Econometrics **January 7, 2023**

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

$$y_i = w_i\beta\left(z_i\right) + h\left(z_i\right) + \epsilon_i \tag{3.1}$$

where

- w_i, z_i are treated as **fixed**
- ϵ_i is **independent**, with $\mathbb{E}\left[\epsilon_i\right] = 0, \mathbb{E}\left[\epsilon_i^2\right] = \sigma_i^2$
- 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

$$L_\mu\left[\beta(\cdot)\right] = \int \beta(z)\mathrm{d}\mu(z)$$

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

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

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

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

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. $L_\mu\left[\beta(\cdot)\right]$ is easy to estimate (Crump et al., 2006; Imbens and Wager, 2019) which gives the worst-case bias

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

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

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

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

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

  - 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 $\sum_{i=1}^{n} a_i w_i \beta(z_i) - \int \beta(z) \mathrm{d}\mu(z) = 0$, 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

3.1.2 Moving-the-goalpost Approach

3.1.3 Constant-treatment-effect Approach

Armstrong et al. (2020) adopt this approach, focusing on the case where $h(\cdot)$ is a high dimensional linear function, and the penalty function is an l_p norm of the coefficients.

Basic setting: Homoskedastic Gaussian errors

First, consider

$$Y = w\beta + Z\gamma + \epsilon \tag{3.5}$$

where

- $\beta \in \mathbb{R}$ is the constant treatment effect to be estimated
- $\gamma \in \Gamma$ is the control coefficients, subject to the restriction (i.e., the function class $\mathcal{H}$)

$$\Gamma = \Gamma(C) = \left\{\gamma \in \mathcal{G} : \mathrm{Pen}(\gamma) \leq C\right\} \tag{3.6}$$

  where $\mathrm{Pen}(\cdot)$ is a seminorm[1] on some linear subspace $\mathcal{G}$ of $\mathbb{R}^k$.
- $w = (w_1, \cdots, w_n)' \in \mathbb{R}^n$ and $Z = \left(z_1', \cdots, z_n'\right)' \in \mathbb{R}^{n\times k}$ are defined as before
- $\epsilon \sim \mathcal{N}\left(0, \sigma^2 I_n\right)$ is assumed **normal and homoskedastic**, with σ^2 known

For estimation, the goal is to construct estimators and CIs for β:

- estimator $\hat{\beta}$: consider the worst-case performance over the parameter space $\mathbb{R}\times\Gamma$ under the **MSE** criterion

$$R_{MSE}\left(\hat{\beta};\Gamma\right) = \sup_{\beta\in\mathbb{R},\gamma\in\Gamma} \mathbb{E}_{\beta,\gamma}\left[\left(\hat{\beta}-\beta\right)^2\right]$$

- for CIs, we have 2 requirements:

  A <u>**coverage**</u>: A $100\cdot(1-\alpha)\%$ CI with half-length $\hat{\chi} = \hat{\chi}(Y, X)$ is an interval $\left\{\hat{\beta} \pm \hat{\chi}\right\}$ s.t.

$$\inf_{\beta\in\mathbb{R},\gamma\in\Gamma} \mathrm{P}_{\beta,\gamma}\left(\beta \in \left\{\hat{\beta} \pm \hat{\chi}\right\}\right) \geq 1-\alpha$$

  B <u>**length**</u>: the exepcted length of a CI $\mathbb{E}_{\beta,\gamma}\left[2\hat{\chi}\right]$ should be as short as possible

  notice that length-optimized CIs are **not** necessarily centered at an MSE-centered $\hat{\beta}$.

Linear estimators and CIs

[1]Seminorm satisfies **triangle inequality** $\mathrm{Pen}\left(\gamma+\tilde{\gamma}\right) \leq \mathrm{Pen}(\gamma)$ and **homogeneity** $\mathrm{Pen}\left(c\gamma\right) = |c|\mathrm{Pen}(\gamma), \forall c$, but **NOT** necessarily positive definite ($\mathrm{Pen}(\gamma) = 0$ does not imply $\gamma = 0$). Essentially, any convex set Γ that is symmetric satisfies this definition.

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.