ECON 251

Discussion Section

Week 10 Solutions

1. In the simple linear regression $Y = \beta_0 + \beta_1 X + U$ we say that the
   - *outcome* is mismeasured if we only observe $Y^* \coloneqq Y + e$ with $E[e|X] = 0$
   - *treatment* is mismeasured if we only see $X^* = X + e$ with $E[e|X] = E[e|U] = 0$

   Derive the Ordinary Least Squares (OLS) estimator $\hat{b}_1^{OLS}$ when we specify the model as $Y^* = b_0 + b_1 X + U$ or $Y = b_0 + b_1 X^* + U$ relative to the true data-generating process, and show[1] that *classical* measurement error matters if it exists in the treatment variable (attenuation bias in slope estimator) but not in the outcome variable (no bias in OLS).

$$\operatorname*{plim}_{N\to\infty} \hat{b}_1^{OLS} \coloneqq \operatorname*{plim}_{N\to\infty} \frac{\widehat{\mathrm{Cov}}(X, Y^*)}{\widehat{\mathrm{Var}}(X)} = \frac{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Cov}}(X, Y^*)}{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Var}}(X)} = \frac{\mathrm{Cov}(X, Y^*)}{\mathrm{Var}(X)}$$

$$= \frac{\mathrm{Cov}(X, Y + e)}{\mathrm{Var}(X)} = \frac{\mathrm{Cov}(X, Y) + \mathrm{Cov}(X, e)}{\mathrm{Var}(X)} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} = \beta_1$$

$$\operatorname*{plim}_{N\to\infty} \hat{b}_1^{OLS} \coloneqq \operatorname*{plim}_{N\to\infty} \frac{\widehat{\mathrm{Cov}}(X^*, Y)}{\widehat{\mathrm{Var}}(X^*)} = \frac{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Cov}}(X^*, Y)}{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Var}}(X^*)} = \frac{\mathrm{Cov}(X^*, Y)}{\mathrm{Var}(X^*)}$$

$$= \frac{\mathrm{Cov}(X + e, Y)}{\mathrm{Var}(X + e)} = \frac{\mathrm{Cov}(X, Y) + \mathrm{Cov}(e, Y)}{\mathrm{Var}(X) + \mathrm{Var}(e)}$$

$$= \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X) + \mathrm{Var}(e)} \cdot \frac{\mathrm{Var}(X)}{\mathrm{Var}(X)} = \beta_1 \cdot \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_e^2} \right)$$

   Note that $\mathrm{Cov}(X, e) = 0$ follows from $E[e|X] = 0$ and the law of iterated expectations, and $\mathrm{Cov}(e, Y) = 0$ follows from $E[e|X] = E[e|U] = 0$.

[1] I use a different method than Q3 in HW3 asks you to use, and I want you to be careful about this distinction! :)

2. Canonical example of Instrumental Variables (IV) = distance to college as an instrument for possibly endogenous education variation in Mincer (1974) log earnings[2] regressions.
    a) Discuss proposed IVs for this equation, and the logic of what we are trying to do.
    b) Prove that the method of moments (MM) estimator for the returns to education is consistent under the IV assumptions: relevance, exogeneity, and exclusion.
    c) Discuss when the asymptotic bias of this estimator of education returns is worse than the MM/OLS estimator under usual exogeneity conditions.

3. Sensitivity analysis = one answer to intellectual nihilism imposed by ubiquitous omitted variables lurking within natural variation where we attempt to discern "truth"
    a) Define sensitivity analysis in terms of estimands (the true slopes we target), estimators, and the assumptions we may believe about the natural variation we use to construct an estimate of the treatment effect.
    b) How do we "know" that smoking tobacco causes *some* increase in the probability of getting lung cancer? (Hint: use genetic arguments to rule out selection bias.)
    c) Why does the sensitivity analysis argument above not apply directly to general regressions specifying relationships between outcomes and treatments?
    d) Describe the new method in Cinelli and Hazlet (2020) to overcome this issue.

[2] See Heckman et al. (2003) for a history of the Mincer regression, theoretical derivation, and some extensions.