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.

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

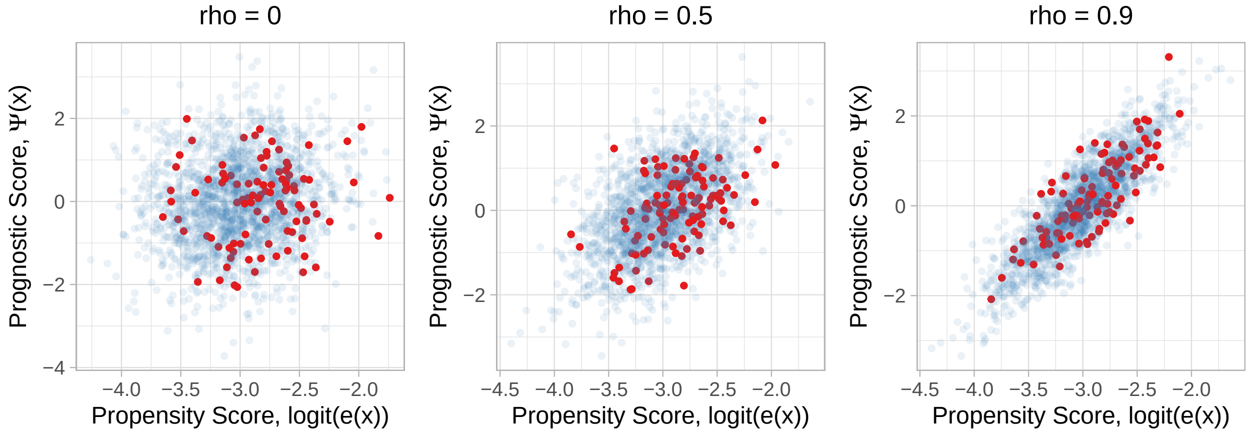
where the true propensity and prognositic scores are given by the linear combinations

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

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`

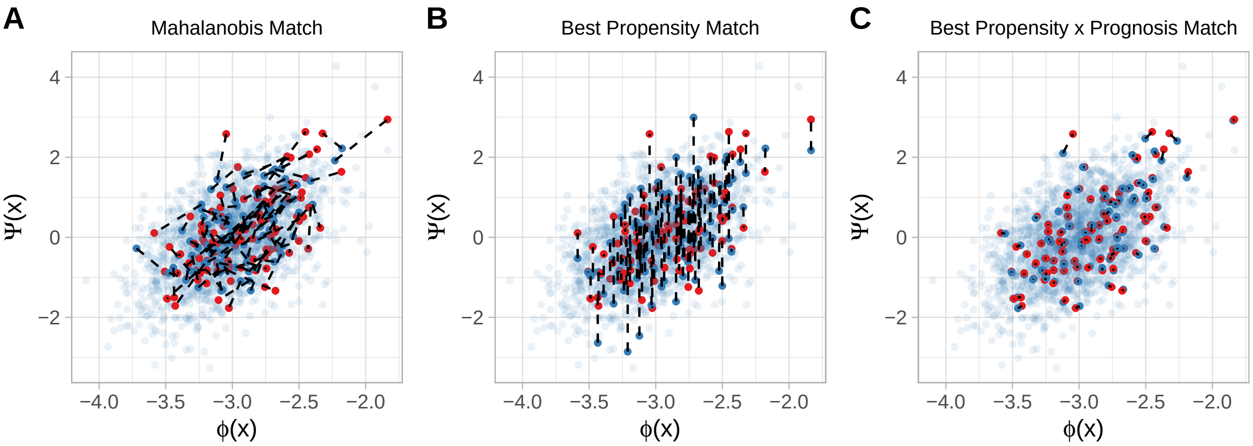
Fisher-Mill Visualizations

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

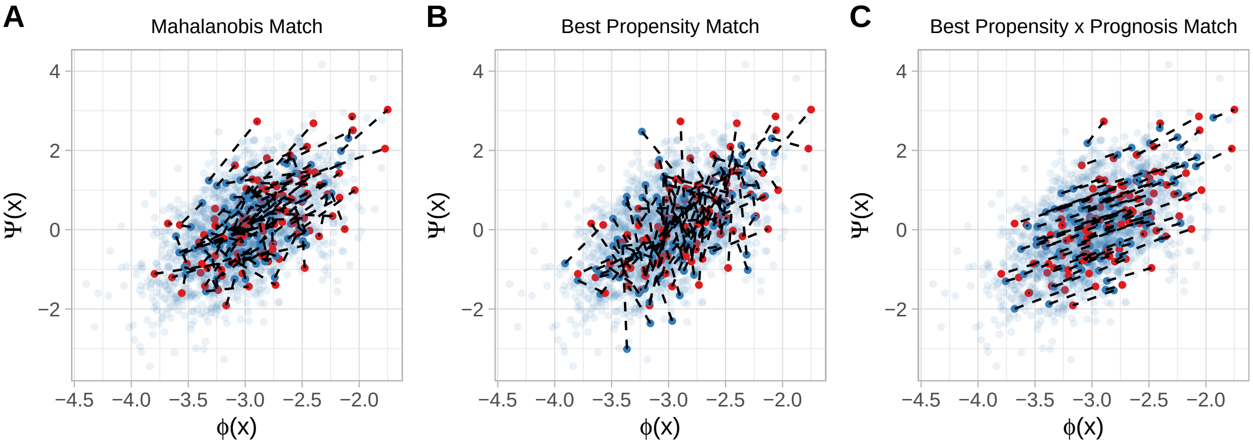


But there's something sinister going on here. Suppose we fit the best possible model for propenisity and prognostic score. That is, our coefficients for X1 and X2 in our propensity and prognostic models are exactly correct, but we leave out U in the model because we never observed it. The plots below 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
4
2
0
−2
Ψ(x)
−4.5
−4.0
−3.5
−3.0
−2.5
−2.0
ϕ(x)

Results with no SITA violation

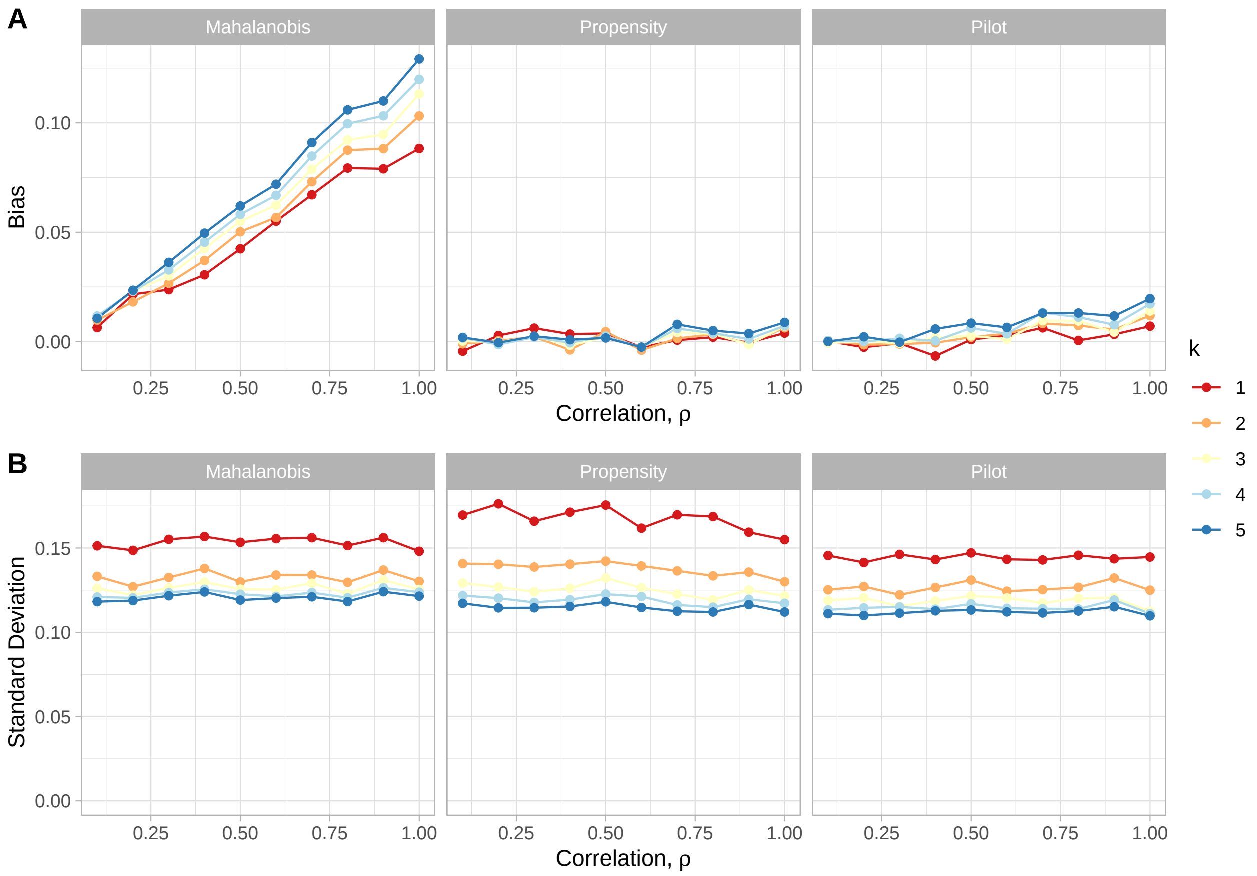


Figure 1 shows the bias and variance of $1 : k$ matching for each method on the same simulated data set as k increases (colored lines) and the correllation of propensity and prognosis, ρ increases from left to right. When $\rho = 0$, the treatment effect is completely unconfounded, since treatment is entirely determined (up to randomness) by variation in X_{i1}, and outcome (under the control assignment) is entirely determined by variation in X_{i2}. When $\rho = 1$, the data is highly confounded, since outcome and treatment assignment are both determined solely by variation in X_{i1}.

A first observation is that the bias from Mahalanobis distance matching is large, and it increases with the correllation between prognosis and treatment assignment (Figure 1A). This is probably because, as ρ approaches 1, only a single covariate, X_{i1} is important to match on, yet Mahalanobis distance considers 9 other covariates which are entirely random noise unimportant to the problem. It thus selects worse matches with respect to the true covariate of interest. This problem is exacerbated when the number of uninformative covariates is increased (Supplementary Figure 3).

Figure 1B shows the protective effect of pilot matching against the variance of the estimator. Propensity score matching has highest variance, and this is worse when ρ is close to zero. This is because, when ρ is small, the propensity score contains no information about prognosis under the control assignment. This causes the propensity score method to match observations which may be very different in terms of their potential outcome under the control assignment, giving poor prognostic balance and thus larger variance in the estimate of the causal effect. While Mahalanobis distance performs slightly better, it is again choosing lower quality matches because of the uninformative covariates in the data. The lowest variance is achieved by pilot matching, since this algorithm optimizes for prognostic balance as well as propensity score balance.

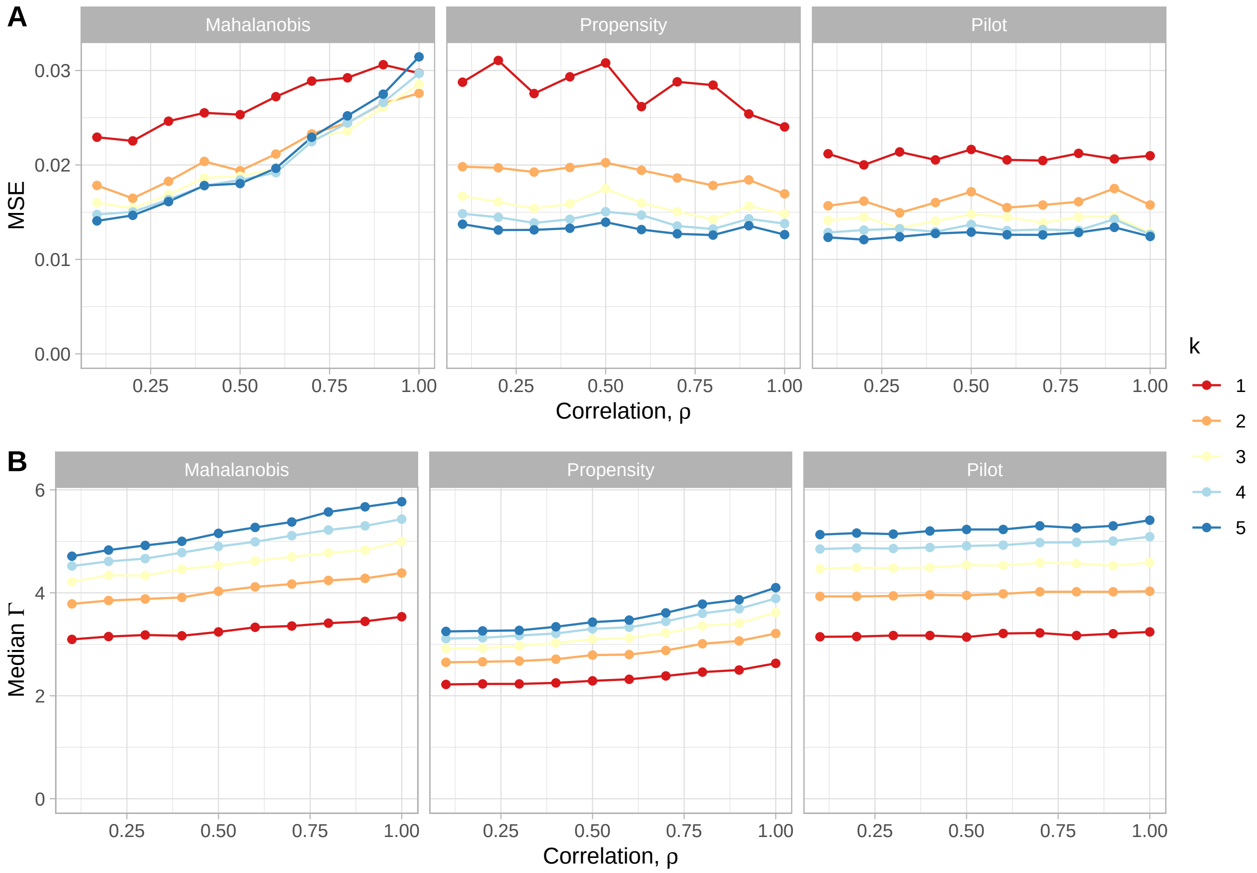
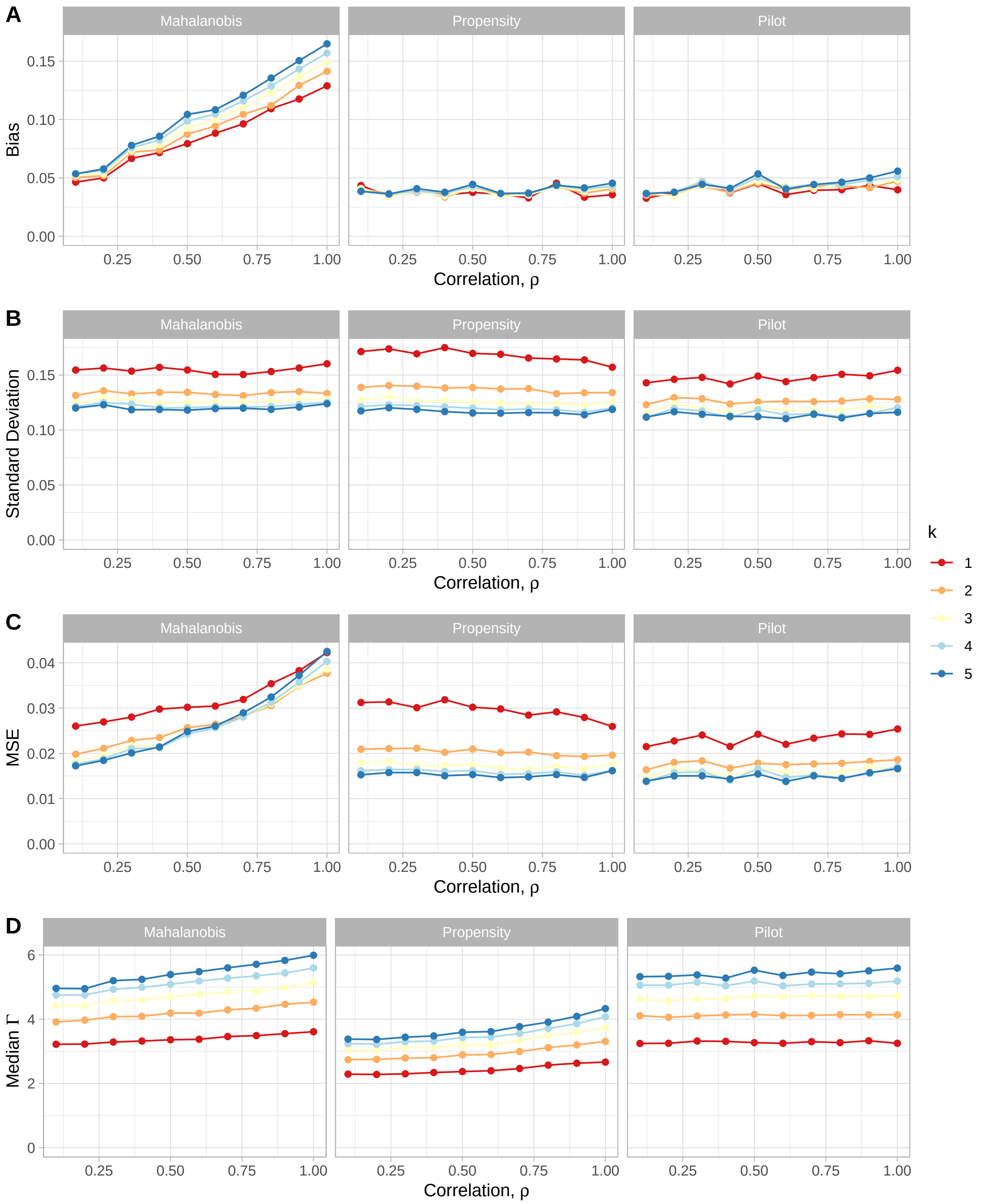
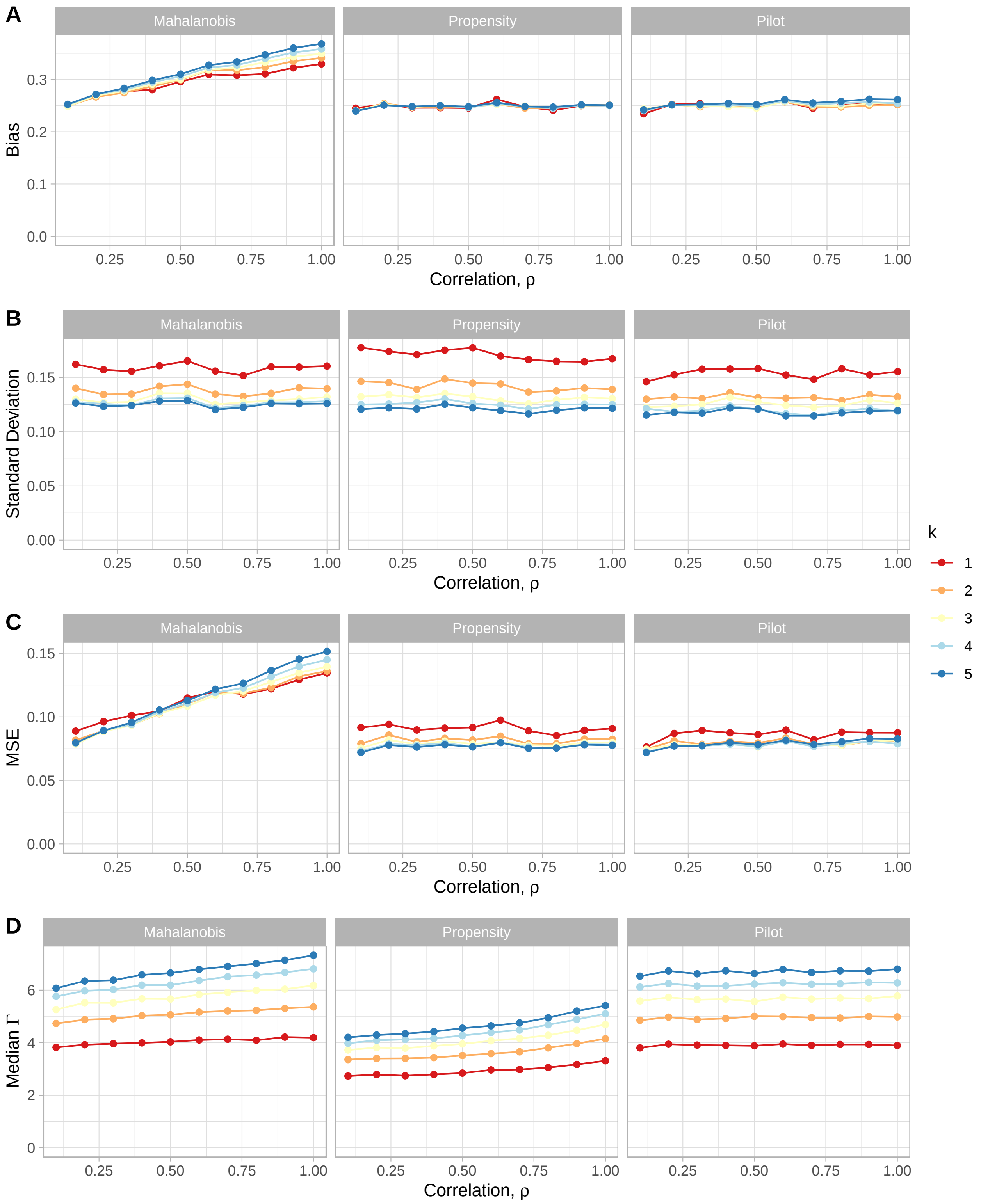


Figure 2 shows the mean squared error and median gamma sensitivity from the same set of simulations as above. Figure 2A demonstrates that pilot matching does well compared to the other two methods in terms of mean squared error, because of the protection against bias and variance shown in figure 1. Figure 2B shows the protective effect of prognostic balance against unobserved confounding. Propensity score matching gives lowest Γ values, especially when ρ is small. When this happens, the propensity score contains no prognostically relevant covariates, so the matches produced have poor prognostic balance. As ρ increases, the prognostic score and the propensity score become highly correllated, so matching on propensity score starts to give some prognostic balance as well as propensity balance. While the Gamma sentistivity of Mahalanobis distance matching seems promising when ρ is large, most of this is likely due to the large amount of bias in the effect estimate from this method (Figure 1A). Because pilot matching directly attempts to achieve prognostic balance in the matched sets, the pilot matching method has better protection against unobserved confounding, without giving a highly biased estimator.

nu = 0.2


A
Mahalanobis
Propensity
Pilot
Bias
0.00
0.05
0.10
0.15
0.25
0.50
0.75
1.00
Correlation, ρ
B
Mahalanobis
Propensity
Pilot
Standard Deviation
0.00
0.05
0.10
0.15
0.25
0.50
0.75
1.00
Correlation, ρ
C
Mahalanobis
Propensity
Pilot
MSE
0.00
0.01
0.02
0.03
0.04
0.25
0.50
0.75
1.00
Correlation, ρ
D
Mahalanobis
Propensity
Pilot
Median Γ
0
2
4
6
0.25
0.50
0.75
1.00
Correlation, ρ
k
1
2
3
4
5

nu = 0.5


A
Mahalanobis
Propensity
Pilot
Bias
0.0
0.1
0.2
0.3
0.25
0.50
0.75
1.00
Correlation, ρ
B
Mahalanobis
Propensity
Pilot
Standard Deviation
0.00
0.05
0.10
0.15
0.25
0.50
0.75
1.00
Correlation, ρ
C
Mahalanobis
Propensity
Pilot
MSE
0.00
0.05
0.10
0.15
0.25
0.50
0.75
1.00
Correlation, ρ
D
Mahalanobis
Propensity
Pilot
Median Γ
0
2
4
6
0.25
0.50
0.75
1.00
Correlation, ρ
k
1
2
3
4
5