

Introduction and Outline

The first three parts study a common problem: estimating the average treatment effect on the treated (ATT) when treatment is not randomly assigned. Treated units reveal their outcomes under treatment, but their outcomes without treatment remain unobserved. Each part constructs this missing counterfactual mean from untreated units with comparable pre-treatment characteristics. The sequence follows the distinction between regression, propensity-score methods, and their combinations in Imbens (2004), *Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review*.

Part 1: Outcome Regression

Outcome regression estimates the mean untreated outcome conditional on observed characteristics, then averages these predictions over the treated population. The causal interpretation requires that, conditional on those characteristics, treated and untreated units have the same mean potential outcome without treatment. Overlap requires untreated comparisons wherever treated units occur. These assumptions identify the counterfactual; consistency additionally requires an estimated regression that approaches the relevant conditional mean. The lecture distinguishes identification from estimation and uses a common numerical example to show why the unadjusted difference in group means can be misleading.

Part 2: Inverse Probability Weighting

The propensity score is the probability of receiving treatment conditional on observed characteristics. For the ATT, weighting untreated units by the propensity odds, $p(X)/(1 - p(X))$, reproduces the treated population's covariate composition. This supplies a second representation of the same missing mean under the same identifying assumptions. Whereas outcome regression models untreated outcomes, weighting models treatment assignment. The lecture examines why these weights recover the target, how their sample counterparts are normalized, and why limited overlap produces large weights and imprecise estimates. A correct propensity model is required for consistency of the weighting estimator.

Part 3: Doubly Robust Estimation

Doubly robust estimation combines the regression prediction with a correction formed from weighted control residuals. If the outcome regression is correct, these residuals have conditional mean zero. If the propensity score is correct, the weighted residuals recover the regression's average prediction error for the treated. Consequently, under the identifying assumptions and regularity conditions, the estimator is consistent when either working model is correct. When both are wrong, its bias depends on their joint errors. The lecture derives this correction and applies it to the same example. Double robustness concerns model specification; it does not establish unconfoundedness or supply missing overlap.

Connection to Difference-in-Differences

The doubly robust Callaway–Sant’Anna DiD estimator extends this construction to treatment adopted at different dates. It estimates effects for each treatment cohort and date using never-treated or not-yet-treated comparisons, before aggregating those effects.

Conditional parallel trends supplies the missing untreated change, with overlap and no anticipation also required. The outcome model predicts untreated changes and the propensity model predicts cohort membership within the eligible comparison sample. This connects these first three parts to [Callaway and Sant’Anna \(2021\)](#), using the two-period doubly robust DiD construction of [Sant’Anna and Zhao \(2020\)](#).

For the first three parts, the corresponding reading in Imbens (2004) is Section II on estimands and identification, followed by Sections III.A, III.C, and III.D on regression, propensity-score methods, and mixed methods.