Applied Econometrics
Prof. Leo Feler
Quiz 1

Name: Key

1. OLS minimizes the sum of squared residuals. Following is your equation:

$$y = X'\beta + \varepsilon,$$

where y, X, β, ε are all in vector form. Minimize the sum of squared residuals to obtain the OLS estimate of β. What's the intuition for the OLS estimate of β?

$$\varepsilon = Y - X\beta$$

$$\varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$$

$$\varepsilon'\varepsilon = (Y' - \beta'X')(Y - X\beta)$$

$$\varepsilon'\varepsilon = Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$$

$$\varepsilon'\varepsilon = Y'Y - 2Y'X\beta + \beta'X'X\beta$$

$$\frac{\partial}{\partial\beta} = 0 - 2X'Y + 2X'X\beta$$

$$0 = -2X'Y + 2X'X\beta$$

$$X'Y = X'X\beta$$

$$(X'X)^{-1}X'Y = (X'X)^{-1}X'X\beta$$

$$\beta = (X'X)^{-1}X'Y = \frac{Cov(\tilde{X}, Y)}{Var(\tilde{X})}$$

where $\tilde{X}$ is the residuals from regressing X_i on all the other X_K.

$X_1 = X_2 \quad X_3 \quad X_4 \quad \ldots$

$\tilde{X}$ = resid

The intuition is that, if our assumptions hold, then these estimates of the true population coefficients will be BLUE; unbiased and with the lowest variance

2. There are three assumptions that need to hold in order for the OLS estimate of β to be unbiased and efficient. What are these three assumptions, and what do they mean in your own words?

1. $\varepsilon_i \sim \text{iid}(0, \sigma^2)$ Expected value of residual is zero, residuals have a constant variance

2. $E(\varepsilon_i | X_{iK}) = 0$ for all K. The residuals are independent of the observables

3. X has full column rank, or no X_i is a linear combination of another X_i.

3. The OLS estimate of β can be biased in the following four cases. Briefly describe each one.

a. Omitted variables.

True: $Y_i = \alpha + \gamma S_i + \theta A_i + \varepsilon_i$

estimated: $Y_i = \tilde{\alpha} + \tilde{\gamma} S_i + \tilde{\varepsilon}_i$

$$E(\tilde{\gamma}) = \gamma + \underbrace{\theta \cdot \frac{Cov(S,A)}{Var(S)}}_{\text{bias}}$$

Direction of bias depends on sign on θ and $Cov(S,A)$

b. Bad controls.

Controls induce bias in our OLS estimate if we have reverse causation or joint causation.

EX/ educ causes wages and type of occupation. We estimate: $wage = \beta_0 + \beta_1 educ + \beta_2 occ + \varepsilon$

Because education jointly determines something on both sides of the equation, part of the effect of educ on wages gets partialled out when we control for one of these channels, occupation, biasing our OLS estimate downwards.

c. Misspecification of functional form.

Misspecification of functional form is a type of omitted variable bias. Even though the variable is not omitted, we are missing a key term that captures the shape of that effect, so that

$$E(\tilde{\gamma}) = \gamma + \underbrace{\theta \cdot \frac{Cov(S,S^2)}{Var(S)}}_{\text{bias}}$$

True: $Y_i = \alpha + \gamma S_i + \theta S_i^2 + \varepsilon_i$

estimated: $Y_i = \tilde{\alpha} + \tilde{\gamma} S_i + \tilde{\varepsilon}_i$

d. Measurement error in y or X.

We are only concerned if the error is in the variable of interest.

If it is in the Y variable, it only matters if the error is correlated with our X_i, which is a form of omitted variable bias that we need to account for by adding in better controls.

If the error is in X, this noise causes attenuation bias, which means our estimates are biased towards zero.

Consider a perfect relationship (corr=1) where wage = 1 · educ

Noise in our estimate of educ will reduce this correlation, giving us, say, wage = .9 · educ

Eventually, so much noise that educ is random will give us a coefficient of zero.