCV = VCV$, $\beta'X'CVCX\beta = \beta'X'CX\beta = 0$, $VCVCX\beta = VCX\beta = 0$.

PROOF.

(a) Using Corollary 10.3.5 and Proposition 10.3.6, $CVC = CVT^-(I-A) = CTT^-(I-A) = (I-A)'T^-TT^-(I-A) = (I-A)'T^-(I-A) = C$.
(b) A similar argument gives the second equality: $CVC_0V = CVT_0^-(I-A_0)V = C(I-A_0)V = T^-(I-A)(I-A_0)V = T^-(I-A)V = CV$.
(c) The equalities in (c) follow from (a) and the fact that $CX = T^-(I-A)X = 0$.

□

Note that Lemma 10.3.7c leads directly to Theorem 10.3.2a. We now establish the conditions necessary for Theorem 10.3.2b.

Lemma 10.3.8.

(a) $V(C_0-C)V(C_0-C)V = V(C_0-C)V$,
(b) $\beta'X'(C_0-C)V(C_0-C)X\beta = \beta'X'(C_0-C)X\beta = \beta'X'C_0X\beta$,
(c) if $X\beta \in C(X_0,V)$, then $V(C_0-C)V(C_0-C)X\beta = V(C_0-C)X\beta$.

PROOF. From parts (a) and (b) of Lemma 10.3.7,

$$\begin{aligned} V(C_0-C)V(C_0-C)V &= VC_0VC_0V - VC_0VCV - VCVC_0V + VCVCV \\ &= VC_0V - VCV - VCV + VCV \\ &= VC_0V - VCV \\ &= V(C_0-C)V. \end{aligned}$$

To show (b), we need only show that

$$X'(C_0-C)V(C_0-C)X = X'(C_0-C)X.$$

Since $CX = 0$, this is equivalent to showing $X'C_0VC_0X = X'C_0X$. The result is immediate from part (a) of Lemma 10.3.7.

To show (c), we need to show that

$$V(C_0-C)V(C_0-C)X\beta = V(C_0-C)X\beta.$$

Since $CX=0$, it is enough to show that $VC_0VC_0X\beta - VCVC_0X\beta = VC_0X\beta$. With $VC_0VC_0X\beta = VC_0X\beta$, we only need $VCVC_0X\beta = 0$. Since $X\beta \in C(T_0)$, by assumption, $(I-A_0)X\beta = T_0\gamma$ for some γ,

$$\begin{aligned} VCVC_0X\beta &= VCVT_0^-(I-A_0)X\beta = VCT_0T_0^-(I-A_0)X\beta \\ &= VCT_0T_0^-T_0\gamma = VCT_0\gamma = VC(I-A_0)X\beta \\ &= VT^-(I-A)(I-A_0)X\beta = VT^-(I-A)X\beta = 0. \end{aligned}$$ □

To establish Theorem 10.3.2c, use the following lemma:

Lemma 10.3.9.

(a) $VCV(C_0-C)V = 0$,
(b) $VCV(C_0-C)X\beta = 0$ if $X\beta \in C(X_0,V)$,
(c) $V(C_0-C)VCX\beta = 0$,
(d) $\beta'X'(C_0-C)VCX\beta = 0$.

PROOF. For part (a), $VCV(C_0-C)V = VCVC_0V - VCVCV = VCV - VCV$. Parts (c) and (d) follow because $CX=0$; also, (b) becomes the condition $VCVC_0X\beta = 0$, as was shown in the proof of Lemma 10.3.8. □

10.4 Least Squares Consistent Estimation

Definition 10.4.1. An estimate $\tilde{\beta}$ of β is said to be *consistent* if $X\tilde{\beta}$ is a consistent estimate of $X\beta$. $\hat{\beta}$ is said to be a *least squares consistent estimate of* β if for any other consistent estimate $\tilde{\beta}$ and any $Y \in C(X,V)$,

$$(Y-X\hat{\beta})'(Y-X\hat{\beta}) \le (Y-X\tilde{\beta})'(Y-X\tilde{\beta}).$$

How does this differ from the usual definition of least squares? In the case where $C(X) \subset C(V)$, it hardly differs at all. Any estimate $\tilde{\beta}$ will have $X\tilde{\beta} \in C(X)$, and, as we observed earlier, when $C(X) \subset C(V)$, any estimate of $X\beta$ satisfying condition (i) of consistency (Definition 10.2.1) also satisfies condition (ii). The main difference between consistent least squares and least squares is that we are restricting ourselves to consistent estimates of $X\beta$. As we saw earlier, estimates that are not consistent are just not reasonable, so we are not losing anything. (Recall Example 10.2.2, in which the least squares estimate was not consistent.) The other difference between consistent least squares estimation and regular least squares is that the current definition restricts Y to $C(X,V)$. In the case where $C(X) \subset C(V)$, this restriction would not be necessary because a least squares estimate will have to be a least squares consistent estimate. In the general case, we need to actually use condition (ii) of consistency. Condition (ii) was based on the fact that $e \in C(V)$ a.s. Since e cannot be observed, we used the related fact that $Y \in C(X,V)$ a.s., and made condition (ii) apply only when $Y \in C(X,V)$.

Theorem 10.4.2a. If AY is a CLUE and $(I-A)r \perp C(X) \cap C(V)$ for any $r \in C(X,V)$, then any $\hat{\beta}$ satisfying $AY = X\hat{\beta}$ is a least squares consistent estimate.

PROOF. Let $\tilde{\beta}$ be any consistent estimate,

$$\begin{aligned}(Y-X\tilde{\beta})'(Y-X\tilde{\beta}) &= (Y-AY+AY-X\tilde{\beta})'(Y-AY+AY-X\tilde{\beta}) \\ &= Y'(I-A)'(I-A)Y+(AY-X\tilde{\beta})'(AY-X\tilde{\beta}) \\ &\quad +2Y'(I-A)'(AY-X\tilde{\beta}).\end{aligned}$$

It is enough to show that

$$Y'(I-A)'(AY-X\tilde{\beta}) = 0 \text{ for } Y \in C(X,V).$$

Note that $AY - X\tilde{\beta} \in C(X)$. Also observe that since AY and $X\tilde{\beta}$ are consistent, $AY - X\tilde{\beta} = (Y-X\tilde{\beta})-(I-A)Y \in C(V)$ for $Y \in C(X,V)$. Thus $AY - X\tilde{\beta} \in C(X) \cap C(V)$. For any $Y \in C(X,V)$, we have

$$(I-A)Y \perp C(X) \cap C(V);$$

so $Y'(I-A)'(AY-X\tilde{\beta}) = 0$. □

We now give a formula for a least squares consistent estimate. Choose V_0 with orthonormal columns such that $C(V_0) = C(X) \cap C(V)$. Also choose V_1 with orthonormal columns such that $C(V_1) \perp C(V_0)$ and $C(V_0,V_1) = C(V)$. It is easily seen that $MY - MV_1\hat{\gamma}$ is a CLUE, where $\hat{\gamma} = [V_1'(I-M)V_1]^{-1}V_1'(I-M)Y$. Observe that

$$C(X,V) = C(X,V_0,V_1) = C(X,(I-M)V_1).$$

To put the computations into a somewhat familiar form, consider the analysis of covariance model

$$Y = [X,(I-M)V_1]\begin{bmatrix}\beta \\ \gamma\end{bmatrix} + e, \quad \mathrm{E}(e) = 0. \tag{1}$$

We are interested in the least squares estimate of $\mathrm{E}(Y)$, so the error vector e is of no interest except that $\mathrm{E}(e) = 0$. The least squares estimate of $\mathrm{E}(Y)$ is $MY + (I-M)V_1\hat{\gamma}$, where $\hat{\gamma} = [V_1'(I-M)V_1]^{-1}V_1'(I-M)Y$. It turns out that $MY - MV_1\hat{\gamma}$ is a CLUE for $X\beta$ in model (10.1.1).

First, $MY - MV_1\hat{\gamma}$ is unbiased, because $\mathrm{E}(MY) = X\beta$ and $\mathrm{E}(\hat{\gamma}) = 0$. $\mathrm{E}(\hat{\gamma})$ is found by replacing Y with $X\beta$ in the formula for $\hat{\gamma}$, but the product of $(I-M)X\beta = 0$.

Second, it is clear that for any Y

$$MY - MV_1\hat{\gamma} \in C(X).$$

Finally, it is enough to show that if $Y \in C(V)$, then $MY - MV_1\hat{\gamma} \in C(V)$. If $Y \in C(V)$, then $Y \in C(X, (I-M)V_1) = C(X,V)$. Therefore, the least squares estimate of $\mathrm{E}(Y)$ in model (1) is Y itself. We have two characterizations of the least squares estimate, and equating them gives

$$MY + (I-M)V_1\hat{\gamma} = Y$$

or

$$MY - MV_1\hat{\gamma} = Y - V_1\hat{\gamma}.$$

Now, it is clear that $V_1\hat{\gamma} \in C(V)$; so if $Y \in C(V)$, then $MY - MV_1\hat{\gamma} \in C(V)$.

Proposition 10.4.3. If $\hat{\beta}$ is an estimate with $X\hat{\beta} = MY - MV_1\hat{\gamma}$, then $\hat{\beta}$ is a least squares consistent estimate.

PROOF. By Theorem 10.4.2a, it is enough to show that

$$(Y - MY + MV_1\hat{\gamma}) \perp C(V_0)$$

for any $Y \in C(X,V)$. Let $w \in C(V_0)$; then, since $Mw = w$,

$$\begin{aligned} w'(Y - MY + MV_1\hat{\gamma}) &= w'Y - w'MY + w'MV_1\hat{\gamma} \\ &= w'Y - w'Y + w'V_1\hat{\gamma} \\ &= w'V_1\hat{\gamma} \\ &= 0. \end{aligned}$$

□

Theorem 10.4.2b. If $\hat{\beta}$ is a least squares consistent estimate, then $(Y - X\hat{\beta}) \perp C(X) \cap C(V)$ for any $Y \in C(X,V)$; and if $\tilde{\beta}$ is any other least squares consistent estimate, $X\hat{\beta} = X\tilde{\beta}$ for any $Y \in C(X,V)$.

PROOF. Let A be the matrix determined by $AY = MY - MV_1\hat{\gamma}$ for all Y. We show that $AY = X\hat{\beta}$ for any $Y \in C(X,V)$ and are done.

We know, by Definition 10.4.1, that

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) = (Y - AY)'(Y - AY) \quad \text{for } Y \in C(X,V).$$

As in the proof of Theorem 10.4.2a, we also know that, for $Y \in C(X,V)$,

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) = (Y - AY)'(Y - AY) + (AY - X\hat{\beta})'(AY - X\hat{\beta}).$$

Therefore,

$$(AY - X\hat{\beta})'(AY - X\hat{\beta}) = 0;$$

hence

$$AY - X\hat{\beta} = 0$$

or

$$AY = X\hat{\beta} \quad \text{for } Y \in C(X,V). \qquad \square$$

Together, Theorems 10.4.2a and 10.4.2b give an "if and only if" condition for $\hat{\beta}$ to be a least squares CLUE.

In the future, any CLUE of $X\beta$, say AY, that satisfies $(I-A)r \perp C(X) \cap C(V)$ for any $r \in C(X,V)$ will be referred to as a least squares CLUE of $X\beta$. The most important result on least squares CLUEs is:

Theorem 10.4.4. If in model (10.1.1), $C(VV_0) \subset C(V_0)$, then a least squares CLUE of $X\beta$ is a best CLUE of $X\beta$ (and hence a BLUE of $X\beta$).

PROOF. Let AY be a least squares CLUE and BY any other CLUE. We need to show that, for any vector ρ,

$$\mathrm{Var}(\rho' AY) \leq \mathrm{Var}(\rho' BY).$$

We can decompose $\mathrm{Var}(\rho' BY)$,

$$\begin{aligned} \mathrm{Var}(\rho' BY) &= \mathrm{Var}(\rho'(B-A)Y + \rho' AY) \\ &= \mathrm{Var}(\rho'(B-A)Y) + \mathrm{Var}(\rho' AY) + 2\mathrm{Cov}(\rho'(B-A)Y, \rho' AY). \end{aligned}$$

Since variances are nonnegative, it suffices to show that

$$0 = \mathrm{Cov}(\rho'(B-A)Y, \rho' AY) = \sigma^2 \rho'(B-A)VA'\rho.$$

First, we will establish that it is enough to show that the covariance is zero when $\rho \in C(X) \cap C(V)$. Let M_0 be the perpendicular projection matrix onto $C(X) \cap C(V)$. Then $\rho = M_0\rho + (I-M_0)\rho$ and

$$\begin{aligned} \rho'(B-A)VA'\rho &= \rho'(B-A)VA'M_0\rho + \rho'(B-A)VA'(I-M_0)\rho \\ &= \rho' M_0(B-A)VA'M_0\rho + \rho'(I-M_0)(B-A)VA'M_0\rho \\ &\quad + \rho' M_0(B-A)VA'(I-M_0)\rho \\ &\quad + \rho'(I-M_0)(B-A)VA'(I-M_0)\rho. \end{aligned}$$

It turns out that all of these terms except the first is zero. Since AY and BY are CLUEs, unbiasedness and observation (d) in Section 2 give $C(AV) \subset C(X) \cap C(V)$ and $C(BV) \subset C(X) \cap C(V)$. By orthogonality,

$$VA'(I-M_0)\rho = 0 \quad \text{and} \quad \rho'(I-M_0)(B-A)V = 0;$$

so

$$\rho'(B-A)VA'\rho = \rho' M_0 (B-A)VA' M_0 \rho.$$

Henceforth, assume $\rho \in C(X) \cap C(V)$.

To obtain the final result, observe that since AY is a least squares CLUE, any column of $(I-A)V$ is orthogonal to ρ; so

$$V(I-A)'\rho = 0 \quad \text{and} \quad V\rho = VA'\rho.$$

Since, by assumption, $C(VV_0) \subset C(V_0)$, we also have $V\rho \in C(X) \cap C(V)$. The covariance term is $\rho'(B-A)VA'\rho = \rho'(B-A)V\rho$. Since AY and BY are unbiased and $V\rho \in C(X)$, $(B-A)V\rho = 0$; hence

$$\rho'(B-A)VA'\rho = 0. \qquad \square$$

As mentioned in the statement of the theorem, a best CLUE is a BLUE. That occurs because all BLUEs have the same variance and there is a BLUE that is a CLUE.

As would be expected, when $C(X) \subset C(V)$, a least squares CLUE will equal MY for all $Y \in C(V)$. When $C(X) \subset C(V)$, then $C(X) = C(V_0)$; so $MV_1 = 0$ and $MY - MV_1\hat{\gamma} = MY$.

The following theorem characterizes when ordinary least squares estimates are BLUEs.

Theorem 10.4.5. The following conditions are equivalent:

(a) $C(VX) \subset C(X)$,
(b) $C(VV_0) \subset C(V_0)$ and $X'V_1 = 0$,
(c) MY is a BLUE for $X\beta$.

PROOF. Let X_1 be a matrix with orthonormal columns such that $C(X) = C(V_0, X_1)$ and $V_0'X_1 = 0$. Also let

$$V = [V_0, V_1] \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} V_0' \\ V_1' \end{bmatrix};$$

so $V = V_0 B_{11} V_0' + V_1 B_{22} V_1' + V_0 B_{12} V_1' + V_1 B_{21} V_0'$. By symmetry, $B_{12} = B_{21}'$. Recall that V_0 and V_1 also have orthonormal columns.

$a \Rightarrow b$

Clearly $C(VX) \subset C(V)$; so if $C(VX) \subset C(X)$, we have $C(VX) \subset C(V_0)$. It is easily seen that

$$C(VX) \subset C(V_0) \text{ if and only if } C(VV_0) \subset C(V_0) \text{ and } C(VX_1) \subset C(V_0).$$

We show that $C(VX_1) \subset C(V_0)$ implies that $X_1'V = 0$; hence $X_1'V_1 = 0$ and $X'V_1 = 0$. First $VX_1 = V_1 B_{22} V_1' X_1 + V_0 B_{12} V_1' X_1$; we show that both terms are zero.

Consider $VV_0 = V_0B_{11} + V_1B_{21}$; since $C(VV_0) \subset C(V_0)$, we must have $V_1B_{21} = 0$. By symmetry, $B_{12}V_1' = 0$ and $V_0B_{12}V_1'X_1 = 0$. To see that $VX_1 = V_1B_{22}V_1'X_1 = 0$, observe that since $C(VX_1) \subset C(V_0)$, it must be true that $C(V_1B_{22}V_1'X_1) \subset C(V_0)$. However, $C(V_1B_{22}V_1'X_1) \subset C(V_1)$ but $C(V_0)$ and $C(V_1)$ are orthogonal, so $V_1B_{22}V_1'X_1 = 0$.

$b \Rightarrow a$

If $X'V_1 = 0$, then $X_1'V_1 = 0$ and $X_1'V = 0$. Write $X = V_0B_0 + X_1B_1$ so $VX = VV_0B_0 + VX_1B_1 = VV_0B_0$. Thus,

$$C(VX) = C(VV_0B_0) \subset C(VV_0) \subset C(V_0) \subset C(X).$$

$b \Rightarrow c$

If $C(VV_0) \subset C(V_0)$, then $MY - MV_1\hat{\gamma}$ is a BLUE. Since $X'V_1 = 0$, $MV_1 = 0$ and MY is a BLUE.

$c \Rightarrow b$

If AY and BY are BLUEs, then $AY = BY$ for $Y \in C(X,V)$. As in Proposition 10.2.3, there exists a BLUE, say AY, such that $AV = VA'$. Since MY is a BLUE and $C(V) \subset C(X,V)$, $AV = MV$ and $MV = VM$. Finally, $VV_0 = VMV_0 = MVV_0$, so $C(VV_0) = C(MVV_0) \subset C(X)$. Since $C(VV_0) \subset C(V)$, we have $C(VV_0) \subset C(V_0)$.

From Theorem 10.4.4 we know that a least squares CLUE is a BLUE; hence $MY = MY - MV_1\hat{\gamma}$ for $Y \in C(X,V) = C(X,V_1)$. Since

$$\hat{\gamma} = \left[V_1'(I-M)V_1\right]^{-1} V_1'(I-M)Y,$$

we must have $0 = MV_1\left[V_1'(I-M)V_1\right]^{-1} V_1'(I-M)V_1 = MV_1$. Thus $0 = X'V_1$. □

In the proof of $a \Rightarrow b$, it was noted that $V_1B_{21} = 0$. That means we can write $V = V_0B_{11}V_0' + V_1B_{22}V_1'$. For ordinary least squares estimates to be BLUEs, $C(V)$ must admit an orthogonal decomposition into a subspace contained in $C(X)$ and a subspace orthogonal to $C(X)$. Moreover, the error term e in model (10.1.1) must have $e = e_0 + e_1$, where $\mathrm{Cov}(e_0, e_1) = 0$, $\mathrm{Cov}(e_0) = V_0B_{11}V_0'$, and $\mathrm{Cov}(e_1) = V_1B_{22}V_1'$. Thus, with probability 1, the error can be written as the sum of two orthogonal vectors, both in $C(V)$, and one in $C(X)$. The two vectors must also be uncorrelated.

Exercise 10.6 Show that $V_1'(I-M)V_1$ is invertible.

Answer: The columns of V_1 form a basis, so $0 = V_1b$ iff $b = 0$. Also $(I-M)V_1b = 0$ iff $V_1b \in C(X)$, but $V_1b \in C(X)$ iff $b = 0$ by choice of V_1. Thus, $(I-M)V_1b = 0$ iff $b = 0$; hence $(I-M)V_1$ has full column rank and $V_1'(I-M)V_1$ is invertible.

Exercise 10.7 Show that if $MY - MV_1\hat{\gamma}$ is a BLUE of $X\beta$, then $C(VV_0) \subset C(V_0)$.

Hint: Multiply on the right by V_0 after showing that

$$\begin{aligned}\Big[M - MV_1\left[V_1'(I-M)V_1\right]^{-1} V_1'(I-M)\Big]V \\ = V\Big[M - (I-M)V_1\left[V_1'(I-M)V_1\right]^{-1} V_1'M\Big].\end{aligned}$$

We include a result that allows one to find the matrix V_1, and thus find least squares CLUEs.

Proposition 10.4.6. $r \perp C(X) \cap C(V)$ if and only if $r \in C(I-M, I-M_V)$.

PROOF. If $r \in C(I-M, I-M_V)$, then write $r = (I-M)r_1 + (I-M_V)r_2$.

Let $w \in C(X) \cap C(V)$ so that $w = M_V w = Mw$. We need to show that $w'r = 0$. Observe that

$$\begin{aligned} w'r &= w'(I-M)r_1 + w'(I-M_V)r_2 \\ &= w'M(I-M)r_1 + w'M_V(I-M_V)r_2 \\ &= 0. \end{aligned}$$

The vector space here is, say, $\mathbf{R}^n$. Let $r[C(X) \cap C(V)] = m$. From the above result, $C(I-M, I-M_V)$ is orthogonal to $C(X) \cap C(V)$. It is enough to show that the rank of $C(I-M, I-M_V)$ is $n-m$. If this is not the case, there exists a vector $w \neq 0$ such that $w \perp C(I-M, I-M_V)$ and $w \perp C(X) \cap C(V)$.

Since $w \perp C(I-M, I-M_V)$, we have $(I-M)w = 0$ or $w = Mw \in C(X)$. Similarly, $w = M_V w \in C(V)$; so $w \in C(X) \cap C(V)$, a contradiction. $\square$

To find V_1, one could use Gram–Schmidt to first get an orthonormal basis for $C(I-M, I-M_V)$, then extend this to $\mathbf{R}^n$. The extension is a basis for $C(V_0)$. Finally, extend the basis for $C(V_0)$ to an orthonormal basis for $C(V)$. The extension is a basis for $C(V_1)$. A basis for $C(X_1)$ can be found by extending the basis for $C(V_0)$ to a basis for $C(X)$.

Exercise 10.8 Give the general form for a BLUE of $X\beta$ in model (10.1.1).

Hint: Add something to a particular BLUE.

Exercise 10.9 From inspecting Figure 10.2, give the least squares CLUE for Example 10.2.2. Do not do any matrix manipulations.

Remark. Suppose that we are analyzing the model $Y = X\beta + e$, $\mathrm{E}(e) = 0$, $\mathrm{Cov}(e) = \Sigma(\theta)$, where $\Sigma(\theta)$ is some nonnegative definite matrix depending on a vector of unknown parameters θ. The special case where $\Sigma(\theta) = \sigma^2 V$ is what we have been considering so far. It is clear that if θ is known, our current theory gives BLUEs. If it happens to be the case that for any value of θ the BLUEs are identical, then the BLUEs are known even though θ may not be. This is precisely what we have been doing with $\Sigma(\theta) = \sigma^2 V$. We have found BLUEs for any value of σ^2, and they do not depend on σ^2. Another important example of this occurs when $C(\Sigma(\theta)X) \subset C(X)$ for any θ. In this case, least squares estimates are BLUEs for any θ, and least squares estimates do not depend on θ, so it does not matter that θ is unknown. The split plot design model is one in which the covariance matrix depends on two parameters, but for any value of those parameters, least squares estimates are BLUEs. Such models are examined in the next chapter.

Exercise 10.10 Show that ordinary least squares estimates are best linear unbiased estimates in the model $Y = X\beta + e$, $\mathrm{E}(e) = 0$, $\mathrm{Cov}(e) = V$ if the columns of X are eigenvectors of V.

Exercise 10.11 Use Definition B.31 and Proposition 10.4.6 to show that $M_{C(X)\cap C(V)} = M - M_W$ where $C(W) = C[M(I - M_V)]$.

10.5 Perfect Estimation and More

One of the interesting things about the linear model

$$Y = X\beta + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma^2 V, \tag{1}$$

is that when $C(X) \not\subset C(V)$ you can learn things about the parameters with probability 1. We identify these estimable functions of the parameters and consider the process of estimating those parameters that are not perfectly known.

Earlier, to treat $C(X) \not\subset C(V)$, we obtained estimates and tests by replacing V in model (1) with T where $C(X) \subset C(T)$. Although we showed that the procedure works, it is probably not the first thing one would think to do. In this section, after isolating the estimable functions of $X\beta$ that can be known perfectly, we replace model (1) with a new model $\tilde{Y} = V_0\gamma + e$, $\mathrm{E}(e) = 0$, $\mathrm{Cov}(e) = \sigma^2 V$ in which $C(V_0) \subset C(V)$ and $\tilde{Y}$ is just Y minus a perfectly known component of $X\beta$. I find it far more intuitive to make adjustments to the model matrix X, something we do regularly in defining reduced and restricted models, than to adjust the covariance matrix V. Additional details are given in Christensen and Lin (2010).

EXAMPLE 10.5.1. Consider a one-sample model $y_i = \mu + \varepsilon_i$, $i = 1,2,3$, with uncorrelated observations but in which the second and third observations have variance 0. If this model is correct, you obviously should have $\mu = y_2 = y_3$ with probability 1. The matrices for model (1) are

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

All of our examples in this section use this same Y vector.

The first thing to do with models having $C(X) \not\subset C(V)$ is to see whether they are even plausible for the data. In particular, Lemma 1.3.5 implies that $\Pr[(Y - X\beta) \in$

$C(V)] = 1$ so that $Y \in C(X,V)$ a.s. This should be used as a model-checking device. If $Y \notin C(X,V)$, you clearly have the wrong model.

EXAMPLE 10.5.1 CONTINUED. In this example, if $y_2 \neq y_3$ we obviously have the wrong model. It is easily seen that

$$C(X,V) = C\left(\begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}\right)$$

and if $y_2 \neq y_3$, $Y \notin C(X,V)$.

To identify the estimable parameters that can be known perfectly, let Q be a full rank matrix with

$$C(Q) = C(V)^{\perp}.$$

Note that if $C(X) \not\subset C(V)$, then $Q'X \neq 0$. Actually, the contrapositive is more obvious, if $Q'X = 0$ then

$$C(X) \subset C(Q)^{\perp} = \left[C(V)^{\perp}\right]^{\perp} = C(V).$$

The fact that $Q'X \neq 0$ means that the estimable function $Q'X\beta$ is nontrivial. Note also that $C(X) \not\subset C(V)$ implies that V must be singular. If V were nonsingular, then $C(V) = \mathbf{R}^n$ and $C(X)$ has to be contained in it.

We can now identify the estimable functions that are known perfectly. Because $\Pr[(Y - X\beta) \in C(V)] = 1$, clearly $\Pr[Q'(Y - X\beta) = 0] = 1$ and $Q'Y = Q'X\beta$ a.s. Therefore, whenever $C(X) \not\subset C(V)$, there are nontrivial estimable functions of β that we can learn without error. Moreover, $\mathrm{Cov}(Q'Y) = \sigma^2 Q'VQ = 0$ and only linear functions of $Q'Y$ will have 0 covariance matrices, so only linear functions of $Q'X\beta$ will be estimated perfectly.

Exercise 10.13 Show that $\mathrm{Cov}(B'Y) = 0$ iff $B' = B_*'Q'$ for some matrix B_*.

Hint: First, decompose B into the sum of two matrices B_0 and B_1 with $C(B_0) \subset C(V)$ and $C(B_1) \perp C(V)$. Then use a singular value decomposition of V to show that $B_0'VB_0 = 0$ iff $B_0 = 0$.

If $Q'X$ has full column rank, we can actually learn all of $X\beta$ without error. In that case, with probability 1 we can write

$$X\beta = X\left(X'QQ'X\right)^{-1}\left[X'QQ'X\right]\beta = X(X'QQ'X)^{-1}X'QQ'Y.$$

For convenience, define A so that $AY \equiv X(X'QQ'X)^{-1}X'QQ'Y$, in which case $X\beta = AY$ a.s. In particular, it is easy to see that $\mathrm{E}[AY] = X\beta$ and $\mathrm{Cov}[AY] = 0$. We now illustrate these matrix formulations in some simple examples to show that the matrix results give obviously correct answers.

EXAMPLE 10.5.1 CONTINUED. Using the earlier forms for X and V,

$$Q = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

It follows that

$$Q'X = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and $Q'Y = Q'X\beta$ reduces to

$$\begin{bmatrix} y_2 \\ y_3 \end{bmatrix} = \mu J_2 \quad \text{a.s.}$$

Obviously, for this to be true, $y_2 = y_3$ a.s. Moreover, since $Q'X$ is full rank, upon observing that $X'QQ'X = 2$ we can compute

$$\mu J_3 = X\beta = AY = [(y_2 + y_3)/2]J_3 \quad \text{a.s.}$$

EXAMPLE 10.5.2. This is a two-sample problem, with the first two observations from sample one and the third from sample two. Again, observations two and three have 0 variance. Clearly, with probability 1, $\mu_1 = y_2$ and $\mu_2 = y_3$. The key matrices are

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Since $\beta = [\mu_1, \mu_2]'$ in the two-sample problem and

$$Q'X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

we have

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = Q'X\beta = Q'Y = \begin{bmatrix} y_2 \\ y_3 \end{bmatrix} \quad \text{a.s.}$$

In particular, with probability 1,

$$\begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \end{bmatrix} = X\beta = AY = \begin{bmatrix} y_2 \\ y_2 \\ y_3 \end{bmatrix}.$$

EXAMPLE 10.5.3. This is a one-sample problem similar to Example 10.5.1 except that now the first two observations have variance 1 but the third has variance 0. The key matrices are

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

With $Q'X = 1$ we get $Q'X\beta = \mu$ equaling $Q'Y = y_3$ with probability 1. Moreover, since $X'QQ'X = 1$ we can easily compute

$$\mu J_3 = X\beta = AY = y_3 J_3.$$

The next two examples do not have $Q'X$ with full rank, so they actually have something to estimate.

EXAMPLE 10.5.4. Consider a two-sample problem similar to Example 10.5.2 except that now the first two observations have variance 1 but the third has variance 0. The key matrices are

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

With $\beta = [\mu_1, \mu_2]'$ and

$$Q'X = [0 \quad 1],$$

we get

$$\mu_2 = [0 \quad 1] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = Q'X\beta = Q'Y = y_3 \quad \text{a.s.}$$

Clearly, $y_3 = \mu_2$ a.s. but μ_1 would be estimated with the average $(y_1 + y_2)/2$. More on this later.

EXAMPLE 10.5.5 Finally, in the two-sample problem we move the second observation from the first group to the second group. This time

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

with

$$Q'X = [0 \quad 1].$$

Again, with probability 1,

$$\mu_2 = [0 \quad 1] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = Q'X\beta = Q'Y = y_3,$$

so $y_3 = \mu_2$ a.s. But this time, only y_1 would be used to estimate μ_1, and y_2 is of no value for estimating the means. However, y_2 could be used to estimate an unknown variance via $(y_2 - \mu_2)^2 = (y_2 - y_3)^2$. Similar results on estimating the variance apply in all the examples except Example 10.5.4.

To get perfect estimation of *anything*, we need $C(X) \not\subset C(V)$. To get perfect estimation of *everything* we need $C(X) \cap C(V) = \{0\}$. In other words, to get perfect estimation of $X\beta$ we need $Q'X$ with full column rank and to get $Q'X$ with full column rank, we need $C(X) \cap C(V) = \{0\}$.

Proposition 10.5.6. For X of full column rank, $Q'X$ is of full column rank if and only if $C(X) \cap C(V) = \{0\}$.

PROOF. This is a special case of Lemma 10.5.7. □

Although $C(X) \cap C(V) = \{0\}$ is actually a necessary and sufficient condition for perfect estimation of $X\beta$, with the methods we have illustrated it is not obviously sufficient. For our current method, we need $Q'X$ to have full column rank, which obviously will not happen if X is not full rank. Fortunately, we can always simply choose X to have full column rank. In addition, we close this section with the mathematics needed to deal with arbitrary X.

Now consider models in which some aspect of $X\beta$ is known but some aspect is not. In particular, we know that $Q'X\beta$ is known, but how do we estimate the rest of $X\beta$? As discussed above, we must now consider the case where $C(X) \cap C(V) \neq \{0\}$. Write $\beta = \beta_0 + \beta_1$ with $\beta_0 \in C(X'Q)$ and $\beta_1 \perp C(X'Q)$. We show that $X\beta_0$ is known, so that we need only estimate $X\beta_1$ to learn all that can be learned about $X\beta$. These methods make no assumption about $r(X)$.

In fact, β_0 is known, not just $X\beta_0$. Let $P_{X'Q}$ be the ppo onto $C(X'Q)$. By the definition of β_0 as part of an (unique) orthogonal decomposition, with probability 1,

$$\beta_0 = P_{X'Q}\beta = X'Q[Q'XX'Q]^- Q'X\beta = X'Q[Q'XX'Q]^- Q'Y.$$

Since the perpendicular projection operator does not depend on the choice of generalized inverse, neither does β_0.

Now we show how to estimate $X\beta_1$. Let V_0 be such that $C(V_0) = C(X) \cap C(V)$. Proposition 10.4.6 can be used to find V_0. Note that $\beta_1 \perp C(X'Q)$ iff $Q'X\beta_1 = 0$ iff $X\beta_1 \perp C(Q)$ iff $X\beta_1 \in C(V)$ iff $X\beta_1 \in C(V_0)$ iff $X\beta_1 = V_0\gamma$ for some γ. Since $X\beta_0$ is fixed and known, it follows that $\mathrm{E}(Y - X\beta_0) = X\beta_1 \in C(V_0)$ and $\mathrm{Cov}(Y - X\beta_0) = \sigma^2 V$, so we can estimate $X\beta_1$ by fitting

$$Y - X\beta_0 = V_0\gamma + e, \quad \mathrm{E}(e) = 0, \quad \mathrm{Cov}(e) = \sigma^2 V, \tag{2}$$

and taking

$$X\hat{\beta}_1 \equiv V_0\hat{\gamma} = V_0(V_0'V^- V_0)^- V_0 V^-(Y - X\beta_0),$$

cf. Section 2. Under normality, tests are also relatively easy to construct.

EXAMPLE 10.5.4 CONTINUED. Using the earlier versions of X, V, Q, and $Q'X$, observe that

$$C(V_0) = C\left(\begin{bmatrix}1\\1\\0\end{bmatrix}\right), \qquad C(X'Q) = C\left(\begin{bmatrix}0\\1\end{bmatrix}\right).$$

It follows that with $\beta = [\mu_1, \mu_2]'$,

$$\beta_0 = \begin{bmatrix}0\\ \mu_2\end{bmatrix}, \qquad \beta_1 = \begin{bmatrix}\mu_1\\ 0\end{bmatrix}.$$

Thus, since we already know that $\mu_2 = y_3$ a.s., $X\beta_0 = [0,0,\mu_2]' = [0,0,y_3]'$ and $X\beta_1 = [\mu_1, \mu_1, 0]'$. Finally, model (2) reduces, with probability 1, to

$$Y - X\beta_0 = \begin{bmatrix}y_1\\ y_2\\ y_3-\mu_2\end{bmatrix} = \begin{bmatrix}y_1\\ y_2\\ 0\end{bmatrix} = \begin{bmatrix}1\\1\\0\end{bmatrix}\gamma + e.$$

Recalling that $X\beta_1 \equiv V_0\gamma$, it is easily seen in this example that the BLUE of $\mu_1 \equiv \gamma$ is $(y_1+y_2)/2$.

This theory applied to Example 10.5.1 is quite degenerate, but it still works.

EXAMPLE 10.5.1 CONTINUED. Using the earlier versions of X, V, Q, and $Q'X$, observe that

$$C(V_0) = C\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\right), \qquad C(X'Q) = C\left(\begin{bmatrix}1 & 1\end{bmatrix}\right).$$

It follows that

$$\beta_0 = \mu, \qquad \beta_1 = 0.$$

Since we already know that $\mu = y_2 = y_3$ a.s.,

$$X\beta_0 = \begin{bmatrix}\mu\\ \mu\\ \mu\end{bmatrix} = \begin{bmatrix}y_2\\ y_2\\ y_2\end{bmatrix} = \begin{bmatrix}y_3\\ y_3\\ y_3\end{bmatrix} = \begin{bmatrix}y_2\\ y_2\\ y_3\end{bmatrix} \quad \text{a.s.}$$

and $X\beta_1 = [0,0,0]'$. Finally, model (2) reduces, with probability 1, to

$$Y - X\beta_0 = \begin{bmatrix}y_1-\mu\\ y_2-\mu\\ y_3-\mu\end{bmatrix} = \begin{bmatrix}y_1-y_2\\ 0\\ 0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}\gamma + e,$$

which provides us with one degree of freedom for estimating σ^2 using either of $(y_1 - y_i)^2$, $i = 2,3$.

The results in the early part of this section on perfect estimation of $X\beta$ required X to be of full rank. That is never a very satisfying state of affairs. Rather than assuming X to be of full rank and considering whether $Q'X$ is also of full rank, the more general condition for estimating $X\beta$ perfectly is that $r(X) = r(Q'X)$. Moreover, with $A \equiv X(X'QQ'X)^{-}X'QQ'$, we always have perfect estimation of $AX\beta$ because $AX\beta$ is a linear function of $Q'X\beta = Q'Y$ a.s. but for perfect estimation of $X\beta$ we need

$$X\beta = AX\beta = AY \quad \text{a.s.}$$

for any β which requires A to be a projection operator onto $C(X)$. This added generality requires some added work.

Lemma 10.5.7

(a) $r(X) = r(Q'X)$ iff for any b, $Q'Xb = 0$ implies $Xb = 0$.
(b) $r(X) = r(Q'X)$ iff $C(X) \cap C(V) = \{0\}$.

PROOF.
Proof of (a): Recall that $r(Q'X) = r(X)$ iff $r[\mathscr{N}(Q'X)] = r[\mathscr{N}(X)]$. Since the null spaces have $\mathscr{N}(X) \subset \mathscr{N}(Q'X)$, it is enough to show that $\mathscr{N}(Q'X) = \mathscr{N}(X)$ is equivalent to the condition that for any b, $Q'Xb = 0$ implies $Xb = 0$ and, in particular, it is enough to show that $\mathscr{N}(Q'X) \subset \mathscr{N}(X)$ is equivalent to the condition. But by the very definition of the null spaces, $\mathscr{N}(Q'X) \subset \mathscr{N}(X)$ is equivalent to the condition that for any b we have $Q'Xb = 0$ implies that $Xb = 0$.

Proof of (b): Note that for any b,

$$Q'Xb = 0 \quad \text{iff} \quad Xb \perp C(Q) \quad \text{iff} \quad Xb \in C(Q)^{\perp} = \left[C(V)^{\perp}\right]^{\perp} = C(V),$$

so

$$Q'Xb = 0 \quad \text{iff} \quad Xb \in C(V) \quad \text{iff} \quad Xb \in C(X) \cap C(V).$$

If follows immediately that if $C(X) \cap C(V) = \{0\}$, then $Q'Xb = 0$ implies $Xb = 0$ and $r(X) = r(Q'X)$. It also follows immediately that since $Q'Xb = 0$ is equivalent to having $Xb \in C(X) \cap C(V)$, the condition that $Q'Xb = 0$ implies $Xb = 0$ means that the only vector in $C(X) \cap C(V)$ is the 0 vector. □

Proposition 10.5.8 If $r(X) = r(Q'X)$, the matrix $A \equiv X(X'QQ'X)^{-}X'QQ'$ is a projection operator onto $C(X)$.

PROOF. By its definition we clearly have $C(A) \subset C(X)$, so it is enough to show that $AX = X$.

Let $M_{Q'X}$ be the ppo onto $C(Q'X)$. Note that for any b, $Xb - AXb \in C(X)$. Moreover, from the definitions of A and $M_{Q'X}$,

$$Q'Xb - Q'AXb = Q'Xb - M_{Q'X}Q'Xb = 0.$$

Writing

$$0 = Q'Xb - M_{Q'X}Q'Xb = Q'X\left[I - (X'QQ'X)^{-}X'QQ'X\right]b,$$

by the condition $r(X) = r(Q'X)$ and Lemma 10.5.8a, we have

$$0 = X\left[I - (X'QQ'X)^{-}X'QQ'X\right]b = Xb - AXb,$$

hence $X = AX$. $\square$

Exercise 10.13 Show that the results in this section do not depend on the particular choice of Q.

Exercise 10.14 Let $C(V_0) = C(X) \cap C(V)$, $C(X) = C(V_0, X_1)$, $C(V) = C(V_0, V_1)$ with the columns of V_0, V_1, and X_1 being orthonormal. Show that the columns of $[V_0, V_1, X_1]$ are linearly dependent.

Hint: Write $V_0 b_0 + V_1 b_1 + X_1 b_2 = 0$ and show that $b_i = 0$, $i = 0, 1, 2$. In particular, write

$$0.5V_0 b_0 + V_1 b_1 = -(0.5V_0 b_0 + X_1 b_2),$$

$0.5V_0 b_0 + V_1 b_1 \in C(V)$ and $-(0.5V_0 b_0 + X_1 b_2) \in C(X)$ so the vector is in $C(V_0) = C(X) \cap C(V)$.