Simulations with SITA Violation

Rachael Caelie (Rocky) Aikens

2/13/2020

Set Up with SITA violations

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. For this set of simulations, we also add a weak, unobserved confounder, U.

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

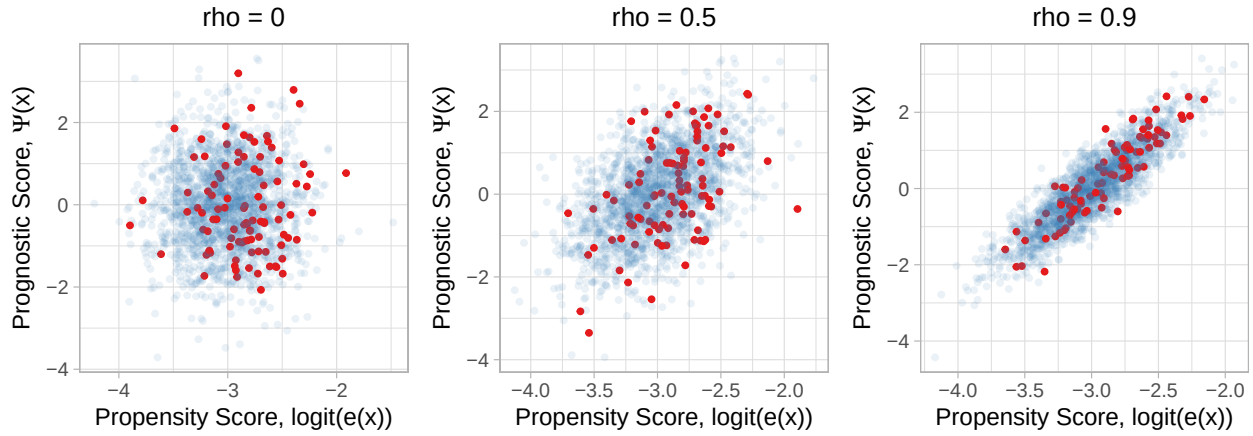
where the true propensity and prognositic scores are given by the linear combinations

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

The constant, c in the propensity score formula was chosen such that there were approximately 100 treated observations in each dataset. For the simulations reported in the main figures of the paper, we let $c = 3$. We consider $p = 10$, $\rho = 0, 0.1, \ldots, 0.9, 1.0$, and $k = 1, \ldots, 10$. Each simulation consisted of a dataset of size $n = 2000$ and was repeated $N = 1000$ 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`

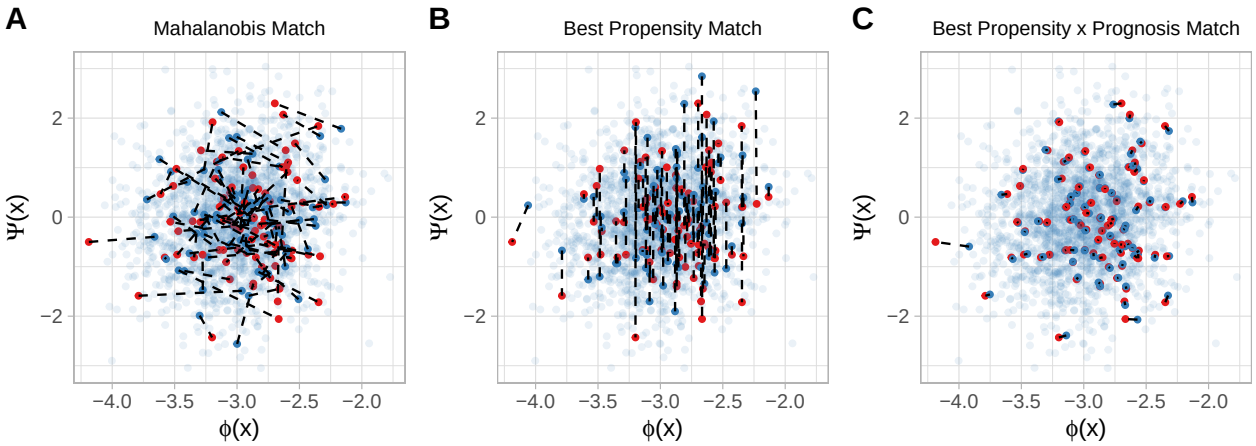
Fisher-Mill Visualizations

This set-up is a little weird. Here are some Fisher-Mill plots of the resulting data:

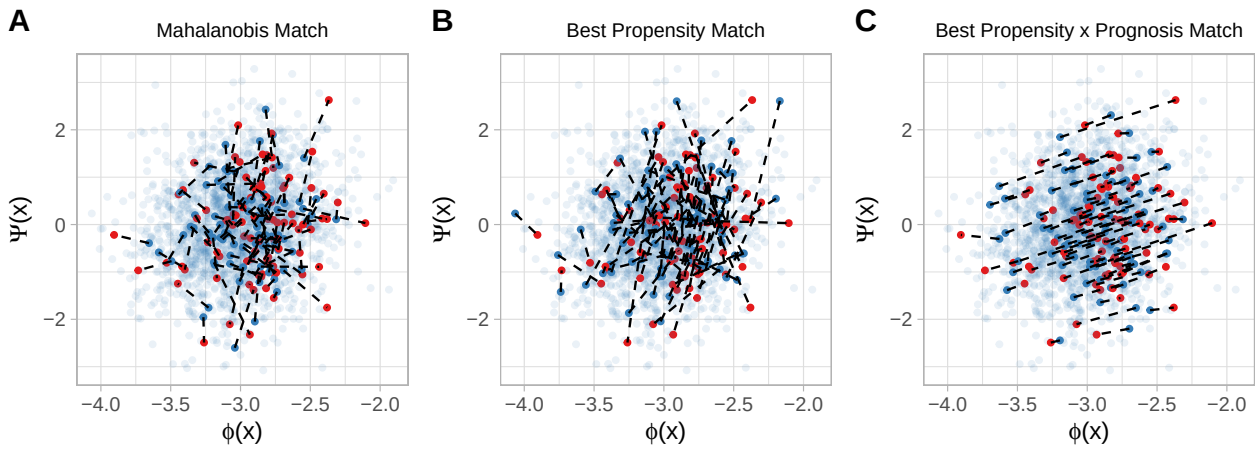


But there's something sinister going on here. Suppose we fit the best possible model for propenisity and prognostic score. That is, our coefficients for X1 and X2 in our propensity and prognostic models are exactly correct, but we leave out U in the model because we never observed it. The plots below show the matches we might choose in this scenario under Mahalanobis, Propensity, and Joint Propensity and Prognostic score matching. As you can see, we are missing some amount of variation between matched individuals.

If we made Fisher Mill plots from our scores and matched on them, this what we'd see:



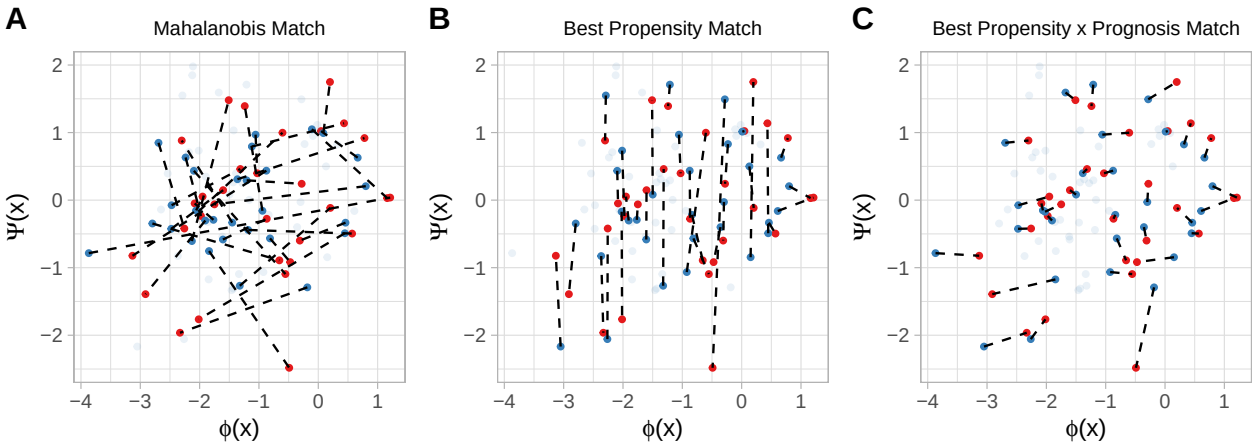
But of course, the Fisher-Mill plots we make and the matches we select are missing the unobserved confounder, U. If we made true Fisher-Mill plots based on the data generating functions, we'd see that the above matching scheme is not doing as well as we thought.


A
Mahalanobis Match
B
Best Propensity Match
C
Best Propensity x Prognosis Match
Ψ(x)
ϕ(x)
2
0
−2
−4.0
−3.5
−3.0
−2.5
−2.0

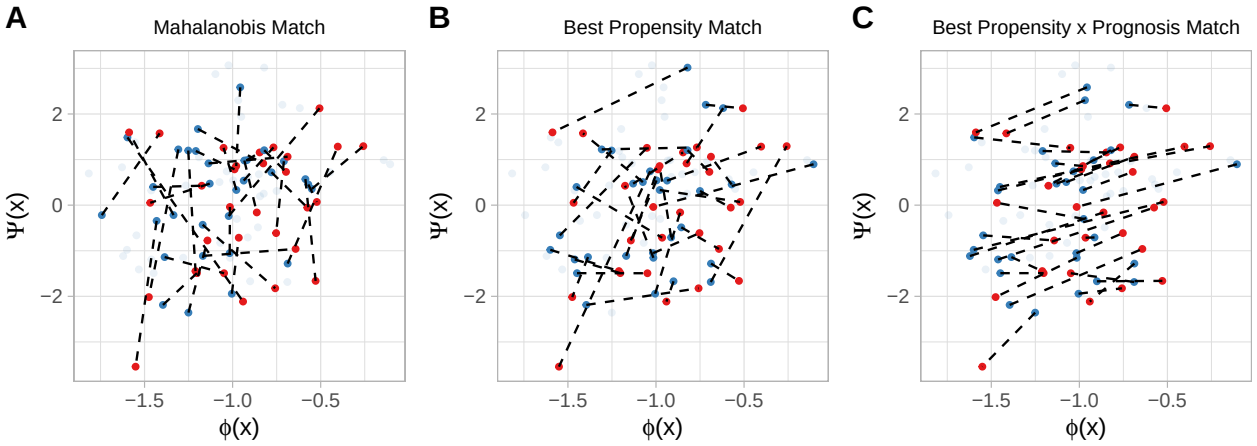
More intuition for gamma and confounding

To try and understand this better, I thought it might help to think about a smaller sample size. The sample below has 100 observations, and I've tinkered with the probability of treatment so that there are about 20-30 treated individuals.

Again, let's suppose our propensity and prognostic models are fit perfectly on the observed covariates. If we made Fisher Mill plots from our scores and matched on them, this what we'd see:

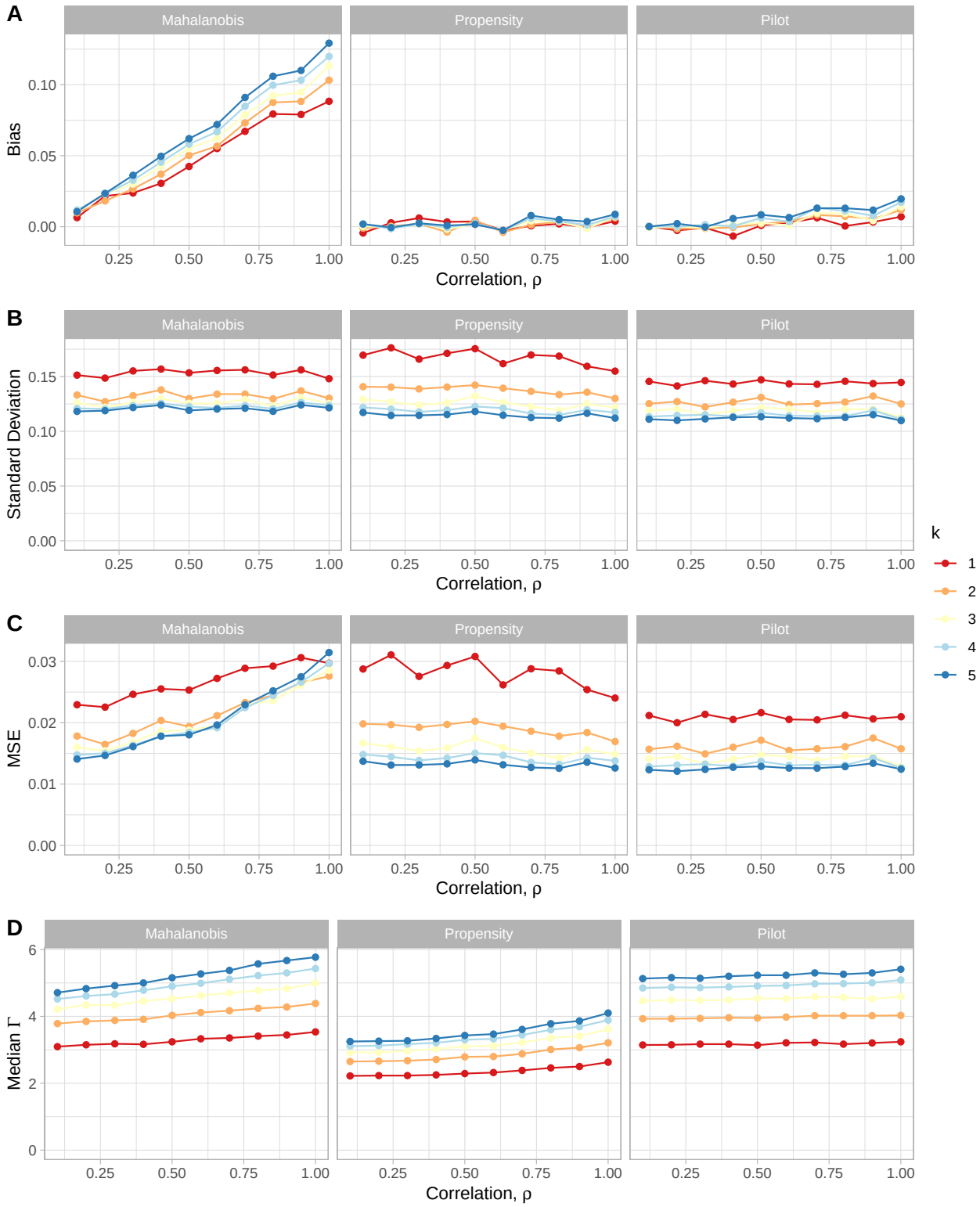


But of course, the Fisher-Mill plots we make and the matches we select are missing the unobserved confounder, U. If we made true Fisher-Mill plots based on the true data generating functions (with U), this is how we're actually doing.

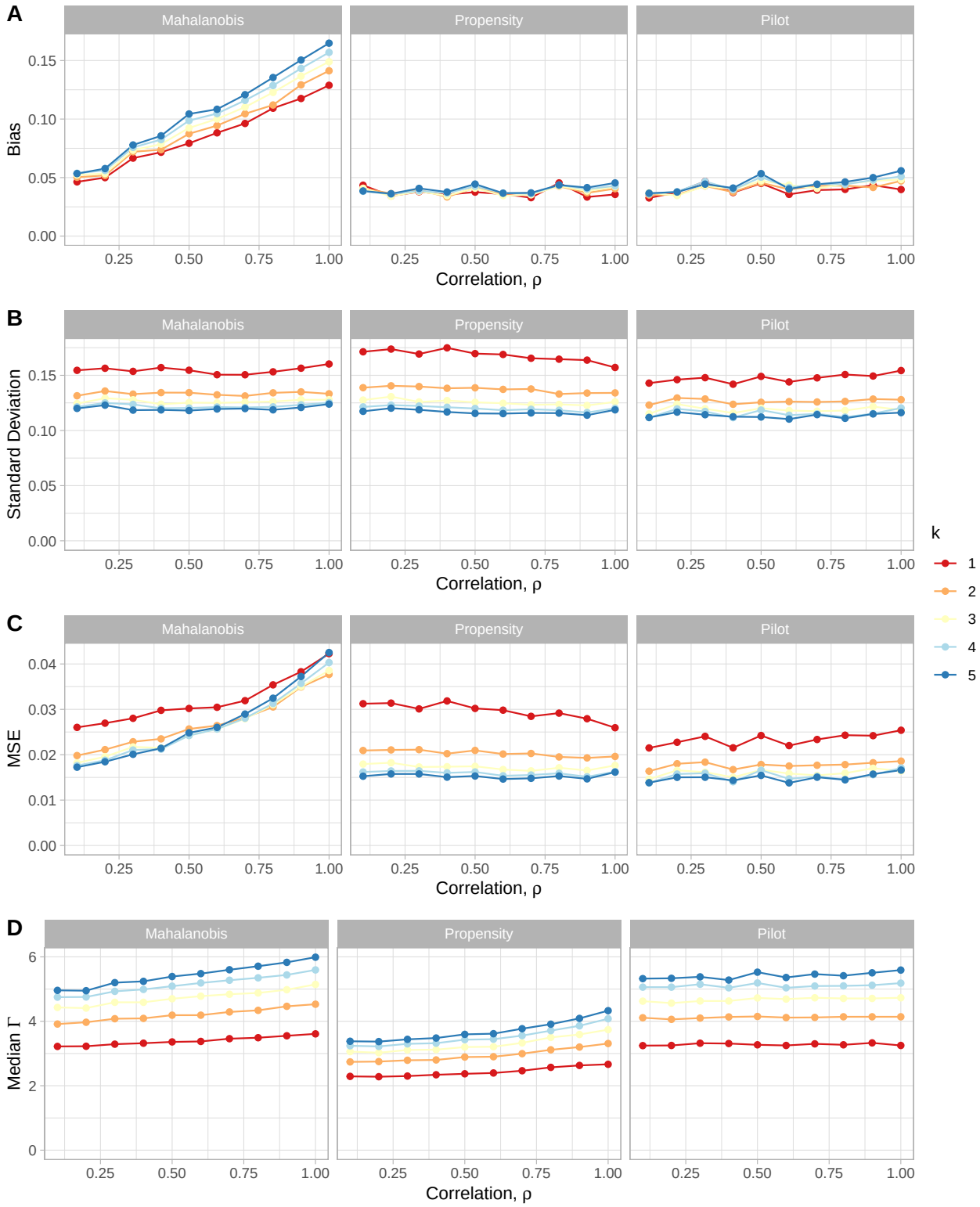


Yeah, I'm not having any brilliant insights.

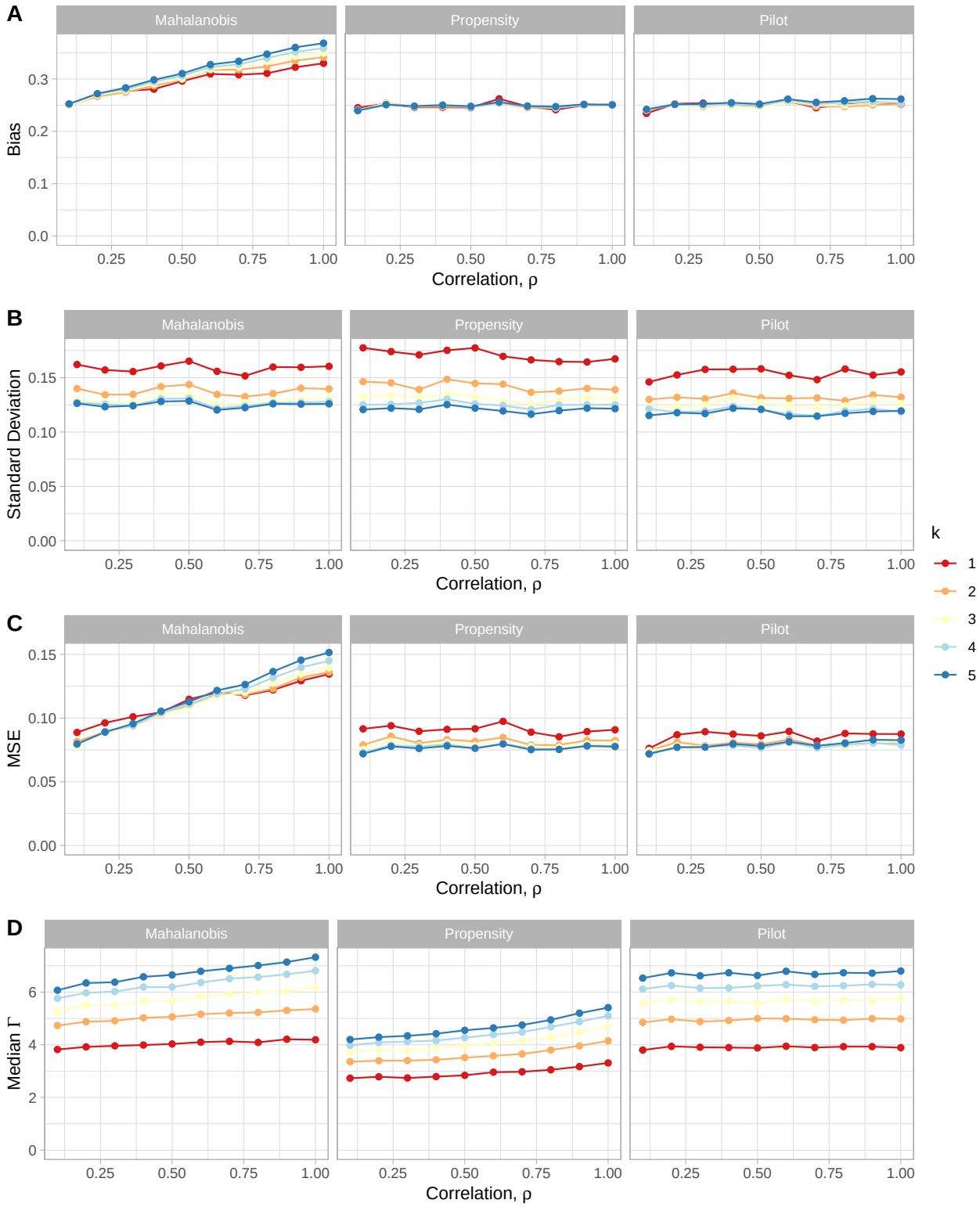
Results with no SITA violation



$\mathbf{tau = 1,\ nu = 0.2}$

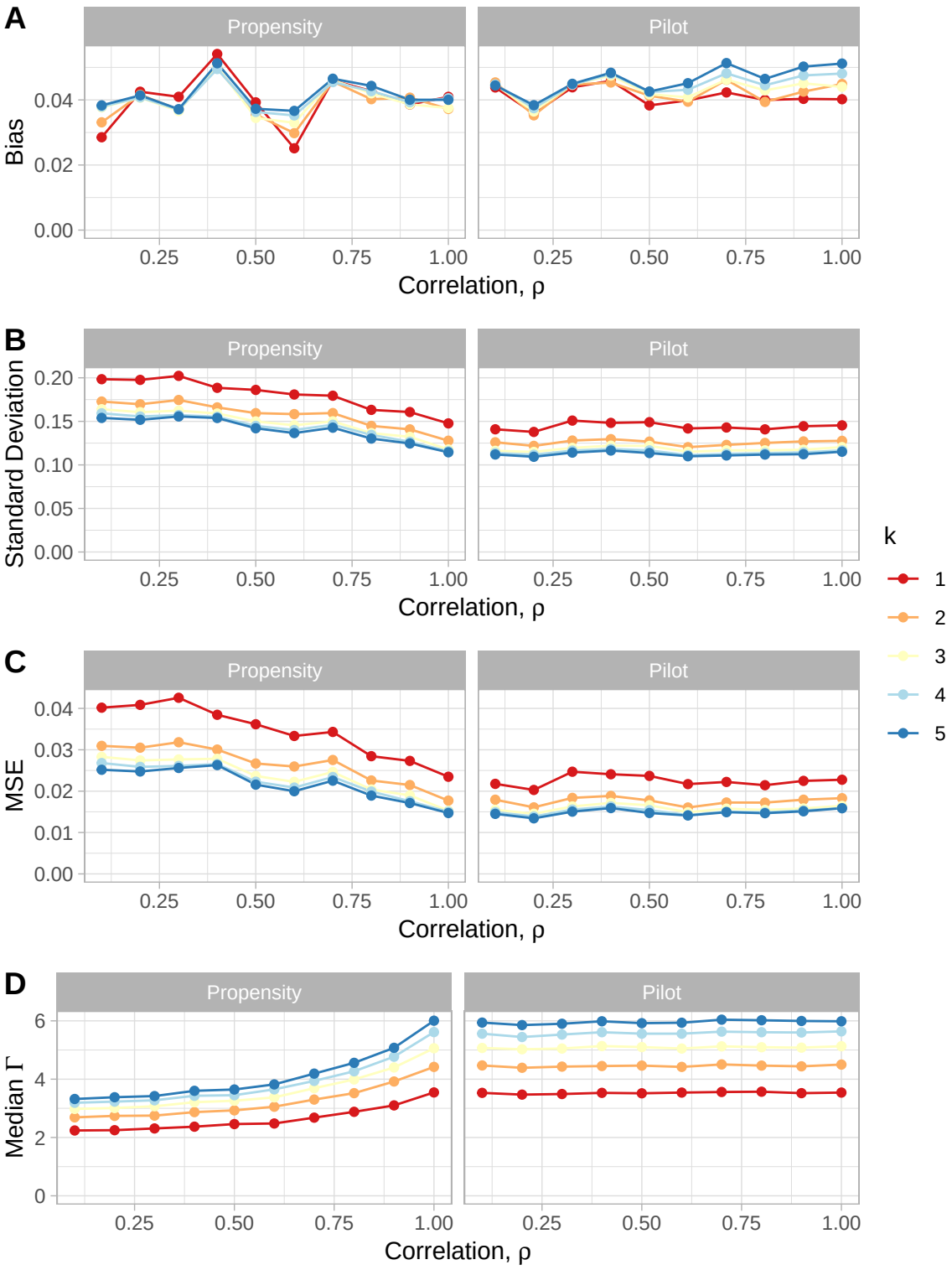


$\mathbf{tau = 1,\ nu = 0.5}$



tau = 1, nu = 0.2, model correctly specified

Below are the results we'd see if we correctly specified the models for propensity and prognosis. I've left out the results for Mahalanobis distance because there was a bug in the Mahalanobis distance matching code for these simulations.



$\mathbf{tau = 0,\ nu = 0.2}$