Evaluation of Programs and Policies

DR. ADRIANA KUGLER
CHIEF ECONOMIST
U.S. DEPARTMENT OF LABOR

JUNE 2012

Policy Questions require Causal Answers

- The most interesting questions in terms of policy are causal questions.
- These questions allow us to examine whether the effects of particular interventions or programs or policies are as predicted:
  - Effect of training on employment and earnings,
  - Effect of eligibility assessments on UI claims,
  - Effects of conditional cash transfers on child labor,
  - Effects of experience rating on reduced injuries,
  - Effect of minimum wages on employment.

“What if” Type Questions

- Causal questions involve “What if” type of statements.
- Thus, to answer these causal questions we must be able to construct counter-factual outcomes for each situation – i.e., the outcome had the person being in the alternative situation.
- However, we can never observe the same person in the alternative scenario we are interested in.

Approximating Counter-factuals

- Thus, the best we can do is to approximate or construct a convincing counter-factual.
- Usually, to approximate the alternative scenario or counter-factual, the group of people exposed to a program/policy is contrasted to the group of people not exposed to the program/policy.
- The problem, of course, is that people select or are selected into situations and the two groups of individuals/employers are likely to be different. This generates what we know as “Selection Bias.”

How to deal with Selection Bias?

- Today, I want to talk about various techniques used to construct convincing counter-factuals.
- These techniques, thus, overcome selection bias and allow to more convincingly answer policy questions or questions about the effectiveness of programs than if we simply compared individuals exposed and not exposed to a policy/program.

Impact Evaluation Methods

- There are a number of methods or techniques used to construct counter-factuals, which allow to overcome selection biases and to answer causal questions:

1. Randomized Control Trial (RCTs)
2. Matching Methods
3. Difference-in-differences (Diff-'n-diff)
4. Instrumental Variables (IV's)
5. Regression Discontinuity Designs (RDDs)

Randomized Trials

- The best or ideal way to construct a counter-factual, so that two groups of people are as close as possible under two different policy scenarios/treatments, is to toss a coin and to randomly assign one group to treatment and to leave another group without treatment.

- The ideal way to construct a counter-factual is, thus, similar to the idea of randomly assigning patients when running an experiment on the effectiveness of medical treatments.

Randomization Solves Selection Bias

- Essentially, one can compare the average outcome for those randomly assigned to a program or treatment to the average outcome for those assigned out of the program or to the control group.

- Thus, randomized trials solve the most important problem that arises in empirical research – selection bias, since individuals are otherwise identical in the two groups in terms of their characteristics (observable and unobservable).

Randomization and Regression

- Randomized data can be analyzed using regression analysis by simply running a regression of the outcome on an indicator or dummy variable that indicates if the person was randomly assigned to the program.
- Even in experiments, sometimes the two groups are not, other things equal. That is, the two groups may differ other than just due to their assignment to the program. For example, the average characteristics of the individuals may be different and thus it will make sense to control for these differences in a regression.

Potential Problems with Randomization

1. As mentioned above, there may be potential imbalances in observables in spite of randomization. Solution: control for observables.

2. Differential attrition of those randomly assigned and not assigned to the program. Solution: impute outcomes for those who attrite (Krueger, 1999).

3. Reassignment of individuals to the program. Solution: rely on initial assignment not actual take up of the program.

Matching Methods

- Randomizing is not always possible either because programs are mandatory or because it is not possible to convince the program providers to randomly assign people. Matching is an alternative way to construct the counter-factual.

- Matching methods essentially proceed in two steps:
  1. Pair treated observations with similar non-treated observations in terms of observable characteristics, i.e., find clones!
  2. Then, compare the average outcomes of the treated and paired non-treated observations.

Matching Methods

- The key is how to find the clones, i.e., how to pair or match observations:
  a) Stratification Matching – find individuals with similar characteristics within a range of values.
  b) Nearest Neighbor Matching – find individuals with the closest characteristics.
  c) Radius Matching – find individuals with characteristics which are similar or close "enough" (within a radius).
  d) Kernell Matching – use all individuals as comparisons but give more weight to those close by and less weight to those farther away in terms of characteristics.

Potential Problems with Matching

- Matching requires a <u>common support</u>, i.e., for individuals exposed/not exposed to the program to have the same range of characteristics. However, sometimes, this will not be the case, which means that one will not have matches for many individuals in the program.
- Matching works best when one can control for observable characteristics pre-dating entry into a program.
- Also, matching controls for observable characteristics but does not control for unobservable traits, such as motivation or drive.

Differences-in-differences

- When , the the policy or program varies at a more aggregate level (e.g., health care benefits for pregnant women or minimum wages), the important thing will be to control for unobserved variables at the state or cohort or year level.
- In these cases, it is possible to control for these unobservables by comparing the average outcome for those in the state or cohort affected by the policy to those in states or cohorts not affected by the policy before and after the policy.

Difference-in-differences

- The underlying assumption behind difference-in-differences is that outcome trends would be the same in states/cohorts with the policy and in states/cohorts without the policy in the absence of treatment.

- This is known as the common or parallel trend assumption and it can be tested by simply comparing the trends of those affected and not affected by the policy/program were the same before the policy or program came in effect.

Difference-in-differences and Regression

- As with other methodologies, one can use regression analysis to estimate difference-in-difference effects.

- It is easy to add characteristics and to use individual data:

$$Y_{it} = \gamma_s + \lambda_t + \beta D_s \text{ x } Post_t + \delta X_{ist} + \varepsilon_{ist}$$

- Here the model includes two main effects for states and year and an interaction term that marks observations for the covered states interacted with a Post-policy/program dummy.

Potential Problems with Diff's-'n-Diff's

- Anticipation of policies: If policies are anticipated, then the effect we may be capturing would not be the causal effect of the policy but just the change in behavior due to future expected changes. Can check if a dummy/indicator variable ahead of the policy is significant.
- Endogeneous treatment/control groups: a potential problem comes up if the composition of the those affected by the program/policy changes in response to the policy/program (e.g., migration, pregnancy, labor force participation). Thus, best if use status which cannot be manipulated or status prior to introduction of intervention.

Instrumental Variables

- Like difference-in-differences, instrumental variables help to get at the counter-factual term.
- In the case of instrumental variables, the treatment and counter-factual groups are constructed by looking at individuals who are the same except that some are exposed and some are not exposed to an intervention simply because of some external factor beyond their control (e.g., a natural disaster, date of birth).

Instrumental Variables

- Instrumental variables methods can be used if one has access to a factor called an instrument or instrumental variable such that:
  1. The instrument is correlated with the policy/program of interest.
  2. The instrument is uncorrelated with unobservable factors – this is known as the exclusion restriction.
- IV methods can be estimated using two regressions, where the first regression is a regression of the policy on the instrumental variable and the second is a regression of the outcome on the predicted policy variable.

Potential Problems with IV’s

- The main potential problem with IV’s is to find a truly exogenous or external factor that is not related to unobservables.
- The second problem with IV’s arises if the IV is weakly related to the policy or program of interest. This is known as the weak instrument problem.

Regression Discontinuity Design

- Like other methods I have mentioned so far, regression discontinuity designs try to get at the counter-factual and get rid of selection bias.
- Regression discontinuity or RD gets at the counter-factual and eliminates selection bias by using precise knowledge about the mechanisms or rules determining the treatment.

Regression Discontinuity Design

- There are two types of regression discontinuity designs:

1. Sharp RD – assignment to the program is a deterministic function of a characteristic.

2. Fuzzy RD – the likelihood of being assigned to the program depends on the value of a characteristic.

Regression Discontinuity Design

- A feature of RDD, which contrasts with matching strategies based on treatment-control comparisons conditional on similar characteristics, is that there is no value of the characteristic for which we observe both individuals in and outside of the program.

- Thus, the validity of RDD relies on the willingness to construct counter-factuals on the basis of different though close characteristics.

Potential Problems with RDD

- One potential problem, is that those on one side and the other of the determination rule may not be very similar. A solution to this is to focus on those right to the left and right of the determination rule or what is known as the <u>discontinuity sample</u>.

- Another problem is that one may confuse a continuous change in the outcome around the determination rule with a jump.

Examples of Jumps vs. Non-linearities