Constraints on Solutions to the Normal Equations

If $rank(\underset{n\times p}{\boldsymbol{X}}) = r < p$, there are infinitely many solutions to the NE

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}.$$

For $(\boldsymbol{X}'\boldsymbol{X})^-$ any GI of $\boldsymbol{X}'\boldsymbol{X}$, the set of all solutions is

$$\left\{(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{y} + (\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X})\boldsymbol{z} : \boldsymbol{z} \in \mathbb{R}^p\right\}.$$

It is possible to place additional constraints on a solution to the NE so that $\exists$ a unique solution that satisfies the constraints.

Example:

Suppose

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (i = 1, \ldots, t;\ j = 1, \ldots, n_i)$$

where

$$E(\varepsilon_{ij}) = 0 \quad \forall\ i, j.$$

Consider the following constraints on a solution to the NE $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \vdots \\ \hat{\tau}_2 \end{bmatrix}$.

(Note that constraints are on $\hat{\boldsymbol{\beta}}$ not $\boldsymbol{\beta}$.)

Four common choices are

1. $\hat{\tau}_1 = 0$ (set first to zero)

2. $\hat{\tau}_t = 0$ (set last to zero)

3. $\sum_{i=1}^{t} \hat{\tau}_i = 0$ (sum to zero)

4. $\sum_{i=1}^{t} n_i \hat{\tau}_i = 0$ (weighted sum to zero)

Any one of these constraints may be imposed by insisting that a solution to the NE $\hat{\boldsymbol{\beta}}$ satisfies $\boldsymbol{A}\hat{\boldsymbol{\beta}} = \mathbf{0}$ for some matrix $\boldsymbol{A}$ whose rows define linear constraints on $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \vdots \\ \hat{\tau}_2 \end{bmatrix}$.

For example,

1. $\boldsymbol{A} = [0, 1, 0, \ldots, 0]$.

2. $\boldsymbol{A} = [0, 0, 0, \ldots, 1]$.

3. $\boldsymbol{A} = [0, 1, 1, \ldots, 1]$.

4. $\boldsymbol{A} = [0, n_1, n_2, \ldots, n_t]$.

Note that solution $\hat{\beta}$ satisfies both the NE and the constraint equations $\boldsymbol{A}\hat{\boldsymbol{\beta}} = \mathbf{0}$ iff

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} \\ \boldsymbol{A} \end{bmatrix} \hat{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \mathbf{0} \end{bmatrix}.$$

Furthermore, we know that

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y} \iff \boldsymbol{X}\boldsymbol{b} = \boldsymbol{P_X}\boldsymbol{y}.$$

Thus, we have $\hat{\beta}$ a solution to NE satisfying the constraints iff

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{A} \end{bmatrix} \hat{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{P_X}\boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}.$$

We now know that if we want a unique solution to NE and constraint equations, we need $\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{A} \end{bmatrix}$ to be of full-column rank; that is, we need

$$rank\left(\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{A} \end{bmatrix}\right) = p.$$

$\underset{n\times p}{rank(X)} = r < p$. Thus, we know X has r LI rows.

If we can find $s \equiv p - r \quad p \times 1$ vectors $\ni$, when these vectors are combined with r LI rows of X, the set of p vectors is LI, then we can use the s vectors as rows of A to get

$$rank\left(\begin{bmatrix} X \\ A \end{bmatrix}\right) = p.$$

Suppose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ denote r LI rows of $\boldsymbol{X}$ (written as column vectors) or, equivalently, r LI columns of $\boldsymbol{X}'$.

We seek $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s (s = p - r)$ so that

$$\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_s\}$$

is a set of $p = r + s$ LI vectors in $\mathbb{R}^p$.

Then, with $\boldsymbol{A}' = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s]$, $\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{A} \end{bmatrix}$ will have full-column rank.

Obviously, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s$ must be LI and $\boldsymbol{a}_k \notin \mathcal{C}(\boldsymbol{X}') \; \forall \; k = 1, \ldots, s$.

Show by example that these conditions are not sufficient to guarantee

$$rank\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{A}\end{bmatrix}\right) = p.$$

Prove the following results:

Suppose $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r$ LI in $\mathbb{R}^p$.

Suppose $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s$ LI in $\mathbb{R}^p$.

Suppose $r + s = p$. Then

$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_s \text{ LI}$$
$$\iff span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r\} \cap span\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s\} = \{\boldsymbol{0}\}.$$

Note that the condition

$$span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r\} \cap span\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s\} = \{\boldsymbol{0}\}$$

$\Rightarrow (d_1\boldsymbol{a}_1 + \cdots + d_s\boldsymbol{a}_s)'\boldsymbol{\beta}$ is not estimable whenever

$d_1, \ldots, d_s$ are not all 0.

Thus, the constraints that we add, as well as all nontrivial LCs of those constraints, must correspond to nonestimable functions.

Note that $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s$ with the desired properties always exist. Why is that?

The result that we proved can be alternatively stated as follows:

Suppose $rank(\underset{n\times p}{\boldsymbol{X}}) = r < p$. Suppose $rank(\underset{s\times p}{\boldsymbol{A}}) = s$ where $s = p - r$. Then

$$\mathcal{C}(\boldsymbol{X}') \cap \mathcal{C}(\boldsymbol{A}') = \{\boldsymbol{0}\} \iff rank\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{A}\end{bmatrix}\right) = p.$$

Lemma 3.1:

For $\underset{s\times p}{\boldsymbol{A}} \ni rank(\boldsymbol{A}) = s = p - rank(\underset{n\times p}{\boldsymbol{X}})$ and $\mathcal{C}(\boldsymbol{X}') \cap \mathcal{C}(\boldsymbol{A}') = \{\boldsymbol{0}\}$,

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} \\ \boldsymbol{A} \end{bmatrix} \boldsymbol{b} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}$$ is equivalent to

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} \\ \boldsymbol{A}'\boldsymbol{A} \end{bmatrix} \boldsymbol{b} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix},$$ which is equivalent to

$$(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{A}'\boldsymbol{A})\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}.$$

Result 3.6:

Suppose $rank(\underset{n\times p}{\boldsymbol{X}}) = r, rank(\underset{s\times p}{\boldsymbol{A}}) = s = p - r$, and $\mathcal{C}(\boldsymbol{X}') \cap \mathcal{C}(\boldsymbol{A}') = \{\boldsymbol{0}\}$. Then

(i) $\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{A}'\boldsymbol{A}$ is nonsingular.

(ii) $(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is unique solution to $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$ and $\boldsymbol{A}\boldsymbol{b} = \boldsymbol{0}$.

(iii) $(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{A}'\boldsymbol{A})^{-1}$ is GI of $\boldsymbol{X}'\boldsymbol{X}$.

(iv) $\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{X}' = \boldsymbol{0}$.

(v) $\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}' = \boldsymbol{I}$.

Returning to our example,

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (i = 1, \ldots, t;\ j = 1, \ldots, n_i)$$

where

$$E(\varepsilon_{ij}) = 0 \quad \forall\ i, j.$$

$\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_t \end{bmatrix}$. Let's consider the constraint

$$\sum_{i=1}^{t} n_i \hat{\tau}_i = 0.$$

Find $\hat{\beta}$ that satisfies NE and the constraint.

$$X = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_t} & \mathbf{0}_{n_t} & \mathbf{0}_{n_t} & \cdots & \mathbf{1}_{n_t} \end{bmatrix}$$

$$X'X = \begin{bmatrix} n & n_1 & n_2 & \cdots & n_t \\ n_1 & n_1 & 0 & \cdots & 0 \\ n_2 & 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_t & 0 & 0 & \cdots & n_t \end{bmatrix}.$$

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} n\hat{\mu} + n_1\hat{\tau}_1 + \cdots + n_t\hat{\tau}_t \\ n_1\hat{\mu} + n_1\hat{\tau}_1 \\ \vdots \\ n_t\hat{\mu} + n_t\hat{\tau}_t \end{bmatrix}$$

$$\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} y_{\cdot\cdot} \\ y_{1\cdot} \\ \vdots \\ y_{t\cdot} \end{bmatrix}.$$

Equating $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$ and $\boldsymbol{X}'\boldsymbol{y}$ leads to

$$
\begin{aligned}
n\hat{\mu} + n_1\hat{\tau}_1 + \cdots + n_t\hat{\tau}_t &= y_{\cdot\cdot} \\
n_1\hat{\mu} + n_1\hat{\tau}_1 \qquad\qquad\qquad &= y_{1\cdot} \\
&\vdots \\
n_t\hat{\mu} + n_t\hat{\tau}_t \qquad\qquad\qquad &= y_{t\cdot}.
\end{aligned}
$$

This system of equations has an infinite number of solutions.

However, if we insist that

$$n_1\hat{\tau}_1 + \cdots + n_t\hat{\tau}_t = 0,$$

the first equation becomes

$$n\hat{\mu} = y_{\cdot\cdot} \iff \hat{\mu} = \bar{y}_{\cdot\cdot}.$$

Substituting $\hat{\mu} = \bar{y}_{\cdot\cdot}$ in the other equations yields

$$\begin{aligned} n_i\bar{y}_{\cdot\cdot} + n_i\hat{\tau}_i = y_{i\cdot} &\iff \\ \bar{y}_{\cdot\cdot} + \hat{\tau}_i = \bar{y}_{i\cdot} &\iff \\ \hat{\tau}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot} & \end{aligned}$$

For the general case, we can compute

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$