Theorem 2.9. \( a^\top \beta \) is estimable if and only if any one of the following conditions.

(a) \( a \in C(X^\top) \) [note: \( C(X^\top) = C(X^\top X) \)]

(b) \( a^\top (X^\top X) X = a^\top \)

(c) \( E[a^\top \hat{\beta}] = a^\top \beta \), where \( \hat{\beta} \) is an LSE.

**Proof:**

(a) Assume that \( a \in C(X^\top) \). Then \( \exists \ c \ s.t. \ a = X^\top c \).

Consider \( c^\top Y \). \( E[c^\top Y] = c^\top X \beta = a^\top \beta \). By the definition of estimable, \( a^\top \beta \) is estimable.

On the other hand, if \( a^\top \beta \) is estimable, \( \exists \ c \ s.t. \ a^\top \beta = E[c^\top Y] = c^\top X \beta \) for all \( \beta \).

So \( a = X^\top c \), i.e. \( a \in C(X^\top) \).

(b) If \( a^\top (X^\top X) X = a^\top \), then

\[
E[a^\top (X^\top X)^\top Y] = a^\top (X^\top X) X \beta = a^\top \beta \Rightarrow a^\top \beta \text{ is estimable.}
\]

If \( a^\top \beta \) is estimable, by (a), \( a = X^\top d \); therefore

\[
a^\top (X^\top X)^\top X = d^\top X^\top (X^\top X)^\top X = d^\top X^\top X = a^\top
\]

(c) If \( E[a^\top \hat{\beta}] = a^\top \beta \) for all \( \beta \), then

\[
a^\top (X^\top X) X \beta = a^\top \beta \text{ for all } \beta,
\]

which implies that \( a^\top (X^\top X)^\top X = a^\top \). By (b), \( a^\top \beta \) is estimable.

If \( a^\top \beta \) is estimable, by (b), \( \exists \ d \ s.t. \ a = X^\top d \).

So \( E[a^\top \hat{\beta}] = d^\top X^\top (X^\top X)^\top X \beta = d^\top X^\top X \beta = a^\top \beta \frac{X^\top X}{X^\top X} \) (by def of g-inv)
Example: \( Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \ i = 1, 2, j = 1, 2; \ E[\epsilon_{ij}] = 0 \)

\[
Y = \begin{pmatrix}
Y_{11} \\
Y_{12} \\
Y_{21} \\
Y_{22}
\end{pmatrix}, \quad X = \begin{pmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
0 & 1 & 1
\end{pmatrix}, \quad \text{rank}(X) = 2 < 3, \quad \beta = \begin{pmatrix}
\mu \\
\alpha_1 \\
\alpha_2
\end{pmatrix}
\]

\[
X^T X = \begin{pmatrix}
4 & 2 & 2 \\
2 & 2 & 0 \\
2 & 0 & 2
\end{pmatrix}
\]

Consider two generalized inverses:

\[
(X^T X)_1 = \begin{pmatrix}
0 & 0 & 0 \\
0 & 1/2 & 0 \\
0 & 0 & 1/2
\end{pmatrix}, \quad (X^T X)_2 = \begin{pmatrix}
1/2 & -1/2 & 0 \\
-1/2 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

\[
X^T Y = \begin{pmatrix}
Y_{1.} \\
Y_{2.}
\end{pmatrix}, \quad \hat{\beta}_1 = (X^T X)_1 X^T Y = \begin{pmatrix}
0 \\
Y_{1.} \\
Y_{2.}
\end{pmatrix}, \quad \hat{\beta}_2 = (X^T X)_2 X^T Y = \begin{pmatrix}
Y_{1.} \\
Y_{2.} - Y_{1.}
\end{pmatrix}
\]

Claim: \( \mu, \ alpha_1, \alpha_2, \alpha_1 + \alpha_2 \) are not estimable, \( \mu \) and \( \alpha_1 \) and \( \alpha_2 \) are estimable.
$M + a_1 = E[Y_{11}]$, so $M + a_1$ is estimable
$X_1 - X_2 = E[Y_{11} - Y_{21}]$, so $X_1 - X_2$ is estimable.

How to show that a function is not estimable?

E.g. $u = a^T \beta$ where $a = (1, 0, 0)^T$

Suppose that $u$ is estimable. By Theorem 2.9,

$a \in C(X)$, i.e., $\exists d$ s.t.

$$
a = \begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix} = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0
\end{pmatrix} \begin{pmatrix}
d_1 \\
d_2 \\
d_3 \\
d_4
\end{pmatrix} = \begin{pmatrix}
d_1 + d_2 + d_3 + d_4 \\
d_1 + d_2 \\
d_3 + d_4
\end{pmatrix}
$$

Row 2 $\Rightarrow d_1 + d_2 = 0$ $\Rightarrow d_1 + d_2 + d_3 + d_4 = 0$
Row 3 $\Rightarrow d_3 + d_4 = 0$ $\Rightarrow d_1 + d_2 + d_3 + d_4 = 1$

But row 1 $\Rightarrow d_1 + d_2 + d_3 + d_4 = 1$.

The contradiction implies that there is no $d$ s.t. $a = x^T d$, i.e. $a \notin C(X)$. By Theorem 2.9, $u$ is not estimable.

Example: consider the model:

$$Y_{ij} = \alpha_i + \beta_j + \varepsilon_{ij},$$

where $i = 1, \ldots, a$, $j = 1, \ldots, b$, $\varepsilon_{ij} \sim (0, \sigma^2)$.

Derive the necessary and sufficient condition for $\sum C_i \alpha_i + \sum D_j \beta_j$ to be estimable.
The hat matrix

\[ P_x = X(X^TX)^{-1}X^T \]

We have proved that
\[ P_x^2 = P_x \]
\[ P_x^T = P_x \]
\[ P_x = \frac{1}{S} P_x \]

\( P_x \) is a projection matrix.

We mentioned that \( P_x \) is invariant to the choice of \( (X^TX)^{-1} \). Here is the proof:

Let \( A \) and \( B \) be two different g-inverses

\[ (XAX^T - XBX^T)(XAX^T - XBX^T)^T \]
\[ = (XAX^T - XBX^T)(XAX^T - XBX^T)^T \]
\[ = (XAX^T - XBX^T)X(A^TX - B^TX) \]
\[ = (XAX^T - XBX^T)(A^TX - B^TX) \]
\[ = (X - X)(A^TX - B^TX) \]
\[ = 0 \]

which implies that \( XAX^T = XBX^T \), i.e., \( P_x \) is invariant to the choice of \( (X^TX)^{-1} \).
Theorem 2.10 Let \( \hat{\beta} \) be an LSE of \( \beta \) from regression model \( Y=X\beta+E \) with \( E[\beta]=0 \). Consider an estimable function \( a^T\beta \), then

1. \( a^T\hat{\beta} \) is unique, i.e., \( a^T\hat{\beta} \) is invariant to the choice of \( (X^TX)^{-1} \).

2. \( E[a^T\hat{\beta}]=a^T\beta \). When \( X \) is full rank, \( \beta \) is estimable and \( E[\hat{\beta}]=\beta \).

3. \( \text{Var}[a^T\hat{\beta}]=0^2 a^T(X^TX)^{-1} a \), if we further assume \( \text{Var}[\varepsilon]=0^2 I_n \). When \( X \) is full rank, \( \text{Var}[\hat{\beta}]=0^2 (X^TX)^{-1} \).

Proof: Based on the assumptions, an LSE of \( \beta \) is \( \hat{\beta}=(X^TX)^{-1}X^TY \). Note, when \( X \) is full rank, \( (X^TX)=(X^TX)^{-1} \), we have

\[
a^T\hat{\beta}=a^T(X^TX)^{-1}X^TY
\]

Uniqueness of \( a^T\beta \). Because \( a^T\beta \) is estimable,
\[
a \in C(X^T).
\]
Thus \( \exists c \) s.t. \( a=X^Tc \)

Therefore,
\[
a^T\hat{\beta}=c^TX(X^TX)^{-1}X^TY=c^TPxY,
\]
which is invariant to the choice of \( (X^TX)^{-1} \); because \( Px=X(X^TX)^{-1}X^T \) is invariant to the choice of \( (X^TX)^{-1} \).

Mean: Because \( a^T\beta \) is estimable
\( a \in \text{EC}(X^T) \), i.e. \( \exists c \) s.t. \( a = c^T \).

Thus, \( E[a^T \hat{\beta}] = c^T X (X^T X)^{-1} X^T E[Y] \)

\[ = c^T X (X^T X)^{-1} X^T \beta \]
\[ = c^T \beta \]
\[ = \hat{a}^T \beta \]

**Variance.** Note that \( \hat{a}^T \beta = a^T (X^T X)^{-1} X^T \beta \)
\[ = c^T (X^T X)^{-1} X^T \beta \]
\[ = c^T \beta \]
\[ = c^T \beta \] Thus,
\[ \text{Var}[\hat{a}^T \beta] = c^T P x \text{Var}[X^T] P x C \]
\[ = c^T P x \sigma^2 I n P x C \]
\[ = \sigma^2 c^T P x C \]
\[ = \sigma^2 c^T X (X^T X)^{-1} X^T \beta \]
\[ = \sigma^2 \hat{a}^T (X^T X)^{-1} a \]

**Summary**

1. The expected value of any observation is estimable.
2. Any linear combination of estimable functions is estimable.
3. The function \( a^T \beta \) is estimable iff
   - (a) \( a \in \text{EC}(X^T) \)
   - (b) \( a^T = a^T (X^T X)^{-1} X^T \)
   - (c) \( a^T \beta \) is unbiased for \( a^T \beta \).
4. \( X \beta \) is estimable, and any linear function of \( X \beta \) is estimable.
5. $X^T X \beta$ is estimable, and any linear function of $X^T X \beta$ is estimable.

6. There are exactly $r = \text{rank}(X)$ linearly independent estimable functions of $\beta$. 
2.6 Optimality of LSE

Theorem 2.11 (Gauss-Markov) Let \( \hat{\beta} \) be an LSE from linear regression model \( Y = X\beta + \varepsilon \) with \( \varepsilon \sim N(0, \sigma^2 I_n) \) [i.e., \( E[\varepsilon] = 0, \text{Var}[\varepsilon] = \sigma^2 I_n \)]. Consider an estimable function \( a^T \beta \). Then \( a^T \hat{\beta} \) is the unique estimate with minimum variance among the class of unbiased linear estimators of \( a^T \beta \), i.e., the LSE is the best linear unbiased estimator.

Proof

Consider any linear unbiased estimator \( b^T Y \).
Since \( b^T Y \) is unbiased,
\[
a^T \beta = b^T E[Y] = b^T X \beta \quad \text{for all } \beta,
\]
which implies that \( b^T X = a^T \).

\[
\text{Var}[b^T Y] = \text{Var}[b^T Y - a^T \hat{\beta} + a^T \hat{\beta}]
\]
\[
= \text{Var}[b^T Y - b^T \hat{\beta} + b^T \hat{\beta}]
\]
\[
= \text{Var}[b^T Y - b^T \hat{\beta} X + b^T \hat{\beta} X]
\]
\[
= \text{Var}[b^T (I - \hat{X}) Y + b^T \hat{\beta} X]
\]
\[
= \text{Var}[b^T (I - \hat{X}) Y] + \text{Var}[b^T \hat{\beta} X] + 2 \text{Cov}(b^T (I - \hat{X}) Y, b^T \hat{\beta} X)
\]

\( - \gamma \) holds

\[
\iff b^T Y = a^T \hat{\beta}
\]

\[
= \text{Var}[b^T (I - \hat{X}) Y] + \text{Var}[b^T \hat{\beta} X] + 2 b^T (I - \hat{X}) \sigma^2 I_n \hat{X} b
\]
\[
= \text{Var}[b^T (I - \hat{X}) Y] + \text{Var}[b^T \hat{\beta} X] + \frac{\sigma^2}{a^T \beta}
\]
\[
\geq \text{Var}(a^T \hat{\beta})
\]