Notes: Double/Debiased Machine Learning

Ziyang Gong

Date: July 23, 2021

1 Intutation

Consider the partical linear regression model

$$\begin{aligned} Y_i = D_i\theta + g(\mathbf{X}_i) + U_i, \quad E\left(U_i \mid \mathbf{X}_i, D_i\right) = 0 \\ D_i = m(\mathbf{X}_i) + V_i, \quad E\left(V_i \mid \mathbf{X}_i\right) = 0 \end{aligned} \tag{1}$$

where $\mathbf{X}_i \in \mathbb{R}^p$, $i = 1, 2, \ldots, n$. Suppose

$$\begin{aligned} m(\mathbf{X}_i) =& \gamma_1 x_{i,1} + \gamma_2 x_{i,2} + \gamma_3 x_{i,3} + U_i \\ g(\mathbf{X}_i) =& \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + V_i \end{aligned} \tag{2}$$

and

$$n = 1000, \quad p = 200, \quad \boldsymbol{\gamma} = (3, 2, 1)', \quad \boldsymbol{\beta} = (1, 2, 3)'$$

1.1 Regularization Bias

1. Non Orthogonal: Suppose $\hat{g}$ are estimated using the auxiliary samples indexed by I^c, thus
$$\check{\theta} = \left(\frac{1}{n}\sum_{i\in\mathrm{I}} D_i^2\right)^{-1} \frac{1}{n}\sum_{i\in\mathrm{I}} D_i\left(Y_i - \hat{g}\left(X_i\right)\right) \tag{3}$$
2. Orthogonal: Suppose $\hat{g}$ and $\hat{m}$ are estimated using the auxiliary samples indexed by I^c, thus
$$\check{\theta} = \left(\frac{1}{n}\sum_{i\in\mathrm{I}} \widehat{V}_i D_i\right)^{-1} \frac{1}{n}\sum_{i\in\mathrm{I}} \hat{V}_i\left(Y_i - \hat{g}\left(X_i\right)\right) \tag{4}$$
where $\hat{V} = D - \hat{m}(\mathbf{X})$.

1.2 Overfitting Bias

1.3 Score Functions

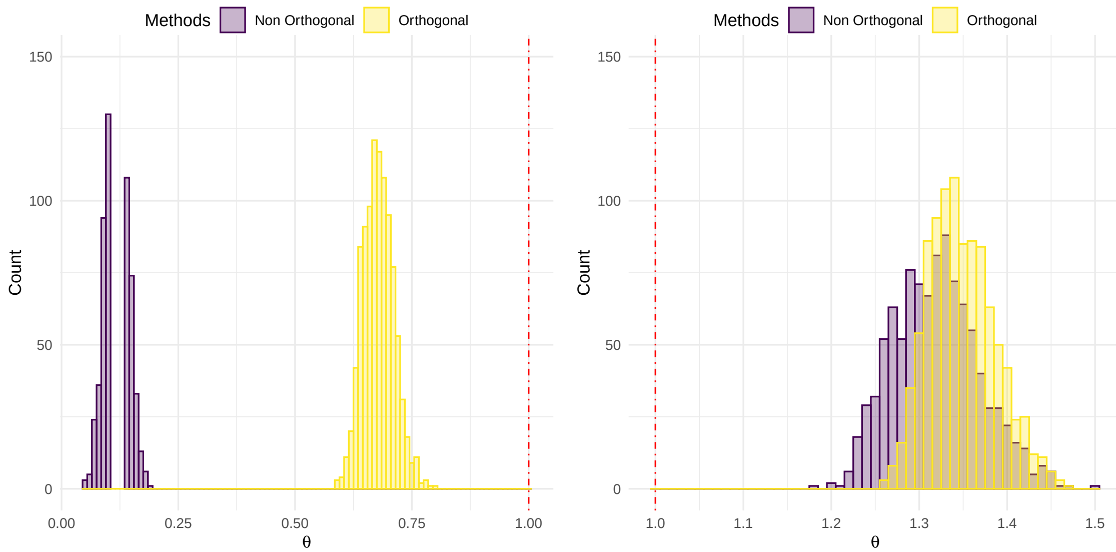
Definition 1.1 (Neymaxn Orthogonal Score Function)**.**

Remark. Suppose the score function is the linearity function of in the parameter θ, that,

$$\psi\left(\mathbf{W}; \theta, \boldsymbol{\eta}\right) = \psi_a\left(\mathbf{W}; \boldsymbol{\eta}\right)\theta + \psi_b\left(\mathbf{W}; \boldsymbol{\eta}\right) \tag{5}$$

Hence the estimator can be written as

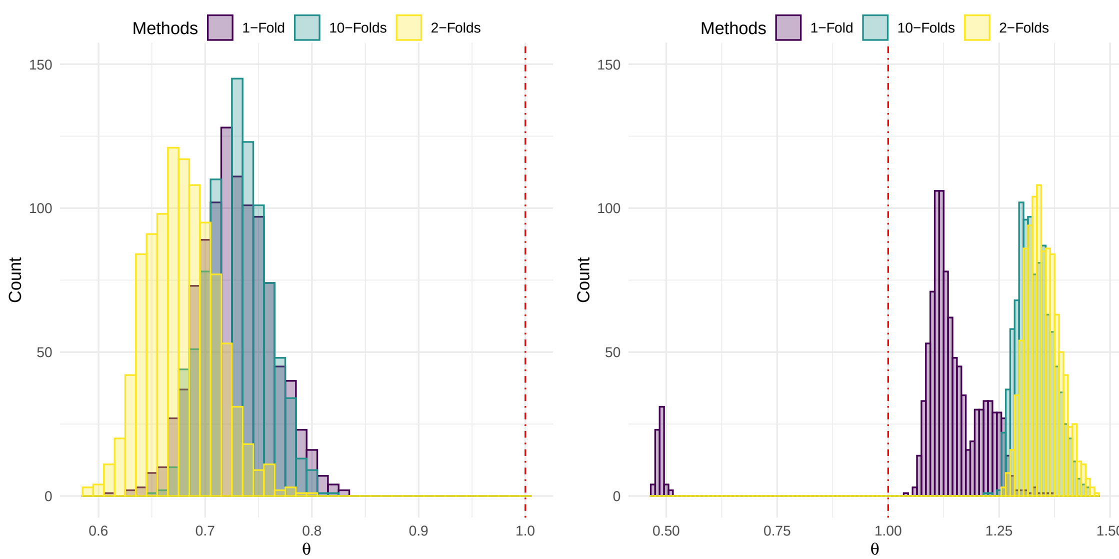
$$\check{\theta} = -\frac{\mathbb{E}_N\left[\psi_b\left(\mathbf{W}; \boldsymbol{\eta}\right)\right]}{\mathbb{E}_N\left[\psi_a\left(\mathbf{W}; \boldsymbol{\eta}\right)\right]} \tag{6}$$



(a) LASSO

(b) Random Forest

Figure 1: Bias Raised by Regularization of Machine Learning



(a) LASSO

(b) Random Forest

Figure 2: Bias Raised by Overfitting of Machine Learning

1. IV-type: The score function of IV-type is

$$\psi(\mathbf{W};\theta,\boldsymbol{\eta}) := [Y - D\theta - g(\mathbf{X})]\,[D - m(\mathbf{X})] \tag{7}$$

where $\boldsymbol{\eta} = (g, m)$. Suppose $\hat{g}$ and $\hat{m}$ are estimated using the auxiliary samples indexed by I^c, Thus

$$\check{\theta} = \left(\frac{1}{n}\sum_{i\in\mathrm{I}} \widehat{V}_i D_i\right)^{-1} \frac{1}{n}\sum_{i\in\mathrm{I}} \hat{V}_i\left(Y_i - \hat{g}(X_i)\right) \tag{8}$$

where $\hat{V} = D - \hat{m}(\mathbf{X})$.

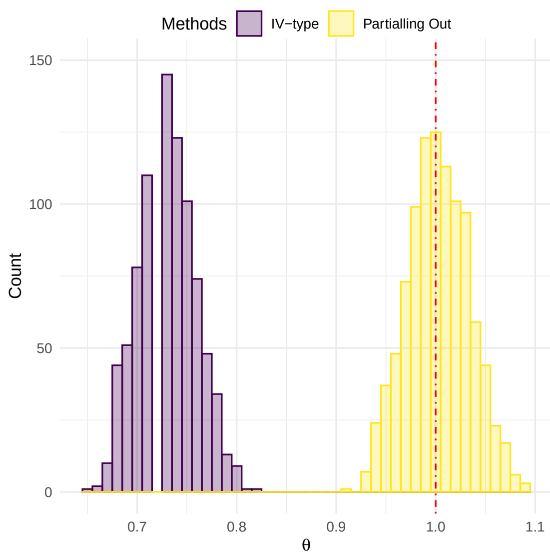
2. Partialling-Out: The score function of partialling-out is

$$\psi(\mathbf{W};\theta,\boldsymbol{\eta}) := [Y - \ell(\mathbf{X}) - \theta(D - m(\mathbf{X}))]\,[D - m(\mathbf{X})] \tag{9}$$

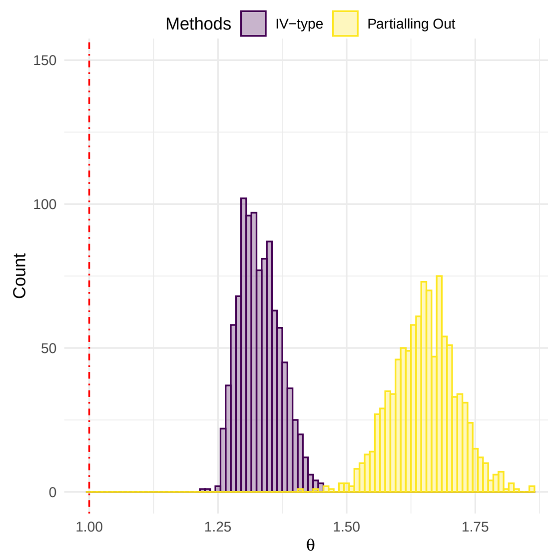
where $\boldsymbol{\eta} = (\ell, m)$. Suppose $\hat{\ell}$ and $\hat{m}$ are estimated using the auxiliary samples indexed by I^c, Thus

$$\check{\theta} = \left(\frac{1}{n}\sum_{i\in\mathrm{I}} \widehat{V}_i^2\right)^{-1} \frac{1}{n}\sum_{i\in\mathrm{I}} \hat{V}_i\left(Y_i - \hat{\ell}(X_i)\right) \tag{10}$$

where $\hat{V} = D - \hat{m}(\mathbf{X})$.



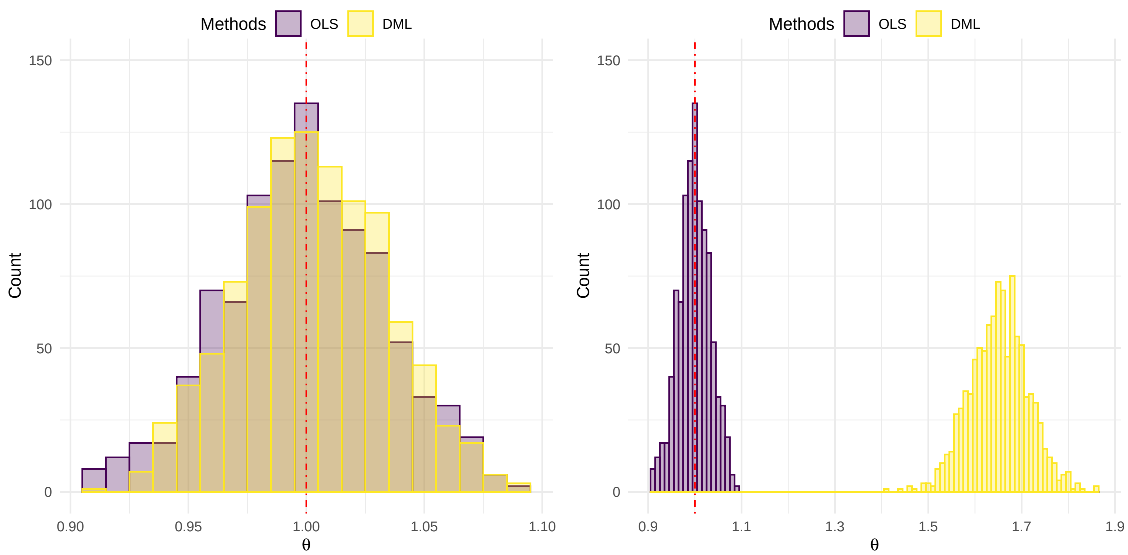
(a) LASSO



(b) Random Forest

Figure 3

1.4 Complexity of Paramters



(a) LASSO

(b) Random Forest

Figure 4