Double Machine Learning and Bad Controls: A Cautionary Tail

Itamar Caspi and Paul Hünermund

Causal Data Science Meeting 2020

November 11, 2020

Double Machine What?

This Project in a Nutshell

- TBA

A Bird's Eye View of the Literature

- Machine learning and econometrics
  - Mullainathan and Spiess (2017, JEP); Athey and Imbens (2019, ARE).
- High-dimensional inference
  - Belloni et al. (2014, JEP); Chernozhukov et al. (2017, AER); Angrist and Frandsen (2019).
- Causal inference (DAGs) and econometrics
  - Hünermund and Bareinboim (2019); Imbens (2019).

Outline

- High-dimensional Inference
- Good and Bad Controls
- Simulation Setup and Results
- Empirical Illustration

High-dimensional Inference

High-dimensional causal inference

- Consider the following linear regression model

$$Y = \alpha D + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon,$$

  Where Y is the outcome, D is a binary treatment, and $X_1, \ldots, X_p$ are control variables

- Our goal is to estimate the causal effect $D \to Y$, denoted by α.
- When p is large with respect to n, we're in a high-dimensional world. What could go wrong?
  - Overfitting when p close to n (sample size).
  - Infeasibility when $p > n$.

The Naïve Approach: Variable Selection Using Lasso

- Lasso (Tibshirani, 1996) is often used as variable selection tool. In short, Lasso minimizes

$$\min \text{RSS}, \qquad \text{s.t.} \quad |\beta_1| + \cdots + |\beta_p| \leq k$$

  I.e., Lasso puts a "budget constraint" on the sum of the absolute values of beta.

- The solution for $\boldsymbol{\beta}$ is sparse in the sense that it yields a small number of non-zero entries.

- What could go wrong?

  - Lasso selects X's that best predict Y (and only Y) → potential for omitted variable bias.

A Better Approach: Double Lasso

- Belloni et al. (2014, REStud):
  1. Lasso $Y \sim X$ and save selected X's $\rightarrow X_1$.
  2. Lasso $D \sim X$ and save selected X's $\rightarrow X_2$.
  3. Take union of the X's in 1 and 2.
  4. OLS $Y \sim D + X_1 + X_2$ and precede as usual.
- Note that steps 1+2 are about prediction and 4 is about causal and statistical inference.

Double Machine Learning

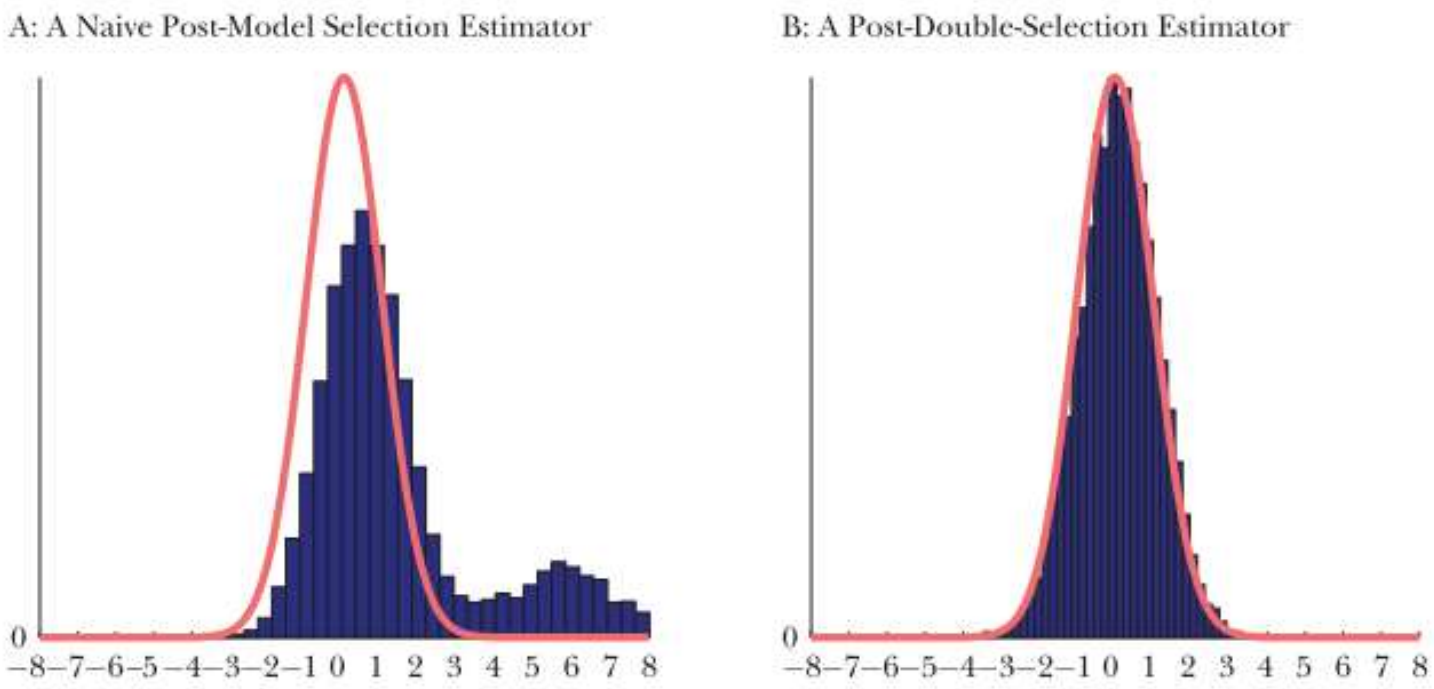
- Why settle for Lasso?
- Chernozhukov et al. (2017): use any machine learning model in steps 1+2.
  - For example, use random forests for $Y \sim X$, and neural nets for $D \sim X$.
- Caveat: If you want statistical inference to behave well in step 4, you'll need to use sample splitting.

Does it Work? Yes (But Wait)

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)

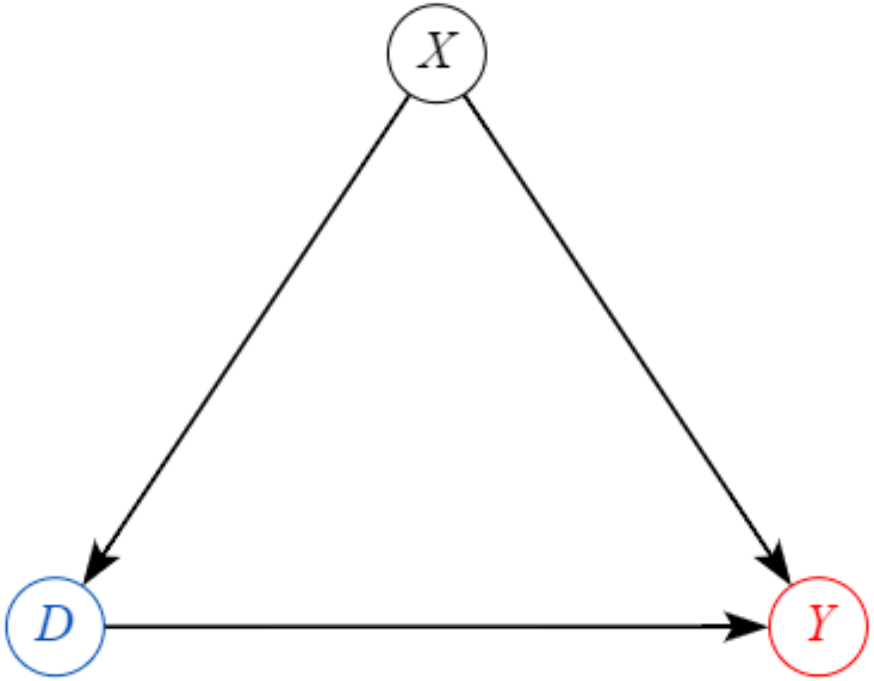


Source: Belloni et al. (2014, JEP).

Good and Bad Controls

Good Control: Confoundedness

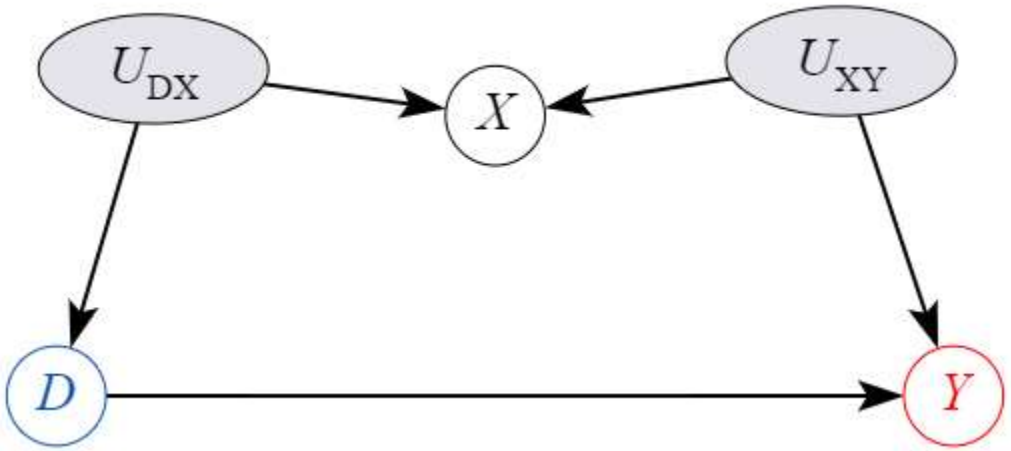
Here, X is a confounder. Controlling for X will reveal the true causal effect $D \rightarrow Y$



Note: DAGs made using causalfusion.net.

Bad Control: Collider

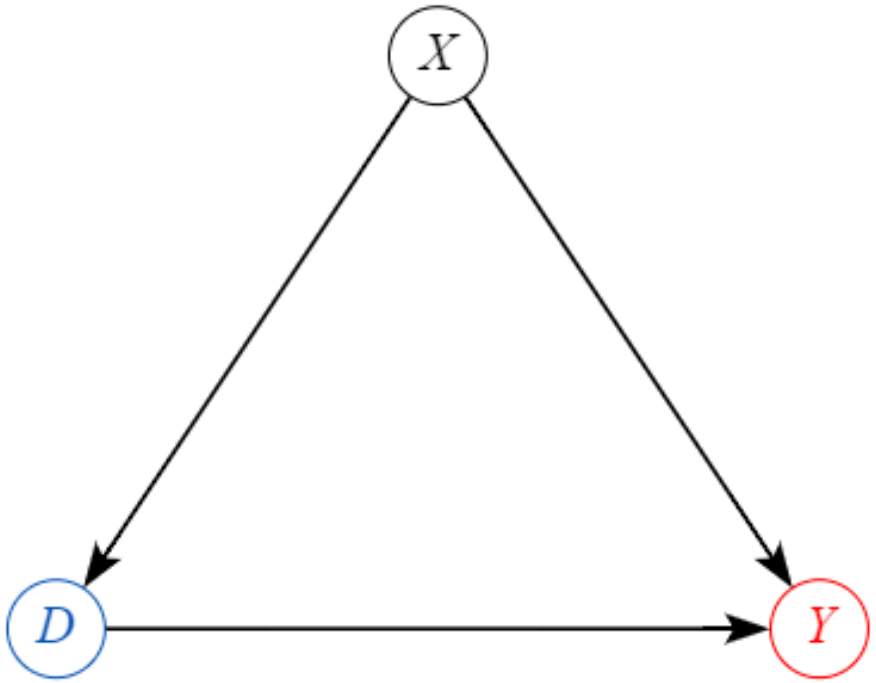
Here, X is a collider. Controlling for X will bias the true causal effect $D \rightarrow Y$

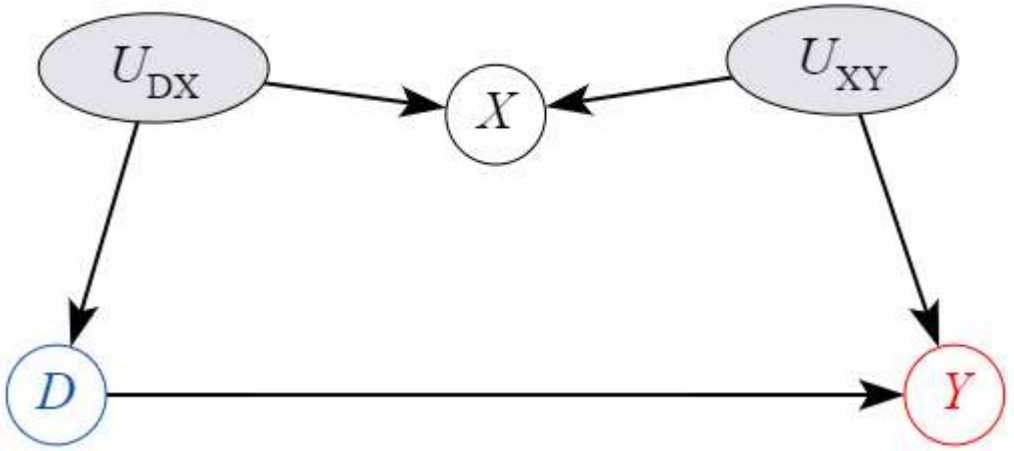


Note: DAGs made using causalfusion.net.

Confounder, Collider, and Prediction

Importatnly, in both cases, X is a (potentially) a great predictor of D and Y!





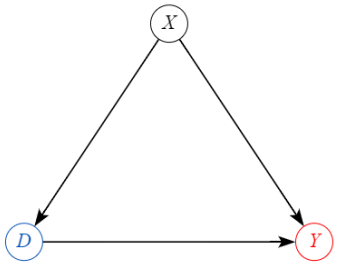
Note: DAGs made using causalfusion.net.

Simulation Study

Simulation Setup

- Simulate two data generating processes: good and bad controls
- True causal effect = 0
- Sample size = 100
- 1,000 replications
- Perform statistical inference using three methods:
  - Naïve Lasso
  - Double Lasso
  - t-test

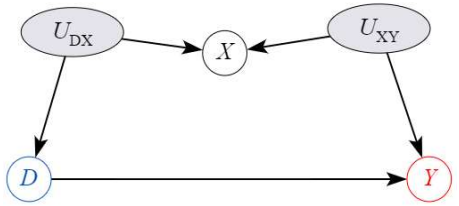
Two Data Generating Processes



MODEL #1: Good Control

$$X = U_X$$
$$D = 0.8X + 0.2U_D$$
$$Y = 0.2X + U_Y$$

$$U_X, U_D, U_Y, \sim i.i.d.N(0,1)$$



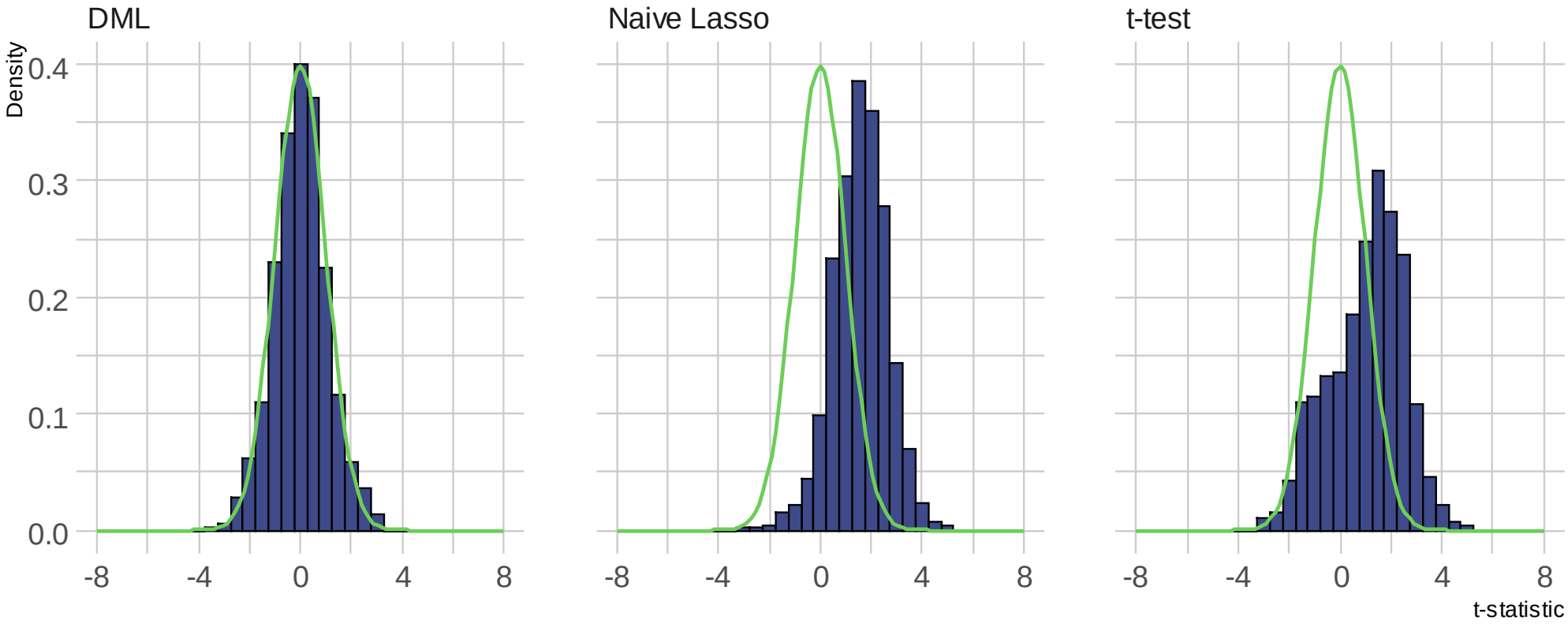
MODEL #2: Bad Control

$$X = 0.8U_{DX} + 0.2U_{XY} + 0.6U_X$$
$$D = U_{DX}$$
$$Y = U_{XY}$$

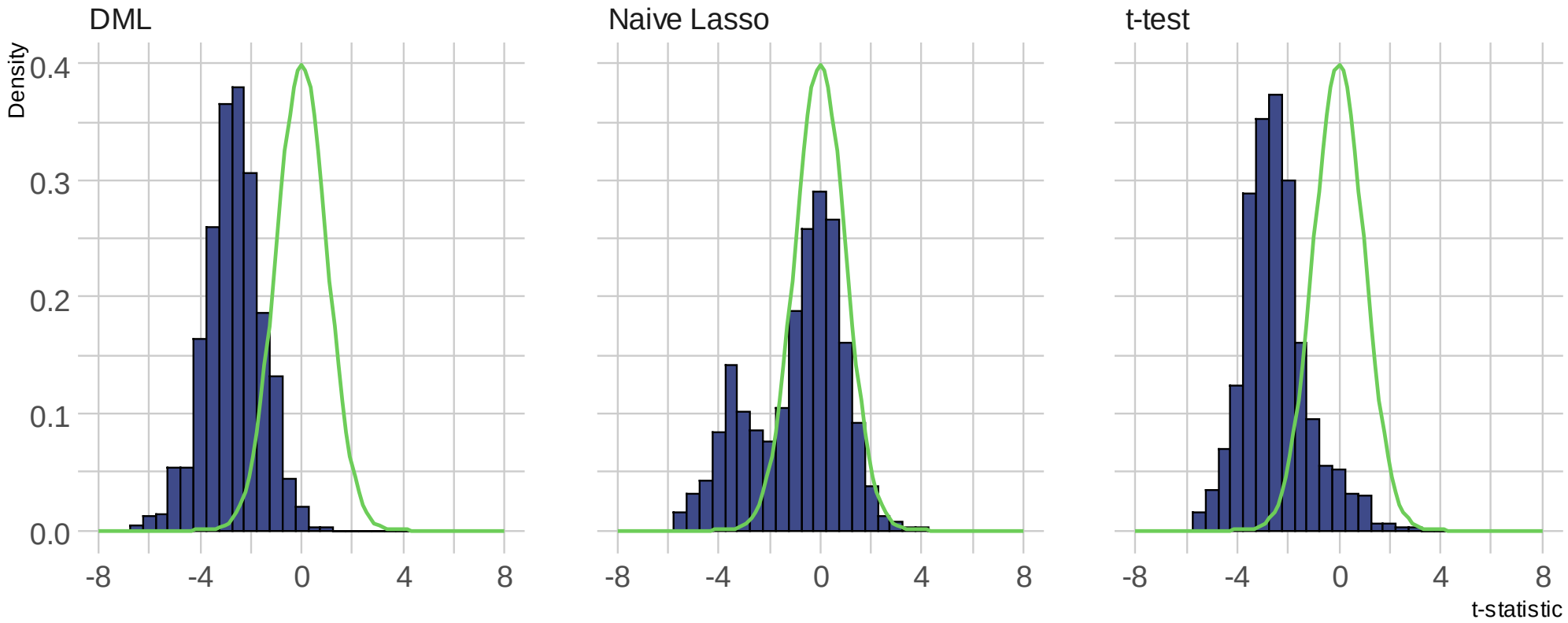
$$U_X, U_{DX}, U_{XY} \sim i.i.d.N(0,1)$$

In both cases:
$\mathrm{Corr}(X, Y) = 0.8$ and $\mathrm{Corr}(X, D) = 0.2$

Case #1: Good Control

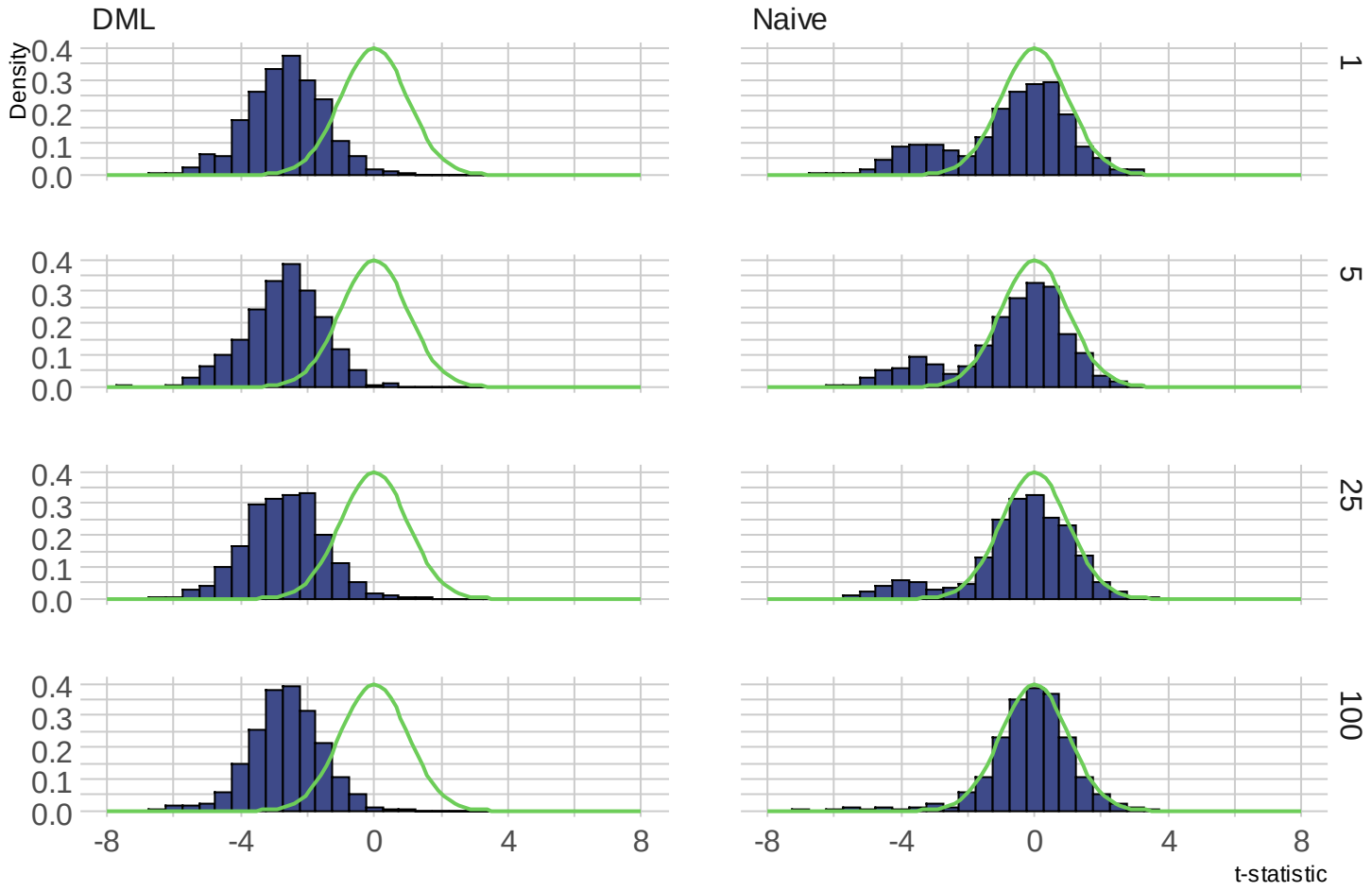


Case #2: Bad Control



Extension: Multivariate Regression

- Same DGP as before, only now we add irrelevant X's to the regression, where each X is iid standard normal.



Key Takeaway

- When in a high-dimensional setting, need to watch out from lurking bad controls.
- Automatic variable (or model) selection methods which rank models based on their prediction quality cannot distinguish between good and bad controls.
- Deconfounding cannot be automated.

Empirical Illustration

Journal of Economic Literature 2017, 55(3), 789–865
https://doi.org/10.1257/jel.20160995

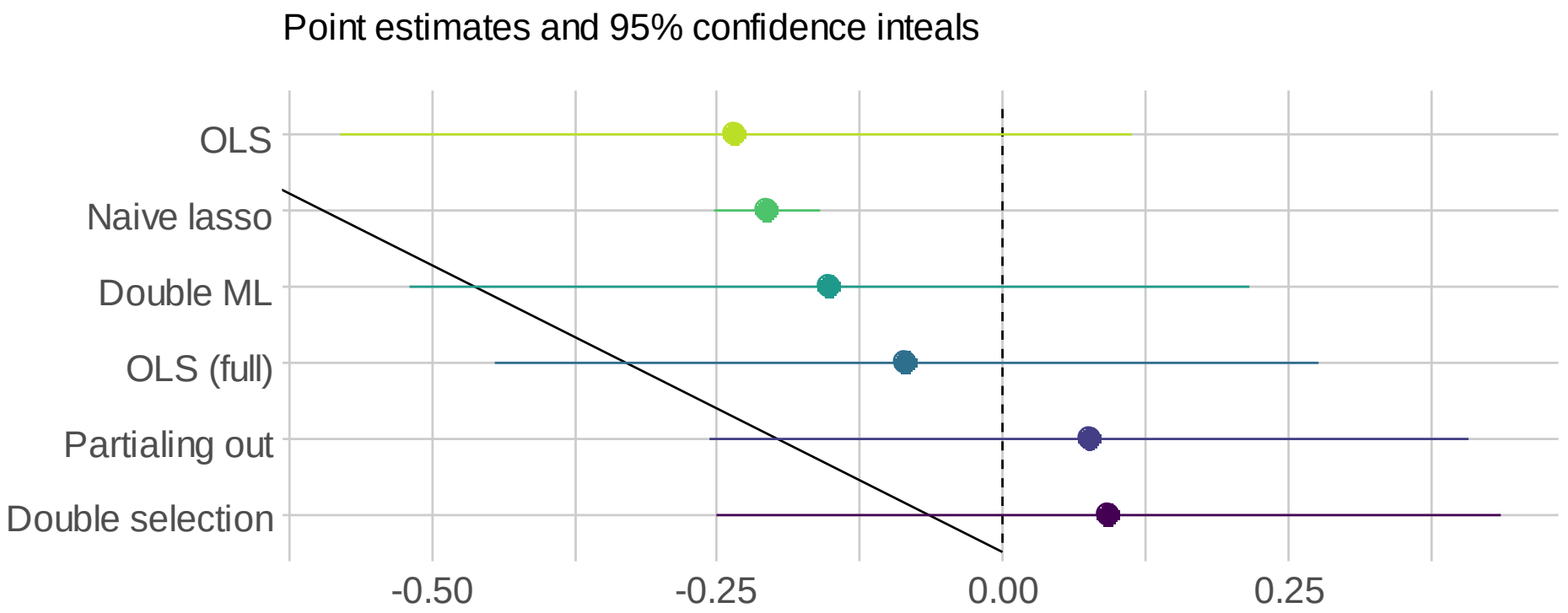
The Gender Wage Gap: Extent, Trends, and Explanations†

Francine D. Blau and Lawrence M. Kahn*

Using Panel Study of Income Dynamics (PSID) microdata over the 1980–2010 period, we provide new empirical evidence on the extent of and trends in the gender wage gap, which declined considerably during this time. By 2010, conventional human capital variables taken together explained little of the gender wage gap, while gender differences in occupation and industry continued to be important. Moreover, the gender pay gap declined much more slowly at the top of the wage distribution than at the middle or bottom and by 2010 was noticeably higher at the top. We then survey the literature to identify what has been learned about the explanations for the gap. We conclude that many of the traditional explanations continue to have salience. Although human-capital factors are now relatively unimportant in the aggregate, women's work force interruptions and shorter hours remain significant in high-skilled occupations, possibly due to compensating differentials. Gender differences in occupations and industries, as well as differences in gender roles and the gender division of labor remain important, and research based on experimental evidence strongly suggests that discrimination cannot be discounted. Psychological attributes or noncognitive skills comprise one of the newer explanations for gender differences in outcomes. Our effort to assess the quantitative evidence on the importance of these factors suggests that they account for a small to moderate portion of the gender pay gap, considerably smaller than, say, occupation and industry effects, though they appear to modestly contribute to these differences. (*JEL* I26, J16, J24, J31, J71)

Occupation DAG

Estimated Gender Wage Gap by Method



Note: 2009 wave.

Concluding Remarks

- TBA

Thanks ☺

Comments are welcome!

Itamar.caspi@boi.org.il

phu.si@cbs.dk

References

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. Annual Review of Economics, 11, 685-725.

Angrist, J., & Frandsen, B. (2019). Machine labor (No. w26584). National Bureau of Economic Research.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014a). Inference on treatment effects after selection among high-dimensional controls. The Review of Economic Studies, 81(2), 608-650.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014b). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2), 29-50.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5), 261-65.

Cinelli, C., Forney, A., & Pearl, J. (2020). A Crash Course in Good and Bad Controls.

Hünermund, P., & Bareinboim, E. (2019). Causal inference and data-fusion in econometrics. arXiv preprint arXiv:1912.09104.

Imbens, G. (2019). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics (No. w26104). National Bureau of Economic Research.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach.