An Overview of Programme Evaluation Methods

2012 ESRC Research Methods Festival

Adam M. Rosen

UCL, CeMMAP, and IFS

July 4th 2012

Introduction

- We focus on evaluation methods in microeconomic policy analysis.
  - such as welfare, training, wage subsidy, and tax-credit programmes.
- Our goal in this talk is to give an overview of some of these methods, in particular those based on randomized trials and IV methods.
- No single method is universally "best".
- We will highlight
  - the assumptions needed to justify each method.
  - the interpretation of resulting estimates.

The Setup

- We model a population of individuals each of whom either does or does not receive treatment.
  - If individual i receives treatment $d_i = 1$.
  - If individual i does not receive treatment $d_i = 0$.
- Each individual has outcome y_i, e.g. income, employment status, measure of health, ...
- Individuals may also have other characteristics, x_i. We omit these from the model to simplify notation, but all statement could be made conditional on observable x_i.

The Goal

- We observe a random sample of observations (y_i, x_i, d_i), $i = 1, ..., n$.
- From this data we want to learn "the effect" of treatment d_i.
- There are many different way we could think about measuring this effect.
- In practice we may also wish to consider how treatment affects people with different x_i differently.
- Here I will draw primarily on material from Manski (1989, 1995).

The Challenge: A Missing Data Problem

- The main challenge is that for each individual we only observe their outcome for the treatment they received.
- If individual i receives treatment, we observe the corresponding outcome, but not the outcome s/he would have had without treatment.
- We thus use y_i^0 and y_i^1 to denote individual i's *potential* outcomes from non-receipt and receipt of treatment, respectively.
- The *realized* outcome y_i corresponds to treatment received, d_i:

$$y_i = d_i y_i^1 + (1 - d_i)\, y_i^0.$$

- The other potential outcome is unobserved, and is referred to as the *counterfactual* outcome.

The Challenge: A Missing Data Problem

- The individual treatment effect is the difference in potential outcomes $\alpha_i = y_i^1 - y_i^0$.
- Since we only observe either y_i^1 or y_i^0, we don't know α_i.
- Typically we seek to learn certain features of the joint distribution of α_i in the population, such as:
  - The Average Treatment Effect (ATE): $E\left[y_i^1 - y_i^0\right]$.
  - The Average Treatment Effect on the Treated: $E\left[y_i^1 - y_i^0 | d_i = 1\right]$.
  - The Average Treatment Effect on the Non-Treated: $E\left[y_i^1 - y_i^0 | d_i = 0\right]$.
  - and others...

The Average Treatment Effect

- Consider first the conditional ATE.

$$E\left[y_i^1 - y_i^0\right] = E\left[y_i^1\right] - E\left[y_i^0\right]$$

- Using the Law of Iterated Expectations $E\left[y_i^1\right]$, $E\left[y_i^0\right]$ can be decomposed as

$$\begin{aligned} E\left[y_i^1\right] &= E\left[y_i^1 | d_i = 1\right] \times p + E\left[y_i^1 | d_i = 0\right] \times (1-p), \\ E\left[y_i^0\right] &= E\left[y_i^0 | d_i = 1\right] \times p + E\left[y_i^0 | d_i = 0\right] \times (1-p). \end{aligned}$$

- Consider first $E\left[y_i^1\right]$. When $d_i = 1$, the observed outcome $y_i = y_i^1$.

The Average Treatment Effect

- Therefore $E\left[y_i^1 | d_i = 1\right] = E\left[y_i | d_i = 1\right]$.
- Similarly $E\left[y_i^0 | d_i = 0\right] = E\left[y_i | d_i = 0\right]$.
- Each can thus be consistently estimated by sample counterparts.
- However, $E\left[y_i^1 | d_i = 0\right]$ and $E\left[y_i^0 | d_i = 1\right]$ are means of counterfactual outcomes.
- To recover credible estimates for these, we need to bring in additional information.
    - Such information can be used to justify approaches such as matching, diff-in-diffs, and instrumental variable methods.

"Worst-Case" Bounds

- First, let's see what we get without addition further information/assumptions.
- Consider again $E\left[y_i^1\right] = E\left[y_i^1 | d_i = 1\right] \times p + E\left[y_i^1 | d_i = 0\right] \times (1-p)$.
- Suppose that outcomes can only take values in the unit interval $[0,1]$.
- Then $E\left[y_i^1 | d_i = 0\right]$ must also lie in $[0,1]$.
- We can use this to bound $E\left[y_i^1\right]$:

$$E\left[y_i | d_i = 1\right] \times p \leq E\left[y_i^1\right] \leq E\left[y_i | d_i = 1\right] \times p + (1-p).$$

"Worst-Case" Bounds

- The same reasoning can be used to bound $E\left[y_i^0\right]$.
- This in turn leads to bounds on the ATE:

$$LB \leq E\left[y_i^1 - y_i^0\right] \leq UB$$

- Where

$$\begin{aligned} LB &= E\left[y_i | d_i = 1\right] \times p - \left(E\left[y_i | d_i = 0\right] \times (1-p) + p\right), \\ UB &= E\left[y_i | d_i = 1\right] \times p + (1-p) - E\left[y_i | d_i = 0\right] \times (1-p). \end{aligned}$$

- The bounds are easy to estimate given estimates for $E\left[y_i | d_i = 1\right]$, $E\left[y_i | d_i = 0\right]$, p.
- The worst-case bounds provide a useful starting point but will always cross zero. That is, they can never distinguish the sign of the ATE..

"Worst-Case" Bound Estimates

- Bound estimates $\widehat{LB}$ and $\widehat{UB}$ can be constructed using the previous formulas with

$$\hat{p} = n^{-1} \sum_{i=1}^{n} d_i,$$

$$\hat{E}_n \left[y_i | d_i = 1 \right] = \left(\hat{p} n \right)^{-1} \sum_{i=1}^{n} y_i d_i,$$

$$\hat{E}_n \left[y_i | d_i = 0 \right] = \left(\left(1 - \hat{p} \right) n \right)^{-1} \sum_{i=1}^{n} y_i \left(1 - d_i \right).$$

Random Assignment

- When treatment is randomly assigned the problem is easily solved.
- Random assignment typically means potential outcomes $\left(y_i^1, y_i^0\right)$ are independent of treatment assignment d_i.
- This implies $E\left[y_i^t|d_i = 1\right] = E\left[y_i^t|d_i = 0\right]$ for either $t = 0, 1$, which is all that is required for the ATE.
- This may be credible in experimental setups, for example in clinical trials or certain experimental cash transfer programmes.
- It is a very powerful assumption, and thus very useful when credible, but rarely holds with data in economics.
- There can for example be problems such as selection or attrition.

Random Assignment

- Suppose indeed that assignment is random so that $E\left[y_i^t | d_i = 1\right] = E\left[y_i^t | d_i = 0\right]$.
- Consider again $E\left[y_i^1\right]$:

$$E\left[y_i^1\right] = E\left[y_i | d_i = 1\right] \times p + E\left[y_i^1 | d_i = 0\right] \times (1-p) .$$

- Under random assignment $E\left[y_i^1 | d_i = 0\right] = E\left[y_i^1 | d_i = 1\right]$, so that

$$\begin{aligned} E\left[y_i^1\right] &= E\left[y_i | d_i = 1\right] \times p + E\left[y_i^1 | d_i = 1\right] \times (1-p) \\ &= E\left[y_i | d_i = 1\right] \times p + E\left[y_i | d_i = 1\right] \times (1-p) \\ &= E\left[y_i | d_i = 1\right] \end{aligned}$$

- Thus under this assumption $\hat{E}_n\left[y_i | d_i = 1\right]$ provides a consistent estimator for $E\left[y_i^1\right]$.
- Likewise $\hat{E}_n\left[y_i | d_i = 0\right]$ now provides a consistent estimator for $E\left[y_i^0\right]$ and we can use

$$\widehat{ATE} = \hat{E}_n\left[y_i | d_i = 1\right] - \hat{E}_n\left[y_i | d_i = 0\right] .$$

Introduction of Instruments

- As previously mentioned random assignment does not hold in most settings.
- An alternative way to improve upon the worst-case analysis is to use instumental variables (IVs).
- There are in fact many different IV assumptions that can be made.
- The general idea is that the instrument
  1. must be correlated with treatment assignment, but
  2. must not otherwise have an effect on outcomes (perhaps conditional on some covariates).
- The IV thus induces *exogenous* variation in assignment.
- IVs are useful, but are sometimes challenging to find.

Introduction of Instruments

- Again, there are many types of IVs.
- To begin suppose we have an instrument z_i such that for any two possible values z^*, z^{**}.
- Thus potential outcomes are mean independent of the realization of z_i.
- Recall the bounds on $E\left[y_i^1\right]$. Applying the same approach we can bound $E\left[y_i^1|z_i = z\right]$ for any value of z:

$$\begin{aligned} E\left[y_i|d_i = 1, z_i = z\right] \times p &\leq E\left[y_i^1|z_i = z\right] \\ &\leq E\left[y_i|d_i = 1, z_i = z\right] \times p + (1-p) . \end{aligned}$$

- By (*) above we know $E\left[y_i^1|z_i = z\right] = E\left[y_i^1\right]$, hence

$$E\left[y_i|d_i = 1, z_i = z\right] \times p \leq E\left[y_i^1\right] \leq E\left[y_i|d_i = 1, z_i = z\right] \times p + (1-p) .$$

Introduction of Instruments

$$E\left[y_i|d_i=1,z_i=z\right]\times p\leq E\left[y_i^1\right]\leq E\left[y_i|d_i=1,z_i=z\right]\times p+(1-p).$$

- The above holds for all possible values of z.
- Each value of z delivers a lower bound and an upper bound for $E\left[y_i^1\right]$ as above:
$$LB^1(z)=E\left[y_i|d_i=1,z_i=z\right]\times p,$$
$$UB^1(z)=E\left[y_i|d_i=1,z_i=z\right]\times p+(1-p).$$
- Since $LB^1(z)\leq E\left[y_i^1\right]\leq UB^1(z)$ for all z it follows that $E\left[y_i^1\right]$ obeys the *intersection bounds*
$$\max_z LB^1(z)\leq E\left[y_i^1\right]\leq\min_z UB^1(z).$$
- Similar reasoning yields intersection bounds for $E\left[y_i^0\right]$:
$$\max_z LB^0(z)\leq E\left[y_i^0\right]\leq\min_z UB^0(z).$$

Introduction of Instruments

- The intersection bounds above can be combined to yield tighter bounds on the ATE than the worst-case bounds.
- How tight the bounds are will depend on how much treatment assignment varies with z, consequently how much z affects $LB^1(z), LB^0(z), UB^1(z), UB^0(z)$.
- In general, without somehow strengthening the IV assumption, this will not yield identification, i.e. there will not be a point estimator, though the bounds can be quite tight.
- It is even possible that they in fact cross, which indicates that the IV assumption is incorrect!