

# Covariate Adjustment in Randomized Trial

Steve Yadowsky

A frequent question that comes up is whether covariate adjustment will likely increase or reduce the estimation error of causal inferences made from a randomized trial. Freedman [1] gives an example of how applying regression with an invalid structural model can create bias and invalid standard errors. Lin [3] shows that in the asymptotic large-sample regime (when the number of samples  $n$  grows,  $n \rightarrow \infty$ ), these issues can easily be fixed by using Huber-White robust standard errors in a model with treatment-covariate interactions, so that the regression adjusted estimator has equal, or lower, asymptotic variance.

In this short note, we study the finite sample behavior in a simple model where both the randomized and regression adjusted estimators are justified. In this model, we give conditions under which the regression adjustment would lead to a higher variance estimator than the unadjusted regression, in terms of the dimension and strength of the effect of the covariates on the outcome. When the dimension of the covariates stays fixed as  $n \rightarrow \infty$ , the adjusted estimator will have no worse variance than the unadjusted estimator. However, for fixed sample sizes, or if  $\lim_{n \rightarrow \infty} d/n > 0$ , then the adjusted estimator may have higher variance.

## The data distribution and effect estimators

Let  $Y_i$  be the observed outcome in a randomized trial with treatment  $W \in \mathbb{R}$  and observed pre-treatment covariates  $Z \in \mathbb{R}^d$ . Suppose that the following structural model holds:

$$Y(w) = \tau w + \beta^\top Z + \epsilon,$$

where  $Y(w)$  is the counterfactual outcome at treatment level  $w$ ,  $\beta \in \mathbb{R}^d$  is fixed (but unknown) and  $\epsilon \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ . Additionally, assume the standard consistency assumption,  $Y = Y(W)$ , and let the treatment be randomized,  $W \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ , and the covariates drawn from a population with  $X \stackrel{\text{iid}}{\sim} \mathcal{N}(0, I_d)$ . We will use  $Z = (X^\top W)^\top$  and  $\gamma = (\beta^\top \tau)^\top$  to simplify notation.

Given a sample of  $n$  observations  $\{(Y_i, W_i, X_i)\}_{i=1}^n$  with  $n \geq p$ , consider the following two consistent estimators of  $\tau$ . First,

$$\hat{\tau}_{\text{unadj}} = \frac{\frac{1}{n} \sum_{i=1}^n W_i Y_i}{\frac{1}{n} \sum_{i=1}^n W_i^2}, \quad (0.1)$$

is the unadjusted OLS estimator of the treatment on the outcome. Second,

$$\hat{\tau}_{\text{adj}} = \begin{bmatrix} 0 & 1 \end{bmatrix} \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i Y_i \quad (0.2)$$

is the least squares estimator that adjusts for baseline covariates. Plugging in the model for  $Y_i = \tau W_i + \beta^\top Z_i + \epsilon_i$  gives the expanded forms

$$\hat{\tau}_{\text{unadj}} = \tau + \frac{\frac{1}{n} \sum_{i=1}^n (\beta^\top X_i + \epsilon_i) W_i}{\frac{1}{n} \sum_{i=1}^n W_i^2},$$

and

$$\begin{aligned}\hat{\tau}_{\text{adj}} &= \begin{bmatrix} 0 & 1 \end{bmatrix} \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i (Z_i^\top \gamma + \epsilon_i) \\ &= \tau + \begin{bmatrix} 0 & 1 \end{bmatrix} \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i \epsilon_i.\end{aligned}$$

### Statistical properties

Noting that  $\mathbb{E}[\epsilon_i \mid Z_i] = 0$  shows that both of these estimators are unbiased. Next, we turn to their variance. Both we derive from the law of total variance,  $\text{Var}(A) = \text{Var}(\mathbb{E}[A \mid B]) + \mathbb{E}[\text{Var}(A \mid B)]$ .

First, for the variance of  $\sqrt{n}(\hat{\tau}_{\text{unadj}} - \tau)$ , we note that  $\text{Var}(\beta^\top X_i + \epsilon_i) = \|\beta\|_2^2 + 1$ , and apply the law of total variance conditionally on  $\{W_i\}_{i=1}^n$ , to get

$$\begin{aligned}\text{Var}(\sqrt{n}(\hat{\tau}_{\text{unadj}} - \tau)) &= n\mathbb{E}[\text{Var}(\hat{\tau}_{\text{unadj}} - \tau \mid \{W_i\}_{i=1}^n)] + n\text{Var}(\mathbb{E}[\hat{\tau}_{\text{unadj}} - \tau \mid \{W_i\}_{i=1}^n]) \\ &= \mathbb{E} \left[ \frac{\frac{1}{n} \sum_{i=1}^n (\|\beta\|_2^2 + 1) W_i^2}{\left( \frac{1}{n} \sum_{i=1}^n W_i^2 \right)^2} \right] \\ &= (\|\beta\|_2^2 + 1) \frac{n}{n-2}.\end{aligned}$$

For the variance of  $\sqrt{n}(\hat{\tau}_{\text{adj}} - \tau)$ , we apply a similar trick, conditioning on  $\{Z_i\}_{i=1}^n$ ,

$$\text{Var}(\sqrt{n}(\hat{\tau}_{\text{adj}} - \tau)) = \begin{bmatrix} 0 & 1 \end{bmatrix} \text{Var} \left( \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i \epsilon_i \right) \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Noting that this is simply the bottom right element of  $\text{Var}(\sqrt{n}(\frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i \epsilon_i)$ , we work out this variance to simplify notation.

$$\begin{aligned}\text{Var} \left( \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i \epsilon_i \right) &= \mathbb{E} \left[ \text{Var} \left( \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i \epsilon_i \mid \{Z_i\}_{i=1}^n \right) \right] \\ &\quad + \text{Var} \left( \mathbb{E} \left( \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i \epsilon_i \mid \{Z_i\}_{i=1}^n \right) \right) \\ &= \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n Z_i I_{d+1} Z_i^\top \right) \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \right] \\ &= \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top \right)^{-1} \right] \\ &= \frac{n}{n-d-2} I_{d+1},\end{aligned}$$

where the last equality follows from the fact that  $(\sum_{i=1}^n Z_i Z_i^\top)^{-1}$  is a  $d+1$ -dimensional inverse Wishart matrix with  $n$  degrees of freedom [2]. Therefore  $\text{Var}(\sqrt{n}(\hat{\tau}_{\text{adj}} - \tau)) = (1 - (d+2)/n)^{-1}$ .

## Conclusions

Notice that the variance of the adjusted least squares is greater than the unadjusted estimator whenever

$$\|\beta\|_2^2 \leq \frac{d}{n-2-d}.$$

If  $d$  stays fixed as  $n \rightarrow \infty$ , then  $\frac{d}{n-2-d} \rightarrow 0$ , so the variance of the adjusted estimator will always have asymptotic variance less than or equal to the unadjusted estimator, consistent with the observations of Lin [3], and strictly less if  $\liminf_{n \rightarrow \infty} \|\beta\|_2 > 0$ . However, in finite samples, it can have larger variance, especially if  $\|\beta\|_2$  is small. The model with additional regression adjustment for interaction terms  $XW$  would have  $d$  more covariates, so that the variance would be  $(1 - (2d + 2)/n)^{-1}$ .

## References

- [1] D. A. Freedman. On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40(2):180–193, 2008.
- [2] L. Haff. An identity for the wishart distribution with applications. *Journal of Multivariate Analysis*, 9(4):531–544, 1979.
- [3] W. Lin. Agnostic notes on regression adjustments to experimental data: Reexamining freedman’s critique. *Annals of Applied Statistics*, 7(1):295–318, 2013.