Selection on Observables
Lecture 11 - Introduction to Causal Inference

Kevin Li

Controlling for Confounders

In lecture 1, we briefly discussed how one way to identify causal effects is to control/account for all confounding variables.

▶ This is usually not a recommended strategy - it is almost impossible to control for every possible confounder. Many confounders are unobserved and not measurable as well.

However, if no other previous method is possible, we can fall back on controlling confounders - called selection on observables.

▶ Idea of controlling for confounders: if all confounders are accounted for, then treatment is as-if randomly assigned/exogenous (as explained in lecture 1).

▶ Thus, to identify causal effects, we should compare treated vs. control units with the same confounder values.

Assumptions of Selection on Observables

We need to meet two assumptions in order to conduct selection on observables designs.

1. **Conditional Independence/Exogeneity/Ignorability**: basically, you have controlled for all confounders. More technically, it means treatment is as-if randomly assigned when holding confounders constant.

2. **Common Support**: essentially, no matter the confounder values, any unit always has a chance of being assigned to either treatment or control. In other words, under no confounder values are you guaranteed (100%) to be in treatment or control.

Conditional independence is very hard to meet - you need every possible confounder (including ones that cannot be measured).

Regression and Specification Issues

One (simple and common) way to control for confounders is regression:

$$Y_i = \hat{\alpha} + D_i\hat{\tau} + \mathbf{X}_i^\top\hat{\boldsymbol{\beta}} + \hat{\varepsilon}_i$$

Linear regression assumes a linear relationship between independent and dependent variables.

▶ For binary variables (like D_i), this does not matter. But for continuous variables (like many confounders), this can be a problematic assumption. Not all confounders have linear relationships with the outcome.

▶ You might ask? Well we could just add a quadratic/log transformation. Issue - we have to choose what relationship to use. If we choose wrong, our causal estimates will be wrong.

Regression and Heterogeneity Issues

When we hold confounders constant, treatment is exogenous. Thus, $\tau_{ATE}|\mathbf{X}$ (effect given confounders $\mathbf{X}$) is identified. To get the overall τ_{ATE}, we weight all of $\tau_{ATE}|\mathbf{X}$ for each value of $\mathbf{X}$ together:

$$\tau_{ATE} = \underbrace{\int (\tau_{ATE}|\mathbf{X}) \cdot d \Pr(\mathbf{X})}_{\text{weighted average of } \tau|\mathbf{X}}$$

Issue: Angrist (1998) show that OLS estimate of τ equals:

$$\tau_{OLS} = \int (\tau_{ATE}|\mathbf{X}) \cdot d \underbrace{\frac{\text{Var}(D_i|\mathbf{X}_i)\Pr(\mathbf{X}_i)}{\sum_{\mathbf{X}^c}\text{Var}(D_i|\mathbf{X}^c)\Pr(\mathbf{X}^c)}}_{\text{weight}}$$

The weight of τ_{OLS} is clearly not equal to τ_{ATE}. Thus, if there is any treatment heterogeneity, OLS will not estimate the ATE.

Solution to Regression Issues

So OLS regression assumes linear relationships between $\mathbf{X}$ and Y, and also is not accurately estimating the ATE when there is treatment heterogeneity.

How to solve this issue?

1. Use an adjusted regression framework, like **Lin (2013) fully interacted estimator**. This solves the heterogeneity/weighting problem, but does not completely solve the linear relationships problem.

2. Use a non-parametric (not equation/model based) method, like **matching** or **weighting**.

We will focus on matching, the most common and intuitive method in modern causal inference.

Intuition of Matching

$$\tau_i = Y_i(1) - Y_i(0)$$

For treated individuals, $Y_i(0)$ is the unobserved counterfactual. Matching is about estimating $Y_i(0)$.

Selection on observables says we should control for confounders by comparing treated and untreated individuals with the same confounder values.

Solution: Matching. For each treated unit i in our data, find an untreated unit j with the same/similar confounder values.

▶ Use unit j's observed untreated outcome $Y_j(0)$ to estimate the missing $Y_i(0)$ counterfactual of the treated unit.

▶ Then we can estimate τ_i for all treated units. Average them all together, and we get the τ_{ATT}.

Distance (Nearest-Neighbour) Matching

How do we find a untreated unit with "similar" confounder values. **Distance Matching** is a solution to this issue.

1. Graph all units confounder values on a multi-dimensional graph. (ex. 2 confounders, 2 axis, plot all units confounders values).
2. For each treated unit, find the untreated unit whose confounder values is **closest** by multi-dimensional distance. That is the match.

Usually, we use mahalanobis distance (a standardised form of standard euclidean distance).

Issue: curse of dimensionality - basically, the more confounders we have, the less close matches we will have. That means we will be comparing units with different confounder values, causing bias.

Propensity Score Matching

Propensity score matching is about matching units together that have similar likelihoods of being treated, according to their confounder values.

▶ Remember, confounders cause selection into treatment - so we should be able to predict the probability of a unit being treated by their confounder values.

This likelihood of being treated is called a **propensity score** π. They are estimated with a logistic regression:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}$$

Then, we match treated to untreated units based on the closest propensity score estimates $\hat{\pi}$.

Testing Assumptions in Matching

In matching, we can test our conditional independence/exogeneity assumption with a balance test.

▶ These are very similar to the ones in randomised experiments.

After all, once we are done with the matching process, we should have each treated unit matched to a similar untreated unit. Thus, we have a treated group and a control group.

We can then check if key confounder values between these two groups are similar using a regression:

$$X_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$

▶ β_1 is the difference in X_i value between treated and control groups. If it is insignificant, we can conclude the assumption is met.

Other Estimation Methods

1. **Genetic Matching**: basically distance matching, but we also estimate a weights matrix $\mathbf{W}$ that puts different weights on different confounders to achieve the best balance between treatment and control. Performs better than propensity score/distance matching.

2. **Inverse Probability Weighting**: uses propensity scores to weight observations differently to remove selection bias. Uncovers the ATE.

3. **Lin (2013) Regression Adjustment**: Adjusts regression with interactions to solve the heterogeneity issue of OLS. Uncovers the ATE.

4. **Doubly-Robust**: Combines weighting with regression adjustment (hence doubly), shown to be more robust and reliable (also the estimator used in Callaway Sant'Anna DiD estimator).