Text Exercise 2

Jorge Pineño

Theoretical Questions

Let the complete model be given by:

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon,$$

with the OLS estimator from the restricted model:

$$b_R = (X_1'X_1)^{-1}X_1'y,$$

and define:

$$P = (X_1'X_1)^{-1}X_1'X_2.$$

(a) Prove that $\mathbb{E}(b_R) = \beta_1 + P\beta_2$

We begin with the expression for b_R:

$$b_R = (X_1'X_1)^{-1}X_1'y.$$

Substitute $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ into the above:

$$b_R = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon).$$

Distribute:

$$b_R = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon.$$

Taking expectations and using $\mathbb{E}[\varepsilon] = 0$:

$$\mathbb{E}(b_R) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 = \beta_1 + P\beta_2.$$

(b) Prove that $\mathrm{Var}(b_R) = \sigma^2(X_1'X_1)^{-1}$

Using the expression for b_R:

$$b_R = (X_1'X_1)^{-1}X_1'y = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon),$$

we keep ε, the only part that contributes to variance:

$$\mathrm{Var}(b_R) = \mathrm{Var}\left((X_1'X_1)^{-1}X_1'\varepsilon\right).$$

Assuming that $\mathrm{Var}(\varepsilon) = \sigma^2 I$, and knowing that $\mathrm{Var}(A\varepsilon) = A\,\mathrm{Var}(\varepsilon)\,A'$, we get:

$$\mathrm{Var}(b_R) = (X_1'X_1)^{-1}X_1' \cdot \sigma^2 I \cdot X_1(X_1'X_1)^{-1} = \sigma^2(X_1'X_1)^{-1}.$$

(c) Prove that $b_R = b_1 + Pb_2$

Let the OLS estimator from the full model be:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = (X'X)^{-1}X'y, \quad \text{where } X = (X_1\ X_2).$$

We know that:

$$b_R = (X_1'X_1)^{-1}X_1'y.$$

Substituting $y = X_1b_1 + X_2b_2 + \varepsilon$, we get:

$$b_R = (X_1'X_1)^{-1}X_1'(X_1b_1 + X_2b_2 + \varepsilon) = b_1 + Pb_2 + (X_1'X_1)^{-1}X_1'\varepsilon.$$

However, when the full model is estimated using OLS, the residuals are orthogonal to the regressors, so the last term disappears. Therefore:

$$b_R = b_1 + Pb_2.$$

(d) Find the regressions of `Age`, `Educ`, and `Parttime` on `Female`

From Lecture 2.1, the simple regressions of the omitted variables on a constant and `Female` are:

$$\begin{aligned} \text{Age} &= 40.05 - 0.11 \cdot \text{Female} + e, \\ \text{Educ} &= 2.26 - 0.49 \cdot \text{Female} + e, \\ \text{Parttime} &= 0.20 + 0.25 \cdot \text{Female} + e. \end{aligned}$$

Thus, each variable is regressed on a constant and the `Female` dummy. The full matrix P includes both the intercept and the slope coefficients.

(e) Compute the matrix P

Collecting the coefficients from part (d), the matrix P is:

$$P = \begin{pmatrix} 40.05 & 2.26 & 0.20 \\ -0.11 & -0.49 & 0.25 \end{pmatrix}.$$

where:

- The first row corresponds to the intercepts.
- The second row corresponds to the coefficients on `Female`.

(f) Verify numerically that $b_R \approx b_1 + Pb_2$

From the full regression of `log(Wage)` on `Female`, `Age`, `Educ`, and `Parttime`, we have:

$$\begin{aligned} b_1 &= -0.041 \quad \text{(coefficient on Female)} \\ b_2 &= \begin{pmatrix} 0.031 \\ 0.233 \\ -0.365 \end{pmatrix} \quad \text{(coefficients on Age, Educ, Parttime).} \end{aligned}$$

To verify the relation $b_R = b_1 + Pb_2$, we use only the second row of P, which corresponds to the coefficients on `Female`:

$$\begin{pmatrix} -0.11 & -0.49 & 0.25 \end{pmatrix}$$

Thus:

$$\begin{aligned} Pb_2 &= (-0.11)(0.031) + (-0.49)(0.233) + (0.25)(-0.365) \\ &= -0.00341 - 0.11417 - 0.09125 \\ &= -0.20883 \\ b_R &= b_1 + Pb_2 = -0.041 + (-0.20883) = -0.24983 \approx -0.25 \end{aligned}$$

This matches the estimated coefficient from the simple regression of `log(Wage)` on `Female`, which was:

$$b_R = -0.25$$

Thus, the identity $b_R = b_1 + Pb_2$ holds numerically.