

Double Robustness

Rachael Caelie (Rocky) Aikens

4/10/2020

Double Robustness and Fisher Mill

I've been thinking a little about a visualization for double robustness. Let's consider a data generating process much like our usual:

$$\begin{aligned} X_i &\sim_{iid} \text{Normal}(0, I_p), \\ T_i &\sim_{iid} \text{Bernoulli}\left(\frac{1}{1 + \exp(-\phi(X_i))}\right), \\ Y_i &= \tau T_i + \Psi(X_i) + \epsilon_i, \\ \epsilon_i &\sim_{iid} N(0, \sigma^2), \end{aligned}$$

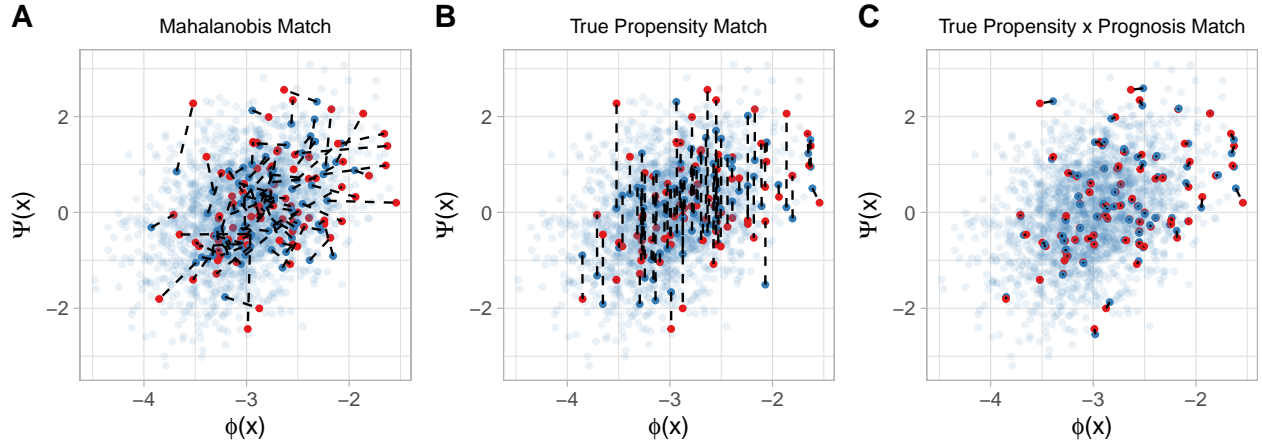
Except that the true propensity score now depends on two variables: X_1 and X_3 ,

$$\begin{aligned} \phi(X_i) &= X_{i1}/3 + X_{i3} - c, \\ \Psi(X_i) &= \rho X_{i1} + \sqrt{(1 - \rho^2)} X_{i2}, \end{aligned}$$

This way we can think a little bit about what happens when the model isn't perfectly specified.

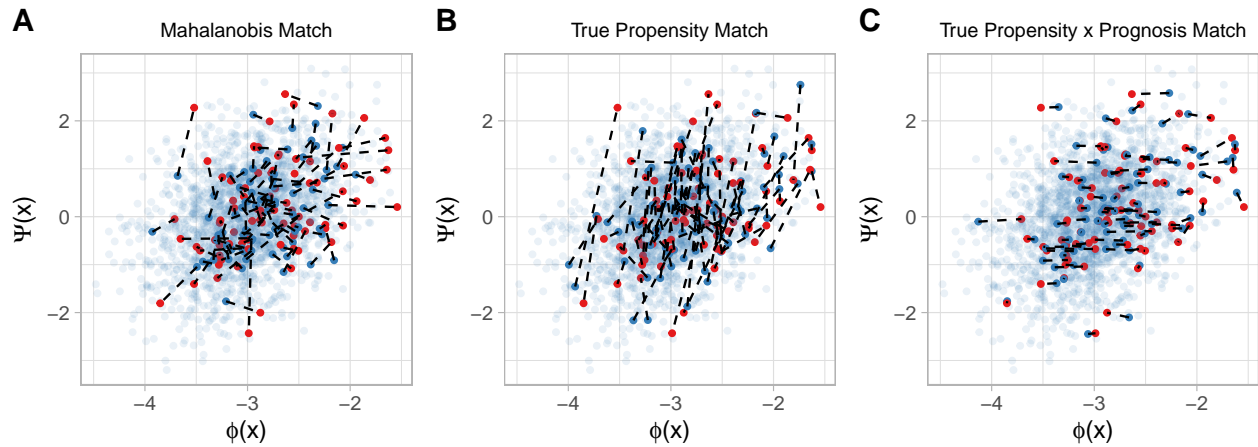
Baseline: All models are correct

As usual, below is the plot of the true optimal matches in each scenario:



Propensity score mis-specified, Prognostic score correct

Now, let's suppose that we have obtained a perfect prognostic score model, but our propensity score model is missing information. In particular, let's assume that we have perfectly estimated the effect of X_1 on treatment assignment, but we have entirely missed X_3 in our model. Below we see how the propensity score matches are a lot more variable in quality, but matching jointly on the prognostic score helps ameliorate some of those problems.



Prognostic score mis-specified, Propensity score correct

And below, let's suppose the opposite scenario: our model for propensity is perfect, but our model for prognosis is missing the contribution of X_1 .

