

# Introduction to Matching and Propensity Score Reweighting Estimators in Observational Studies

Justin McCrary  
Berkeley Law, NBER

Conference on Empirical Legal Studies  
November 4, 2011

$Y$ : Outcome of interest

$T$ : Treatment Assignment Indicator, 0=control, 1=treatment

$X$ : Vector of covariates influencing both  $T$  and  $Y$

Sampling schemes: randomly selected persons or groups of persons

$Y$ : Outcome of interest

$T$ : Treatment Assignment Indicator, 0=control, 1=treatment

$X$ : Vector of covariates influencing both  $T$  and  $Y$

Sampling schemes: randomly selected persons or groups of persons

## UNOBSERVED VARIABLES AND CONCEPTUAL OBJECTS

$Y(1)$  and  $Y(0)$ : Counterfactual outcomes

$\tau(X) \equiv \mathbb{E}[Y(1) - Y(0)|X]$ : the conditional treatment effect

$\theta \equiv \mathbb{E}[\tau(X)|T = 1]$ : example target parameter (there are many)

$p(X) \equiv \mathbb{E}[T|X]$ : *propensity score*, or conditional treatment prob.

## OBSERVED DATA

## CONTEXT

$Y$ : Outcome of interest

$T$ : Treatment Assignment Indicator, 0=control, 1=treatment

$X$ : Vector of covariates influencing both  $T$  and  $Y$

Sampling schemes: randomly selected persons or groups of persons

## UNOBSERVED VARIABLES AND CONCEPTUAL OBJECTS

$Y(1)$  and  $Y(0)$ : Counterfactual outcomes

$\tau(X) \equiv \mathbb{E}[Y(1) - Y(0)|X]$ : the conditional treatment effect

$\theta \equiv \mathbb{E}[\tau(X)|T = 1]$ : example target parameter (there are many)

$p(X) \equiv \mathbb{E}[T|X]$ : *propensity score*, or conditional treatment prob.

## GOAL

Estimate  $\theta$  or an analogous parameter such as  $\mathbb{E}[\tau(X)]$ .

## OBSERVED DATA

## CONTEXT

$Y$ : Outcome of interest

$T$ : Treatment Assignment Indicator, 0=control, 1=treatment

$X$ : Vector of covariates influencing both  $T$  and  $Y$

Sampling schemes: randomly selected persons or groups of persons

## UNOBSERVED VARIABLES AND CONCEPTUAL OBJECTS

$Y(1)$  and  $Y(0)$ : Counterfactual outcomes

$\tau(X) \equiv \mathbb{E}[Y(1) - Y(0)|X]$ : the conditional treatment effect

$\theta \equiv \mathbb{E}[\tau(X)|T = 1]$ : example target parameter (there are many)

$p(X) \equiv \mathbb{E}[T|X]$ : *propensity score*, or conditional treatment prob.

## GOAL

Estimate  $\theta$  or an analogous parameter such as  $\mathbb{E}[\tau(X)]$ .

## PROBLEM

Not obvious what is the best way to do so

Many competing methods and many asymptotic theoretic results

However, implications for empirical work not obvious (more later)

## OBSERVED DATA

## CONTEXT

$Y$ : Outcome of interest

$T$ : Treatment Assignment Indicator, 0=control, 1=treatment

$X$ : Vector of covariates influencing both  $T$  and  $Y$

Sampling schemes: randomly selected persons or groups of persons

## UNOBSERVED VARIABLES AND CONCEPTUAL OBJECTS

$Y(1)$  and  $Y(0)$ : Counterfactual outcomes

$\tau(X) \equiv \mathbb{E}[Y(1) - Y(0)|X]$ : the conditional treatment effect

$\theta \equiv \mathbb{E}[\tau(X)|T = 1]$ : example target parameter (there are many)

$p(X) \equiv \mathbb{E}[T|X]$ : *propensity score*, or conditional treatment prob.

## GOAL

Estimate  $\theta$  or an analogous parameter such as  $\mathbb{E}[\tau(X)]$ .

## PROBLEM

Not obvious what is the best way to do so

Many competing methods and many asymptotic theoretic results

However, implications for empirical work not obvious (more later)

## TWO KEY ASSUMPTIONS FOR ALL EXISTING METHODS

Conditional Independence

An Overlap Condition

## OBSERVED DATA

## CONTEXT

$Y$ : Outcome of interest

$T$ : Treatment Assignment Indicator, 0=control, 1=treatment

$X$ : Vector of covariates influencing both  $T$  and  $Y$

Sampling schemes: randomly selected persons or groups of persons

## UNOBSERVED VARIABLES AND CONCEPTUAL OBJECTS

$Y(1)$  and  $Y(0)$ : Counterfactual outcomes

$\tau(X) \equiv \mathbb{E}[Y(1) - Y(0)|X]$ : the conditional treatment effect

$\theta \equiv \mathbb{E}[\tau(X)|T = 1]$ : example target parameter (there are many)

$p(X) \equiv \mathbb{E}[T|X]$ : *propensity score*, or conditional treatment prob.

## GOAL

Estimate  $\theta$  or an analogous parameter such as  $\mathbb{E}[\tau(X)]$ .

## PROBLEM

Not obvious what is the best way to do so

Many competing methods and many asymptotic theoretic results

However, implications for empirical work not obvious (more later)

## TWO KEY ASSUMPTIONS FOR ALL EXISTING METHODS

Conditional Independence

An Overlap Condition

} defined on next slide

## Conditional Independence

- Conditional on covariates, treatment assignment is a *coin flip*
- Formally,  $(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$
- Example: Moving to Opportunity (MTO) Program



## Conditional Independence

- Conditional on covariates, treatment assignment is a *coin flip*
- Formally,  $(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$
- Example: Moving to Opportunity (MTO) Program

## Overlap

- The treatment and control groups have to be “similar enough” that we will have some control and some treatment for each possible configuration of the covariates
- Formally, this assumption comes in 2 primary flavors:

## Conditional Independence

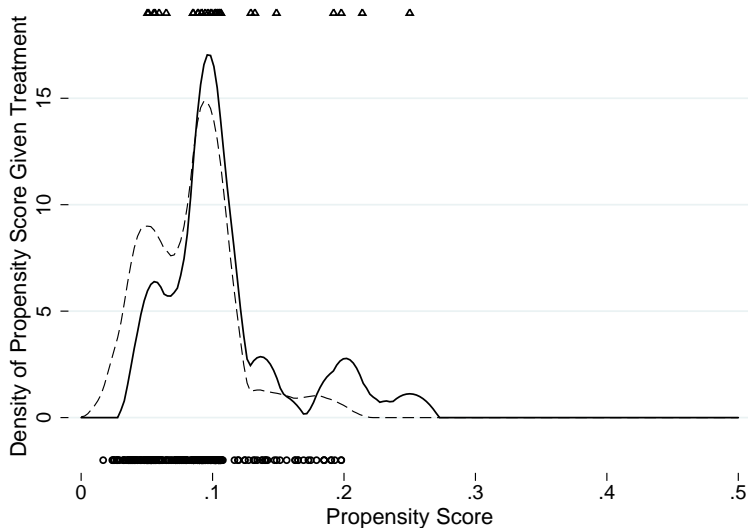
- Conditional on covariates, treatment assignment is a *coin flip*
- Formally,  $(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$
- Example: Moving to Opportunity (MTO) Program

## Overlap

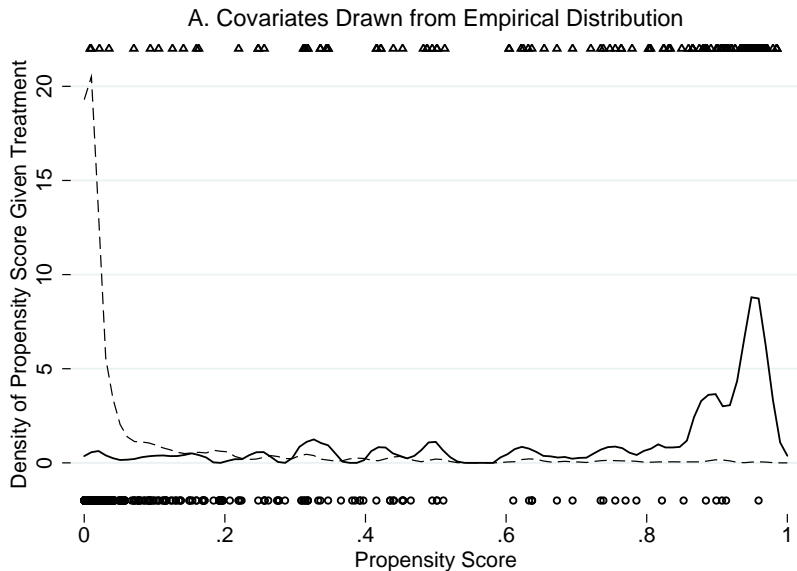
- The treatment and control groups have to be “similar enough” that we will have some control and some treatment for each possible configuration of the covariates
- Formally, this assumption comes in 2 primary flavors:
  1. “Strict Overlap”: There exists a  $c > 0$  such that, for almost every  $x$  in the support of  $X$ ,  $c < p(x) < 1 - c$ .
  2. “Overlap”: Allows  $c = 0$

# Good Overlap: Current Population Survey

## “Effect” of Being African American on Log Wages



# Bad Overlap: National Supported Work Demonstration Effect of Job Training on Annual Earnings



# APPROACHES

- ▶ Regression
  - ▶ Easy to motivate from perspective of constant treatment effect
  - ▶ May involve extrapolations beyond the range of the data
  - ▶ Works well if the relationship between  $Y$  and  $X$  is smooth; less well if the relationship is rough and thus hard to approximate
  - ▶ Research process of settling on a model involves constantly observing the treatment effect estimate you are going to get
- ▶ Matching and Reweighting
  - ▶ Easy to motivate even if treatment effects are heterogeneous
  - ▶ Makes it obvious when we are extrapolating beyond the range of the data
  - ▶ Works well if the relationship between  $T$  and  $X$  is smooth; less well if rough
  - ▶ Research process of settling on a model involves searching for balance in covariates
  - ▶ Commitment device to help avoid (hopefully subconscious) specification searching for the result that is more likely to get published

## APPROACHES: MATCHING

- ▶ Matching on covariates
  - ▶ **Idea:** For each unit  $i$  in the treatment group, find the unit  $j$  in the control group with the “closest” covariate vector. Impute counterfactual outcome  $\hat{Y}(0)_i$  of  $Y_j$ . Then  $Y_i - \hat{Y}(0)_i$  is averaged over the treated observations.
  - ▶ **Metric:** Requires that the user specify a metric for when two vectors are close. *If my first and second variables are highly correlated, that affects my conclusions regarding how “close” two observations are in terms of covariates. (Use Mahalanobis metric.)*
  - ▶ **Curse of Dimensionality:** As the number of covariates increases from 1 to 2 to 3 to ... it becomes *extremely* difficult to find good matches. *Consider a unit sphere inscribed in the unit cube. In 3D, the sphere has volume of about 4.2 and the cube has volume of 8 (ratio of 52%). In 4D, the ratio is 31%. 5D=16%, 6D=8%, and so on. All the mass is eventually “in the corners”.*
- ▶ Matching on the propensity score (Rosenbaum and Rubin 1983)
  - ▶ **Idea:** Avoid the curse of dimensionality and the thorny issue of a metric by matching on the single number  $p(X)$
  - ▶ **Core critique:** Put all of the preceding difficulties into the choice of functional form of the logit model

# APPROACHES: REWEIGHTING

## ► Reweighting

- **Back to basics:** What was the problem in the first place? The treatment and the control group are dissimilar in terms of covariates.

- **Idea:** Reweight the sample so that they are not dissimilar.

- **4 lines of code:**

```
logit T X1 X2 X3
predict pX
gen W=T+(1-T)*pX/(1-pX)
reg Y T [aw=W]
```

- Note: The weight  $W$  you want to apply depends on the parameter you are trying to estimate. These slides assume TOT.
- **Core critique:** Same as matching on the propensity score—the bodies are buried in the logit model.

- ▶ Too many choices! **REMARKS: MATCHING ON  $p(X)$**
- ▶ Pair matching (lowest bias, but highly variable)
  - ▶  $k$ th nearest neighbor matching
  - ▶ Caliper matching
  - ▶ Kernel matching
  - ▶ Local linear matching
  - ▶ Ridge matching
- ▶ Aside from pair matching, all of these approaches require specifying a “tuning parameter”.  $k$ th nearest neighbor matching requires specification of the number of neighbors (i.e.,  $k$ ). Caliper, kernel, and local linear matching all require specifying a “bandwidth”, i.e., an area over which propensity score units are going to be considered “close”. Ridge matching further requires choosing a “ridge parameter”.
- ▶ Please, please, please do NOT be tempted to “cross-validate”
- ▶ Standard errors? For years people used the bootstrap, but Abadie and Imbens (2008) showed it does not work for at least some matching estimators. They recently have been developing an approach that should work, and are even working on software, but this is in progress.



## REMARKS: REWEIGHTING

- ▶ If the default is matching on the propensity score, then we have already settled on the propensity score.
- ▶ **In for a penny, in for a pound?** The justification for matching on the propensity score, rather than on covariates, was that you believe your propensity score model. Reweighting tries to leverage that information.
- ▶ What is the exact formula? **The right formula normalizes the weights. Please, please, please do NOT directly implement the formula given in standard references such as Wooldridge's textbook. For mysterious reasons, econometricians tend to write down a slightly more elegant version of the estimator that performs terribly in practice. See Imbens (2004). BDM (2011) provide a detailed discussion.**
- ▶ **Standard errors?** Quite easy. First, if  $n > 300$  or so, the standard errors from `reg Y T [aw=W]` need no adjustment. Second, if  $n < 300$  (or if you suffer from *FEAR OF PEER REVIEWERS*), there is a tedious but straightforward adjustment (Busso, Dinardo, and McCrary (2011)). Third, the bootstrap works. This is easier.

When Do Cities Shrink the Police Force?

135 Cities, 1960-2010

 $T$  indicates the police force shrank this yearlogit  $T$   $X_1$   $X_2$   $X_3$   $X_4$ ; predict  $pX$ ; gen  $W=T+(1-T)*pX/(1-pX)$ ;mvreg  $X_1$ - $X_4$   $T$ ; mvreg  $X_1$ - $X_4$   $T$  [aw= $W$ ];psmatch2  $T$ , outcome( $X_1$ - $X_4$ ) pscore( $pX$ )

## Assessing Balance

Covariate	Raw	Reweighting	Pair Matching
Murder $_{t-1}$	-0.0375 (0.0074)	0.0004 (0.0070)	-0.0054 (0.0100)
Murder $_{t-2}$	-0.0056 (0.0074)	0.0001 (0.0071)	-0.0007 (0.0101)
Robbery $_{t-1}$	-0.0272 (0.0050)	0.0006 (0.0047)	-0.0028 (0.0068)
Robbery $_{t-2}$	-0.0245 (0.0051)	0.0001 (0.0047)	-0.0003 (0.0069)

## FURTHER REMARKS: EFFICIENCY

- ▶ Huge amount of discussion in the literature regarding the “most efficient estimator”
- ▶ What is an efficient estimator? Intuitively, a method with a small standard error. Formally, an efficient estimator has the smallest possible asymptotic variance in the class of consistent estimators.
- ▶ Turns out that if overlap is good, reweighting nearly always has a smaller asymptotic variance than matching on the propensity score
- ▶ The exceptions are generally pathologies
- ▶ This should not prevent you from estimating reweighting and matching on the same data set
- ▶ What to do when overlap is bad?

## FURTHER REMARKS: OVERLAP

- ▶ When overlap is bad, there is an argument that neither matching nor reweighting will work
- ▶ With bad overlap likely the only “effective” method is regression
  - ▶ regression will forecast beyond range of data
  - ▶ feature, or bug?
- ▶ If you nonetheless are using matching or reweighting in that context, be aware that asymptotic approximations break down and standard errors can be completely unreliable
- ▶ No observational distinction between strict overlap and (regular) overlap
- ▶ When is it safe to go in the water? Not obvious. Best practice is to run a small-scale simulation study to determine the extent to which overlap problems are a first- or second-order problem