Econometrics Discussion Section 2: Experiments

John Green

Spring 2024

Experiments

- Why do economists like *experiments*?
  - Experiments provide *exogeneous variation* via the treatment variable: they solve the endogeneity problem and let us estimate a causal effect (eg the "treatment effect")
  - A "randomized control trial" (RCT) is the gold standard: assign treatment D randomly, and compare $E[Y|D = 1, X] - E[Y|D = 0, X]$ to get the treatment effect of D on Y
- *Potential outcomes* framework identifies for each individual:
  - Y_1: the outcome if treated
  - Y_0: the outcome if untreated
  - The *treatment effect* is $Y_1 - Y_0$
- But of course, for each individual we only observe one outcome! So we need to be clever to estimate this...

Average treatment effect

- Through OLS we can recover the average treatment effect, or ATE
- If we have an RCT then our model is simply:

$$Y = \beta_0 + \beta_1 D_i + \beta_2 X_i \epsilon_i \tag{1}$$

where X_i are control variables and $\hat{\beta}_1 = E[Y|D=1] - E[Y|D=0]$ is the ATE (differences estimator)

- Controlling for X_i is always helpful, but is crucial if the assignment was randomized based on some controls (why?)
- If we expect the effect to also depend on X_i, we need to include interactions

Potential problems

- Issues with the experiment:
  - Imperfect randomization (causes correlation with error term)
  - Partial compliance (use an IV)
  - Attrition (sample selection bias)
  - Experimental effects
- External validity threats:
  - Non-representative sample
  - Non-representative treatment
  - GE effects

Quasi-experiments

- Economic experiments are rare:
    - Expensive
    - Time-consuming
    - Might be unethical
- So, we often look for a *natural experiment* or *quasi-experiment*
- In a natural experiment the treatment is assigned semi-randomly by some mechanism besides an explicit RCT design
    - Treatment is semi-random (minimum wage changes in NJ, not in NY)
    - Some other factor Z influences receipt of treatment D, so is an instrument
    - Example: section 8 housing vouchers

Natural experiment estiamtors

- Diff-in-diff: compare two groups in (at least) two time periods where one group gets a treatment
- IV Estimation
- Regression discontinuity (sharp)
- Fuzzy discontinuity

Big data and machine learning

- "Big data" are just data sets with many (millions) of observations n or many (thousands) of variables k
  - Might even be the case that $k > n$ and OLS is not possible
- Forget about causation and just think about prediction
  - Train a model on *training data* and test it on *test data*

Prediction

- Standardize all our data to be mean 0 and variance 1 so that the intercept drops out, and our model is:

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + U_i$$

- The "oracle prediction" is the best possible predictor, ie that which minimizes the mean squared percentage error on the out-of-sample data:

$$\text{MSPE} = \mathbb{E}\left[Y^{OOS} - \hat{Y}\left(X^{OOS}\right)\right]^2$$

Oracle error

- Our standard OLS model *does not* give us the oracle prediction because we have to estimate our coefficients, which introduce additional uncertainty

$$\text{MSPE}_{OLS} \cong \left(1 + \frac{k}{n}\right) \sigma_u^2$$

  - So MSPE_{OLS} is decreasing in sample size but increasing in predictors
- We get around this by allowing for bias in our coefficient estimates
- Do so with a shrinkage estimator (like James-Stein) which accepts some bias in exchange for higher variance, which gives us a better OOS prediction:

$$\hat{\beta}^{JS} = c\hat{\beta}$$

Bias-variance tradeoff

- In general, we face a tradeoff between bias and variance
    - Remember that OLS is unbiased!
- Flexible model has low bias, but it will be more sensitive to noise and have a higher variance
- If number of regressors is large, then an increase in bias which reduces variance may reduce the MSPE and thus be "worth it"