Estimating Estimable Functions of $\boldsymbol{\beta}$

The Response Depends on $\boldsymbol{\beta}$ Only through $\boldsymbol{X\beta}$

- In the Gauss-Markov or Normal Theory Gauss-Markov Linear Model, the distribution of $\boldsymbol{y}$ depends on $\boldsymbol{\beta}$ only through $\boldsymbol{X\beta}$, i.e.,

$$\boldsymbol{y} \sim (\boldsymbol{X\beta}, \sigma^2\boldsymbol{I}) \quad \text{or} \quad \boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$$

- If $\boldsymbol{X}$ is not of full column rank, there are infinitely many vectors in the set $\{\boldsymbol{b} : \boldsymbol{Xb} = \boldsymbol{X\beta}\}$ for any fixed value of $\boldsymbol{\beta}$.

- Thus, no matter what the value of $\mathrm{E}(\boldsymbol{y})$, there will be infinitely many vectors $\boldsymbol{b}$ such that $\boldsymbol{Xb} = \mathrm{E}(\boldsymbol{y})$ when $\boldsymbol{X}$ is not of full column rank.

- The response vector $\boldsymbol{y}$ can help us learn about $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta}$, but when $\boldsymbol{X}$ is not of full column rank, there is no hope of learning about $\boldsymbol{\beta}$ alone unless additional information about $\boldsymbol{\beta}$ is available.

Treatment Effects Model

Researchers randomly assigned a total of six experimental units to two treatments and measured a response of interest.

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, 2;\ j = 1, 2, 3$$

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \\ \mu + \tau_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \end{bmatrix}$$

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \end{bmatrix}$$

Treatment Effects Model (continued)

- In this case, it makes no sense to estimate $\boldsymbol{\beta} = [\mu, \tau_1, \tau_2]'$ because there are multiple (infinitely many, in fact) choices of $\boldsymbol{\beta}$ that define the same mean for $\boldsymbol{y}$.

- For example,

$$\begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 6 \end{bmatrix}, \text{ or } \begin{bmatrix} 999 \\ -995 \\ -993 \end{bmatrix}$$

  all yield same $\boldsymbol{X\beta} = \mathrm{E}(\boldsymbol{y})$.

- When multiple values for $\boldsymbol{\beta}$ define the same $\mathrm{E}(\boldsymbol{y})$, we say that $\boldsymbol{\beta}$ is *non-estimable*.

Estimable Functions of β

- A linear function of β, $\boldsymbol{C\beta}$, is said to be *estimable* if there is a linear function of $\boldsymbol{y}$, $\boldsymbol{Ay}$, that is an unbiased estimator of $\boldsymbol{C\beta}$. Otherwise, $\boldsymbol{C\beta}$ is said to be *non-estimable*.

- Note that $\boldsymbol{Ay}$ is an unbiased estimator of $\boldsymbol{C\beta}$ if and only if

$$\begin{aligned} \mathrm{E}(\boldsymbol{Ay}) = \boldsymbol{C\beta} \ \forall\ \boldsymbol{\beta} \in \mathbb{R}^{\mathrm{p}} \quad &\iff \quad \boldsymbol{AX\beta} = \boldsymbol{C\beta} \ \forall\ \boldsymbol{\beta} \in \mathbb{R}^{p} \\ &\iff \quad \boldsymbol{AX} = \boldsymbol{C}. \end{aligned}$$

- This says that we can estimate $\boldsymbol{C\beta}$ as long as $\boldsymbol{C\beta} = \boldsymbol{AX\beta} = \boldsymbol{A}\mathrm{E}(\boldsymbol{y})$ for some $\boldsymbol{A}$, i.e., as long as $\boldsymbol{C\beta}$ is a linear function of $\mathrm{E}(\boldsymbol{y})$.

- The bottom line is that we can always estimate $\mathrm{E}(\boldsymbol{y})$ and all linear functions of $\mathrm{E}(\boldsymbol{y})$; all other linear functions of β are non-estimable.

Treatment Effects Model (continued)

$$E(\boldsymbol{y}) = \boldsymbol{X\beta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \\ \mu + \tau_2 \end{bmatrix} \Longrightarrow$$

$$\begin{aligned} [1,0,0,0,0,0]\boldsymbol{X\beta} &= [1,1,0]\boldsymbol{\beta} = \mu + \tau_1 \\ [0,0,0,1,0,0]\boldsymbol{X\beta} &= [1,0,1]\boldsymbol{\beta} = \mu + \tau_2 \\ [1,0,0,-1,0,0]\boldsymbol{X\beta} &= [0,1,-1]\boldsymbol{\beta} = \tau_1 - \tau_2 \end{aligned}$$

are estimable functions of β.

Estimating Estimable Functions of β

- If $\boldsymbol{C}\boldsymbol{\beta}$ is estimable, then there exists a matrix $\boldsymbol{A}$ such that $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{X}$ and $\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{A}\mathrm{E}(\boldsymbol{y})$ for any $\boldsymbol{\beta} \in \mathbb{R}^p$.

- It makes sense to estimate $\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{A}\mathrm{E}(\boldsymbol{y})$ by

$$\begin{aligned}
\boldsymbol{A}\widehat{\mathrm{E}(\boldsymbol{y})} &= \boldsymbol{A}\hat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} \\
&= \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{C}\hat{\boldsymbol{\beta}}.
\end{aligned}$$

- $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ is called the Ordinary Least Squares (OLS) estimator of $\boldsymbol{C}\boldsymbol{\beta}$.

- Note that although the "hat" is on $\boldsymbol{\beta}$, it is $\boldsymbol{C}\boldsymbol{\beta}$ that we are estimating.

Invariance of $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ to the Choice of $\hat{\boldsymbol{\beta}}$

- Although there are infinitely many solutions to the normal equations when $\boldsymbol{X}$ is not of full column rank, $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ is the same for all normal equation solutions $\hat{\boldsymbol{\beta}}$ whenever $\boldsymbol{C}\boldsymbol{\beta}$ is estimable.

- To see this, suppose $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are any two solutions to the normal equations. Then

$$\begin{aligned}
\boldsymbol{C}\hat{\boldsymbol{\beta}}_1 &= \boldsymbol{A}\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 = \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 \\
&= \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 = \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{y} \\
&= \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 = \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 \\
&= \boldsymbol{A}\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 = \boldsymbol{C}\hat{\boldsymbol{\beta}}_2.
\end{aligned}$$

Treatment Effects Model (continued)

- Suppose our aim is to estimate $\tau_1 - \tau_2$.
- As noted before,

$$\boldsymbol{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \\ \mu + \tau_2 \end{bmatrix} \Longrightarrow$$

$$[1, 0, 0, -1, 0, 0]\boldsymbol{X}\boldsymbol{\beta} = [0, 1, -1]\boldsymbol{\beta} = \tau_1 - \tau_2.$$

- Thus, we can compute the OLS estimator of $\tau_1 - \tau_2$ as

$$[1, 0, 0, -1, 0, 0]\hat{\boldsymbol{y}} = [0, 1, -1]\hat{\boldsymbol{\beta}},$$

where $\hat{\boldsymbol{y}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$ and $\hat{\boldsymbol{\beta}}$ is any solution to the normal equations.

Treatment Effects Model (continued)

The normal equations in this case are

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}' \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}' \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix}$$

$$\iff \begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} y_{\cdot\cdot} \\ y_{1\cdot} \\ y_{2\cdot} \end{bmatrix}.$$

Treatment Effects Model (continued)

$$\hat{\boldsymbol{\beta}}_1 \equiv \begin{bmatrix} \bar{y}_{\cdot\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot} \end{bmatrix} \text{ and } \hat{\boldsymbol{\beta}}_2 \equiv \begin{bmatrix} 0 \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix}$$

are each solutions to the normal equations because

$$\begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} \bar{y}_{\cdot\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot} \end{bmatrix} = \begin{bmatrix} y_{\cdot\cdot} \\ y_{1\cdot} \\ y_{2\cdot} \end{bmatrix} = \begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix}.$$

Thus, the OLS estimator of $\boldsymbol{C\beta} = [0, 1, -1]\boldsymbol{\beta} = \tau_1 - \tau_2$ is

$$\boldsymbol{C}\hat{\boldsymbol{\beta}}_1 = [0, 1, -1] \begin{bmatrix} \bar{y}_{\cdot\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot} \end{bmatrix} = \bar{y}_{1\cdot} - \bar{y}_{2\cdot} = [0, 1, -1] \begin{bmatrix} 0 \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix} = \boldsymbol{C}\hat{\boldsymbol{\beta}}_2.$$

Treatment Effects Model (continued)

Let $(\boldsymbol{X}'\boldsymbol{X})_1^- = \begin{bmatrix} 1/6 & 0 & 0 \\ 0 & 1/6 & -1/6 \\ 0 & -1/6 & 1/6 \end{bmatrix}$ and $(\boldsymbol{X}'\boldsymbol{X})_2^- = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$.

- It is straightforward to verify that $(\boldsymbol{X}'\boldsymbol{X})_1^-$ and $(\boldsymbol{X}'\boldsymbol{X})_2^-$ are each generalized inverses of $\boldsymbol{X}'\boldsymbol{X}$.
- It is also easy to show that $\hat{\beta}_1 = (\boldsymbol{X}'\boldsymbol{X})_1^- \boldsymbol{X}'\boldsymbol{y}$ and $\hat{\beta}_2 = (\boldsymbol{X}'\boldsymbol{X})_2^- \boldsymbol{X}'\boldsymbol{y}$.

Treatment Effects Model (continued)

$$\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}'$$

$$= \begin{bmatrix} 0 & 1/3 & 0 \\ 0 & 1/3 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \\ 0 & 0 & 1/3 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Treatment Effects Model (continued)

Thus

$$\widehat{\mathrm{E}(\boldsymbol{y})} = \hat{\boldsymbol{y}} = \boldsymbol{P_X y} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix}$$

is our OLS estimator of

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \\ \mu + \tau_2 \end{bmatrix}.$$

Treatment Effects Model (continued)

Also, we can see that the OLS estimator of

$$\tau_1 - \tau_2 = [0, 1, -1] \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} = [1, 0, 0, -1, 0, 0] \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix}$$

$$= [1, 0, 0, -1, 0, 0] \begin{bmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \\ \mu + \tau_2 \end{bmatrix} = [1, 0, 0, -1, 0, 0]\mathrm{E}(\boldsymbol{y}) \text{ is}$$

Treatment Effects Model (continued)

$$
\begin{aligned}
[1,0,0,-1,0,0]\widehat{\mathrm{E}(\boldsymbol{y})} &= [1,0,0,-1,0,0]\hat{\boldsymbol{y}} \\
&= [1,0,0,-1,0,0]\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \end{bmatrix} \\
&= \bar{y}_{1\cdot} - \bar{y}_{2\cdot}
\end{aligned}
$$

The Gauss-Markov Theorem

Under the Gauss-Markov Linear Model, the OLS estimator $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ of an estimable linear function $\boldsymbol{c}'\boldsymbol{\beta}$ is the unique *Best Linear Unbiased Estimator* (BLUE) in the sense that $\text{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}})$ is strictly less than the variance of any other linear unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$ and all $\sigma^2 \in \mathbb{R}^+$.

- The Gauss-Markov Theorem says that if we want to estimate an estimable linear function $\boldsymbol{c}'\boldsymbol{\beta}$ using a linear estimator that is unbiased, we should always use the OLS estimator.

- In our simple example of the treatment effects model, we could have used $y_{11} - y_{21}$ to estimate $\tau_1 - \tau_2$. It is easy to see that $y_{11} - y_{21}$ is a linear estimator that is unbiased for $\tau_1 - \tau_2$, but its variance is clearly larger than the variance of the OLS estimator $\bar{y}_{1\cdot} - \bar{y}_{2\cdot}$ (as guaranteed by the Gauss-Markov Theorem).

