$e = Y - \hat{Y}$ $\cdot$ $\cdot$ $Y \epsilon R^n$

$\cdot$

$\hat{Y}$

$\nearrow$

vector space spanned by the columns of X.

dimension of this space is rank(X).

The residual vector

$$e = Y - \tilde{Y} = (I - P_X)Y$$

is in the space orthogonal to the space spanned by the columns of X. It has dimension

$$n - rank(X).$$

Partition of a total sum of squares

Squared length of Y is $\sum_{i=1}^{n} y_i^2 = Y^T Y$

Squared length of the residual vector is

$$\begin{aligned} \sum_{i=1}^{n} e_i^2 &= e^T e \\ &= [(I - P_X)Y]^T (I - P_X)Y \\ &= Y^T (I - P_X)Y \end{aligned}$$

Squared length of $\hat{Y} = P_X Y$ is

$$\begin{aligned} \sum_{i=1}^{n} \hat{Y}_i^2 &= \hat{Y}^T \hat{Y} \\ &= (P_X Y)^T (P_X Y) \\ &= Y^T (P_X)^T P_X Y \quad \text{since } P_X \text{ is symmetric} \\ &= Y^T P_X P_X Y \quad \text{since } P_X \text{ is idempotent} \\ &= Y^T P_X Y \end{aligned}$$

We have

$$\begin{aligned} Y^T Y &= Y^T (P_X + I - P_X) Y \\ &= Y^T P_X Y + Y^T (I - P_X) Y \ . \end{aligned}$$

ANOVA

Source of Variation	Degrees of Freedom	Sums of Squares
model (uncorrected)	rank(X)	$\hat{Y}^T \hat{Y} = Y^T P_X Y$
residuals	n-rank(X)	$e^T e = Y^T (I - P_X) Y$
total (uncorrected)	n	$Y^T Y = \sum_{i=1}^{n} y_i^2$

<u>Result 3.6</u> For the linear model

$$E(Y) = X\beta \text{ and } Var(Y) = \Sigma,$$

the OLS estimator $\hat{Y} = Xb = P_X Y$ for $X\beta$ is

(i) unbiased, i.e., $E(\hat{Y}) = X\beta$

(ii) a linear function of Y

(iii) has variance-covariance matrix

$$Var(\hat{Y}) = P_X \Sigma P_X$$

This is true for any solution

$$b = (X^T X)^- X^T Y$$

to the normal equations.

Proof:

(ii) is trivial, since $\hat{Y} = P_X \text{Y}$

(iii) follows from result 2.1.(ii)

(i)

$$\begin{aligned} E(\hat{\text{Y}}) &= E(P_X \text{Y}) \\ &= P_X E(Y) \text{ from result 2.1.(i)} \\ &= P_X X\beta \\ &= X\beta \text{ since } P_X X = X \end{aligned}$$

Comments:

- $\hat{\text{Y}} = X\text{b} = P_X \text{Y}$ is said to be a linear unbiased estimator for $E(\text{Y}) = X\beta$
- For the Gauss-Markov model, $Var(\text{Y}) = \sigma^2 I$ and

$$\begin{aligned} Var(\hat{\text{Y}}) &= P_X(\sigma^2 I)P_X \\ &= \sigma^2 P_X P_X \\ &= \sigma^2 P_X \\ &= \sigma^2 \underline{X(X^T X)^- X^T} \end{aligned}$$

↑ this is sometimes called the "hat" matrix.

Questions

Is $\hat{\text{Y}} = X\text{b} = P_X \text{Y}$ the "best" estimator for $E(\text{Y}) = X\beta$?

Is $\hat{\text{Y}} = X\text{b} = P_X \text{Y}$ the "best" estimator for $E(Y) = X\beta$ in the class of linear, unbiased estimators?

What other linear functions of β, say

$$\text{c}^T\beta = c_1\beta_1 + c_2\beta_2 + \cdots + c_k\beta_k,$$

have OLS estimators that are invariant to the choice of

$$\text{b} = (X^T X)^- X^T \text{Y}$$

that solves the normal equations?

Estimable Functions

Some estimates of linear functions of the parameters have the same value, regardless of which solution to the normal equations is used

- These are called estimable functions
- An example is $E(\text{Y}) = X\beta$

Check that $X\text{b}$ has the same value for each solution to the normal equations obtained in Example 3.2, i.e.,

$$X\text{b} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \end{bmatrix}$$

Estimable Functions

<u>Defn 3.6:</u> For a linear model

$$E(\mathrm{Y}) = X\beta \quad \text{and} \quad Var(\mathrm{Y}) = \Sigma$$

we will say that

$$\mathrm{c}^T\beta = c_1\beta_1 + c_2\beta_2 + \cdots + c_k\beta_k$$

is <u>e</u>stimable if there exists a linear unbiased estimator $\mathrm{a}^T\mathrm{Y}$ for $\mathrm{c}^T\beta$, i.e., for some non-random vector a, we have $E(\mathrm{a}^T\mathrm{Y}) = \mathrm{c}^T\beta$.

<u>Example 3.2.</u> Blood coagulation times

Diet 1	Diet 2	Diet 3
$Y_{11} = 62$	$Y_{21} = 71$	$Y_{31} = 72$
$Y_{12} = 60$		$Y_{32} = 68$
		$Y_{33} = 67$

The "Effects" model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

can be written as

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

Examples of estimable functions

<u>$\mu + \alpha_1$</u>

Choose $\mathrm{a}^T = (\frac{1}{2}\ \frac{1}{2}\ 0\ 0\ 0\ 0)$. Then,

$$\begin{aligned} E(\mathrm{a}^T\mathrm{Y}) &= E(\frac{1}{2}Y_{11} + \frac{1}{2}Y_{12}) \\ &= \frac{1}{2}E(Y_{11}) + \frac{1}{2}E(Y_{12}) \\ &= \frac{1}{2}(\mu + \alpha_1) + \frac{1}{2}(\mu + \alpha_1) \\ &= \mu + \alpha_1 \end{aligned}$$

Choose $\mathrm{a}^T = (1\ 0\ 0\ 0\ 0\ 0)$ and note that $E(\mathrm{a}^T\mathrm{Y}) = E(Y_{11}) = \mu + \alpha_1$.

<u>$\mu + \alpha_2$</u>

Choose $\mathrm{a}^T = (0\ 0\ 1\ 0\ 0\ 0)$. Then,

$$\mathrm{a}^T\mathrm{Y} = Y_{21}$$

and

$$E(\mathrm{a}^T\mathrm{Y}) = E(Y_{21}) = \mu + \alpha_2.$$

<u>$\mu + \alpha_3$</u>

Choose $\mathrm{a}^T = (0\ 0\ 0\ 1\ 0\ 0)$. Then,

$$E(\mathrm{a}^T\mathrm{Y}) = E(Y_{31}) = \mu + \alpha_3$$

$\underline{\alpha_1 - \alpha_2}$

Note that

$$\begin{aligned} \alpha_1 - \alpha_2 &= (\mu + \alpha_1) - (\mu + \alpha_2) \\ &= E(Y_{11}) - E(Y_{21}) \\ &= E(Y_{11} - Y_{21}) \\ &= E(\mathrm{a}^T \mathrm{Y}) \end{aligned}$$

where

$$\mathrm{a}^T = (1 \ 0 \ \text{-}1 \ 0 \ 0 \ 0)$$

$\underline{2\mu + 3\alpha_1 - \alpha_2}$

Note that

$$\begin{aligned} 2\mu + 3\alpha_1 - \alpha_2 &= 3(\mu + \alpha_1) - (\mu + \alpha_2) \\ &= 3E(Y_{11}) - E(Y_{21}) \\ &= E(3Y_{11} - Y_{21}) \\ &= E(\mathrm{a}^T \mathrm{Y}) \end{aligned}$$

where

$$\mathrm{a}^T = (3 \ 0 \ \text{-}1 \ 0 \ 0 \ 0)$$

Quantities that are <u>not</u> estimable include

$$\mu, \ \alpha_1, \ \alpha_2, \ \alpha_3, \ 3\alpha_1, \ \alpha_1 + \alpha_2$$

To show that a linear function of parameters,

$$c_0\mu + c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3$$

is <u>not</u> estimable, one must show that there is no non-random vector

$$\mathrm{a}^T = (a_0, \ a_1, \ a_2, \ a_3)$$

for which

$$E(\mathrm{a}^T \mathrm{Y}) = c_0\mu + c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3$$

For α_1 to be estimable we would need to find an a that satisfies

$$\begin{aligned} \alpha_1 &= E(\mathrm{a}^T \mathrm{Y}) \\ &= a_1 E(Y_{11}) + a_2 E(Y_{12}) + a_3 E(Y_{21}) \\ &\quad + a_4 (E(Y_{31}) + a_5 E(Y_{32}) + a_6 E(Y_{33}) \\ &= (a_1 + a_2)(\mu + \alpha_1) + a_3(\mu + \alpha_2) \\ &\quad + (a_4 + a_5 + a_6)(\mu + \alpha_3) \end{aligned}$$

This implies $0 = a_3 = (a_4 + a_5 + a_6)$,

Then $\alpha_1 = (a_1 + a_2)(\mu + \alpha_1)$
which is impossible.

Example 3.1. Yield of a chemical process

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} 1 & 160 & 1 \\ 1 & 165 & 3 \\ 1 & 165 & 2 \\ 1 & 170 & 1 \\ 1 & 175 & 2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

Since X has full column rank, each element of β is estimable.

Consider $B_1 = \mathrm{c}^T\beta$ where $c = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

Since X has full column rank, the unique least squares estimator for β is

$$\mathrm{b} = (X^TX)^{-1}\mathrm{X}^T\mathrm{Y}$$

and an unbiased linear estimator for $\mathrm{c}^T\beta$ is

$$\mathrm{c}^T\mathrm{b} = \underbrace{\mathrm{c}^T(X^TX)^{-1}X^T}_{\text{call this } \mathrm{a}^T}\mathrm{Y}$$

Result 3.7 For a linear model with $E(\mathrm{Y}) = X\beta$ and $Var(Y) = \Sigma$

(i) The expectation of any observation is estimable.

(ii) A linear combination of estimable functions is estimable.

(iii) Each element of β is estimable if and only if rank(X) $= k =$ number of columns.

(iv) Every $\mathrm{c}^T\beta$ is estimable if and only if rank(X) $= k =$ number of columns in X.

Proof:

(i) For $\mathrm{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ with $E(\mathrm{Y}) = X\beta$ we have

$$Y_i = \mathrm{a}_i^T\mathrm{Y} \text{ where } \mathrm{a}_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \text{one in the } i\text{th position}$$

Then

$$\begin{aligned} E(Y_i) = E(\mathrm{a}_i^T\mathrm{Y}) &= \mathrm{a}_i^T E(\mathrm{Y}) \\ &= \underline{a_i^T X}\beta \\ &= \mathrm{c}_i^T\beta \end{aligned}$$

(ii) Suppose $c_i^T\beta$ is estimable. Then, there is an a_i such that $E(\mathrm{a}_i^T\mathrm{Y}) = \mathrm{c}_i^T\beta$.

Now consider a linear combination of estimable functions

$$w_1\mathrm{c}_1^T\beta + w_2\mathrm{c}_2^T\beta + \cdots + w_p\mathrm{c}_p^T\beta$$

Let $\mathrm{a} = w_1\mathrm{a}_1 + w_2\mathrm{a}_2 + \cdots + w_p\mathrm{a}_p$.

Then,

$$\begin{aligned} E(\mathrm{a}^T\mathrm{Y}) &= E(w_1\mathrm{a}_1^T\mathrm{Y} + \cdots + w_p\mathrm{a}_p^T\mathrm{Y}) \\ &= w_1E(\mathrm{a}_1^T\mathrm{Y}) + \cdots + w_pE(\mathrm{a}_p^T\mathrm{Y}) \\ &= w_1\mathrm{c}_1^T\beta + \cdots + w_p\mathrm{c}_p^T\beta \end{aligned}$$

(iii) Previous argument.

(iv) Follows from (ii) and (iii).

Result 3.8. For a linear model with $E(\mathrm{Y}) = X\beta$ and $Var(\mathrm{Y}) = \Sigma$, each of the following is true if and only if $\mathrm{c}^T\beta$ is estimable.

(i) $\mathrm{c}^T = \mathrm{a}^T X$ for some a i.e., c is in the space spanned by the rows of X.

(ii) $\mathrm{c}^T\mathrm{a} = 0$ for every a for which $X\mathrm{a} = 0$.

(iii) $\mathrm{c}^T\mathrm{b}$ is the same for any solution to the normal equations $(X^TX)\mathrm{b} = X^T\mathrm{Y}$, i.e., there is a unique least squares estimator for $\mathrm{c}^T\beta$.

Use Result 3.8. (ii) to show that μ is not estimable in Example 3.2. In that case

$$E(\mathrm{Y}) = X\beta = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

and

$$\mu = \mathrm{c}^T\beta = [1\ 0\ 0\ 0]\beta.$$

Let $\mathrm{d}^T = [1\ \ -1\ \ -1\ \ -1]$, then

$$X\mathrm{d} = 0, \quad \text{but} \quad \mathrm{c}^T\mathrm{d} = 1 \neq 0$$

Hence, μ is not estimable.

Part (ii) of Result 3.8 sometimes provides a convenient way to identify all possible estimable functions of β.

In example 3.2, $X\mathrm{d} = 0$ if and only if

$$\mathrm{d} = w \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \text{ for some scalar } w.$$

Then, $\mathrm{c}^T\beta$ is estimable if and only if

$$0 = \mathrm{c}^T\mathrm{d} = w(c_1 - c_2 - c_3 - c_4)$$

$$\iff \quad c_1 = c_2 + c_3 + c_4.$$

Then,

$$(c_2 + c_3 + c_4)\mu + c_2\alpha_1 + c_3\alpha_2 + c_4\alpha_3$$

is estimable for any $(c_2\ c_3\ c_4)$ and these are the only estimable functions of $\mu, \alpha_1, \alpha_2, \alpha_3$.

Some estimable functions are

$$\mu + \frac{1}{3}(\alpha_1 + \alpha_2 + \alpha_3) \quad (c_2 = c_3 = c_4 = \frac{1}{3})$$

and

$$\mu + \alpha_2 \quad (c_2 = 1\ c_3 = c_4 = 0)$$

but

$$\mu + 2\alpha_2$$

is not estimable.

Defn 3.7: For a linear model with

$$E(\mathrm{Y}) = X\beta \text{ and } Var(\mathrm{Y}) = \Sigma,$$

where X is an $n \times k$ matrix, $C_{r\times k}\beta_{k\times 1}$ is said to be estimable if all of its elements

$$C\beta = \begin{bmatrix} \mathrm{c}_1^T \\ \mathrm{c}_2^T \\ \vdots \\ \mathrm{c}_r^T \end{bmatrix} \beta = \begin{bmatrix} \mathrm{c}_1^T\beta \\ \mathrm{c}_2^T\beta \\ \vdots \\ \mathrm{c}_r^T\beta \end{bmatrix}$$

are estimable.

Result 3.9 For the linear model with $E(\mathrm{Y}) = X\beta$ and $Var(\mathrm{Y}) = \Sigma$, where X is an $n \times k$ matrix, each of the following conditions hold if and only if $C\beta$ is estimable.

(i) $AX = C$ for some matrix A, i.e., each row of C is in the space spanned by the rows of X.

(ii) $C\mathrm{d} = 0$ for any d for which $X\mathrm{d} = 0$.

(iii) $C\mathrm{b}$ is the same for any solution to the normal equations $(X^TX)\mathrm{b} = X^T\mathrm{y}$.

Summary

For a linear model

$$\mathrm{Y} = X\beta + \epsilon$$

with $E(\mathrm{Y}) = X\beta$ and $Var(\mathrm{Y}) = \Sigma$, we have

- Any estimable function has a unique interpretation
- The OLS estimator for an estimable function $C\beta$ is unique
  $$C\mathrm{b} = C\,(X^TX)^-X^T\mathrm{Y}$$
- The OLS estimator for an estimable function $C\beta$ is
  - a linear estimator
  - an unbiased estimator

In the class of linear unbiased estimators for $c^T\beta$, is the OLS estimator the "best?"

Here "best" means smallest expected squared error. Let $t(Y)$ denote an estimator for $c^T\beta$. Then, the expected squared error is

$$
\begin{aligned}
MSE &= E[t(Y) - c^T\beta]^2 \\
&= E[t(Y) - E(t(Y)) + E(t(Y)) - c^T\beta]^2 \\
&= E[t(Y) - E(t(Y))]^2 \\
&\quad + [E(t(Y)) - c^T\beta]^2 \\
&\quad + 2[E(t(Y)) - c^T\beta]E[t(Y) - E(t(Y))] \\
&= E[t(Y) - E(t(Y))]^2 + [E(t(Y)) - c^T\beta]^2 \\
&= Var(t(Y)) + [\text{bias}]^2
\end{aligned}
$$

If we restrict our attention to linear unbiased estimators for $c^T\beta$:

- $E(t(Y)) = c^T\beta$
- $t(Y) = a^TY$ for some a

then, $t(Y) = a^TY$ is the best linear unbiased estimator (blue) for $c^T\beta$ if

$$Var(a^TY) \le Var(d^TY)$$

for any d and any value of β.

<u>Result 3.10</u> (Gauss-Markov Theorem)

For the Gauss-Markov model,

$$E(Y) = X\beta \text{ and } Var(Y) = \sigma^2 I,$$

the OLS estimator of an estimable function $c^T\beta$ is the <u>unique</u> best linear unbiased estimator (blue) of $c^T\beta$.

<u>Proof:</u>

(i) For any solution $b = (X^TX)^-X^TY$ to the normal equations, the OLS estimator for $c^T\beta$ is

$$c^Tb = c^T(X^TX)^-X^Ty$$

which is a <u>linear</u> function of Y.

(ii) From Result 3.8.(i), there exists a vector a such that $c^T = a^TX$. Then

$$
\begin{aligned}
E(c^Tb) &= E(c^T(X^TX)^-X^TY) \\
&= c^T(X^TX)^-X^TE(Y) \\
&= c^T(X^TX)^-X^TX\beta \\
&= a^T\underline{X(X^TX)^-X^T}X\beta
\end{aligned}
$$

↑ projection P_X onto the column space of X

$$
\begin{aligned}
&= a^TX\beta \\
&= c^T\beta
\end{aligned}
$$

Hence, c^Tb is an <u>unbiased</u> estimator.

(iii) Minimum variance in the class of linear unbiased estimators

Suppose $\mathrm{d}^T\mathrm{Y}$ is any other <u>linear unbiased</u> estimator for $\mathrm{c}^T\beta$. Then

$$E(\mathrm{d}^T\mathrm{Y}) = \mathrm{d}^T E(\mathrm{Y}) = \mathrm{d}^T X\beta = \mathrm{c}^T\beta$$

for every β. Hence, $\mathrm{d}^T X = \mathrm{c}^T$ and $\mathrm{c} = X^T\mathrm{d}$.

We must show that

$$Var(\mathrm{c}^T\mathrm{b}) \leq Var(\mathrm{d}^T\mathrm{Y}).$$

First, note that

$$\begin{aligned} Var(\mathrm{d}^T\mathrm{Y}) &= Var(\mathrm{c}^T\mathrm{b} + [\mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b}]) \\ &= Var(\mathrm{c}^T\mathrm{b}) + Var(\mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b}) \\ &\quad +2Cov(\mathrm{c}^T\mathrm{b}, \mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b}) \end{aligned}$$

Then

$$\begin{aligned} Var(\mathrm{d}^T\mathrm{Y}) &\geq Var(\mathrm{c}^T\mathrm{b}) \\ &\quad +2Cov(\mathrm{c}^T\mathrm{b}, \mathrm{d}^T Y - \mathrm{c}^T\mathrm{b}) \\ &= Var(\mathrm{c}^T\mathrm{b}) \end{aligned}$$

because

$$Cov(\mathrm{c}^T\mathrm{b}, \mathrm{d}^T Y - \mathrm{c}^T\mathrm{b}) = 0.$$

To show this first note that

$$\mathrm{c}^T\mathrm{b} = \mathrm{c}^T(X^T X)^- X^T\mathrm{Y}$$

is invariant with respect to the choice of $(X^T X)^-$. Consequently, we can use the Moore-Penrose generalized inverse which is symmteric. (Not every generalized inverse of $X^T X$ is symmetric.)

Then,

$$Cov(\mathrm{c}^T\mathrm{b}, \mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b})$$

$$= Cov(\mathrm{c}^T(X^T X)^- \mathrm{X}^T\mathrm{Y}, [\mathrm{d}^T - \mathrm{c}^T(X^T X)^- X^T]\mathrm{Y})$$

$$= (\mathrm{c}^T(X^T X)^- \mathrm{X}^T)\, Var(\mathrm{Y})[\mathrm{d}^T - \mathrm{c}^T(X^T X)^- X^T]^T$$

$$= [\mathrm{c}^T(X^T X)^- \mathrm{X}^T]\sigma^2 I[\mathrm{d} - \mathrm{X}\underline{(\mathrm{X}^T X)^-}\mathrm{c}]$$

↑ This is where the symmetry of $(X^T X)^-$ is needed.

$$= \sigma^2[\mathrm{c}^T(X^T X)^- \underline{X^T\mathrm{d}} - \mathrm{c}^T\underline{(X^T X)^- X^T X (X^T X)^-}\mathrm{c}]$$

↑ since $X^T\mathrm{d} = \mathrm{c}$

Since $\mathrm{c}^T\mathrm{b}$ is invariant to the choice of b (result 3.8.(iii)), we were able to use the Moore-Penrose inverse for $(X^T X)^-$ which satisfies

$$(X^T X)^-(X^T X)(X^T X)^- = X^T X$$

by definition.

Then,

$$Cov(\mathrm{c}^T\mathrm{b}, \mathrm{d}^T\mathrm{Y} - \mathrm{c}^T b)$$

$$= \sigma^2[\mathrm{c}^T(X^T X)^-\mathrm{c} - \mathrm{c}^T(X^T X)^-\mathrm{c}]$$

$$= 0$$

Consequently,

$$Var(\mathrm{d}^T\mathrm{Y}) \geq Var(\mathrm{c}^T\mathrm{b})$$

and $\mathrm{c}^T\mathrm{b}$ is blue.

(iv) To show that the OLS estimator is the unique blue, note that

$$\begin{aligned} Var(\mathrm{d}^T\mathrm{Y}) &= Var(\mathrm{c}^T\mathrm{b} + [\mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b}]) \\ &= Var(\mathrm{c}^T\mathrm{b}) + Var(\mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b}) \end{aligned}$$

because $Cov(\mathrm{c}^T\mathrm{b}, \mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b}) = 0$.

Then, $\mathrm{d}^T\mathrm{Y}$ is blue if and only if

$$Var(\mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b}) = 0 \ .$$

This is equivalent to

$$\mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b} = \text{ constant.}$$

Since both estimators are unbiased

$$\begin{aligned} E(\mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b}) &= E(\mathrm{d}^T\mathrm{Y}) - E(\mathrm{c}^T\mathrm{b}) \\ &= 0. \end{aligned}$$

Consequently, $\mathrm{d}^T\mathrm{Y} - \mathrm{c}^T\mathrm{b} = 0$ for all Y and $\mathrm{c}^T\mathrm{b}$ is the unique blue.

What if you have a linear model that is <u>not</u> a Gauss-Markov model?

$$\begin{aligned} E(\mathrm{Y}) &= X\beta \\ Var(\mathrm{Y}) &= \Sigma \neq \sigma^2 I \end{aligned}$$

- Parts (i) and (ii) of the proof of result 3.11 do not require

  $$Var(\mathrm{Y}) = \sigma^2 I \ .$$

  Consequently, the OLS estimator for $\mathrm{c}^T\beta$,

  $$\mathrm{c}^T\mathrm{b} = \mathrm{c}^T(X^TX)^-X^T\mathrm{Y}$$

  is a linear unbiased estimator.

- Result 3.8 does not require

  $$Var(\mathrm{Y}) = \sigma^2 I$$

  and the OLS estimator for any estimable quantity,

  $$\mathrm{c}^T\mathrm{b} = \mathrm{c}^T(X^TX)^-X^T\mathrm{Y}\,,$$

  is invariant to the choice of $(X^TX)^-$.

- The OLS estimator $\mathrm{c}^T\mathrm{b}$ may not be blue. There may be other linear unbiased estimators with smaller variance.

Variance of the OLS estimator of an estimable quantity:

$$Var(\mathbf{c}^T\mathbf{b}) = Var(\mathbf{c}^T(X^TX)^-X^T\mathbf{Y})$$
$$= \mathbf{c}^T(X^TX)^-X^T\Sigma X[(X^TX)^-]^T\mathbf{c}$$

For the Gauss-Markov model

$$Var(\mathbf{Y}) = \Sigma = \sigma^2 I$$

and

$$Var(\mathbf{c}^T\mathbf{b})$$
$$= \sigma^2\mathbf{c}^T(X^TX)^-X^TX[(X^TX)^-]^T\mathbf{c}$$
$$= \sigma^2\mathbf{c}^T(X^TX)^-\mathbf{c}$$

Generalized Least Squares (GLS) Estimation:

Defn 3.8: For a linear model with

$$E(\mathbf{Y}) = X\beta \quad \text{and} \quad Var(\mathbf{Y}) = \Sigma,$$

where Σ is positive definite, a generalized least squares estimator for β minimizes

$$(\mathbf{Y} - X\mathbf{b}_{GLS})^T\Sigma^{-1}(\mathbf{Y} - X\mathbf{b}_{GLS})$$

Strategy: Transform $\mathbf{Y}$ to a random vector $\mathbf{Z}$ for which the Gauss-Markov model applies.

The spectral decomposition of Σ yields

$$\Sigma = \sum_{j=1}^{n} \lambda_j \mathbf{u}_j\mathbf{u}_j^T \ .$$

Define

$$\Sigma^{-1/2} = \sum_{j=1}^{n} \frac{1}{\sqrt{\lambda_j}}\mathbf{u}_j\mathbf{u}_j^T$$

and create the random vector

$$\mathbf{Z} = \Sigma^{-1/2}\mathbf{Y}.$$

Then

$$\begin{aligned} Var(\mathbf{Z}) &= Var(\Sigma^{-1/2}\mathbf{Y}) \\ &= \Sigma^{-1/2}\Sigma\,\Sigma^{-1/2} = I \end{aligned}$$

and

$$\begin{aligned} E(\mathbf{Z}) &= E(\Sigma^{-1/2}\mathbf{Y}) \\ &= \Sigma^{-1/2}E(\mathbf{Y}) = \Sigma^{-1/2}X\beta \\ &= W\beta \end{aligned}$$

and we have a Gauss-Markov model for $\mathbf{Z}$, where $W = \Sigma^{-1/2}X$ is the model matrix.

Note that

$$(\mathrm{Z} - \boldsymbol{W}\mathrm{b})^T(\mathrm{Z} - \boldsymbol{W}\mathrm{b})$$

$$= (\Sigma^{-1/2}\mathrm{Y} - \Sigma^{1/2}\boldsymbol{X}\mathrm{b})^T$$
$$(\Sigma^{-1/2}\mathrm{Y}\Sigma^{-1/2}\boldsymbol{X}\mathrm{b})$$

$$= (\boldsymbol{Y} - \boldsymbol{X}\mathrm{b})^T\Sigma^{-1/2}\Sigma^{-1/2}(\boldsymbol{Y} - \boldsymbol{X}\mathrm{b})$$

$$= (\boldsymbol{Y} - \boldsymbol{X}\mathrm{b})^T\Sigma^{-1}(\boldsymbol{Y} - \boldsymbol{X}\mathrm{b})$$

Hence, any GLS estimator for the Y model is an OLS estimator for the Z model.

It must be a solution to the normal equations for the Z model

$$\boldsymbol{W}^T\boldsymbol{W}\mathrm{b} = \boldsymbol{W}^T\mathrm{Z}$$
$$\Leftrightarrow (\boldsymbol{X}^T\Sigma^{-1/2}\Sigma^{-1/2}\boldsymbol{X})\mathrm{b}$$
$$= \boldsymbol{X}^T\Sigma^{-1/2}\Sigma^{-1/2}\mathrm{Y}$$
$$\Leftrightarrow (\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})\mathrm{b} = \boldsymbol{X}^T\Sigma^{-1}\mathrm{Y}$$

These are the generalized least squares estimating equations.

Any solution

$$\begin{aligned} \mathrm{b}_{\mathrm{GLS}} &= (\boldsymbol{W}^T\boldsymbol{W})^-\boldsymbol{W}^T\mathrm{Z} \\ &= (\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})^-\boldsymbol{X}^T\Sigma^{-1}\mathrm{Y} \end{aligned}$$

is called a generalized least squares (GLS) estimator for β.

<u>Result 3.11</u> For a linear model with $E(\mathrm{Y}) = \boldsymbol{X}\beta$ and $Var(\mathrm{Y}) = \Sigma$, the GLS estimator of an estimable function $\mathrm{c}^T\beta$,

$$\mathrm{c}^T\mathrm{b}_{\mathrm{GLS}} = \mathrm{c}^T(\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})^-\boldsymbol{X}^T\Sigma^{-1}\mathrm{Y} \ ,$$

is the unique blue of $\mathrm{c}^T\beta$.

<u>Proof:</u> Since $\mathrm{c}^T\beta$ is estimable, there is an a such that

$$\begin{aligned} \mathrm{c}^T\beta &= E(\mathrm{a}^T\mathrm{Y}) \\ &= E(\mathrm{a}^T\Sigma^{1/2}\Sigma^{-1/2}\mathrm{Y}) \\ &= E(\mathrm{a}^T\Sigma^{1/2}\mathrm{Z}) \end{aligned}$$

Consequently, $\mathrm{c}^T\beta$ is estimable for the Z model. Apply the Gauss-Markov theorem (result 3.10) to the Z model.

Comments

- For the linear model with

$$E(Y) = X\beta \text{ and } Var(Y) = \Sigma,$$

both the OLS and GLS estimators for an estimable function $c^T\beta$ are linear unbiased estimators.

$Var(c^T b_{OLS}) =$
$c^T(X^TX)^- X^T \Sigma X[(X^TX)^-]^T c$

$Var(c^T b_{GLS}) =$
$c^T(X^T\Sigma^{-1}X)^- X^T\Sigma^{-1}X(X^T\Sigma^{-1}X)^- c$

and

$$Var(c^T b_{OLS}) \geq Var(c^T b_{GLS})$$

- For the Gauss-Markov model,

$$c^T b_{GLS} = c^T b_{OLS} \ .$$

- The blue propoerty of $c^T b_{GLS}$ assumes that $Var(Y) = \Sigma$ is known.
- The same results, including Results 3.12, hold for the Aitken model where $E(Y) = X\beta$ and $Var(Y) = \sigma^2 V$ for some known matrix V.

- In practice $Var(Y) = \Sigma$ is usually unknown. An approximation to

$$b_{GLS} = (X^T\Sigma^{-1}X)^- X^T\Sigma^{-1}Y$$

is obtained by substituting a consistent estimator $\hat{\Sigma}$ for Σ.
  - use method of moments or maximum likelihood estimation to obtain $\hat{\Sigma}$

- the resulting estimator
  - is not a linear estimator
  - is consistent but not necessarily unbiased
  - does not provide a blue for estimable functions
  - may have larger mean squared error than the OLS estimator