Double/Debiased Machine Learning

Chris Conlon

May 2, 2023

NYU Stern

Setup

The core problem

$$Y_i = \tau_i \cdot D_i + g_0(X_i) + u_i \qquad \text{with } \mathbb{E}[u_i \mid x_i, D_i] = 0$$

$$D_i = m_0(X_i) + v_i \qquad \text{with } \mathbb{E}[v_i \mid x_i] = 0$$

- D_i is treatment indicator and τ_i is treatment effect
- X_i are covariates ("controls" or "confounders") $\rightarrow$
- We call $m_0(\cdot)$ and $g_0(\cdot)$ nuisance parameters or nuisance functions.

What if we try “machine learning¿’

$$Y_i = \tau_i \cdot D_i + g_0(X_i) + u_i \qquad \text{with } \mathbb{E}[u_i \mid x_i, D_i] = 0$$
$$D_i = m_0(X_i) + v_i \qquad \text{with } \mathbb{E}[v_i \mid x_i] = 0$$

Could alternate steps:

1. Fit random forest of $Y_i - \widehat{\tau}_i \cdot D_i$ on Z_i to get $\widehat{g_0}(Z_i)$.
2. Run OLS of $Y_i - \widehat{g_0}(Z_i)$ on D_i to get $\widehat{\tau}$ or $\widehat{\tau}_i$

This fits the data great but gives terrible estimates of $\widehat{\tau}$!

Why? Frisch-Lovell-Waugh really needs things to be linear!

What goes wrong

- We are trading off bias for variance reduction
- But, we can't trust plug-in estimates when $m_0(\cdot), g_0(\cdot)$ are not linear.
- So bias can be very dangerous now...

What we need to do is orthogonalize things properly (like a nonlinear Frisch-Lovell-Waugh)

- Bias from regularization $\rightarrow$Orthogonalization
- Bias from overfitting $\rightarrow$Sample Splitting

Sample Splitting

- Split the sample into two parts main sample and auxiliary sample.
- On the auxiliary Sample:
  - Estimate $\widehat{g}(X_i)$ from $Y_i = \tau \cdot D_i + g(X_i) + u_i$
  - Estimate $\widehat{m}(X_i)$ from $D_i = m(X_i) + v_i$
- Now on the main sample:
  - Compute the residual: $\widehat{v_i} = D_i - \widehat{m}(X_i)$.
  - Estimate $\hat{\tau} = (\hat{v}' D)^{-1} \hat{v}' (Y - \hat{g}(X))$

More Info

To learn more watch Chernozhukov lecture here:

https://www.youtube.com/watch?v=eHOjmyoPCFU&t=37s

The R package https://docs.doubleml.org/stable/index.html

Thanks!