Recap of Fall 2023

Fall 2023 started with the goal of showing that $\hat{\theta}_{prop}$ in Equation 1 is (or is not) an optimal estimator.

$$\begin{aligned}\hat{\theta}_{prop} = & \\ & n^{-1}\sum_{i=1}^{n} E[g_i \mid X_i] + n^{-1}\sum_{i=1}^{n} \frac{\delta_{1+}}{\pi_{1+}}(E[g_i \mid X_i, Y_{1i}] - E[g_i \mid X_i]) \\ & + n^{-1}\sum_{i=1}^{n} \frac{\delta_{2+}}{\pi_{2+}}(E[g_i \mid X_i, Y_{2i}] - E[g_i \mid X_i]) \\ & + n^{-1}\sum_{i=1}^{n} \frac{\delta_{11}}{\pi_{11}}(g_i - E[g_i \mid X_i, Y_{1i}] - E[g_i \mid X_i, Y_{2i}] + E[g_i \mid X_i]).\end{aligned} \tag{1}$$

While this estimator performs quite well overall, it is outperformed by $\hat{\theta}_{\delta}$ (see Table 1) in a simulation where the number of observations in segments A_{10} differed from A_{01}. Studying this estimator led to the creation of a class of estimators in Equation 2 with specific examples in Table 1.

$$\hat{\theta} = \frac{\delta_{11}}{\pi_{11}} g(Z) + \beta_0(\delta, c_0)E[g(Z) \mid X] + \beta_1(\delta, c_1)E[g(Z) \mid X, Y_1] + \beta_2(\delta, c_2)E[g(Z) \mid X, Y_2]. \tag{2}$$

Table 1: Specific examples of estimators from the larger class in Equation 2.

Estimator	$\beta_0(\delta, c_0)$	$\beta_1(\delta, c_1)$	Implemented
$\hat{\theta}_{prop}$	$\left(1 - \frac{(\delta_{10}+\delta_{11})}{(\pi_{10}+\pi_{11})} - \frac{(\delta_{01}+\delta_{11})}{(\pi_{01}+\pi_{11})} + \frac{\delta_{11}}{\pi_{11}}\right)$	$\left(\frac{\delta_{10}+\delta_{11}}{\pi_{10}+\pi_{11}} - \frac{\delta_{11}}{\pi_{11}}\right)$	✓
$\hat{\theta}_{prop}^{ind}$	$\left(1 - \frac{(\delta_{10})}{(\pi_{10})} - \frac{(\delta_{01})}{(\pi_{01})} + \frac{\delta_{11}}{\pi_{11}}\right)$	$\left(\frac{\delta_{10}}{\pi_{10}} - \frac{\delta_{11}}{\pi_{11}}\right)$	✓
$\hat{\theta}_{c}$	$c_0\left(1 - \frac{(\delta_{10}+\delta_{11})}{(\pi_{10}+\pi_{11})} - \frac{(\delta_{01}+\delta_{11})}{(\pi_{01}+\pi_{11})} + \frac{\delta_{11}}{\pi_{11}}\right)$	$c_1\left(\frac{\delta_{10}+\delta_{11}}{\pi_{10}+\pi_{11}} - \frac{\delta_{11}}{\pi_{11}}\right)$	✓
$\hat{\theta}_{c,ind}$	$c_0\left(1 - \frac{(\delta_{10})}{(\pi_{10})} - \frac{(\delta_{01})}{(\pi_{01})} + \frac{\delta_{11}}{\pi_{11}}\right)$	$c_1\left(\frac{\delta_{10}}{\pi_{10}} - \frac{\delta_{11}}{\pi_{11}}\right)$	✓
$\hat{\theta}_{\delta}$	$c_0\left(\frac{\delta_{11}}{\pi_{11}} - \frac{\delta_{00}}{\pi_{00}}\right)$	$c_1\left(\frac{\delta_{11}}{\pi_{11}} - \frac{\delta_{10}}{\pi_{10}}\right)$	✓

The challenge with working with this class is the β coefficients. These are functional coefficients and so it is difficult for me to understand how to show that a particular class is optimal.

Current Work

As discussed in a related note, we are currently trying to understand the loss of efficiency of using a semiparametric estimator with the model is correctly specified. I am having difficulties with the current simulation setup but this work is contained in `efficiencyloss_semi.tex`.

Potential Ideas to Pursue

- Tsiatis (2006) describes a recursive method to get the optimal semiparametric estimator for non-monotone data. This is difficult and confusing. However, we can pose the estimation as the solution to an integral equation. While people know that one can do this, it has not been done. The real work would be estimating the variance.

- Often in simulations, the optimal value for c_i would be 1 or -1. Since the optimal semiparametric estimator is a double projection (see Chapter 10 of Tsiatis (2006)), I hypothesis that it may be possible to construct a test to see how close to the optimal we are by projecting the the β coefficients onto some contants c.