Bias Variance Tuning

Rachael Caelie (Rocky) Aikens

5/1/2019

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:

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

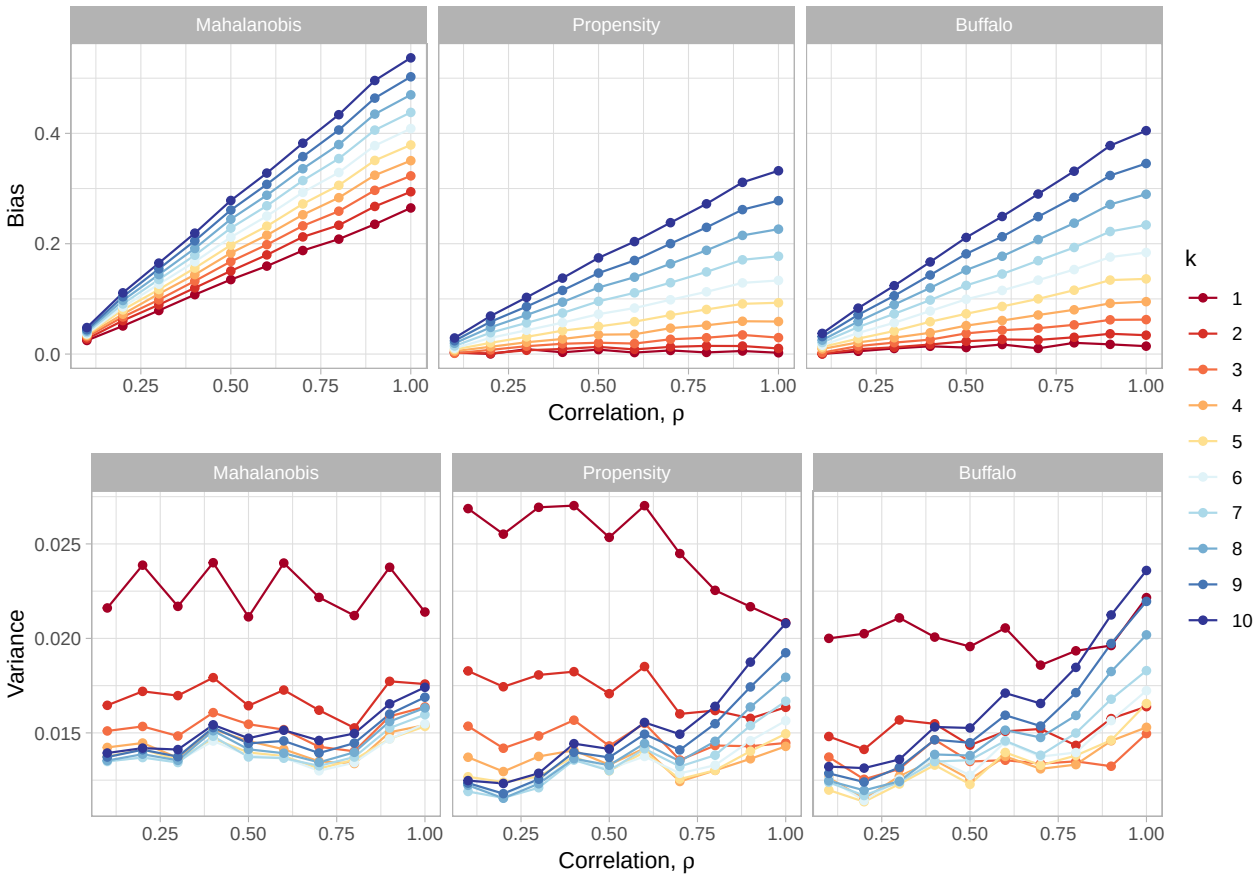
where the true propensity and prognositic scores are given by the linear combinations

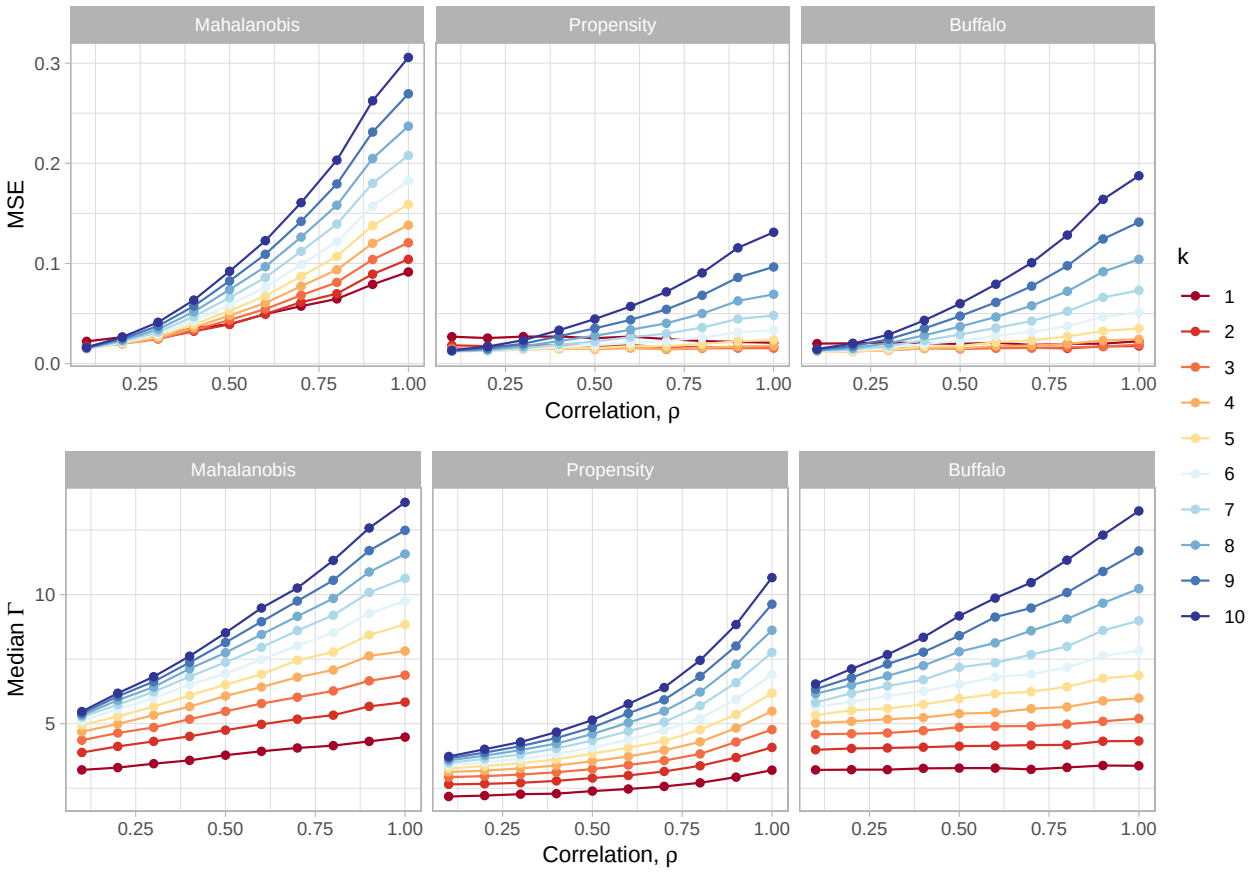
$$
\begin{aligned}
\phi(X_i) &= X_{i1} - 10/3, \\
\Psi(X_i) &= \rho X_{i1} + \sqrt{(1-\rho^2)} X_{i2},
\end{aligned}
$$

so that $\text{Cor}(\phi(X_i), \Psi(X_i)) \propto \rho$. The propensity score formula was chosen such that there were approximately 100 treated observations in each dataset. We consider $p = 10$, $\rho = 0, 0.1, \ldots, 0.9, 1.0$, 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`

What we have already

Below, we see the results of the latest complete batch of N = 1000 simulations, which use the exact set-up above. One issue with this set of results is that the bias is on such a larger scale than the variance that the bias dominates the MSE, and the bias/variance tradeoffs of the different approaches aren't highlighted.

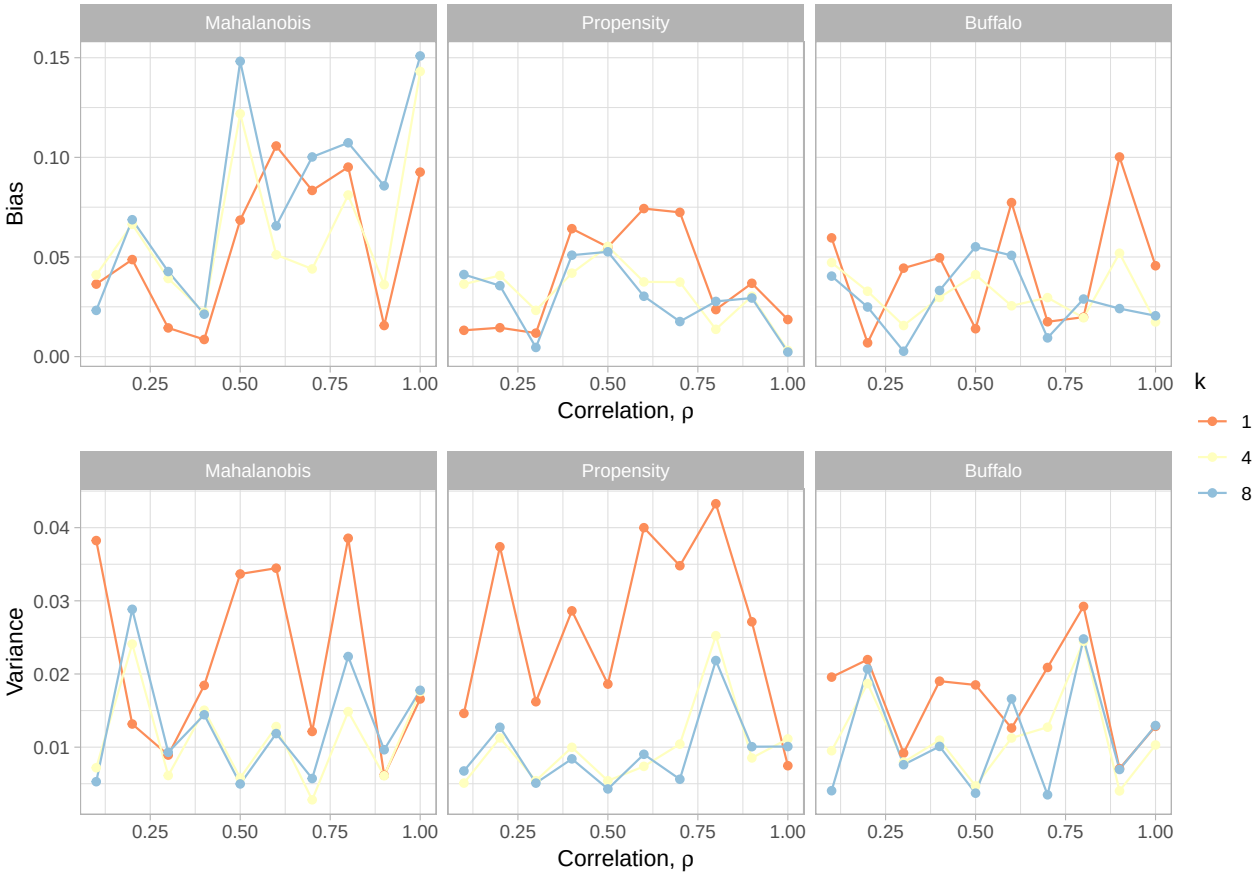

Mahalanobis
Propensity
Buffalo
Bias
0.0
0.2
0.4
0.25
0.50
0.75
1.00
Correlation, ρ
Variance
0.015
0.020
0.025
k
1
2
3
4
5
6
7
8
9
10

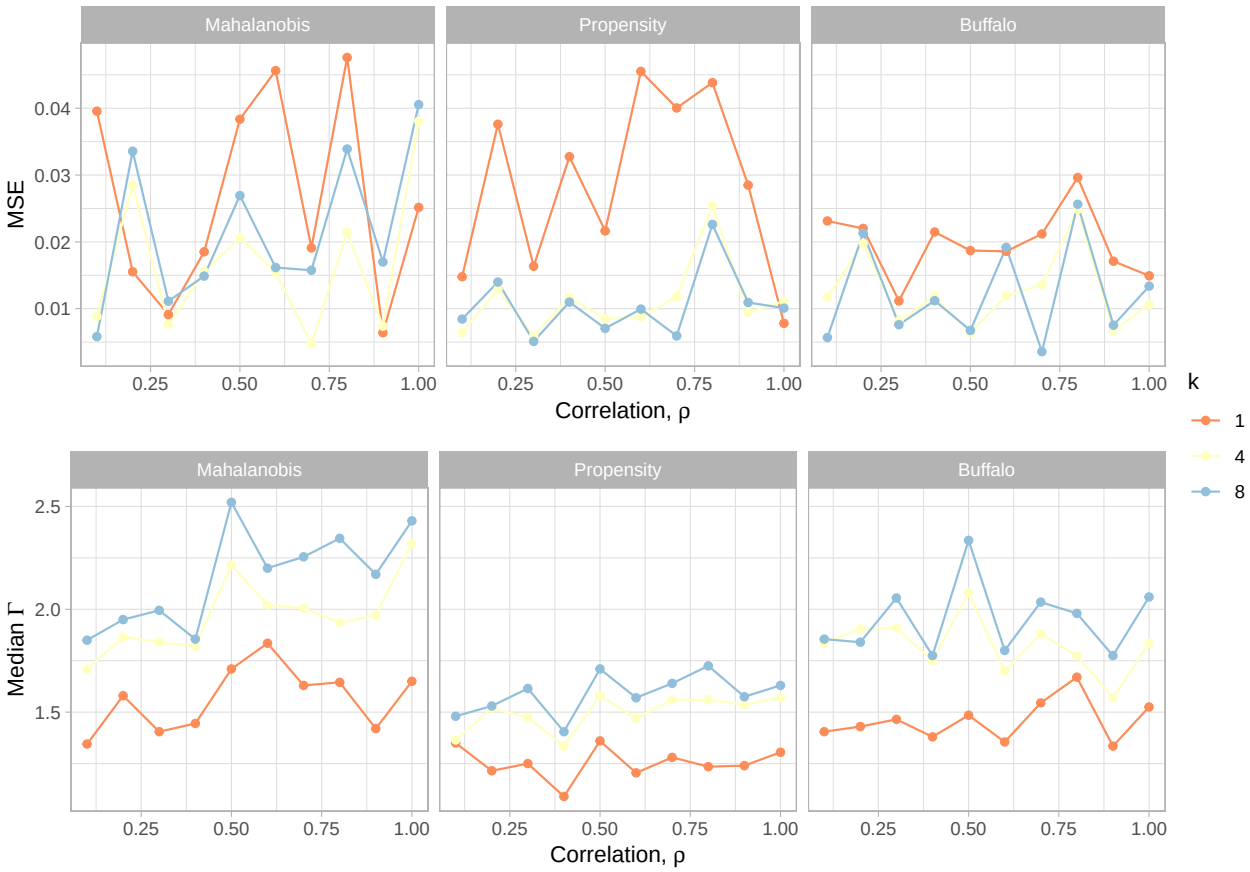

Mahalanobis
Propensity
Buffalo
MSE
0.0
0.1
0.2
0.3
0.25
0.50
0.75
1.00
Correlation, ρ
Median Γ
5
10
k
1
2
3
4
5
6
7
8
9
10

Playing with Tau

Suppose we switch to the following model:

$$\phi(X_i) = X_{i1}/3 - 3,$$
$$\Psi(X_i) = \rho X_{i1} + \sqrt{(1-\rho^2)} X_{i2}, \tau \qquad = 1/2$$





Playing with Sigma

Sigma = 2

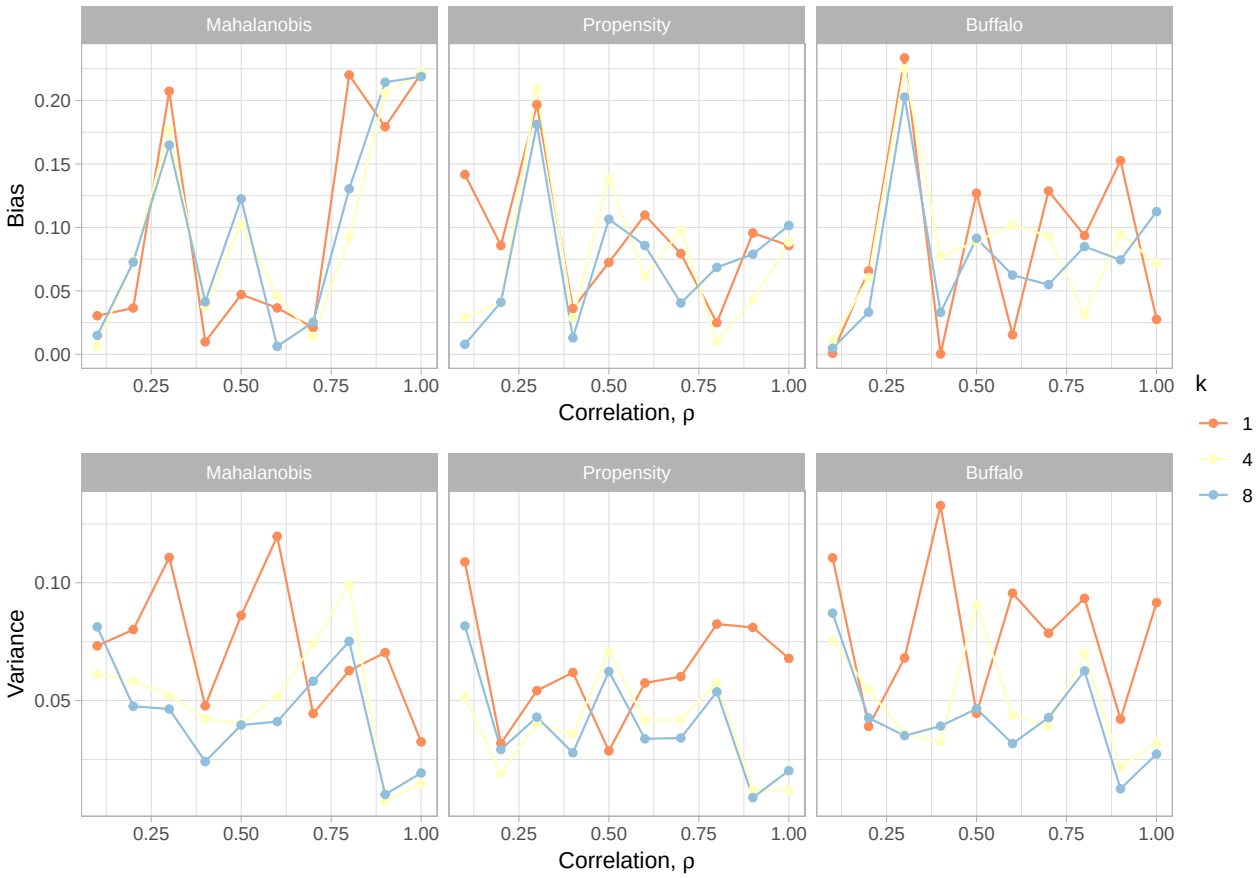
Another solution is increasing the noise. Suppose we have this model:

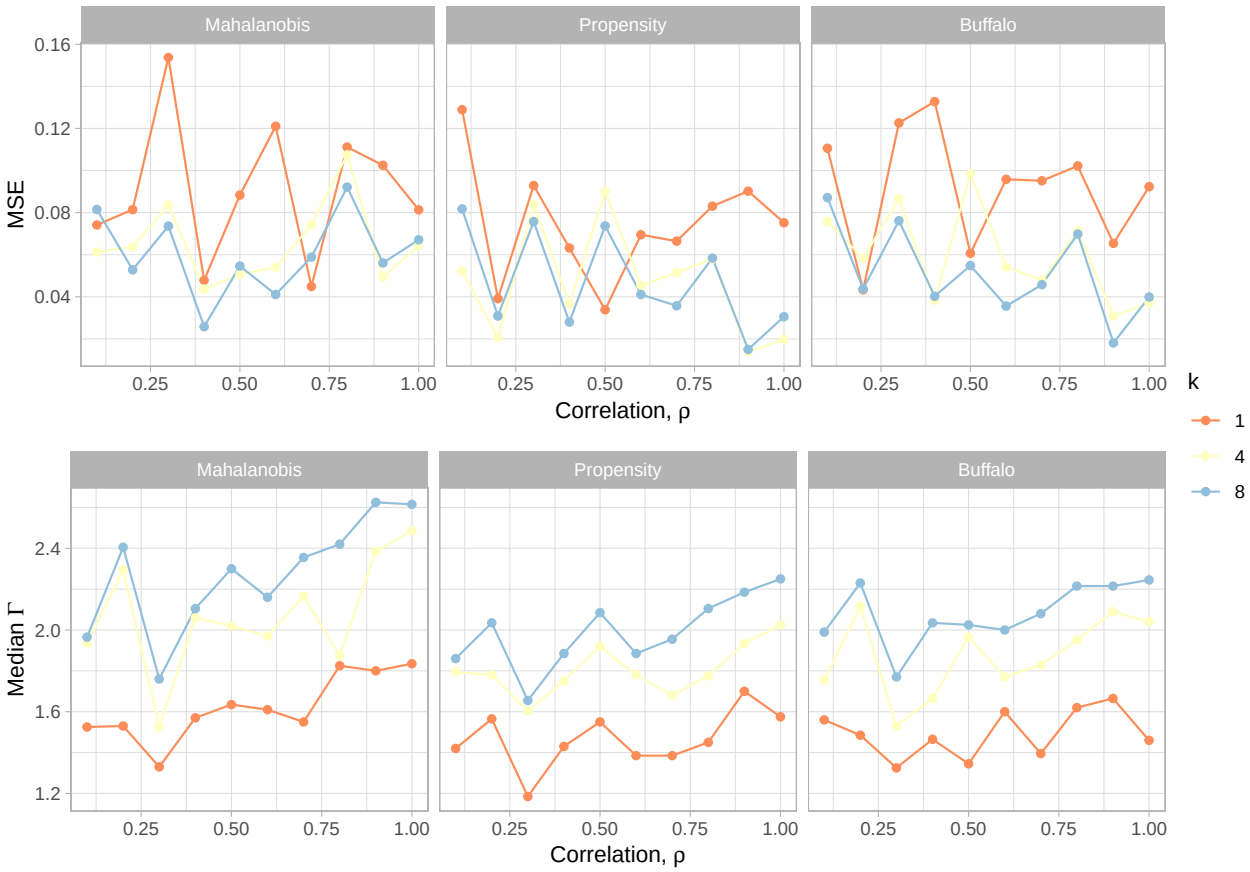
$$
\begin{aligned}
X_i &\sim_{iid} \text{Normal}(0, I_p), \\
T_i &\sim_{iid} \text{Bernoulli}\left(\frac{1}{1+\exp(-\phi(X_i))}\right), \\
Y_i &= \tau T_i + \Psi(X_i) + \epsilon_i, \\
\epsilon_i &\sim_{iid} N(0, \sigma^2),
\end{aligned}
$$

With the following formulae for propensity and prognosis

$$
\begin{aligned}
\phi(X_i) &= X_{i1} - 10/3, \\
\Psi(X_i) &= \rho X_{i1} + \sqrt{(1-\rho^2)} X_{i2},
\end{aligned}
$$

Except that we let $\sigma = 2$.


Mahalanobis
Propensity
Buffalo
Bias
0.00
0.05
0.10
0.15
0.20
0.25
0.50
0.75
1.00
Correlation, ρ
Variance
k
1
4
8



Playing with Propensity

Phi = X1/3 - 4

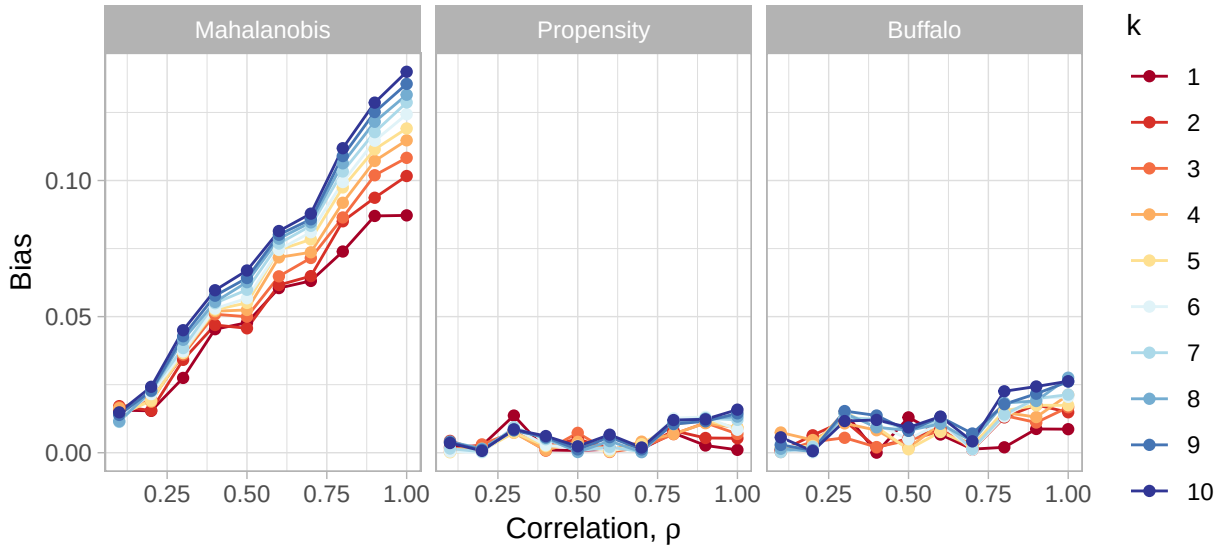
This is the model that Dylan was using when I took over:

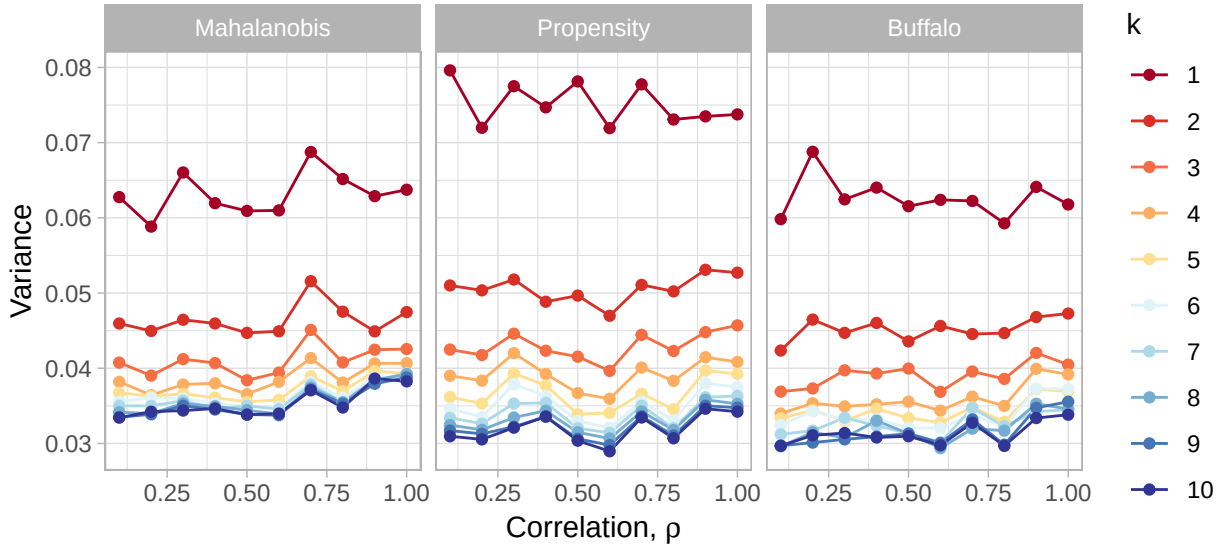
$$\phi(X_i) = X_{i1}/3 - 4,$$
$$\Psi(X_i) = \rho X_{i1} + \sqrt{(1-\rho^2)} X_{i2},$$

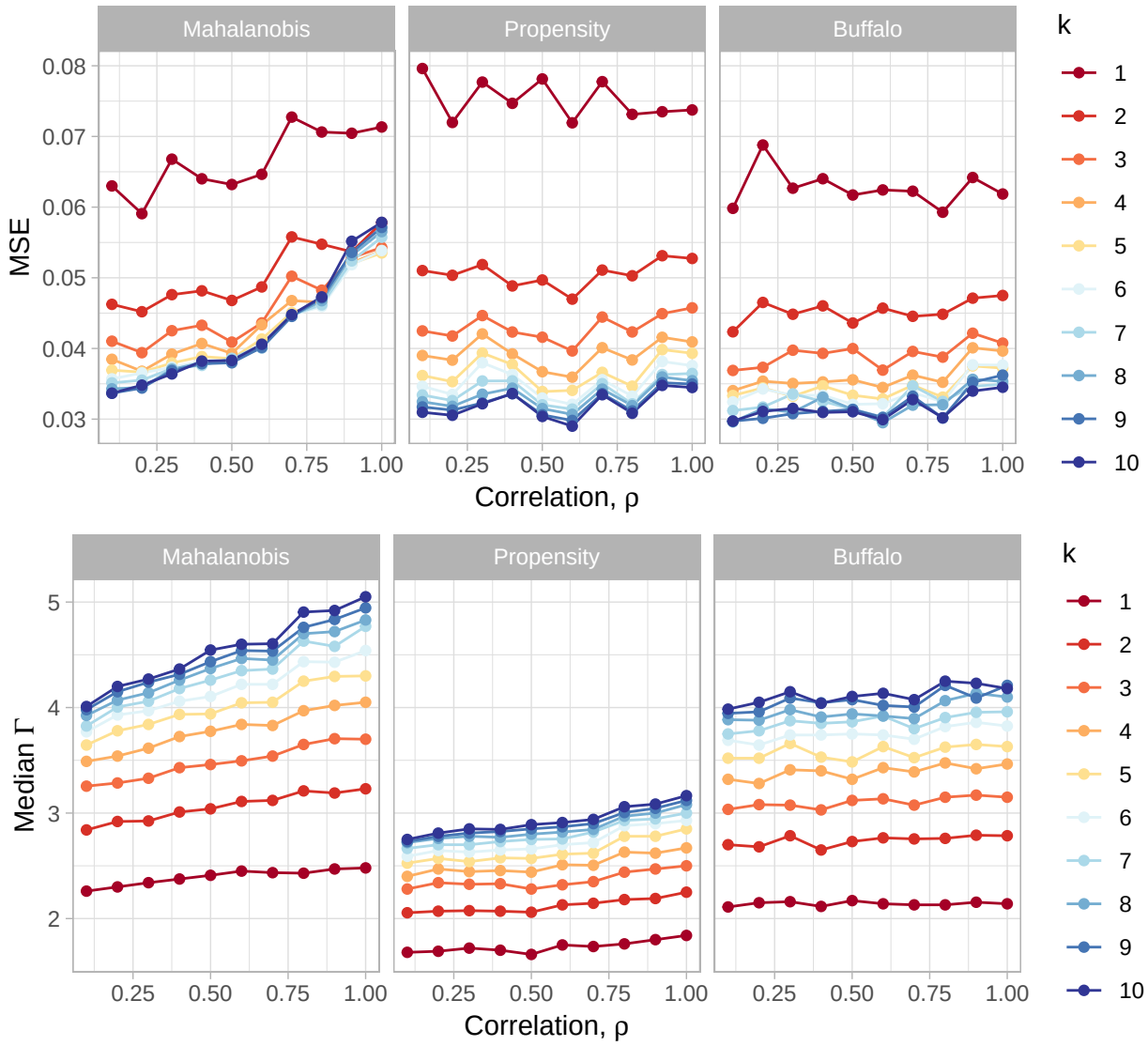
It's nice, because it puts bias and variance on the same approximate scale, but it allows the number of treated individuals to drop to 30-40, rather than 100. Buffalo also does weirdly poorly in terms of gamma compared to mahalanobis.

```
true_tau <- 1

dat <- mutate(dat,
              squared_err = (estimate-true_tau)**2,
              k = as.factor(k))

plt_data <- dat %>%
  group_by(method, k, rho) %>%
  summarize(Bias = abs(mean(estimate) - true_tau),
```

```
            median_gamma = median(gamma),
            Variance = var(estimate),
            MSE = Bias^2 + Variance) %>%
  ungroup() %>%
  mutate(method = recode(method, propensity = "Propensity",
                         mahalanobis = "Mahalanobis",
                         prognostic = "Buffalo"))
```





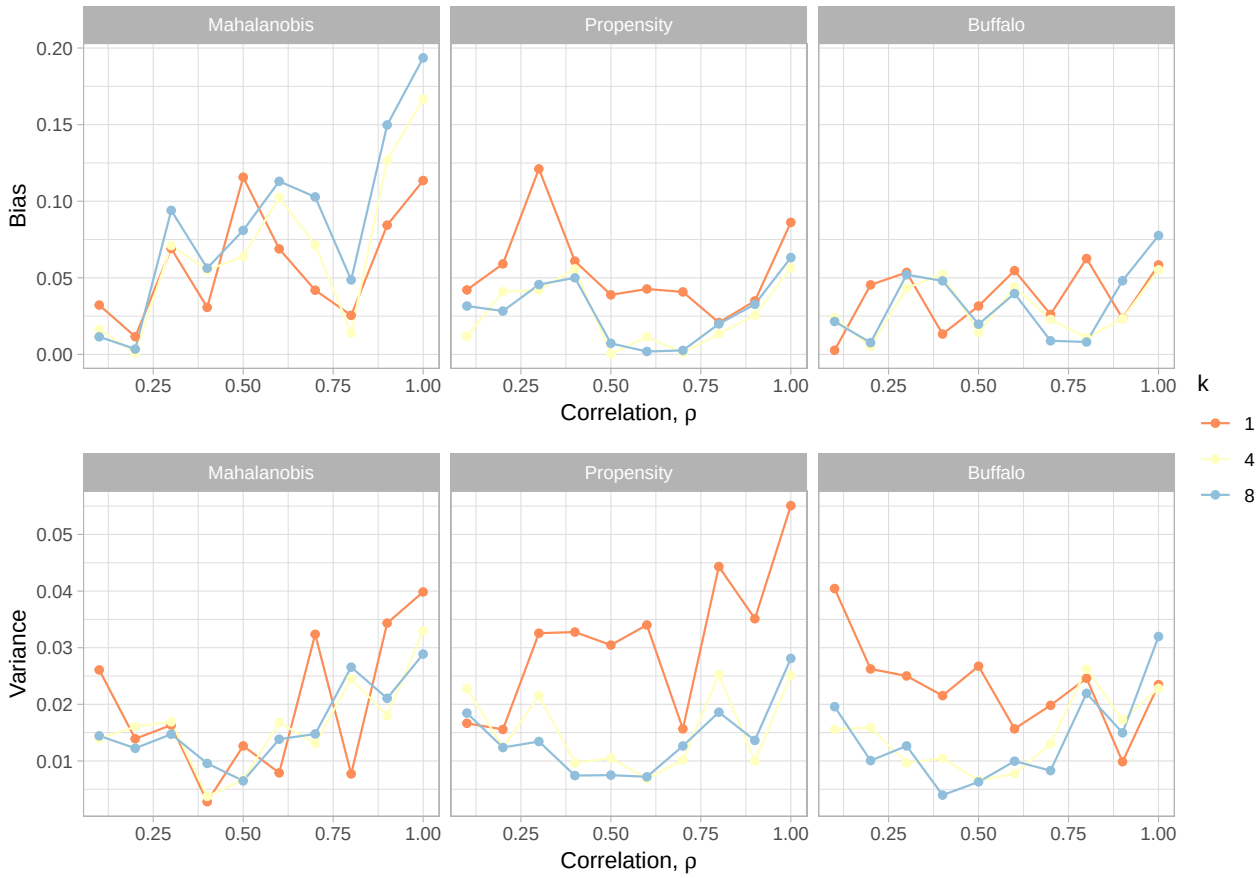


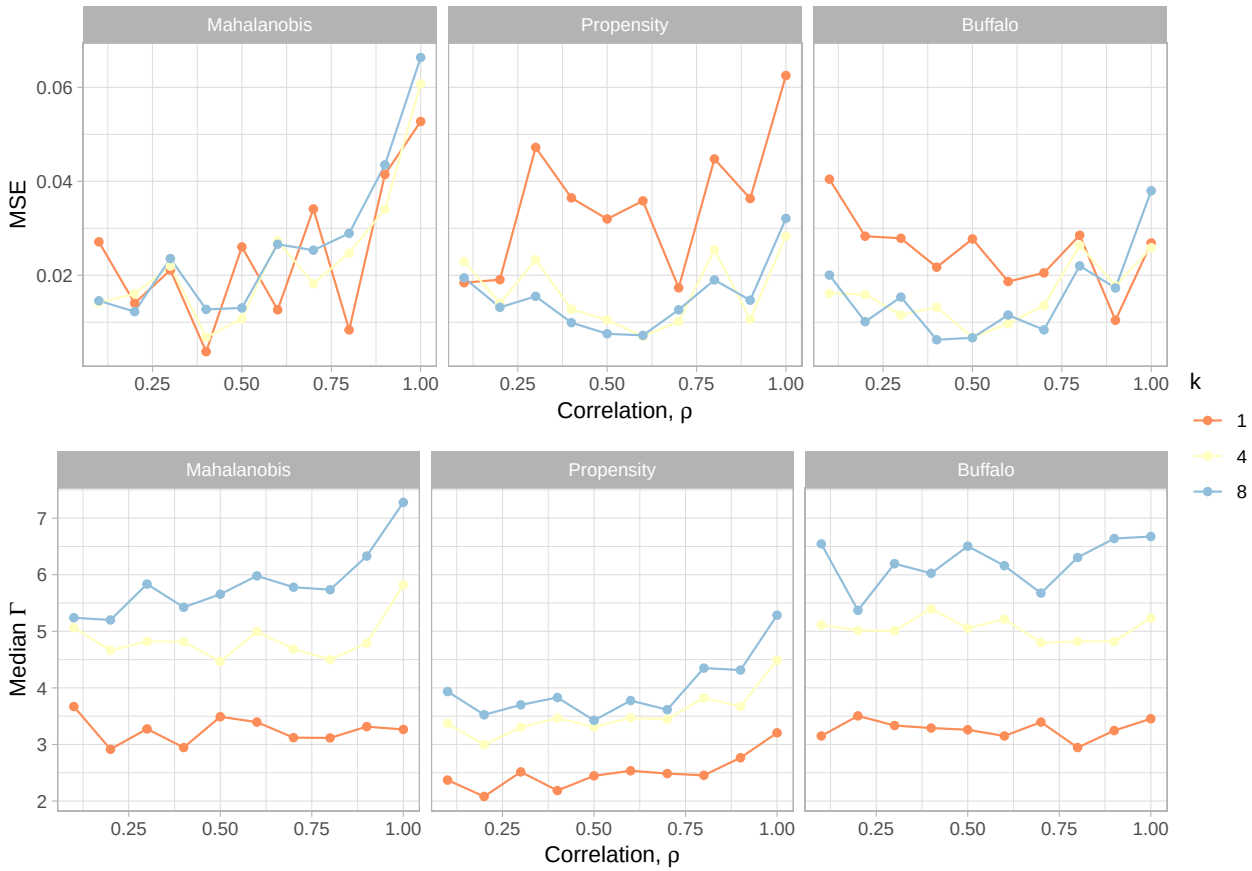
Phi = X1/3 - 3

Suppose we switch to the following model:

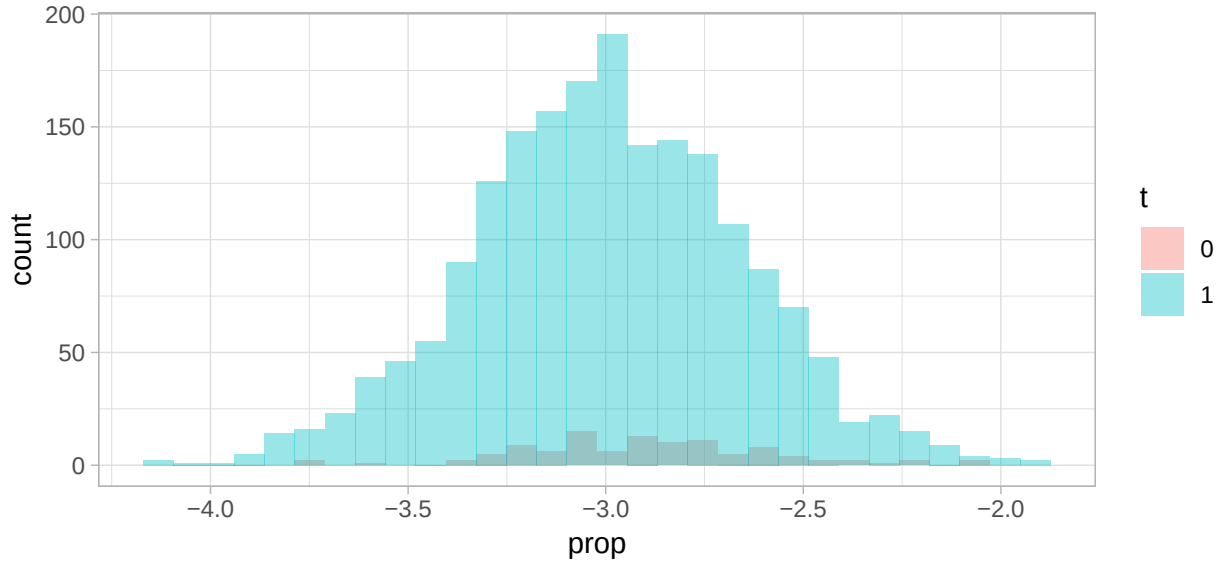
$$\phi(X_i) = X_{i1}/3 - 3,$$
$$\Psi(X_i) = \rho X_{i1} + \sqrt{(1-\rho^2)} X_{i2},$$

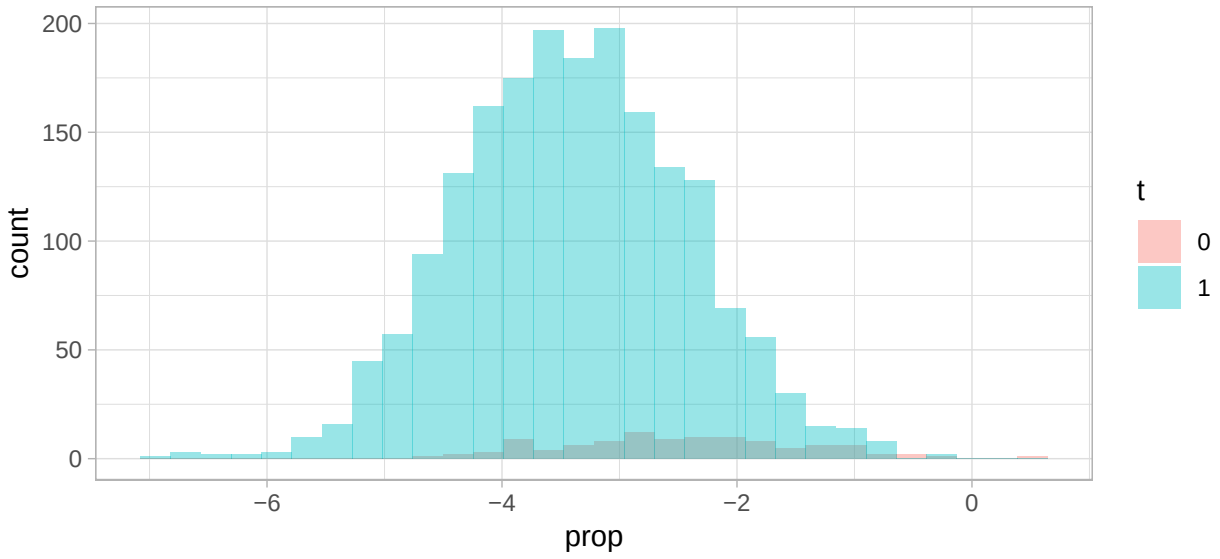
This decreases the importance of X_1 in determining treatment assignment, while keeping the number of treated individuals at approximately 100. Doing this decreases bias while increasing variance.


Mahalanobis
Propensity
Buffalo
Bias
0.20
0.15
0.10
0.05
0.00
0.25
0.50
0.75
1.00
Correlation, ρ
Variance
0.05
0.04
0.03
0.02
0.01
k
1
4
8



One thing here that I'm not crazy about is that when we diminish the importance of X_1 in determining treatment assignment, we get closer and closer to random treatment assignment, which maybe isn't the use case that we're most interested in. Below, I've printed a histogram of ϕ for treated (red) and control (blue) individuals. On the left, we see what happens when $\phi = X_1/3 - 3$, and on the right, we see what happens when $\phi = X1 - 10/3$.





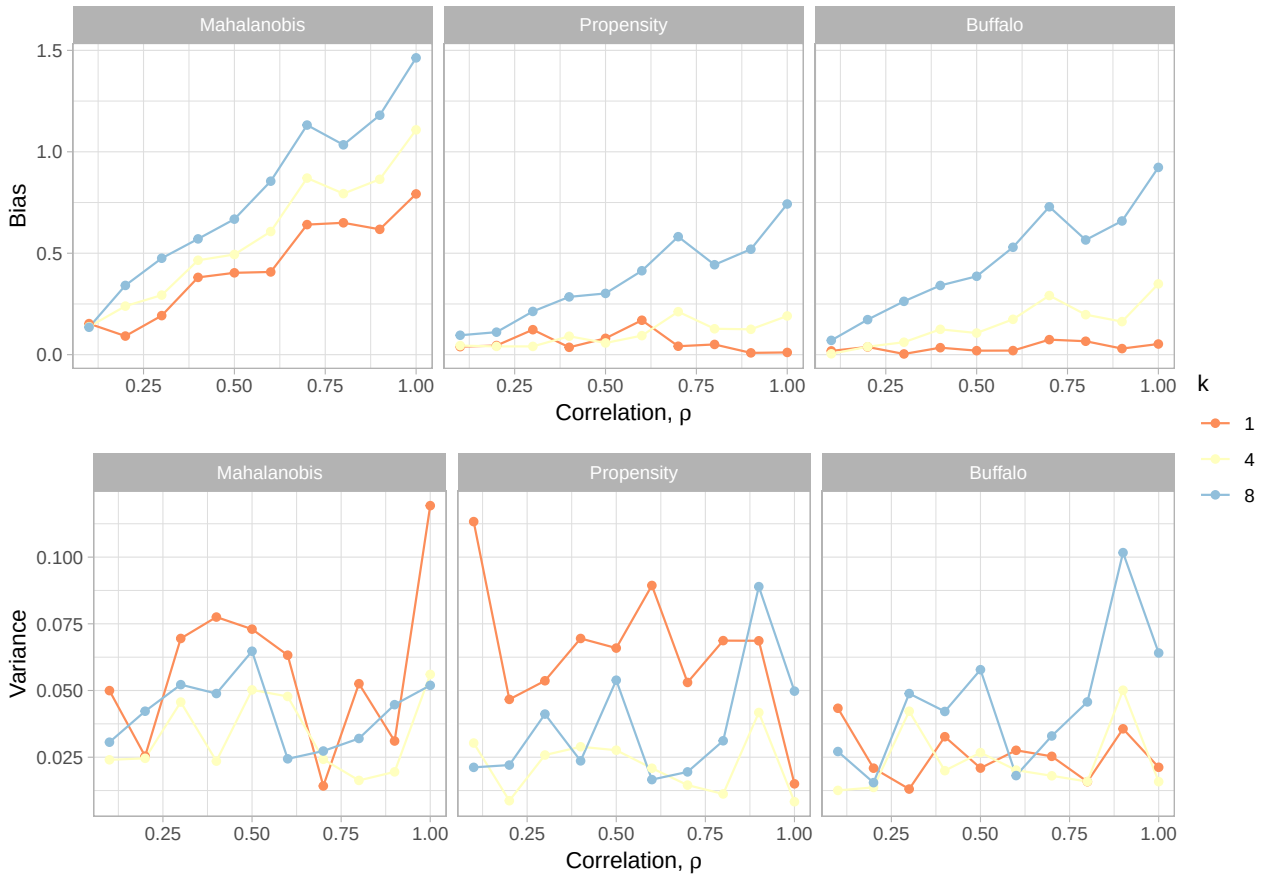
Playing with Prognosis

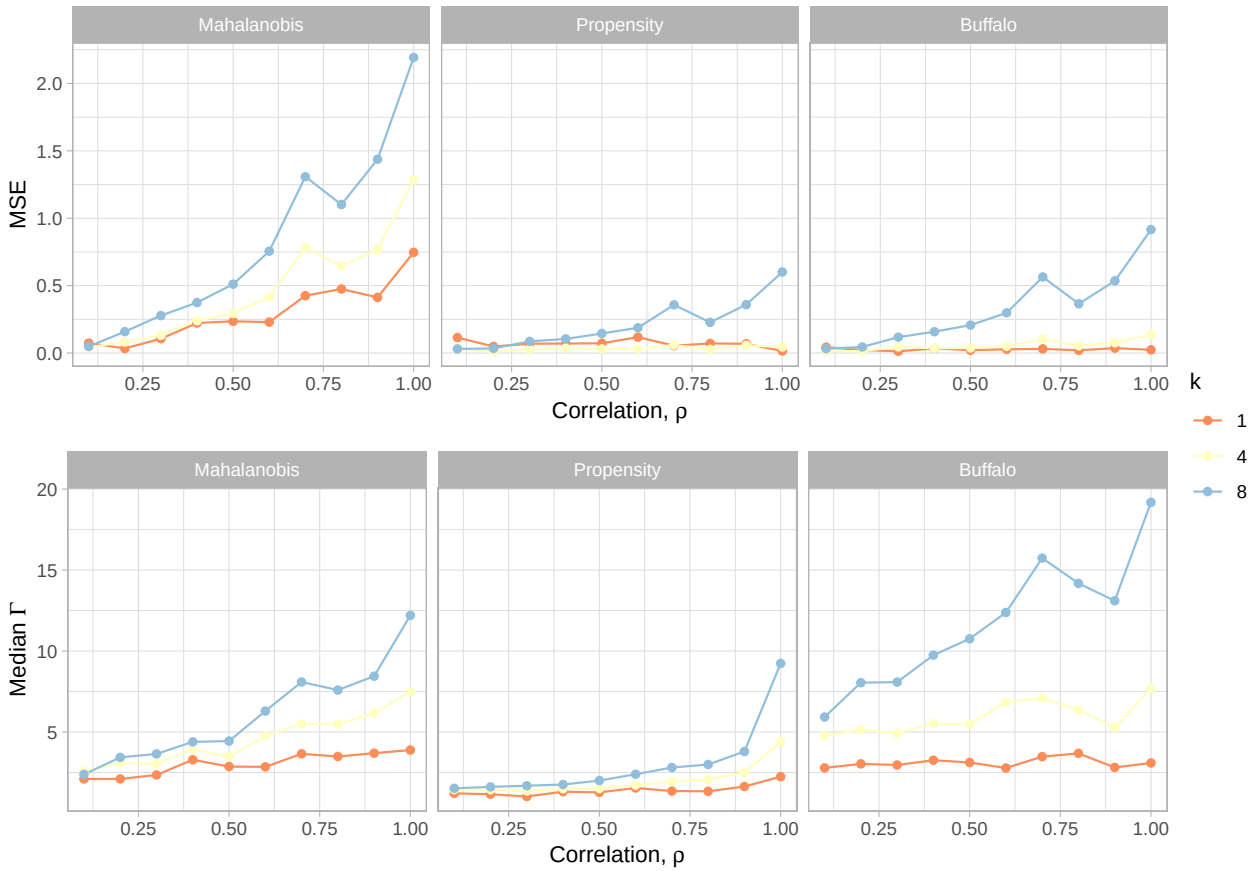
Psi times 3

Suppose we switch to the following model:

$$\phi(X_i) = X_{i1} - 10/3,$$
$$\Psi(X_i) = 3\rho X_{i1} + 3\sqrt{(1-\rho^2)}X_{i2},$$

Then bias and variance both increase, but variance increases more than bias. This makes bias and variance closer to the same scale (but not quite the same). The maximum bias can be about 10 times the maximum variance. Also worth noting: Gamma for the buffalo goes *way* up.


Mahalanobis
Propensity
Buffalo
Bias
1.5
1.0
0.5
0.0
0.25
0.50
0.75
1.00
Correlation, ρ
k
1
4
8
Mahalanobis
Propensity
Buffalo
Variance
0.100
0.075
0.050
0.025
0.25
0.50
0.75
1.00
Correlation, ρ

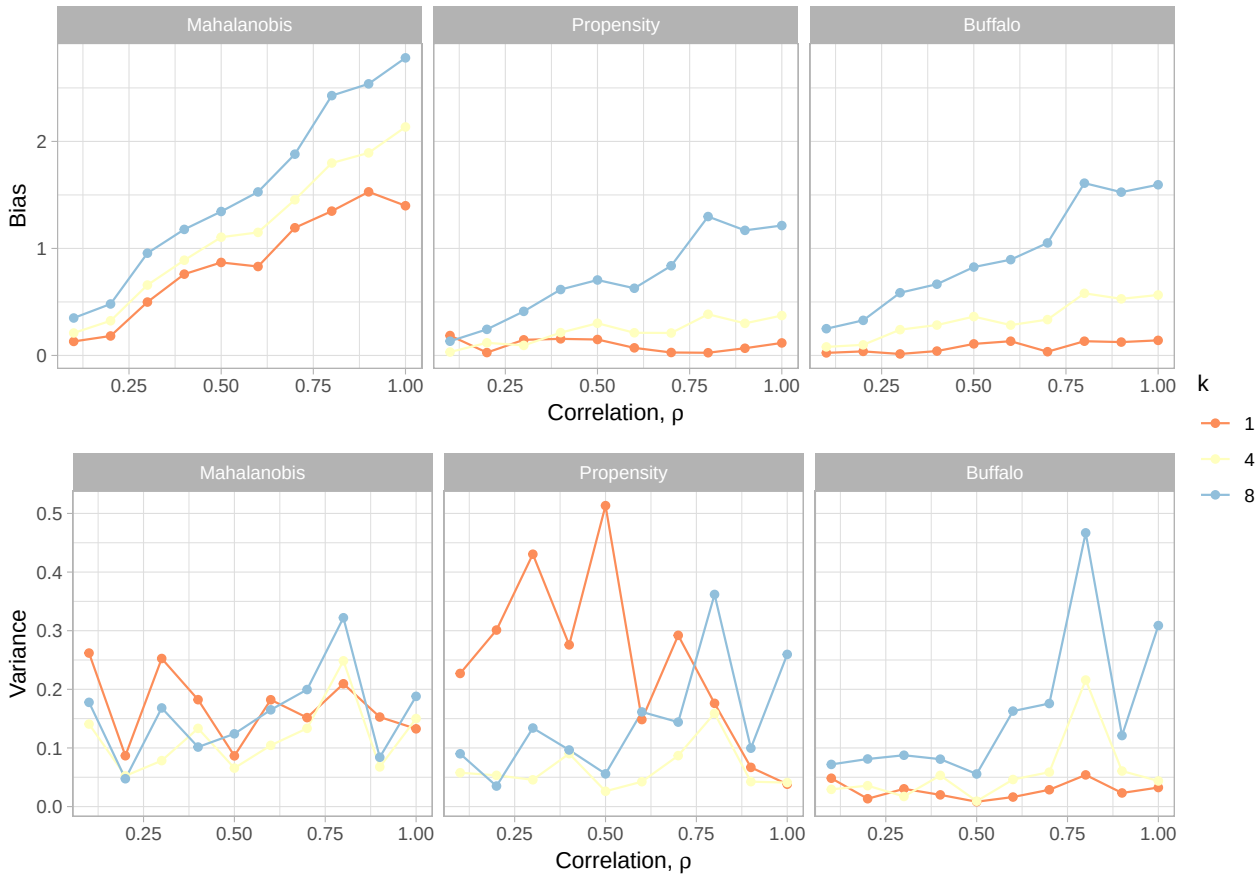


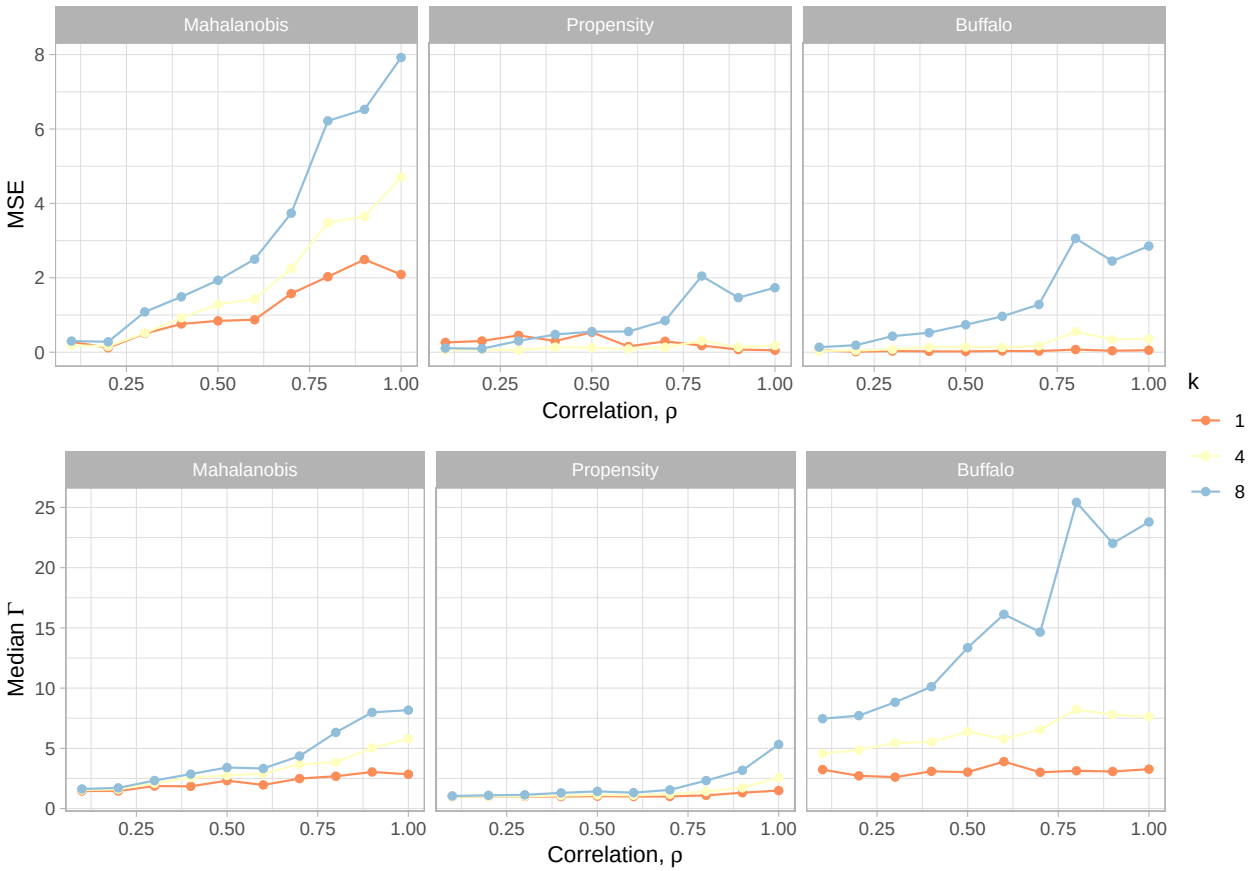
Psi times 6

Now suppose we emphasize these relationships even more:

$$\phi(X_i) = X_{i1} - 10/3,$$
$$\Psi(X_i) = 6\rho X_{i1} + 6\sqrt{(1-\rho^2)}X_{i2},$$

As we might expect, the behavior from the previous experiment is amplified. Bias goes up, but variance goes up more. Now the bias can be about 5 times the variance. Gamma for the buffalo skyrockets.


Mahalanobis
Propensity
Buffalo
Bias
0
1
2
0.25
0.50
0.75
1.00
Correlation, ρ
k
1
4
8
Variance
0.0
0.1
0.2
0.3
0.4
0.5


Mahalanobis
Propensity
Buffalo
MSE
Median Γ
Correlation, ρ
k
1
4
8