Reparameterization in Testing

Example:

Suppose

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (i = 1, \ldots, n),$$

where

$$\varepsilon_1, \ldots, \varepsilon_n \overset{i.i.d.}{\sim} N(0, \sigma^2).$$

Consider testing

$$H_0 : \beta_1 = \beta_2.$$

$$H_0 : \beta_1 = \beta_2 \iff H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0},$$

where

$$\boldsymbol{C} = [1, -1], \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

Let

$$\boldsymbol{x}_j = \begin{bmatrix} x_{j1} \\ \vdots \\ x_{jn} \end{bmatrix} \quad \text{for} \quad j = 1, 2.$$

Then

$$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2].$$

Suppose $\begin{bmatrix} 1 \\ -1 \end{bmatrix} \in \mathcal{C}(\boldsymbol{X}')$ so that

$$H_0 : \beta_1 = \beta_2$$

is testable.

Provide the RNE for the restriction imposed by the null hypothesis.

One way to find the BLUE of $\boldsymbol{\beta}$ subject to $\beta_1 = \beta_2$ is to solve these equations.

Can you think of an easier way?

This is a special case of a general reparametrization strategy.

Suppose

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$$

is testable.

The set of al $\boldsymbol{\beta}$ satisfying H_0 is

$$\Theta_0 = \{\boldsymbol{C}^{-}\boldsymbol{d} + (\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})\boldsymbol{\gamma} : \boldsymbol{\gamma} \in \mathbb{R}^p\}.$$

Thus,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\beta}$ is constrained to $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$ is equivalent to

$$\boldsymbol{y} = \boldsymbol{X}[\boldsymbol{C}^{-}\boldsymbol{d} + (\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})\boldsymbol{\gamma}] + \boldsymbol{\varepsilon}, \quad \boldsymbol{\gamma} \in \mathbb{R}^p$$

$$\iff$$

$$\boldsymbol{y} - \boldsymbol{X}\boldsymbol{C}^{-}\boldsymbol{d} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\gamma} \in \mathbb{R}^p.$$

The model

$$\boldsymbol{y} - \boldsymbol{X}\boldsymbol{C}^{-}\boldsymbol{d} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\gamma} \in \mathbb{R}^{p}.$$

is unconstrained with response vector $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{C}^{-}\boldsymbol{d}$ and design matrix $\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})$.

Thus,

$$\hat{\gamma} = [(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})]^{-}(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})'\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{C}^{-}\boldsymbol{d})$$

solves the unconstrained least squares problem

$$\min_{\boldsymbol{\gamma} \in \mathbb{R}^{p}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{C}^{-}\boldsymbol{d} - \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})\boldsymbol{\gamma}\|^{2}.$$

Now

$$\min_{\boldsymbol{\gamma} \in \mathbb{R}^p} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{C}^{-}\boldsymbol{d} - \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})\boldsymbol{\gamma}\|^2$$
$$\iff \min_{\boldsymbol{\gamma} \in \mathbb{R}^p} \|\boldsymbol{y} - \boldsymbol{X}[\boldsymbol{C}^{-}\boldsymbol{d} + (\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})\boldsymbol{\gamma}]\|^2$$
$$\iff \min_{\boldsymbol{\beta} \in \Theta_0} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2.$$

Thus,

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{C}^{-}\boldsymbol{d} + (\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})\hat{\boldsymbol{\gamma}}$$

solves

$$\min_{\boldsymbol{\beta} \in \Theta_0} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2.$$

Show how this works for our simple example.

In practice, when testing

$$H_0 : \boldsymbol{C\beta} = \boldsymbol{d},$$

we don't often explicitly solve the RNE.

It is more common to reparameterize and carry out an unconstrained maximization for the reparameterized model.

We then use

$$\frac{(\text{SSE}_{\text{Reduced}} - \text{SSE}_{\text{Full}}) / (DF_R - DF_F)}{\text{SSE}_{\text{Full}} / DF_F}$$

as our test statistics, where $\text{SSE}_{\text{Reduced}}$ is the SSE from the reparameterized model with

$$DF_R = n - rank(\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})).$$

We have shown that

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{C}^{-}\boldsymbol{d} - \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}^{-}\boldsymbol{C})\gamma\|^2 = \|\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}\|^2.$$

Thus, SSE for the reparameterized model is $Q(\tilde{\boldsymbol{\beta}})$, the SSE for the constrained model.

Furthermore, it can be shown that

$$\begin{aligned} rank(\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{C}^{-}\boldsymbol{C})) &= rank(\boldsymbol{X}) - rank(\boldsymbol{C}) \\ &= r - q. \end{aligned}$$

Thus, DF for the SSE in the parameterized model is

$$n - r + q$$

so that

$$DF_R - DF_F = n - r + q - (n - r) = q.$$

It follows that

$$\frac{\left(\text{SSE}_{\text{Reduced}} - \text{SSE}_{\text{Full}}\right)/(DF_R - DF_F)}{\text{SSE}_{\text{Full}}/DF_F} = \frac{[Q(\tilde{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}})]/q}{Q(\hat{\boldsymbol{\beta}})/(n-r)}.$$

Thus, the reparameterization strategy is yet another way to arrive at the general linear test or, equivalently, the LRT of

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}.$$