

Topic 3: *Moving the Goalposts* Approach

by Sai Zhang

Key points:

•

Disclaimer: These notes are written by Sai Zhang ([email me](#) or check my [Github page](#)). The main reference for this topic is [Armstrong, Kolesár, and Kwon \(2020\)](#), I thank Prof. Armstrong for his valuable advice.

3.1 Finite Sample Bias-Variance Tradeoffs

3.1.1 Setup

Consider the fixed design regression model

$$y_i = w_i \beta(z_i) + h(z_i) + \epsilon_i \quad (3.1)$$

where

- w_i, z_i are treated as **fixed**
- ϵ_i is **independent**, with $\mathbb{E}[\epsilon_i] = 0, \mathbb{E}[\epsilon_i^2] = \sigma_i^2$
- observation: $\left\{ \left(y_i, w_i, z_i' \right)' \right\}_{i=1}^n$

one example is the case where w_i is **binary**, then

$$\beta(z) = f(1, z) - f(0, z)$$

which is just the ATE conditional on z under the unconfoundedness assumption. This includes the RD design, where z_i is the running variable and w_i is the treatment assignment.

Now, consider for the weighted average treatment effect

$$L_\mu [\beta(\cdot)] = \int \beta(z) d\mu(z)$$

where $\int \mu(z) = 1$ is a **signed** measure (weight, allowing **negative** weights), construct a linear estimator

$$\hat{L}_a = \sum_{i=1}^n a_i y_i$$

where the estimation weights a_i can depend on $\{z_i, w_i, \sigma_i^2\}_{i=1}^n$, but **not** on y_i . Together, the bias of \hat{L}_a for $L_\mu [\beta(\cdot)]$, given the regression function $\beta(\cdot), h(\cdot)$, is

$$\mathbb{E}_{\beta(\cdot), h(\cdot)} [\hat{L}_a] - L_\mu [\beta(\cdot)] = \sum_{i=1}^n a_i [w_i \beta(z_i) + h(z_i)] - \int \beta(z) d\mu(z)$$

and its variance, given the regression function $\beta(\cdot), h(\cdot)$, is just

$$\text{Var}_{\beta(\cdot), h(\cdot)} [\hat{L}_a] = \sum_{i=1}^n a_i^2 \sigma_i^2$$

To bound the bias, assume $h(\cdot)$ is known to belong in a class of functions \mathcal{H} , then two approaches can be adopted, for the regularity of $\beta(\cdot)$ and the choice of $\mu(\cdot)$:

- 1 arbitrary $\beta(\cdot)$, optimizing weights μ by *moving the goalposts*, s.t. $L_\mu [\beta(\cdot)]$ is easy to estimate (Crump et al., 2006; Imbens and Wager, 2019) which gives the worst-case bias

$$\inf_{\mu} \sup_{\beta(\cdot), h(\cdot)} \left| \sum_{i=1}^n a_i [w_i \beta(z_i) + h(z_i)] - \int \beta(z) d\mu(z) \right| \quad \text{s.t. } h(\cdot) \in \mathcal{H}, \int d\mu(z) = 1 \quad (3.2)$$

- 2 assume constant treatment effects, i.e., $\beta(z) = \beta, \forall z$, which means that $L_\mu [\beta(\cdot)] = \beta$ regardless of μ (Armstrong et al., 2020), and the worst-case bias is

$$\sup_{\beta, h(\cdot)} \left| \sum_{i=1}^n a_i [w_i \beta + h(z_i)] - \beta \right| \quad \text{s.t. } h(\cdot) \in \mathcal{H} \quad (3.3)$$

And, the two approaches can be linked as such:

- If $\sum_{i=1}^n a_i w_i = 1$, 3.2 and 3.3 are both equal to

$$\sup_{h(\cdot)} \left| \sum_{i=1}^n a_i h(z_i) \right| \quad \text{s.t. } h(\cdot) \in \mathcal{H} \quad (3.4)$$

- 3.2 automatically equals 3.4
- 3.3 is optimized (w.r.t. μ) by setting μ to place weight $a_i w_i$ on observation i , i.e., $\mu(\mathcal{Z}) = \sum_{i: z_i \in \mathcal{Z}} a_i w_i$, which implies $\sum_{i=1}^n a_i w_i \beta(z_i) - \int \beta(z) d\mu(z) = 0$, hence the equality.
- Otherwise, 3.2 and 3.3 are both infinite:
 - 3.3 can be made arbitrarily large by choosing large enough β
 - 3.2 can be made arbitrarily large by making $\beta(\cdot)$ constant (as in 3.3) and large enough

References

- Timothy B Armstrong, Michal Kolesár, and Soonwoo Kwon. Bias-aware inference in regularized regression models. *arXiv preprint arXiv:2012.14823*, 2020.
- Richard K Crump, V Joseph Hotz, Guido Imbens, and Oscar Mitnik. Moving the goalposts: Addressing limited overlap in the estimation of average treatment effects by changing the estimand, 2006.
- Guido Imbens and Stefan Wager. Optimized regression discontinuity designs. *Review of Economics and Statistics*, 101(2):264–278, 2019.