

# Conducting Empirical Legal Scholarship: The Advanced Course Day 3

Lee Epstein<sup>1</sup>    Andrew D. Martin<sup>2</sup>

<sup>1</sup>Northwestern University School of Law

<sup>2</sup>Washington University School of Law

February 4-6, 2011

# Outline

- 1 Matching Methods for Causal Inference
- 2 Effectively Communicating Research Results III

# Recall: The Fundamental Problem of Causal Inference

# Some Problems with Regression Analysis

- 1 See Greiner paper
- 2 Lots of Assumptions
- 3 Re-re-re-re estimating models . . .
- 4 Large imbalances in pre-treatment covariates

# Imbalance in Pre-Treatment Covariates

```
. tab t nodegree, col chi
```

```
+-----+
| Key   |
+-----+
|       |
| frequency |
| column percentage |
+-----+
```

| t        | Has Degree   | No Degree     | Total         |
|----------|--------------|---------------|---------------|
| No Entry | 43<br>44.33  | 217<br>62.36  | 260<br>58.43  |
| Entry    | 54<br>55.67  | 131<br>37.64  | 185<br>41.57  |
| Total    | 97<br>100.00 | 348<br>100.00 | 445<br>100.00 |

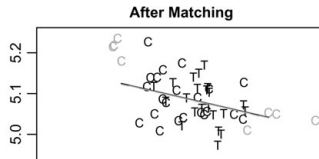
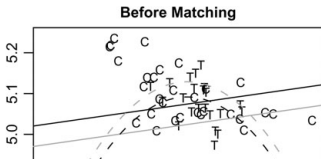
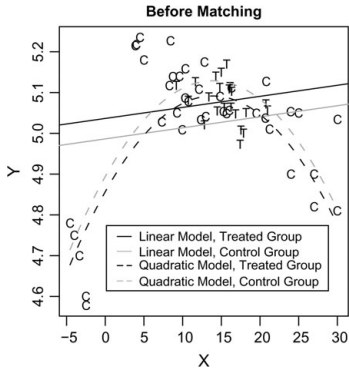
Pearson chi2(1) = 10.1481 Pr = 0.001

**(Title VII) Sex Discrimination**

# The Intuition Behind Matching Methods

- While we may not be able to rerun history, we can match units that are as close as possible **except** on the causal variable of interest.
- Matching is **not** about estimation; it's about preprocessing a dataset so that your estimate of the causal effect will not be model dependent.
  - Preprocessing boils down to "pruning" the dataset to eliminate observations with no "common support" (i.e., no close matches on pre-treatment covariates)

# The Intuition Behind Matching Methods



# Some Types of Matching

## Exact Matching

Match each treated unit (participants) with control units (non-participants) with the same covariate values

## Coarsened Exact Matching (See King, et al.)

Match on "bins" of pre-treatment covariates (e.g., participants and non-participants who attended college rather than on the exact number of years in college).

## Propensity Score Matching

Match on a propensity score (the predicted probability of being treated given the covariates), which acts to balance all the relevant covariates between units that are treated and those that aren't.

# Implementing Propensity Score Matching: The Basics

- Estimate the propensity scores (with the treatment as the response variable), via logit, probit, etc.
- Pick a matching method
- Prune data without common support (i.e., without close matches)
- Match with replacement or not
- Assess balance

# To Stata/psmatch2 Demonstration

# Some Concerns with Propensity Score Matching

- Can't match on unobserved covariates
- Requires large samples
- Requires substantial group overlap

## 1. Acceptable Communication: Create Informative Tables

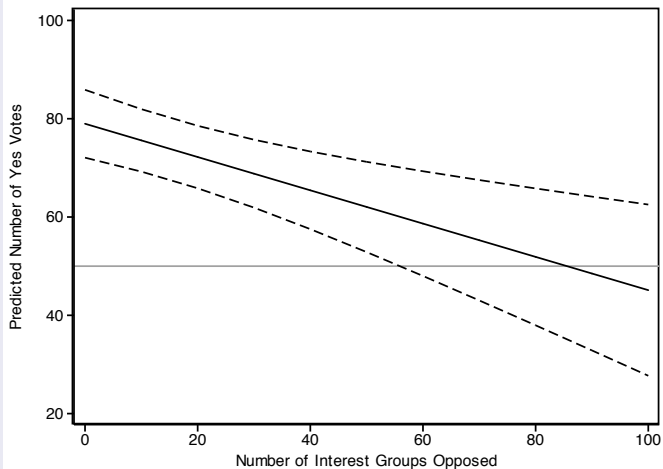
## 2. Good Communication: Stress Substance, Not Statistics

“Under a weak president, a nominee with no interest group opposition will likely receive 79 yea votes. But a nominee opposed by 70 groups can expect only 55 yea votes.”

## 3. Better Communication: Convey Uncertainty

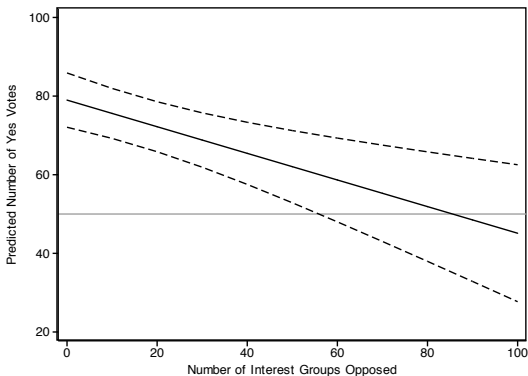
“Under a weak president, a nominee with no interest group opposition will likely receive 79 ( $\pm 7$ ) yea votes. But a nominee opposed by 70 groups can expect only 55 ( $\pm 12$ ) yea votes.”

## 4. Optimal Communication: Graph Substance and Uncertainty Across Values



# Why Graph Substance/Uncertainty Across Values?

“Under a weak president, a nominee with no interest group opposition will likely receive 79 ( $\pm 7$ ) yea votes. But a nominee opposed by 70 groups can expect only 55 ( $\pm 12$ ) yea votes.”



# How to Graph Results/Uncertainty Across Values?

To Stata/Clarify & SPost Demonstration