Selección de Modelos y Regularización

Ciencia de Datos y Econometría Aplicada

Ignacio Sarmiento-Barbieri

Universidad de los Andes

September 7, 2025

Potential outcomes and treatment effects

Potential outcomes and treatment effects

$$$

$43M

Potential outcomes and treatment effects

$43M

Potential outcomes and treatment effects

$43M

$700K

Potential outcomes and treatment effects

$43M
D = 1
Y(1)

$700K
D = 0
Y(0)

Potential outcomes

Potential outcomes and treatment effects

$$$ $

$43M - $700K = $42.3M

Y(1) - Y(0) = Treatment effect

Potential outcomes and treatment effects

$$$ $

\$43M - \$700K = \$42.3M

Y(1) - Y(0) = Treatment effect

counterfactual

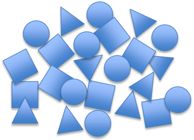
Potential outcomes and treatment effects

$$$ $

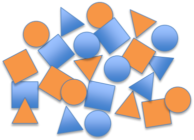
$43M - $700K = $42.3M

Y(1) - Y(0) = Treatment effect

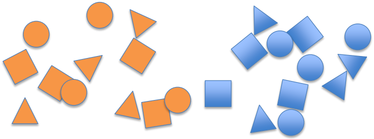
Potential outcomes and treatment effects



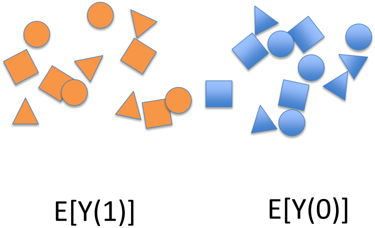
Potential outcomes and treatment effects



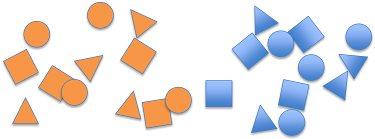
Potential outcomes and treatment effects



Potential outcomes and treatment effects



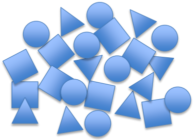
Potential outcomes and treatment effects



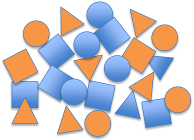
$$E[Y(1)] \quad - \quad E[Y(0)] = E[Y(1) - Y(0)]$$

ATE

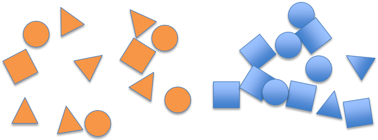
Potential outcomes and treatment effects



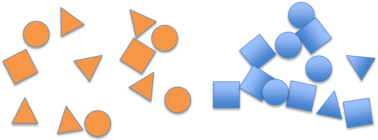
Potential outcomes and treatment effects



Potential outcomes and treatment effects

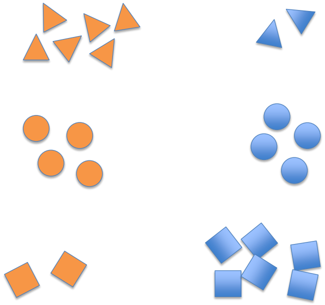


Potential outcomes and treatment effects

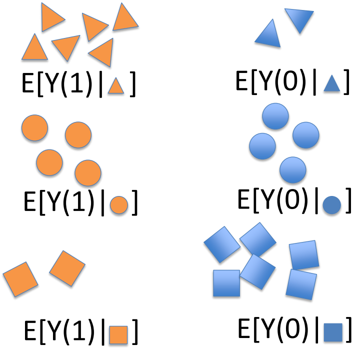


$$E[Y|D=1] - E[Y|D=0] = \overbrace{E[Y(1) - Y(0)|D=1]}^{\text{ATT}} + \underbrace{E[Y(0)|D=1] - E[Y(0)|D=0]}_{\text{selection bias}}$$

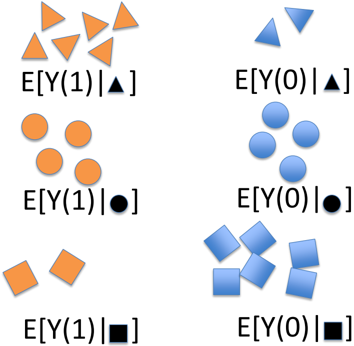
Potential outcomes and treatment effects



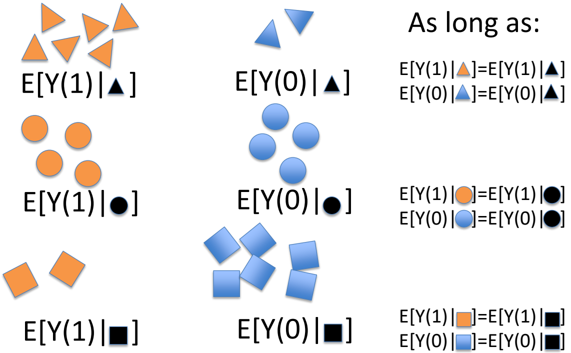
Potential outcomes and treatment effects



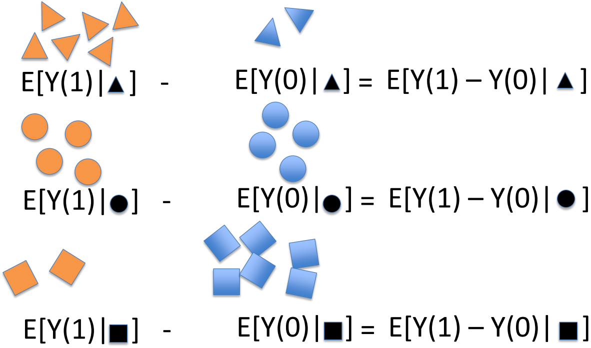
Potential outcomes and treatment effects



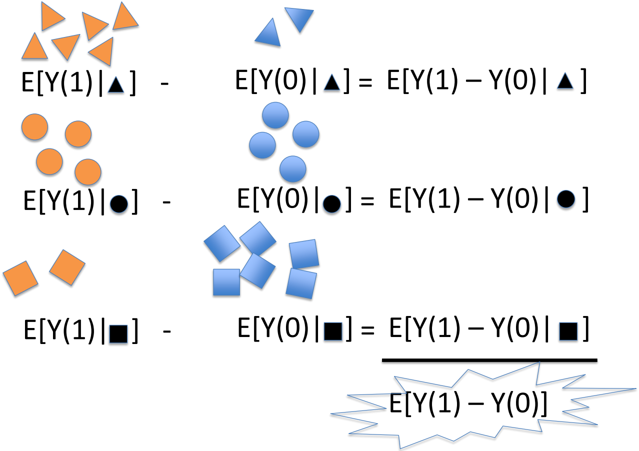
Potential outcomes and treatment effects



Potential outcomes and treatment effects



Potential outcomes and treatment effects



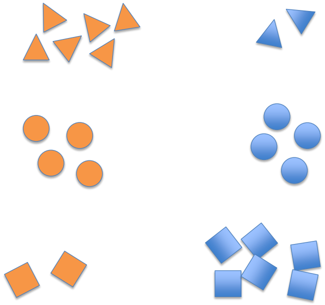
$$E[Y(1)|\blacktriangle] - E[Y(0)|\blacktriangle] = E[Y(1) - Y(0)|\blacktriangle]$$

$$E[Y(1)|\bullet] - E[Y(0)|\bullet] = E[Y(1) - Y(0)|\bullet]$$

$$E[Y(1)|\blacksquare] - E[Y(0)|\blacksquare] = E[Y(1) - Y(0)|\blacksquare]$$

$$E[Y(1) - Y(0)]$$

Potential outcomes and treatment effects

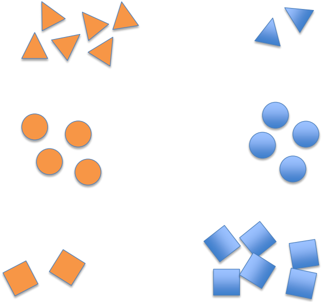


$$Y(1),Y(0) \perp\!\!\!\perp D \,|\, \blacktriangle$$

$$Y(1),Y(0) \perp\!\!\!\perp D \,|\, \bullet$$

$$Y(1),Y(0) \perp\!\!\!\perp D \,|\, \blacksquare$$

Potential outcomes and treatment effects



$$Y(1),Y(0) \perp\!\!\!\perp D \mid \blacktriangle$$

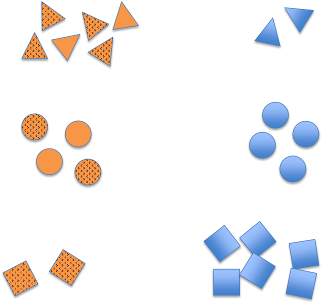
$$Y(1),Y(0) \perp\!\!\!\perp D \mid \bullet$$

$$Y(1),Y(0) \perp\!\!\!\perp D \mid \blacksquare$$

Selection on observables, "unconfoundedness":

$$Y(1),Y(0) \perp\!\!\!\perp D \mid X$$

Potential outcomes and treatment effects

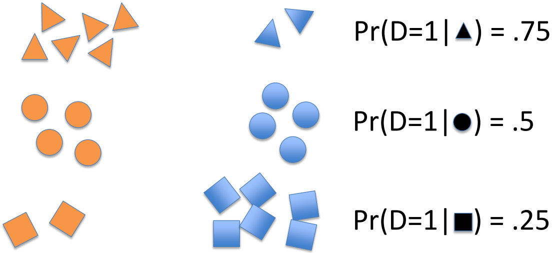


$Y(1), Y(0) \not\perp\!\!\!\perp D \mid \blacktriangle$

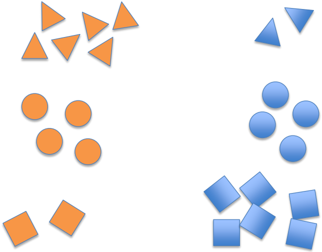
$Y(1), Y(0) \not\perp\!\!\!\perp D \mid \bullet$

$Y(1), Y(0) \not\perp\!\!\!\perp D \mid \blacksquare$

Potential outcomes and treatment effects



Potential outcomes and treatment effects



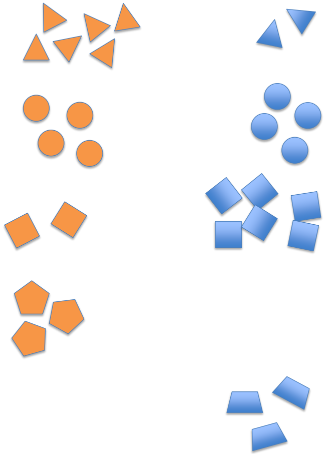
Pr(D=1|▲) = .75

Pr(D=1|●) = .5

Pr(D=1|■) = .25

Common support, "overlap": $0 < \Pr(D=1|X) < 1$

Potential outcomes and treatment effects

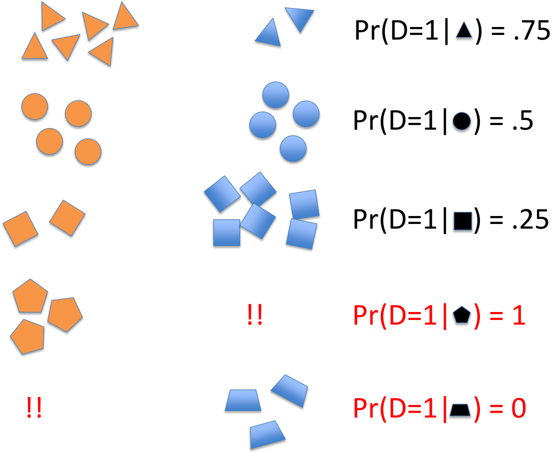


Pr(D=1|▲) = .75

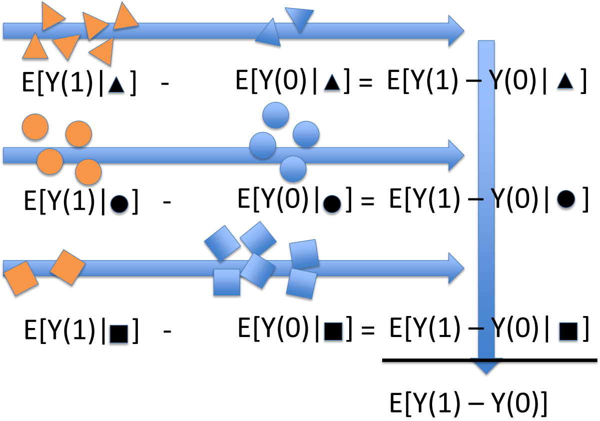
Pr(D=1|●) = .5

Pr(D=1|■) = .25

Potential outcomes and treatment effects



Potential outcomes and treatment effects



Basic causal inference summary

- Target :

$$ATE = E\left[Y_i(1) - Y_i(0)\right] = E\left[\tau_i\right]$$

- Key identifying assumption:

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i | X_i$$

- Estimation:
  - Multiple linear regression (OLS)

$$Y_i = \beta_0 + \tau D_i + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \varepsilon$$

  - Matching
  - Propensity score methods
  - Machine-assisted:
    - Post-Double Selection Lasso

Model Selection When the Goal is Causal Inference

Let's start with the following model

$$y_i = \alpha + \beta D_i + g(X_i) + \zeta_i \quad (1)$$

were

- D_i is the treatment/policy variable of interest,
- X_i is a set controls
- $E[\zeta_i | D_i, X_i] = 0$

Model Selection When the Goal is Causal Inference

- Traditional approach: researcher selects X_i
- Problem: mistakes can occur.
- Same if they use an "automatic" model selection approach.
- It can leave out potentially important variables with small coefficients but non zero coefficients out

Model Selection When the Goal is Causal Inference

- The omission of such variables then generally contaminates estimation and inference results based on the selected set of variables. (e.g. OVB)
- The validity of this approach is delicate because it relies on perfect model selection.
- Because model selection mistakes seem inevitable in realistic settings, it is important to develop inference procedures that are robust to such mistakes.
- Solution here: Lasso

Model Selection When the Goal is Causal Inference

- Using Lasso is useful for prediction
- However, naively using Lasso to draw inferences about model parameters can be problematic.
- Part of the difficulty is that these procedures are designed for prediction, not for inference
- Leeb and Pötscher 2008 show that methods that tend to do a good job at prediction can lead to incorrect conclusions when inference is the main objective
- This observation suggests that more desirable inference properties may be obtained if one focuses on model selection over the predictive parts of the economic problem

Lasso

- We can use Lasso that is slightly modified

$$L(\beta) = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=2}^{p} |\beta_j|\gamma_j \tag{2}$$

  - where $\lambda > 0$ is the penalty level
  - γ_j are *penalty loadings*
    - *penalty loadings* are chosen to insure equivariance of coefficient estimates to rescaling of x_{ij} and can also be chosen to address heteroskedasticity, clustering, and non-gaussian errors

Inference with Selection among Many Controls

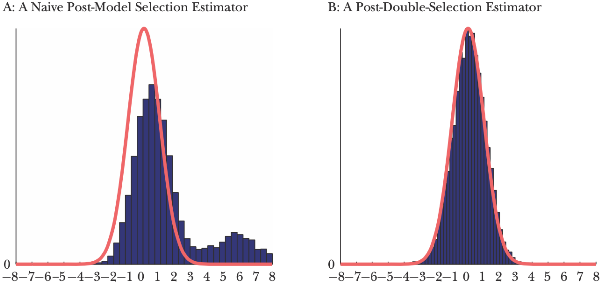
- How do we proceed?

Inference with Selection among Many Controls

- To prevent model selection mistakes, it is important to consider both equations for selection.
- We apply variable selection methods to each of the two reduced form equations and then use all of the selected controls in estimation of β.
- We select
  1. A set of variables that are useful for predicting y_i, say X_{yi}, and
  2. A set of variables that are useful for predicting W_i, say X_{di}.
- We then estimate β by ordinary least squares regression of y_i on W_i and the union of the variables selected for predicting y_i and W_i, contained in X_{yi} and X_{di}.

Inference with Selection among Many Controls

Figure 1
The "Double Selection" Approach to Estimation and Inference versus a Naive Approach: A Simulation from Belloni, Chernozhukov, and Hansen (forthcoming)
(distributions of estimators from each approach)



Source: Belloni, Chernozhukov, and Hansen (forthcoming).
Notes: The left panel shows the sampling distribution of the estimator of α based on the first naive procedure described in this section: applying LASSO to the equation $y_i = d_i + x_i' \theta_y + r_{yi} + \zeta_i$ while forcing the treatment variable to remain in the model by excluding α from the LASSO penalty. The right panel shows the sampling distribution of the "double selection" estimator (see text for details) as in Belloni, Chernozhukov, and Hansen (forthcoming). The distributions are given for centered and studentized quantities.