Estimable Functions and Their Least Squares Estimators

Consider the GLM

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times p}{\boldsymbol{X}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}, \quad \text{where} \quad E(\boldsymbol{\varepsilon}) = \mathbf{0}.$$

Suppose we wish to estimate $\boldsymbol{c}'\boldsymbol{\beta}$ for some fixed and known $\boldsymbol{c} \in \mathbb{R}^p$.

An estimator $t(\boldsymbol{y})$ is an unbiased estimator of the function $\boldsymbol{c}'\boldsymbol{\beta}$ iff

$$E[t(\boldsymbol{y})] = \boldsymbol{c}'\boldsymbol{\beta} \quad \forall\, \boldsymbol{\beta} \in \mathbb{R}^p.$$

An estimator $t(\boldsymbol{y})$ is a linear estimator in $\boldsymbol{y}$ iff

$$t(\boldsymbol{y}) = d + \boldsymbol{a}'\boldsymbol{y}$$

for some known constants $d, a_1, \ldots, a_n$.

A function $\boldsymbol{c}'\boldsymbol{\beta}$ is linearly estimable iff $\exists$ a linear estimator that is an unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$.

Henceforth, we will use estimable as a synonym for linearly estimable.

A function $\boldsymbol{c}'\boldsymbol{\beta}$ is said to be nonestimable if there does not exist a linear estimator that is an unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$.

Result 3.1:

Under the GLM, $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable iff the following equivalent conditions hold:

(i) $\exists\, \boldsymbol{a} \ni E(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{c}'\boldsymbol{\beta} \quad \forall\, \boldsymbol{\beta} \in \mathbb{R}^p$

(ii) $\exists\, \boldsymbol{a} \ni \boldsymbol{c}' = \boldsymbol{a}'\boldsymbol{X} \quad (\boldsymbol{X}'\boldsymbol{a} = \boldsymbol{c})$

(iii) $\boldsymbol{c} \in \mathcal{C}(\boldsymbol{X}')$.

Show conditions (i), (ii), and (iii) are equivalent.

Show that any of the equivalent conditions is equivalent to $\boldsymbol{c}'\boldsymbol{\beta}$ estimable.

Example:

Suppose that when team i competes against team j, the expected margin of victory for team i over team j is $\mu_i - \mu_j$, where $\mu_1, \ldots, \mu_5$ are unknown parameters.

Suppose we observe the following outcomes.

Team	1	beats Team	2	by	7	points
	3		1		3	
	3		2		14	
	3		5		17	
	4		5		10	
	4		1		1	

Determine $\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}$.

Is $\mu_1 - \mu_2$ is estimable?

Is $\mu_1 - \mu_3$ is estimable?

Is $\mu_1 - \mu_5$ is estimable?

Is μ_1 estimable?

Result 3.1 tells us that $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable iff $\exists\ \boldsymbol{a} \ni \boldsymbol{c}'\boldsymbol{\beta} = \boldsymbol{a}'\boldsymbol{X}\boldsymbol{\beta} \quad \forall\ \boldsymbol{\beta} \in \mathbb{R}^p$.

Recall that $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$.

Thus, $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable iff it is a LC of the elements of $E(\boldsymbol{y})$.

This leads to Method 3.1:

LCs of expected values of observations are estimable.

$\boldsymbol{c}'\boldsymbol{\beta}$ is estimable iff $\boldsymbol{c}'\boldsymbol{\beta}$ is a LC of the elements of $E(\boldsymbol{y})$; i.e.,

$$\boldsymbol{c}'\boldsymbol{\beta} = \sum_{i=1}^{n} a_i E(y_i) \quad \text{for some} \quad a_1, \ldots, a_n.$$

Use Method 3.1 to show that $\mu_2 - \mu_4$ is estimable in our previous example.

Method 3.2:

$\boldsymbol{c}'\boldsymbol{\beta}$ is estimable iff $\boldsymbol{c} \in \mathcal{C}(\boldsymbol{X}')$.

Thus, find a basis for $\mathcal{C}(\boldsymbol{X}')$, say $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$, and determine if

$$\boldsymbol{c} = \sum_{i=1}^{r} d_i \boldsymbol{v}_i \quad \text{for some} \quad d_1, \ldots, d_r.$$

Method 3.3:

By Result A.5, we know that $\mathcal{C}(\boldsymbol{X}')$ and $\mathcal{N}(\boldsymbol{X})$ are orthogonal complements in $\mathbb{R}^p$.

Thus,

$$\boldsymbol{c} \in \mathcal{C}(\boldsymbol{X}') \quad \text{iff} \quad \boldsymbol{c}'\boldsymbol{d} = 0 \quad \forall\, \boldsymbol{d} \in \mathcal{N}(\boldsymbol{X}),$$

which is equivalent to

$$\boldsymbol{X}\boldsymbol{d} = \boldsymbol{0} \Rightarrow \boldsymbol{c}'\boldsymbol{d} = 0.$$

Reconsider our previous example.

Use Method 3.3 to show that μ_1 is nonestimable.

Now use method 3.3 to establish that

$$\boldsymbol{c}'\boldsymbol{\beta} = c_1\mu_1 + c_2\mu_2 + c_3\mu_3 + c_4\mu_4 + c_5\mu_5$$

is estimable iff

$$\sum_{i=1}^{5} c_i = 0.$$

The <u>least squares estimator</u> of an estimable function $\boldsymbol{c}'\boldsymbol{\beta}$ is $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is any solution to the NE $(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y})$.

Result 3.2:

If $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable, then $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is the same for all solutions $\hat{\boldsymbol{\beta}}$ to the NE.

Result 3.3:

The least squares estimator of an estimable function $c'\beta$ is a linear unbiased estimator of $c'\beta$.

Consider again our previous example.

Recall that y_1 is a linear unbiased estimator of $\mu_1 - \mu_2$.

Is this the least squares estimator?

Suppose $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $rank(\underset{n \times p}{\boldsymbol{X}}) = p$.

Show that $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable $\forall\, \boldsymbol{c} \in \mathbb{R}^p$.