MOM PS Estimator under Strong Monotonicity

July 6, 2021

Say members of the control group of an RCT have no access to the treatment, but among the treatment group there are compliers and non-compliers—i.e. one-sided non-compliance. (Of course, this could be, e.g., compliance in two different ways, etc.) Then there are two principal strata S, never-takers and compliers. Say the exclusion restriction doesn't hold.

Feller et al. [2017] and Ding and Lu [2017] and others discuss estimating principal effects in one-sided noncompliance (which they call "strong monotonicity") under the assumption of principal ignorability, viz. $Y_C \perp\!\!\!\perp S|\boldsymbol{X}$. This has two problems, to my mind:

1. Sometimes principal ignorability is not plausible
2. If you assume the principal ignorability for Y_T ("strong" principal ignorability) then, as Feller et al. [2017] points out, the principal effects conditional on covariates are equal to the ITT conditional on covariates—the stratum plays no role. But sometimes the role of the strata is what we're trying to estimate!

The approach here takes a different tack, based on a different assumption which is maybe even less plausible—but which, I think, can be relaxed.

1 Setup

Estimating $\mathbb{E}[Y_T|S]$ is straightforward, so for the remainder just consider the "control" group, $Z = 0$, with n (control) subjects and drop the C subscript. The control group is a mixture of compliers and never-takers.

For subject i, $i = 1, \ldots, n$, we have:

- Y_i observed outcome
- $S_i \in \{0, 1\}$ unobserved stratum

- $e(\boldsymbol{X_i}) = Pr(S_i = 1|\boldsymbol{X}_i)$ take as given

Let $\mu_0 = \mathbb{E}[Y|S = 0]$ and $\mu_1 = \mathbb{E}[Y|S = 1]$. The goal is to estimate μ_0 and μ_1.

Assumption:

$$\mathbb{E}[Y_i|\boldsymbol{X}_i, S_i] = \mathbb{E}[Y_i|S_i] = \mu_0 \text{ or } \mu_1 \tag{1}$$

i.e. Y is mean-independent of $\boldsymbol{X}$ conditional on S. This is problematic—why should S contain all information about Y?

Note, it is kinda related to "principal ignorability" of Feller et al. [2017] and Ding and Lu [2017], $Y \perp\!\!\!\perp S|\boldsymbol{X}_i$ but, obviously, different too.

The principal scores will typically be estimated using data from the treatment group. Lemmas 1 & 2 of Feller et al. 2016 imply that $Pr(S = 1|\boldsymbol{X}, Z) = Pr(S = 1|\boldsymbol{X})$, so that's cool. The math below ignores estimation error in principal scores, treating $e(\boldsymbol{X})$ as known. However, as long as principal scores are estimated consistently, the asymptotics work out by Slutsky's theorem.

2 M-Estimator

As a preliminary, note that

$$\begin{aligned}
\mathbb{E}[Y|e(\boldsymbol{X})] &= \mathbb{E}\left\{\mathbb{E}[Y|e(\boldsymbol{X}), S]|e(\boldsymbol{X})\right\} \\
&= \mathbb{E}\left\{\mathbb{E}[Y|S]|e(\boldsymbol{X})\right\} \text{ by (1)} \\
&= \mathbb{E}[\mu_1 S + \mu_0(1 - S)|e(\boldsymbol{X})] \\
&= \mu_1 e(\boldsymbol{X}) + \mu_0(1 - e(\boldsymbol{X}))
\end{aligned}$$

Then we have

$$\mathbb{E}[Y] = \mathbb{E}\mathbb{E}[Y|e(\boldsymbol{X})] = \mu_1 \mathbb{E}e(\boldsymbol{X}) + \mu_0(1 - \mathbb{E}e(\boldsymbol{X}))$$

Next we have

$$\begin{aligned}
\mathbb{E}[Ye(\boldsymbol{X})] &= \mathbb{E}\left\{\mathbb{E}[Ye(\boldsymbol{X})|e(\boldsymbol{X})]\right\} \\
&= \mathbb{E}\left\{e(\boldsymbol{X})\mathbb{E}[Y|e(\boldsymbol{X})]\right\} \\
&= \mathbb{E}\left\{e(\boldsymbol{X})\left[\mu_1 e(\boldsymbol{X}) + \mu_0(1 - e(\boldsymbol{X}))\right]\right\} \\
&= \mu_1 \mathbb{E}[e(\boldsymbol{X})^2] + \mu_0\left(\mathbb{E}[e(\boldsymbol{X})] - \mathbb{E}[e(\boldsymbol{X})^2]\right)
\end{aligned}$$

That gives us four parameters:

$$\begin{aligned}\theta_1 &= \mathbb{E}[e(\boldsymbol{X})] \\ \theta_2 &= \mathbb{E}[e(\boldsymbol{X})^2] \\ \theta_3 &= \mu_0 \\ \theta_4 &= \mu_1\end{aligned}$$

and four estimating equations:

$$\begin{aligned}\sum_i e(\boldsymbol{X_i}) - \theta_1 &= 0 \\ \sum_i e(\boldsymbol{X_i})^2 - \theta_2 &= 0 \\ \sum_i Y_i - \theta_4\theta_1 - \theta_3(1-\theta_1) &= 0 \\ \sum_i Y_i e(\boldsymbol{X_i}) - \theta_4\theta_2 - \theta_3(\theta_1-\theta_2) &= 0\end{aligned}$$

3 Relaxing (1) with regression

There's no particular reason to assume (1) and it seems like it would typically be pretty implausible.

But say you believe the model

$$Y_i = \mu_1 S_i + \mu_0(1-S_i) + \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i \tag{2}$$

Then instead of (1) we could assume something like

$$\mathbb{E}[Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}|\boldsymbol{X}_i, S_i] = \mathbb{E}[Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}|S_i]$$

Under model (2), we have new estimating equations.

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}\left[\mu_1 e(\boldsymbol{X}) + \mu_0(1-e(\boldsymbol{X})) + \boldsymbol{X}\boldsymbol{\beta}\right] \\ \mathbb{E}[Y e(\boldsymbol{X})] &= \mathbb{E}\left[e(\boldsymbol{X})\left(\mu_1 e(\boldsymbol{X}) + \mu_0(1-e(\boldsymbol{X})) + \boldsymbol{X}\boldsymbol{\beta}\right)\right] \\ \mathbb{E}[\boldsymbol{X}Y]\mathbb{E}\left[\boldsymbol{X}\left(\mu_1 e(\boldsymbol{X}) + \mu_0(1-e(\boldsymbol{X})) + \boldsymbol{X}\boldsymbol{\beta}\right)\right]\end{aligned}$$

in other words, no interaction between $\boldsymbol{X}$ and S in (1). Then, if you had an estimate for β, you could just substitute $Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ for Y_i in the estimates. Alternatively, you could use a stacked estimating equation approach and estimate it all together.

4 Simulation

4.1 Design

4.1.1 Data Generation

Three indpendent standard normal covariates, x_k $k = 1, \ldots, 3$. Principal scores are set as:

$$e(\boldsymbol{X_i}) = logit^{-1}\left[\beta(x_{1i} + x_{2i} + x_{3i})\right]$$

Principal stratum $S_i \in \{0, 1\}$ generated as:

$$S_i \sim Bern(e(\boldsymbol{X_i}))$$

where β is a manipuated factor. When β was higher, S was more easily predicted by covariates x_k.

Outcomes are generated as

$$Y_i = 0.5(x_{1i} + x_{2i} + x_{3_i}) + \mu_{01}S_i + \tau_i Z_i + \epsilon_i$$

where Z_i is treatment assignment $Z_i = i\%2$, so that half of subjects are in the treatment group, τ_i is the treatment effect,

$$\tau_i = \begin{cases} \mu_{10} - 0 & \text{if } S_i = 0 \\ \mu_{11} - \mu_{01} & \text{if } S_i = 1 \end{cases}$$

and $\epsilon_i \sim \mathcal{N}(0, 0.2)$ or ϵ_i has a Gumbel distribution with location parameter 0 and scale parameter 0.16 (corresponding to a standard deviation of approximately 0.2). Errors ϵ and covariates x_k are centered, so each has sample mean exactly 0.

4.1.2 Manipulated Factors

We manipulated four factors in the simulation:

- Sample size $n \in \{100, 500, 1000\}$
- $\mu_{01} \in \{0, 0.3\}$
- $\mu_{10} \in \{0, 0.3\}$
- $\beta \in \{0, 0.2, 0.5, 1\}$

$\mu_{11} = 0.3$ in all cases.

When $\mu_{01} = 0$ there is only one mixture component in the control group—i.e. no separation. When $\mu_{01} = \mu_{01} = 0$ the treatment effect is 0.3 for $S = 1$ and 0 for $S = 0$. When $\mu_{01} = \mu_{10} = 0.3$, the treatment effect is 0.3 for $S = 0$ and 0 for $S = 1$. When $\mu_{01} = 0.3$ and $\mu_{10} = 0$, the treatment effect is 0 for both strata, and when $\mu_{01} = 0$ and $\mu_{10} = 0.3$ the effect is 0.3 for both strata.

4.1.3 Analysis Models

Analysis models had access to treatment assignment Z, outcomes Y, principal stratum S for members of the treatment group, and covariates. Even though three covariates, x_1, x_2 and x_3 were included as predictors of principal stratum S and outcome Y in the data generating model, both analysis approaches only had access to the first two covariates, x_1 and x_2. Hence, principal ignorability was violated—S and Y were "confounted" by unobserved variable x_3.

Simulated data were analyzed first with a "stacked equations" /M-Estimation/GEE approach, using the `geex` package in `R` [Saul and Hudgens, 2020, R Core Team, 2020], and then with a more conventional maximum likelihood approach via `rstan` [Stan Development Team, 2020].

Both methods assumed (correctly) that there were no interactions between covariates x_1 and x_2 and treatment assignment (after accounting for S).

In the GEE approach included a set of nine estimating equations. First, a GEE logistic regression fit to the treatment group to estimate principal scores $e(\boldsymbol{X_i}) = logit^{-1}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})$ (three parameters):

$$\sum_i Z_i \left[S_i - e(\boldsymbol{X_i})\right] = 0$$

$$\sum_i Z_i x_{1i} \left[S_i - e(\boldsymbol{X_i})\right] = 0$$

$$\sum_i Z_i x_{2i} \left[S_i - e(\boldsymbol{X_i})\right] = 0$$

Next, four equations to estimate the outcome regression in the treatment

group, including terms for S:

$$\sum_i Z_i \left[Y_i - (\gamma_1 x_{1i} + \gamma_2 x_{2i} + S_i \mu_{11} + (1 - S_i)\mu_{10})\right] = 0$$
$$\sum_i Z_i x_{1i} \left[Y_i - (\gamma_1 x_{1i} + \gamma_2 x_{2i} + S_i \mu_{11} + (1 - S_i)\mu_{10})\right] = 0$$
$$\sum_i Z_i x_{2i} \left[Y_i - (\gamma_1 x_{1i} + \gamma_2 x_{2i} + S_i \mu_{11} + (1 - S_i)\mu_{10})\right] = 0$$
$$\sum_i Z_i S_i \left[Y_i - (\gamma_1 x_{1i} + \gamma_2 x_{2i} + S_i \mu_{11} + (1 - S_i)\mu_{10})\right] = 0$$

Finally, two equations to estimate the mixture model in the control group, for whom S is unobserved:

$$\sum_i (1 - Z_i) \left[Y_i - (\gamma_1 x_{1i} + \gamma_2 x_{2i} + e(\boldsymbol{X_i})\mu_{01} + (1 - e(\boldsymbol{X_i}))\mu_{00})\right] = 0$$
$$\sum_i (1 - Z_i) \left[Y_i - (\gamma_1 x_{1i} + \gamma_2 x_{2i} + e(\boldsymbol{X_i})^2\mu_{01} + (e(\boldsymbol{X_i}) - e(\boldsymbol{X_i})^2)\mu_{00})\right] = 0$$

The average treatment effect for $S = 0$ was $\mu_{10} - \mu_{00}$ and for $S = 1$ was $\mu_{11} - \mu_{01}$.

Next, we estimated effects with MLE, using the built-in optimizer of `rstan`.

In this model we regressed S and Y for the treatment group on an intercept, x_1, and x_2 using a logistic and normal linear specification, respectively. Then, we specified a two-component mixture model for control outcomes.

Code for both models and for the data generating process can be found at `https://github.com/adamSales/psGee`.

4.2 Results

References

Avi Feller, Fabrizia Mealli, and Luke Miratrix. Principal score methods: Assumptions, extensions, and practical considerations. *Journal of Educational and Behavioral Statistics*, 42(6):726–758, 2017.

P. Ding and J. Lu. Principal stratification analysis using principal scores. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:757–777, 2017.

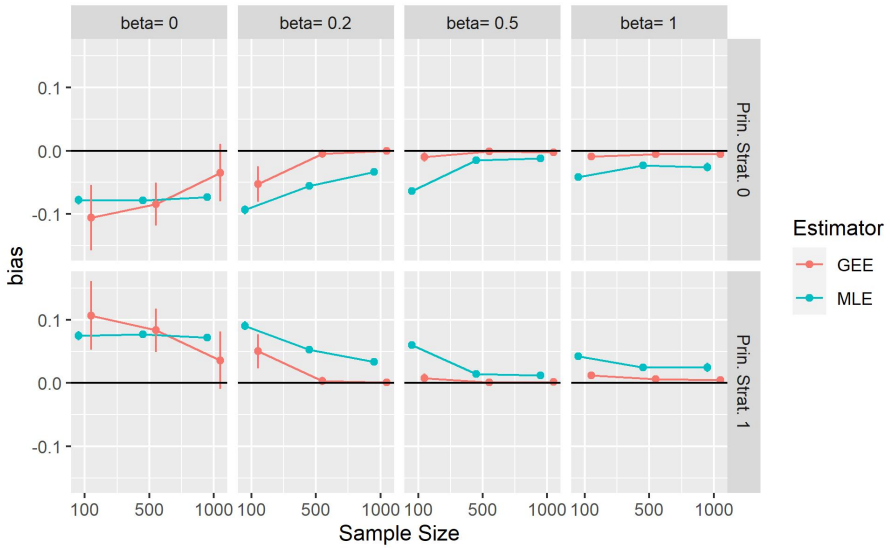

Figure 1: Bias as a function of β, n, and analysis model, pooling over μ_{10} and μ_{01}.

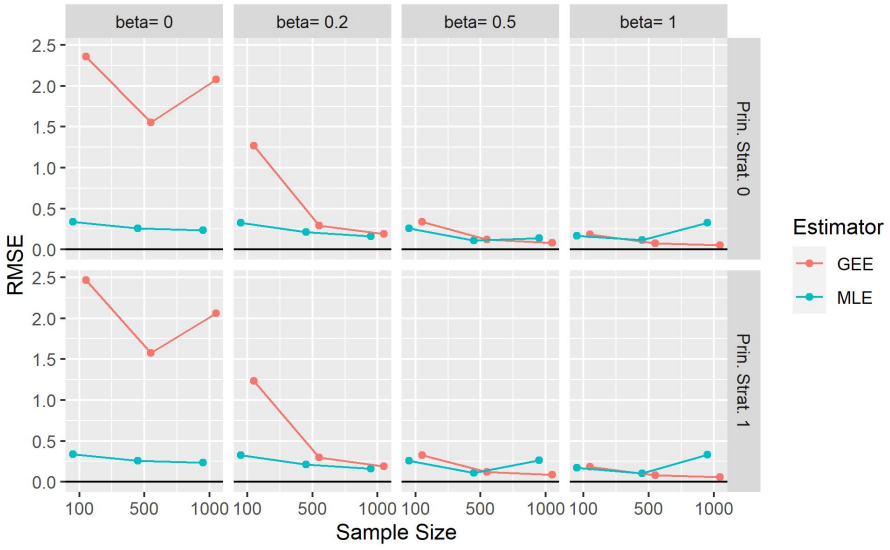

Figure 2: Root mean sequared error (RMSE) as a function of β, n, and analysis model, pooling over μ_{10} and μ_{01}.

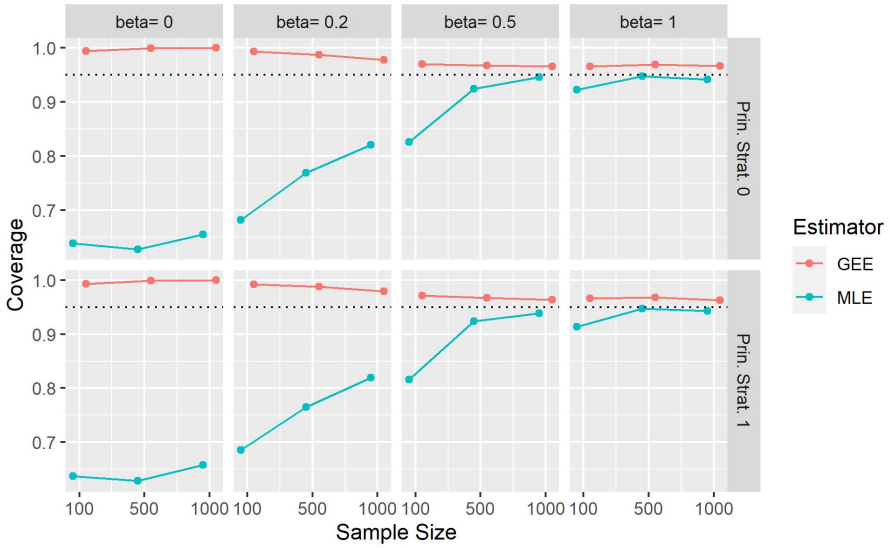

Figure 3: Coverage of nominal 95% confidence intervals as a function of β, n, and analysis model, pooling over μ_{10} and μ_{01}.

Avi Feller, Todd Grindal, Luke Miratrix, and Lindsay C. Page. Compared to what? Variation in the impacts of early childhood education by alternative care type. *The Annals of Applied Statistics*, 10(3):1245 – 1285, 2016. doi: 10.1214/16-AOAS910. URL `https://doi.org/10.1214/16-AOAS910`.

Bradley C. Saul and Michael G. Hudgens. The calculus of m-estimation in R with geex. *Journal of Statistical Software*, 92(2):1–15, 2020. doi: 10.18637/jss.v092.i02.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

Stan Development Team. RStan: the R interface to Stan, 2020. URL `http://mc-stan.org/`. R package version 2.21.2.