ECON 251

Discussion Section

Week 6 Solutions

1. Review omitted variable bias (OVB) formula, and discusses cases when there is zero bias, upward bias, or downward bias after omitting some variable.

True $\quad Y = \beta_0 + \beta_1 \cdot X + \underbrace{\delta \cdot Z + U}_{=E} \quad \text{Cov}(X, U) = 0$

Naïve $\quad Y = b_0 + b_1 \cdot X + E \quad \text{Cov}(X, E) \neq 0$ whenever $\gamma_1 \neq 0$

Auxiliary $\quad Z = \gamma_0 + \gamma_1 \cdot X + V$

$$\Rightarrow \hat{b}_1^{OLS} := \frac{\sum(X - \overline{X})(Y - \overline{Y})}{\sum(X - \overline{X})^2} = \beta_1 + \delta \cdot \gamma_1$$

where the first equality defines the OLS estimator, and second equality is the OVB formula. Note that there is zero bias when $\delta = 0$ (our naïve model is not so naïve) or when $\gamma_1 = 0$ (the omitted variable is uncorrelated with the treatment). Otherwise, there is bias which we can sign given hypotheses about the parameters δ and γ_1.

There is "positive bias" whenever $\delta > 0, \gamma_1 > 0$ or $\delta < 0, \gamma_1 < 0$ (control and auxiliary effects go in the same direction) and "negative bias" if $\delta > 0, \gamma_1 < 0$ or $\delta < 0, \gamma_1 > 0$ (control and auxiliary effects go in different directions).

2. Compare and contrast the OVB formula to selection bias (SB) in simple comparisons when treatment is not randomly assigned using the potential outcomes framework.

Assume that treatment is binary $X \in \{0,1\}$ and recall that

$$ATE \coloneqq E[Y(1) - Y(0)]$$

$$ATT \coloneqq E[Y(1) - Y(0)|X = 1]$$

$$SB \coloneqq E[Y(0)|X = 1] - E[Y(0)|X = 0]$$

where $Y(1), Y(0)$ are potential outcomes (never simultaneously observed). Then

$$\begin{aligned} \Rightarrow \hat{b}_1^{OLS} &\coloneqq \overline{Y}_1 - \overline{Y}_0 \\ &= ATT + SB \\ &= ATE \end{aligned}$$

where first equality defines the estimator, the second is <u>always</u> true, and the last is <u>only</u> true when treatment is randomly assigned, or the independence assumption

$$X \perp Y(1), Y(0)$$

3. In *Stata*, show some of my JMP regressions. Go through interpretation of simple and multiple linear regression tables. Drawing conclusions from t-statistics and p-values on parameter tests. Example of automatic statistics, versus coding up specific test value. Discuss F-statistic and meaning of testing overall model. Machinery and relationships between standard errors, test statistics, and confidence intervals. (Show my hacky code exploiting this to generate event study graphs!)

4. Functional form issues (and possibilities) with *linear* regression model: logarithm of outcome or treatment, binary variables and groups, quadratic forms, interactions.

   In general, the slope on a simple linear regression gives the average change in the outcome given a unit increase in the treatment, holding all other factors fixed (we hope!) See week 3 discussion notes for why taking the logarithm of either variable, or both, leads to a percentage interpretation of the slope. When the treatment is binary, we get the college choice or gains to migration models that we studied extensively so far. When the outcome is binary, we get the linear probability model (LPM) which means that the slope now gives the average change in the probability the outcome equals 1 given a unit increase in the treatment, holding all other factors fixed (inshallah). Additionally, we can use binary variables to compute differences between groups in the value of their outcomes, perhaps with a linear model in some treatment variable. If we believe (or wish to test whether) the effect of treatment varies by group, we can use interactions between treatment and a set of binary variables. Finally, the quadratic model allows the effect of treatment to vary by its level, intensifying or getting weaker as treatment increases. We will study these functional forms and how to interpret them in more detail as the semester progresses! Regardless of how we specify our model, we still need to contemplate the independence or exogeneity assumptions required to interpret our estimates as true causal effects reflecting the change in outcome given an exogenous manipulation in the value of the treatment.