

# Notes: Double/Debiased Machine Learning

Ziyang Gong

Date: July 23, 2021

## 1 Intuition

Consider the partial linear regression model

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

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

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

and

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

### 1.1 Regularization Bias

1. Non Orthogonal: Suppose  $\hat{g}$  are estimated using the auxiliary samples indexed by  $I^c$ , thus

$$\check{\theta} = \left( \frac{1}{n} \sum_{i \in I} D_i^2 \right)^{-1} \frac{1}{n} \sum_{i \in I} D_i (Y_i - \hat{g}(X_i)) \quad (3)$$

2. Orthogonal: Suppose  $\hat{g}$  and  $\hat{m}$  are estimated using the auxiliary samples indexed by  $I^c$ , thus

$$\check{\theta} = \left( \frac{1}{n} \sum_{i \in I} \hat{V}_i D_i \right)^{-1} \frac{1}{n} \sum_{i \in I} \hat{V}_i (Y_i - \hat{g}(X_i)) \quad (4)$$

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

### 1.2 Overfitting Bias

### 1.3 Score Functions

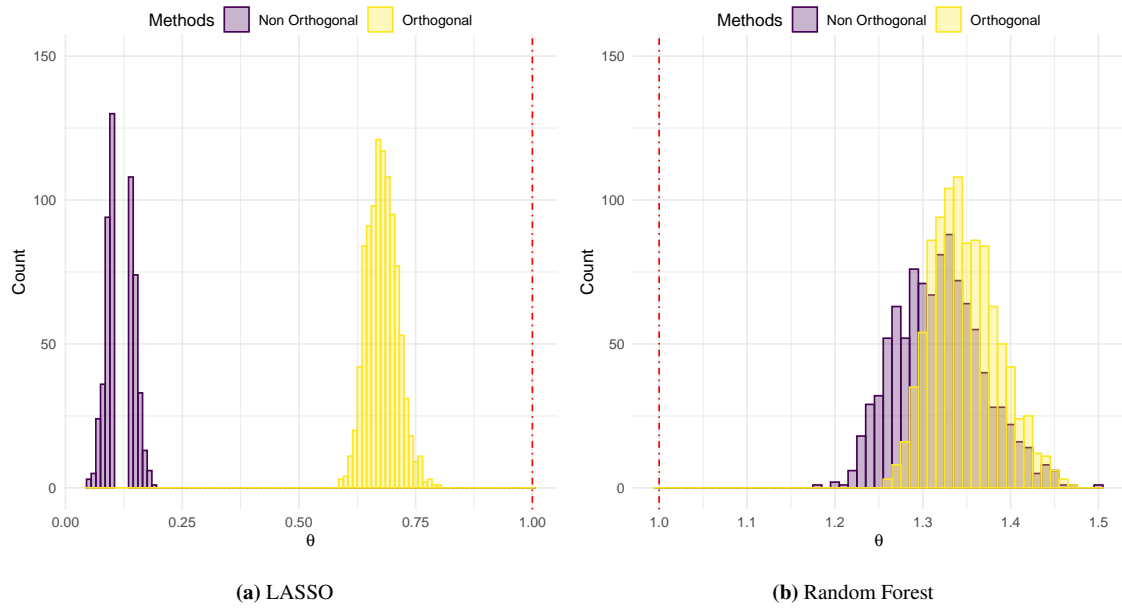
**Definition 1.1** (Neyman's Orthogonal Score Function).

*Remark.* Suppose the score function is the linearity function of in the parameter  $\theta$ , that,

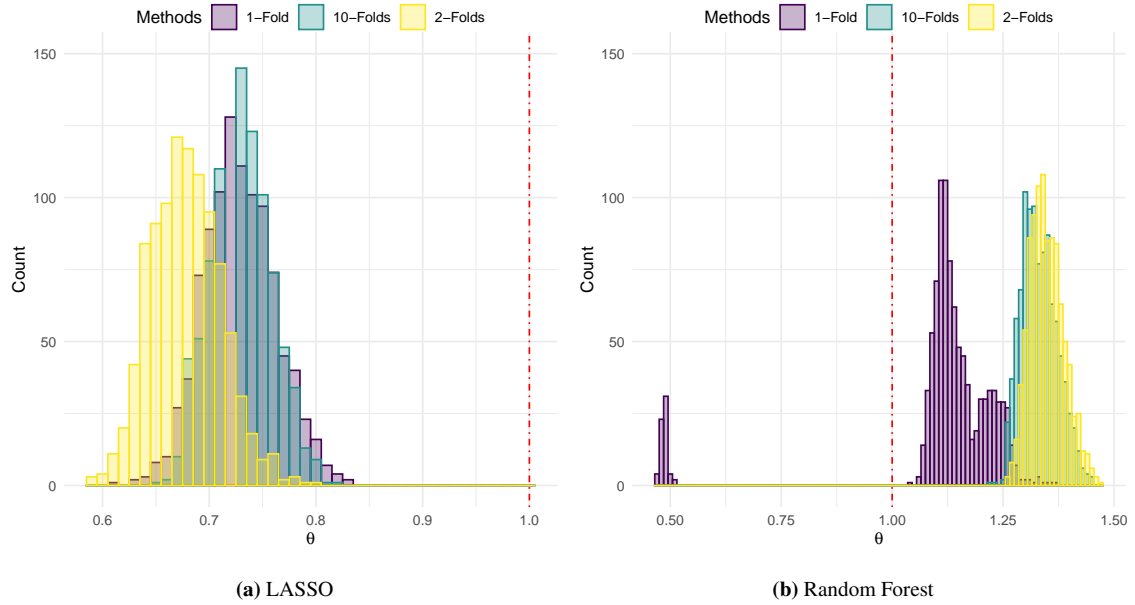
$$\psi(\mathbf{W}; \theta, \boldsymbol{\eta}) = \psi_a(\mathbf{W}; \boldsymbol{\eta}) \theta + \psi_b(\mathbf{W}; \boldsymbol{\eta}) \quad (5)$$

Hence the estimator can be written as

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



**Figure 1: Bias Raised by Regularization of Machine Learning**



**Figure 2: Bias Raised by Overfitting of Machine Learning**

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

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

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

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

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

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

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

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

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

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

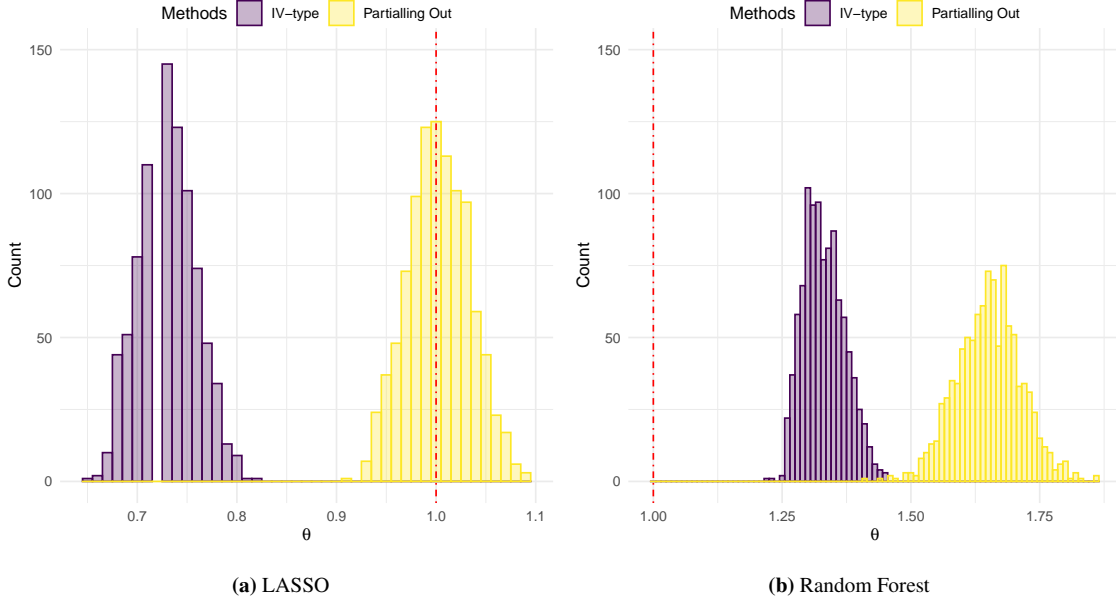
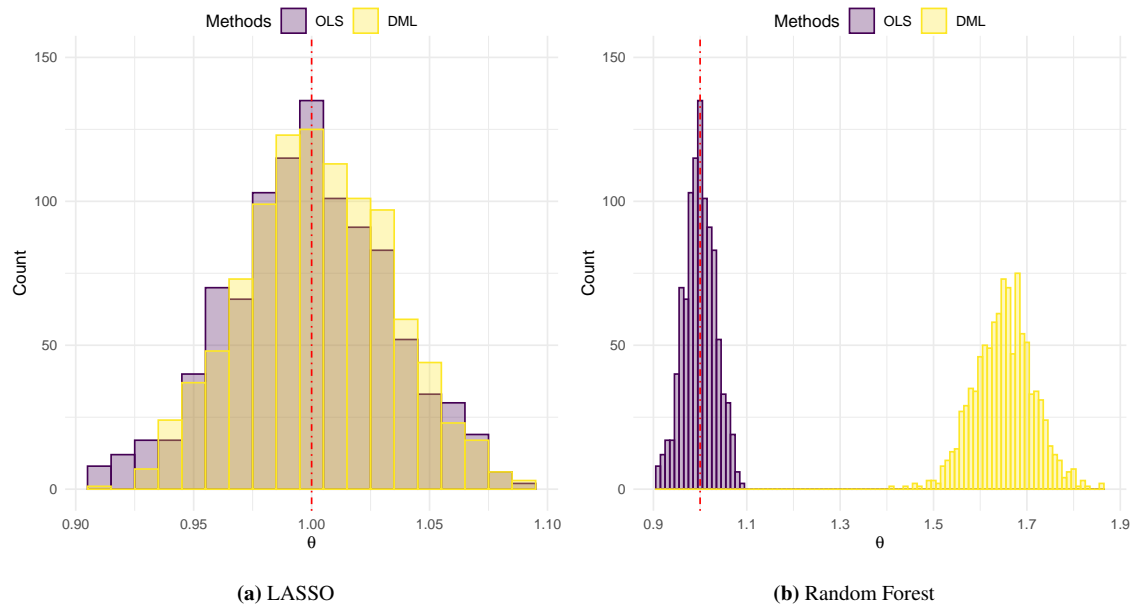


Figure 3

## 1.4 Complexity of Paramters



**Figure 4**