Introduction to Econometrics
PS # 5 - Asymptotic Theory + Multivariate Regression Models

1 Large Sample Properties of OLS

In PS#4 *Q*7.3, we relied on the assumption: $u_i \sim \mathcal{N}(0, \sigma^2)$ to run hypothesis testing on the OLS estimators. Combined with the ZCM and homoskedasticity assumption, this implied that $\hat{\beta}_1 \sim \mathcal{N}\Big(\beta_1, V(\hat{\beta}_1|X_i)\Big)$

Suppose that we have a large sample *i.e.*, $n \to \infty$ (Note: An infinitely large sample is not practical). The model of interest is univariate:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

1. Prove the asymptotic normality of $\hat{\beta}_1$
2. Do we still need to rely on the assumption: $u_i \sim \mathcal{N}(0, \sigma^2)$ to conduct hypothesis testing?

2 Omitted Variable Bias

Consider the following model:

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i \quad (1)$$

1. Interpret the regression coefficients.
2. Suppose you omit the $exper_i$ variable, and run the following regression:
$$wage_i = \gamma_0 + \gamma_1 educ_i + z_i \quad (2)$$
   (a) Do you expect the model (2) to suffer from omitted variable bias?
   (b) Explain whether you would overestimate (positive bias) or underestimate (negative bias) the effect of $educ_i$ on $wages_i$
3. (Use **R**) Perform the regression in (1) using the **wage1.csv** dataset. Report and interpret the results.

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.39054    0.76657  -4.423 1.18e-05 ***
educ         0.64427    0.05381  11.974  < 2e-16 ***
exper        0.07010    0.01098   6.385 3.78e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

4. (Use **R**) Find the $Cov(educ_i, exper_i)$. Using this result, what is the expected sign of the bias?
5. (Use **R**) Perform the regression in model (2) using the same dataset. Report and interpret the results.

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.90485    0.68497  -1.321    0.187
educ         0.54136    0.05325  10.167   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

3 Recap: Zero Conditional Mean

Suppose we are interested in the following model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (3)$$

1. Using the ZCM condition: $E(u_i|X_i) = 0$, does this ensure that $\hat{\beta}_1$ is unbiased?
2. Show that the ZCM condition implies the Zero Covariance (ZC) condition *i.e.*, $Cov(u_i, X_i) = 0$
3. Does the ZC condition imply the ZCM condition?

4 Dummy Variable Trap

Consider the following model:

$$wages_i = \beta_0 + \beta_1 tenure_i + \beta_2 female_i + u_i \tag{4}$$

Where

$$female_i = \begin{cases} 1, & \text{Female} \\ 0, & \text{Male} \end{cases}$$

1. (use **R**) Perform the regression in (3) using the **wage1.csv** dataset. Report and interpret the results.

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.13644    0.24006  25.563  < 2e-16 ***
tenure       0.14875    0.02043   7.279 1.24e-12 ***
female      -2.08651    0.29523  -7.067 5.07e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

2. Suppose we decide to run the following regression instead:

$$wages_i = \gamma_0 + \gamma_1 tenure_i + \gamma_2 male_i + e_i \tag{5}$$

Use estimates from the regression in (3) to find $\hat{\gamma_0}$, $\hat{\gamma_1}$ and $\hat{\gamma_2}$

3. (Use **R**) Create a new variable $male_i$, defined as:

$$male_i = \begin{cases} 1, & \text{Male} \\ 0, & \text{Female} \end{cases}$$

Run the regression in (4). Confirm your results from $\mathcal{Q}$.2.

4. Suppose we run a model including both dummy variables, as follows:

$$wages_i = \omega_0 + \omega_1 tenure_i + \omega_2 female_i + \omega_3 male_i + z_i \tag{6}$$

   (a) Can we estimate both ω_1 and ω_2? [Hint: Think about the baseline group]

   (b) (Use **R**) Estimate the model in (5), what do you find?

5. Suppose we run the model without an intercept, as follows:

$$wages_i = \phi_1 tenure_i + \phi_2 female_i + \phi_3 male_i + \epsilon_i \tag{7}$$

   (a) Do you think we can estimate both ϕ_1 and ϕ_2 in this model?

   (b) Using the estimates from model (5) and (6), find the estimates $\hat{\phi_1}$ and $\hat{\phi_2}$.

   (c) (Use **R**) Run the regression in (7) and validate your results.