ECON 251

Discussion Section

Week 11 Problems

1. We have studied cross-sectional data so far. If we draw a new sample from the population of interest over time, then we have a pooled cross-section dataset (when new units are sampled each period) and a panel or longitudinal dataset (when we track the same units). This means we have two dimensions (units i and time periods t) to consider our research question relating[1] outcome Y_{it} to treatment X_{it} possibly with access to instrument Z_{it}. Discuss the assumptions under which the Random Effects (RE) vs Fixed Effects (FE) estimators consistently estimate the true effect of treatment on the outcome:
    - Random Effects: OLS on $Y_{it} = \theta_t + \beta \cdot X_{it} + U_{it}$ with dummy variables for time periods (possibly interacting with treatment to assess structural change over time)
    - Fixed Effects: OLS on $Y_{it} = \alpha_i + \theta_t + \beta \cdot X_{it} + U_{it}$ with dummy variables for time periods and individual units (if we have a panel) or exogenously defined groups of units (if we have a pooled cross-section).

2. Suppose we have a panel of N units across $T > 1$ periods and estimate β in the two-way fixed effects (TWFE) model $Y_{it} = \alpha_i + \theta_t + \beta \cdot X_{it} + U_{it}$. Why do multiple observations of units over time deal with all time-invariant omitted variables in the error term?

3. Discuss the assumptions needed to consistently estimate β in the two-way fixed effects (TWFE) model $Y_{it} = \alpha_i + \theta_t + \beta \cdot X_{it} + U_{it}$ when $T = 2$ via:
    - OLS on first-differenced variables $\Delta Y_{it} = \Delta\theta_t + \beta \cdot \Delta X_{it} + \Delta U_{it}$
    - OLS on time-demeaned variables $[Y_{it} - \overline{Y}_i] = \beta \cdot [X_{it} - \overline{X}_i] + [U_{it} - \overline{U}_i]$
    - OLS with unit/period dummy variables $Y_{it} = \alpha_i + \theta_t + \beta \cdot X_{it} + U_{it}$

[1] We focus on contemporaneous specifications and ignore more complicated patterns with lagged effects.

4. Compare the probability limit of OLS and IV when the respective exogeneity assumption $\mathrm{Cov}(X,U)=0$ and $\mathrm{Cov}(Z,U)=0$ may fail to be true and the IV relevance assumption $\mathrm{Cov}(Z,X)\neq 0$ is true, but the instrument is weak (so relevance is barely true).

5. Suppose we have access to an instrument Z that is relevant, exogenous, and excluded for treatment X in the linear model $Y=\beta_0+\beta_1 X+U$. Using *Stata*, verify that the slope β_1 in this model can be estimated in three equivalent ways:
   - Ratio of instrument covariance with outcome $\widehat{\mathrm{Cov}}(Z,Y)$ and treatment $\widehat{\mathrm{Cov}}(Z,X)$
   - Ratio of reduced form ($Y=\delta_0+\delta_1 Z+V$) and first stage ($X=\pi_0+\pi_1 Z+W$) slope coefficients estimated via ordinary least squares $\frac{\hat{\delta}_1^{\mathrm{OLS}}}{\hat{\pi}_1^{\mathrm{OLS}}}$
   - Slope of alternative outcome model ($Y=\tilde{\beta}_0+\tilde{\beta}_1\cdot\hat{X}+\widetilde{U}$) where we use predicted treatment $\hat{X}:=\hat{\pi}_0+\hat{\pi}_1 Z$ from first stage regression