Part 8: Treatment Effects

Chris Conlon

March 30, 2020

Applied Econometrics

Intro

Overview

This set of lectures will cover (roughly) the following papers:

Theory:

- Angrist and Imbens (1994)
- Heckman Vytlacil (2005/2007)
- Abadie and Imbens (2006)

And draw heavily upon notes by

- Guido Imbens
- Richard Blundell and Costas Meghir

The Evaluation Problem

- The issue we are concerned about is identifying the effect of a policy or an investment or some individual action on one or more outcomes of interest
- This has become the workhorse approach of the applied microeconomics fields (Public, Labor, etc.)
- Examples may include:
  - The effect of taxes on labor supply
  - The effect of education on wages
  - The effect of incarceration on recidivism
  - The effect of competition between schools on schooling quality
  - The effect of price cap regulation on consumer welfare
  - The effect of indirect taxes on demand
  - The effects of environmental regulation on incomes
  - The effects of labor market regulation and minimum wages on wages and employment

Example: Borjas (1987)

- Consider two countries $(0/1)$ (source and host).

$$\ln w_0 = \alpha_0 + u_0 \quad \text{with } u_0 \sim N(0, \sigma_0^2) \text{ source country}$$
$$\ln w_1 = \alpha_1 + u_1 \quad \text{with } u_1 \sim N(0, \sigma_1^2) \text{ host country}$$

- Now we allow for migration cost of C which he writes in hours: $\pi = \frac{C}{w_0}$.
- Assume workers know everything; you only see u_0 OR u_1 depending on country.
- Correlation in earnings is $\rho = \frac{\sigma_{01}}{\sigma_0 \sigma_1}$.

Example: Borjas (1987)

- Workers will migrate if:

$$(\alpha_1 - \alpha_0 - \pi) + (u_1 - u_0) > 0$$

- Who migrates? Probability of migration. Define $\nu = u_1 - u_0$.

$$\begin{aligned} P = \Pr\left[\nu > (\alpha_0 - \alpha_1 + \pi)\right] &= \Pr\left[\frac{\nu}{\sigma_\nu} > \frac{(\alpha_0 - \alpha_1 + \pi)}{\sigma_\nu}\right] \\ &= 1 - \Phi\left(\frac{(\alpha_0 - \alpha_1 + \pi)}{\sigma_\nu}\right) \equiv 1 - \Phi(z) \end{aligned}$$

- Higher $z \rightarrow$ less migration.

Example: Borjas (1987): How does selection work?

Construct counterfactual wages for workers in source country for those who immigrate:

- For now ignore mean differences $\alpha_0 = \alpha_1 = \alpha$.

$$\begin{aligned} E\left(w_0 \middle| \text{ Immigrate }\right) &= \alpha + E\left(\varepsilon_0 \middle| \frac{\nu}{\sigma_\nu} > z\right) \\ &= \alpha + \sigma_0 E\left(\frac{\varepsilon_0}{\sigma_0} \middle| \frac{\nu}{\sigma_\nu} > z\right) \end{aligned}$$

- Wages depend on:
    1. Mean earnings in the source country
    2. Both error terms (u_0, u_1) through ν
    3. Implicitly, it also depends on the correlation between the error terms.

Example: Borjas (1987): How does selection work?

- If everything is normal, we just run univariate regression $E\left(u_0|\nu\right)=\frac{\sigma_{0\nu}}{\sigma_\nu^2}\nu$:

$$E\left(\frac{u_0}{\sigma_0}|\frac{\nu}{\sigma_\nu}\right)=\frac{1}{\sigma_0}\cdot\frac{\sigma_{0\nu}}{\sigma_\nu^2}\cdot\frac{\sigma_\nu^2}{\sigma_\nu^2}\cdot\nu=\frac{\sigma_{0\nu}}{\sigma_0\sigma_\nu}\frac{\nu}{\sigma_\nu}=\rho_{0\nu}\frac{\nu}{\sigma_\nu}$$

$$\begin{aligned}E\left(w_0|\text{ Immigrate }\right)&=\alpha_0+\sigma_0E\left(\frac{u_0}{\sigma_0}|\frac{\nu}{\sigma_\nu}>z\right)\\&=\alpha_0+\rho_{0\nu}\sigma_0E\left(\frac{\nu}{\sigma_\nu}|\frac{\nu}{\sigma_\nu}>z\right)\\&=\alpha_0+\rho_{0\nu}\sigma_0\left(\frac{\phi(z)}{1-\Phi(z)}\right)\end{aligned}$$

- This hazard rate of the standard normal has a special name Inverse Mills Ratio $E[x|x>z]$.

Example: Borjas (1987): How does selection work?

- A similar expression for those who do immigrate:

$$
\begin{aligned}
E\left(w_1 \middle| \text{ Immigrate } \right) &= \alpha_1 + E\left(u_1 | \frac{\nu}{\sigma_\nu} > z\right) \\
&= \alpha_1 + \rho_{1\nu}\sigma_1 \left(\frac{\phi(z)}{\Phi(-z)}\right)
\end{aligned}
$$

- We can re-write both expressions in terms of the Inverse Mills Ratio

Inverse Mills Ratio

$$
\begin{aligned}
E\left(w_0 \mid \text{ Immigrate }\right) &= \alpha_0 + \rho_{0\nu}\sigma_0\left(\frac{\phi(z)}{1-\Phi(z)}\right) \\
&= \alpha_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\rho - \frac{\sigma_0}{\sigma_1}\right)\left(\frac{\phi(z)}{1-\Phi(z)}\right) \\
E\left(w_1 \mid \text{ Immigrate }\right) &= \alpha_1 + \rho_{1\nu}\sigma_1\left(\frac{\phi(z)}{1-\Phi(z)}\right) \\
&= \alpha_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\frac{\sigma_1}{\sigma_0} - \rho\right)\left(\frac{\phi(z)}{1-\Phi(z)}\right)
\end{aligned}
$$

Where $\rho = \sigma_{01}/\sigma_0\sigma_1$.

Positive Hierarchical Sorting

Let $Q_0 = E\left(u_0|I=1\right), Q_1 = E\left(u_1|I=1\right)$ (expected skill of immigrants).

- Immigrants are positively selected and above average $(Q_0, Q_1) > 0$ and $\frac{\sigma_1}{\sigma_0} > 1$ and $\rho > \frac{\sigma_0}{\sigma_1}$
  - $\frac{\sigma_1}{\sigma_0} > 1$ returns to "skill" are higher in host country.
  - $\rho > \frac{\sigma_0}{\sigma_1}$ correlation between valued skills in both counties is high (similar skills valued in both countries).
- Best and brightest leave because returns to skill are too low in home country.

Negative Hierarchical Sorting

We swap the standard deviations:

- Immigrants are negatively selected and below average $(Q_0, Q_1) < 0$ and $\frac{\sigma_1}{\sigma_0} > 1$ and $\rho > \frac{\sigma_0}{\sigma_1}$
    - $\frac{\sigma_0}{\sigma_1} > 1$ returns to “skill” are lower in host country.
    - $\rho > \frac{\sigma_1}{\sigma_0}$ correlation between valued skills in both counties is high (similar skills valued in both countries).
- Compressed wage structure attracts the low skill types because it provides “insurance” or “subsidizes” low wage workers.

Refugee/Superman Sorting?

- Immigrants are below average at home and above average in host $(Q_0 < 0, Q_1 > 1)$ and $\frac{\sigma_1}{\sigma_0} > 1$:
  - $\rho < \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right)$ being below average in source country makes you above average in host country.
- You are a nerdy intellectual in a country that values physical labor, or are otherwise discriminated against in the labor market.

The missing (fourth) case:

- Mathematically impossible $\rho > \max\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right)$

The Evaluation Problem

- Define an outcome variable Y_i for each individual
- Two potential outcomes for each person $\{Y_i(1), Y_i(0)\}$ depending on whether they receive treatment or not.
- Call $Y_i(1) - Y_i(0) = \beta_i$ the Treatment effect.
- Two major problems:
  - All individuals have different treatment effects (heterogeneity).
  - We don't actually observe any one person's treatment effect ! (Missing Data problem)
- We need strong assumptions in order to recover $f(\beta_i)$ from data.
- Instead we can characterize simpler functions such as $E[\beta_i]$ (ATE) or $E[\beta_i|T_i = 1]$ (ATT) or $E[\beta_i|T_i = 0]$ (ATC) with fewer restrictions.

More Difficulties

What is hard here?

- Heterogeneous effect of β_i in population.
- Selection in treatment may be endogenous. That is T_i depends on $Y_i(1), Y_i(0)$.
- Fisher or Roy (1951) model:

$$Y_i = (Y_i(1) - Y_i(0))T_i + Y_i(0) = \alpha + \beta_i T_i + u_i$$

- Agents usually choose T_i with β_i or u_i in mind.
- Can't necessarily pool across individuals since β_i is not constant.

Structural vs. Reduced Form

- Usually we are interested in one or two parameters of the distribution of β_i (such as the average treatment effect or average treatment on the treated).
- Most program evaluation approaches seek to identify one effect or the other effect. This leads to these as being described as reduced form or quasi-experimental.
- The structural approach attempts to recover the entire joint $f(\beta_i, u_i)$ distribution but generally requires more assumptions, but then we can calculate whatever we need.

Start with Easy Cases

- Let's start with the easy cases: run OLS and see what happens.
- OLS compares mean of treatment group with mean of control group (possibly controlling for other X)

$$\begin{aligned}\beta^{OLS} &= E(Y_i|T_i = 1) - E(Y_i|T_i = 0) \\ &= \underbrace{E[\beta_i|T_i = 1]}_{\text{ATT}} + \left(\underbrace{E[u_i|T_i = 1] - E[u_i|T_i = 0]}_{\text{selection bias}} \right)\end{aligned}$$

- Even in absence of heterogeneity $\beta_i = \beta$ we can still have selection bias.
- $Y_i^0 = \alpha + u_i$ may vary within the population (this is quite common).