Full Matching

Rachael Caelie (Rocky) Aikens

5/15/2019

Set Up

We compare the performance of propensity score matching, Mahalanobis distance matching, and Buffalo Matching (described in the previous section) on simulated data, varying the dimensionality of the problem, the fixed treatment to control ratio during matching, and the correlation between the true propensity and prognostic score. The generative model for all of our simulations is the following:

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

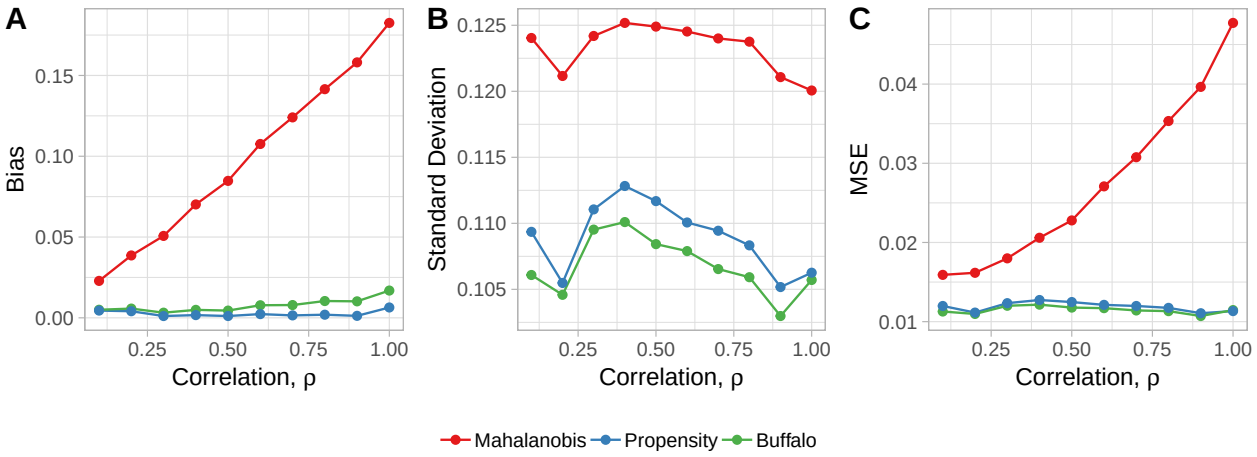
where the true propensity and prognositic scores are given by the linear combinations

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

so that $\text{Cor}(\phi(X_i), \Psi(X_i)) = \rho$. The constant, c, in the propensity score formula was chosen such that there were approximately 100 treated observations in each dataset. We consider $p = 10$, $\rho = 0, 0.1, \ldots, 0.9, 1.0$. Each simulation consisted of a dataset of size $n = 2000$ and was repeated $N = 2000$ times. We fix the treatment effect to be constant with $\tau = 1$ and the noise to be $\sigma = 1$. For a given matching, we estimate ATT and design sensitivity $\tilde{\Gamma}$ using the permutation t-statistic from the package `sensitivtymv`.

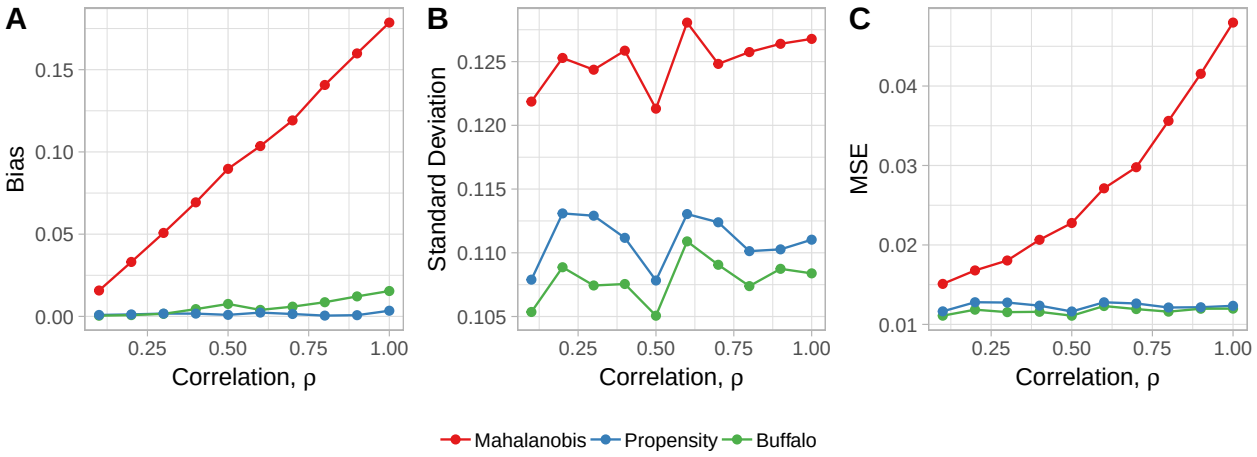
Results

Below, we use the simulation parameters described in the set up to estimate the bias, variance and mse of effect estimates produced from full matching on Mahalanobis distance, propensity score, and prognostic-by-propensity distance via buffalo. Effect estimates (ATE) are based on the fixed effect from a linear mixed model, fit using `lme4`. We did not calculate gamma sensitivity, since this is not possible using this method.



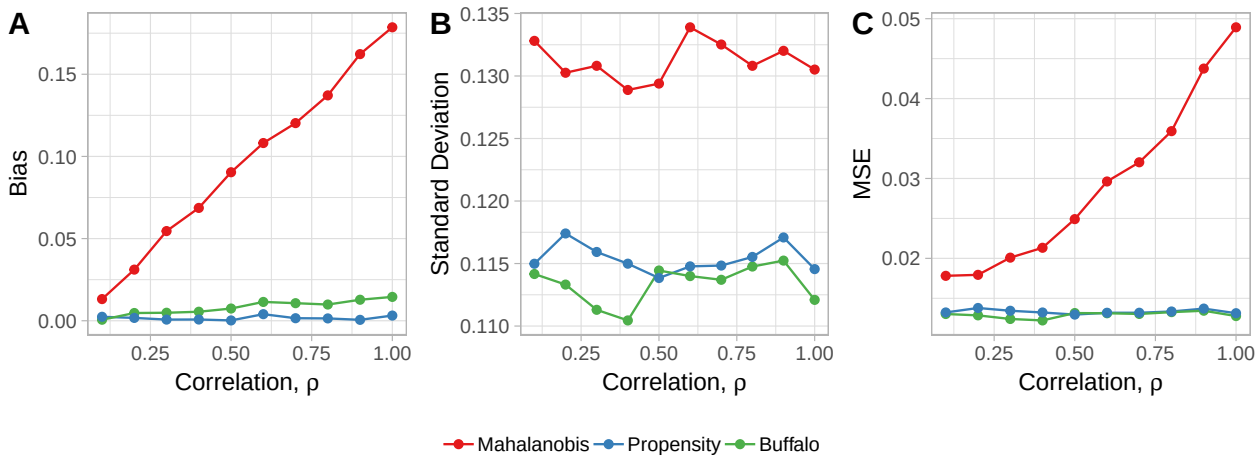
Smaller sample size

Below, we perform the same analysis with a reduced sample size of $n = 1600$. The propensity formula, ϕ was adjusted to $X_{i1}/3 - 2.75$, so that there would still be approximately 100 treated individuals in each sample. Interestingly, the results are pretty much the same.



Smaller sample size

Here, I try again with n=1000. Buffalo is *still* outperforming propensity score in terms of mean squared error, although the difference is small. When the extra gains in gamma sensitivity are considered, this means that Buffalo matching could be a reasonable alternative when the sample size is as low as 1000 (recall, this is a 1:9 treat:control ratio.)


A
Bias
0.15
0.10
0.05
0.00
0.25
0.50
0.75
1.00
Correlation, ρ
B
Standard Deviation
0.135
0.130
0.125
0.120
0.115
0.110
0.25
0.50
0.75
1.00
Correlation, ρ
C
MSE
0.05
0.04
0.03
0.02
0.25
0.50
0.75
1.00
Correlation, ρ
Mahalanobis
Propensity
Buffalo