

Supplementary Figures

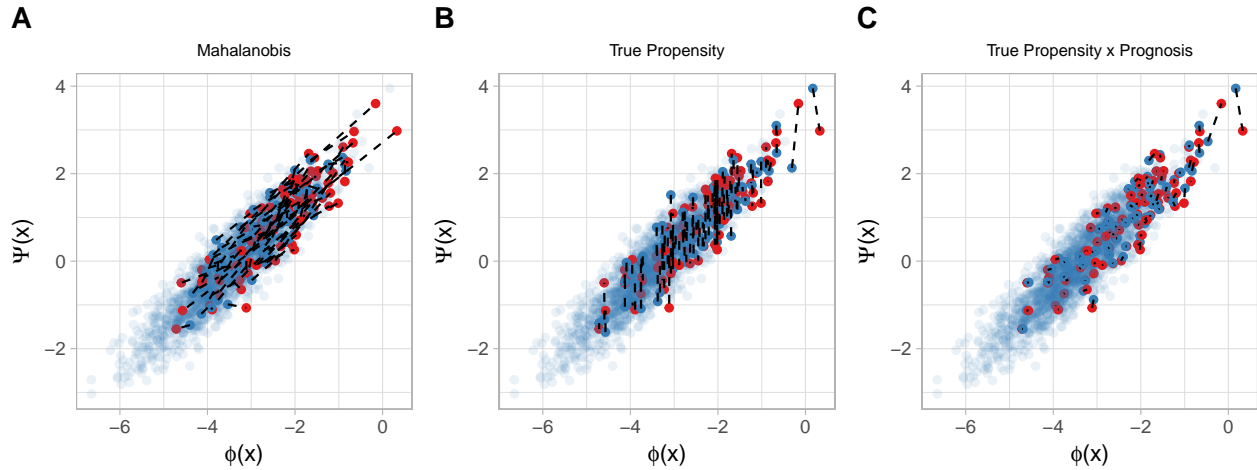
Rachael Caelie (Rocky) Aikens

5/8/2019

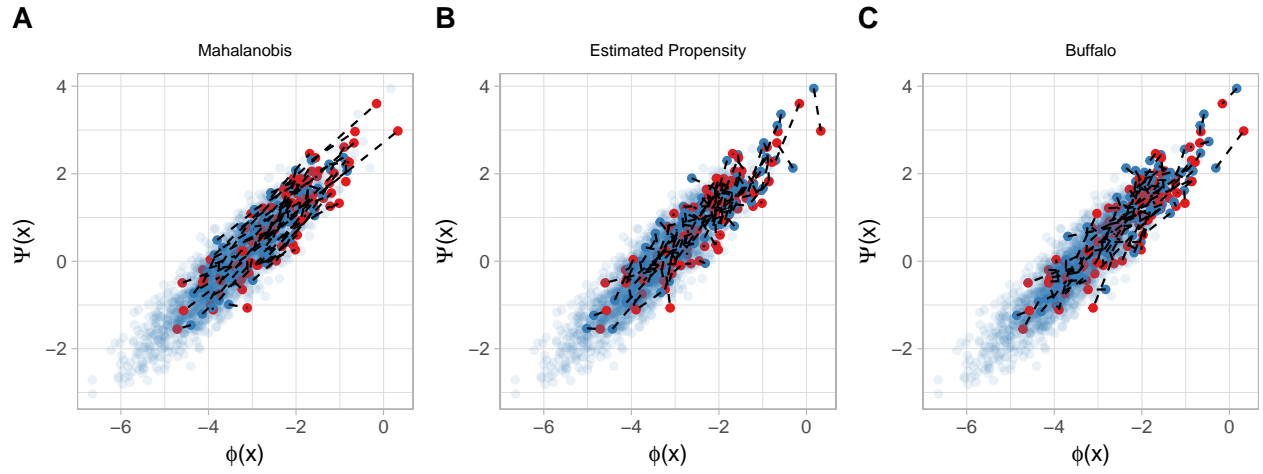
Supplementary Figures 1 and 2: Motivating visualization when prognosis and treatment are highly corellated.

We made the observation that variance actually increased with k when ρ was large. One explanation that we thought of was that, when ρ is large, propensity and prognosis are highly correlated; As we match more and more control individuals to each individual *all* methods are selecting matches which are more distant in prognosis and in propensity. The propensity distance increases the bias of the estimator, while the prognosis distance increases the variance of the estimator. In the visualization plots above, we only show 1:1 matching. This problem is likely to worsen as k increases, explaining an increase in both bias and variance as k increases.

I would predict that this issue is a bigger deal when X_1 is more strongly predictive of treatment assignment (i.e. when treatment assignment is less random and more based on covariates), since this problem is mostly an issue in the regions of limited overlap. This would explain why we saw variance increase with k when $\text{logit}(e(x)) = X_1 - 10/3$, but not when $\text{logit}(e(x)) = X_1/3 - 4/3$.



Now, we take the same data set and imagine that we don't know the true propensity and prognostic scores, and so must build them empirically. For propensity score matching, we build a logistic regression of treatment assignment on all of the variables, using the entire dataset. For buffalo, we use the propensity score logistic regression in addition to a prognostic score, which we fit on a subset of the controls that are chosen to be good matches to the treated individuals based on mahalanobis distance.



Supplementary Figure 3

Bias and MSE for MDM, propensity, and prognostic score matching when $p = 50$.

To do