1 Finite-sample OLS

Assume our model is $y = X\beta + \epsilon$. We want to minimise

$$\begin{aligned} \epsilon'\epsilon &= (y - X\beta)'(y - X\beta) \\ &= y'y - y'X\beta - \beta'X'y + \beta'X'X\beta \\ &= y'y - 2\beta'X'y + \beta'X'X\beta \end{aligned} \tag{1}$$

The necessary condition for the minimum is that the partial derivative of the above with respect to changes in β is zero. That is,

$$\frac{\partial(\epsilon'\epsilon)}{\partial\beta} = -2X'y + 2X'X\beta = 0 \tag{2}$$

From this we get the set of **normal equations**, where we now denote the optimal value of *beta* with b:

$$\begin{aligned} -2X'y + 2X'Xb = 0 \\ \iff \quad 2X'Xb = 2X'y \\ \iff \quad X'Xb = X'y \end{aligned} \tag{3}$$

We can solve for b:

$$b = (X'X)^{-1}X'y \tag{4}$$

If we assume that our inference problem can be formulated as $y = X\beta + \epsilon$ and that ϵ are independently and identically distributed with $\mathbb{E}\,\epsilon = 0$, then β is an unbiased estimator, because

$$\begin{aligned} \mathbb{E}\,b &= \mathbb{E}[(X'X)^{-1}X'y] \\ &= \mathbb{E}[(X'X)^{-1}X'(X\beta + \epsilon)] \\ &= \mathbb{E}[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon] \\ &= \mathbb{E}\,\beta + \mathbb{E}[(X'X)^{-1}X']\,\mathbb{E}\,\epsilon \\ &= \mathbb{E}\,\beta + \mathbb{E}[(X'X)^{-1}X'] \cdot 0 \\ &= \mathbb{E}\,\beta \\ &= \beta \end{aligned} \tag{5}$$

Its variance is

$$\begin{aligned} \mathbb{V}\,b &= \mathbb{E}[(b - \beta)(b - \beta)'] \\ &= \mathbb{E}[((X'X)^{-1}X'\epsilon)((X'X)^{-1}X'\epsilon)'] \\ &= \mathbb{E}[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'\,\mathbb{E}[\epsilon\epsilon']X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\ &= \sigma^2 I(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned} \tag{6}$$

Let's also show that $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is the best unbiased linear estimator for $\boldsymbol{c}'\boldsymbol{\beta}$. The random vector $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is unbiased because

$$\begin{aligned}
\mathbb{E}[\boldsymbol{c}'\hat{\boldsymbol{\beta}}] &= \boldsymbol{c}'\,\mathbb{E}[\hat{\boldsymbol{\beta}}] \\
&= \boldsymbol{c}'\,\mathbb{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}] \\
&= \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathbb{E}[\boldsymbol{y}] \\
&= \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathbb{E}[\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}] \\
&= \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathbb{E}[\boldsymbol{\epsilon}] \\
&= \boldsymbol{c}'\boldsymbol{I}_n\boldsymbol{\beta}+\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\cdot\boldsymbol{0} \\
&= \boldsymbol{c}'\boldsymbol{\beta}
\end{aligned}$$

Let's next show that the variance of the random vector $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is at least as small as that of any other linear unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$. Suppose $\boldsymbol{a}'\boldsymbol{y}$ is an arbitrary linear unbiased estimator of $\boldsymbol{c}'\boldsymbol{\beta}$. Then

$$\begin{aligned}
\boldsymbol{c}'\boldsymbol{\beta} &= \mathbb{E}[\boldsymbol{a}'\boldsymbol{y}] \\
&= \boldsymbol{a}'\,\mathbb{E}[\boldsymbol{y}] \\
&= \boldsymbol{a}'\boldsymbol{X}\boldsymbol{\beta}]
\end{aligned}$$

where it follows that

$$\boldsymbol{c}' = \boldsymbol{a}'\boldsymbol{X}$$

The variance of $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned}
\mathbb{V}[\boldsymbol{c}'\hat{\boldsymbol{\beta}}] &= \boldsymbol{c}'\,\mathbb{V}[\hat{\boldsymbol{\beta}}]\boldsymbol{c} \\
&= \boldsymbol{c}'[\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]\boldsymbol{c} \\
&= \boldsymbol{a}'\boldsymbol{X}[\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]\boldsymbol{X}'\boldsymbol{a} \\
&= \sigma^2\boldsymbol{a}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{a}
\end{aligned}$$

while the variance of $\boldsymbol{a}'\boldsymbol{y}$ is

$$\begin{aligned}
\mathbb{V}[\boldsymbol{a}'\boldsymbol{y}] &= \boldsymbol{a}'(\sigma^2\boldsymbol{I}_n)\boldsymbol{a} \\
&= \sigma^2\boldsymbol{a}'\boldsymbol{a}
\end{aligned}$$

And the difference of the two variances is

$$\begin{aligned}
\mathbb{V}[\boldsymbol{a}'\boldsymbol{y}] - \mathbb{V}[\boldsymbol{c}'\hat{\boldsymbol{\beta}}] &= \sigma^2\boldsymbol{a}'\boldsymbol{a} - \sigma^2\boldsymbol{a}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{a} \\
&= \sigma^2\boldsymbol{a}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{a} \\
&= \sigma^2\boldsymbol{a}'[\boldsymbol{I}_n - \boldsymbol{P}]\boldsymbol{a} \\
&= \sigma^2\boldsymbol{a}'\boldsymbol{M}\boldsymbol{a}
\end{aligned}$$

where $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is familiarly idempotent, meaning its eigenvalues consist solely of zeros and ones. It is also positive semidefinite, and thus

$$\mathbb{V}[\boldsymbol{a}'\boldsymbol{y}] - \mathbb{V}[\boldsymbol{c}'\hat{\boldsymbol{\beta}}] \geq 0$$

1.1 Frisch-Waugh theorem