1. Importance of the Multiplier for Equicorrelated Knockoffs in Knockoff-Assisted Methods

Although the equicorrelated knockoffs method calls for taking $D = \text{diag}(s)$ where $s = 2\lambda_{min}(\Sigma) \wedge 1$ (where $\Sigma = X^T X$), we know that we tend not to be able to invert $2\Sigma - D$ unless we take the multiplier m to be strictly less than 2. I originally ran the simulations with $m = 1.98$ since that resulted in matrices that were very close to what was intended while improving the condition of $2\Sigma - D$ subsantially. However, I later found that power from all the methods using Sarkar and Tang's $\hat{\beta}_1$ and $\hat{\beta}_2$ estimates kept increasing as I decreased m. The problem is that the value of m that you work with for these methods appears to be arbitrary and not what the authors said, even in the supplementary material for their paper. In any case, I re-ran the simulations with $m = 1.9$ and used this going forward. With this setup, I could now replicate the findings from their paper. Power was still increasing for $m = 1.8$, but I decided on $m = 1.9$ since the power did not change that much compared to $m = 1.8$, and still was somewhat close to the original method.

Power is highest for these knockoff-assisted methods if you use $m < 2$ for calculating T_1 from [4] while using the default equicorrelated knockoff (i.e., $m = 2$) to generate T_2. This is because you do not need to invert $2\Sigma - D$ to calculate this vector, so one should calculate D and the knockoff matrix $\tilde{X}$ exactly as from the package to calculate T_2. This would be slower since it generates knockoffs twice. However, tests such as independent BH become significantly less powerful if you use $m < 2$ to calculate T_2.

In these simulations and throughout these notes, n denotes the number of rows of the design matrix X, d denotes the number of columns, and k denotes the number of nonzero β coefficients. (Simulation Settings 1, 2, and 3 use an autoregressive design matrix described in [4]).

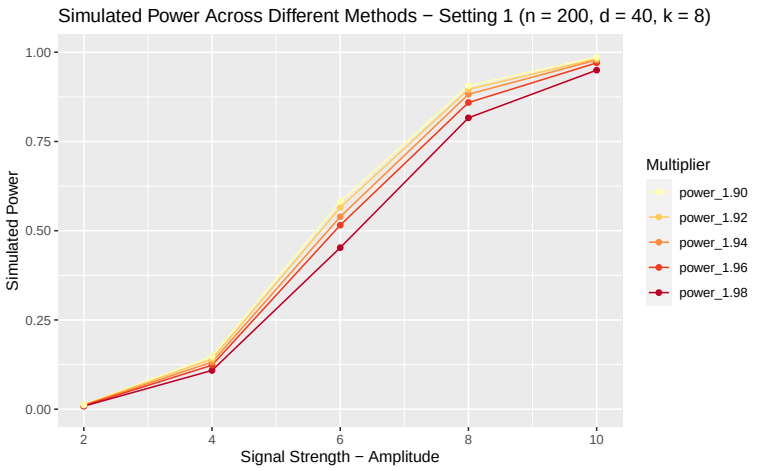

Figure 1. Power graphs for different values of m for the ep-Moment method defined in Section 4.2. Note that power_1.9x denotes the power graph when using $m = 1.9$x.

2. Replication of Sarkar and Tang's Plots

Using the new value of m, I could now replicate the findings from the paper; their methods were now more powerful than doing BH on $\hat{\beta}_2$ by itself. Method 1 and Method 2 are the two methods that Sarkar and Tang introduce in [4], and I simply refer to them as such here.

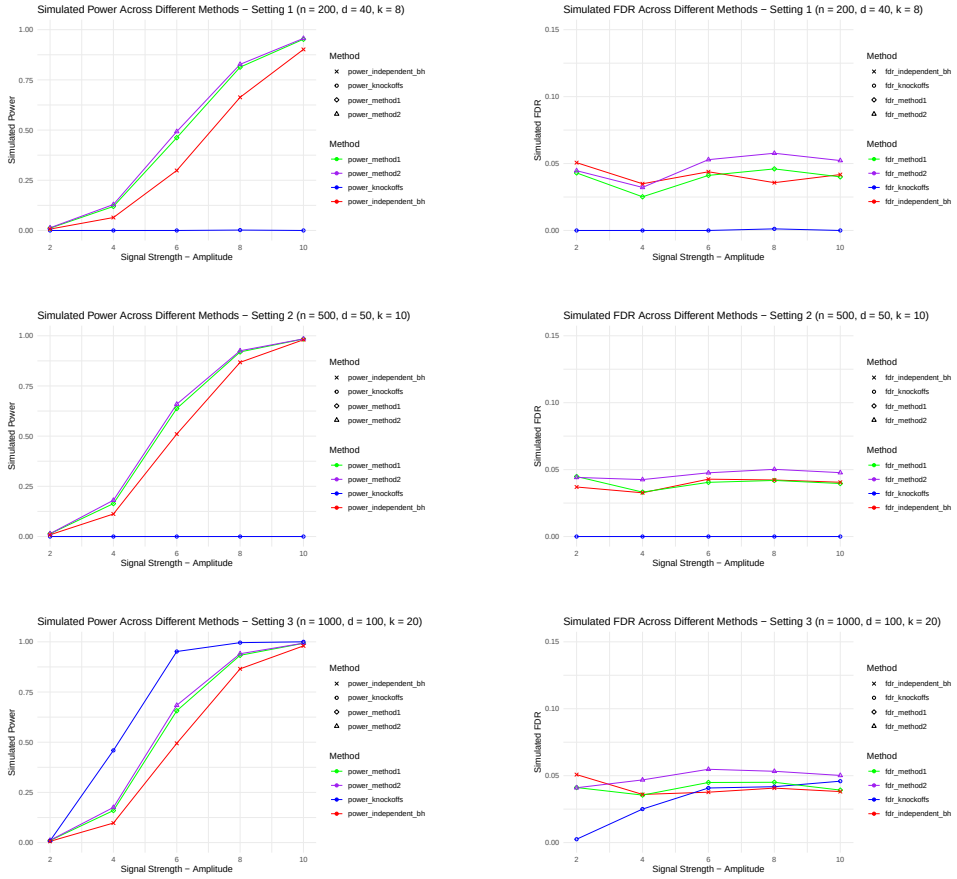

Figure 2. A replication of the simulation data in [4].

3. ep-BH version of Sarkar and Tang Using Different Cutoffs

In [4], we take

$$\tilde{P}_j = \begin{cases} 1, & P_j^{(t_1)} > c \\ P_j^{(t_2)}, & P_j^{(t_1)} \leq c \end{cases} \tag{3.1}$$

where $c = \alpha^{1/2}$ and $P_j^{(t_1)}$ and $P_j^{(t_2)}$ are as defined in [4], coming from independent unbiased estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. We can try other values of c since the original choice may be arbitrary.

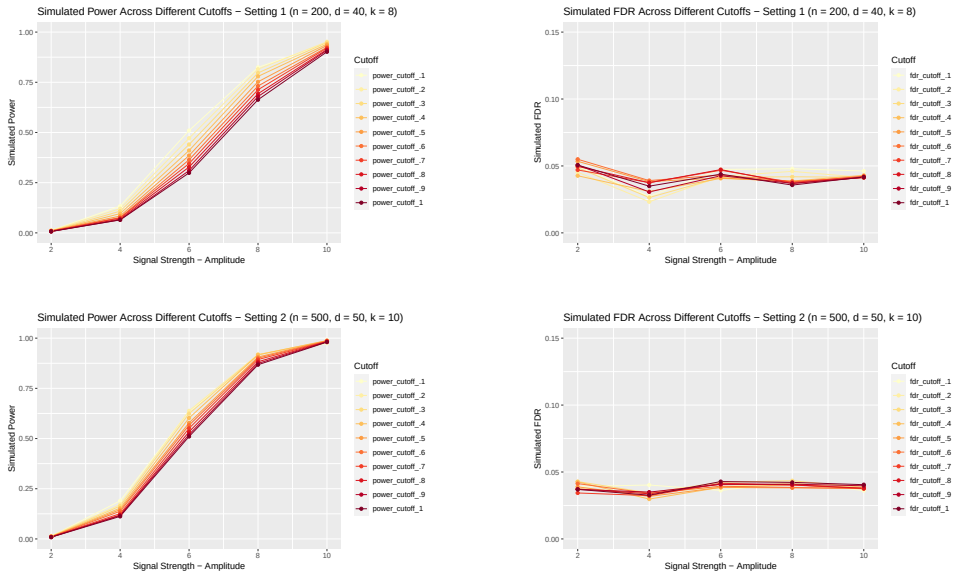

FIGURE 3. Lower cutoffs in Bonferroni-BH type methods seem to provide higher power. power_cutoff_x means that we have the cutoff $c = \mathrm{x}$ in the method in (3.1).

The figure shows that higher cutoffs c are less powerful, which illustrates why BH on $\hat{\beta}_2$ (what I called Independent BH) alone was less powerful than the methods in [4], since independent BH is [4]'s method with $c = 1$.

4. The ep-Moment Method

4.1. **Known Variance.** If the variance parameter σ is known, then we may take

$$\begin{cases} P_j = 2\left(1 - \Phi\left(\dfrac{|\hat{\beta}_{2,j}|}{\sigma_{2,j}}\right)\right), \\ E_j = \dfrac{\left(\dfrac{\hat{\beta}_{1,j}}{\sigma_{1,j}}\right)^{2k}}{\mu_{2k}}, \end{cases}$$

where μ_{2k} is the $2k$th moment of the $N(0,1)$ distribution, k is any positive integer, $\sigma_{1,j} = \sqrt{2\sigma^2((2\Sigma - D)^{-1})_{jj}}$, and $\sigma_{2,j} = \sqrt{2\sigma^2 D_{jj}^{-1}}$.

Theorem 4.1. *The ep-BH procedure with p values as in Section 4.1 and e values as in Section 4.1 has false discovery rate at most $\pi_0\alpha$.*

Proof. Since $\hat{\beta}_2$ has independent components, the P_j's are independent, and since $\hat{\beta}_1$ and $\hat{\beta}_2$ are independent, it follows that $(P_j)_{j\in\{1,\ldots,d\}}$ is independent of $(E_j)_{j\in\{1,\ldots,d\}}$. From Theorem 4 of [2], this procedure has false discovery rate at most $\pi_0\alpha$. □

4.2. **Unknown Variance.** For the rest of these notes, denote $\nu = n - 2d$, where n is the number of rows of the design matrix X and d is the number of columns. If

σ is not known, we run ep-BH on P_j and E_j, for $j = 1, ..., d$, where

$$\begin{cases} P_j = P_j^{(t_2)}, \\ E_j = \dfrac{\left(\dfrac{\hat{\beta}_{1,j}}{\hat{\sigma}_{1,j}}\right)^{2k}}{\tilde{\mu}_k} \end{cases}$$

where k is a positive integer, $\tilde{\mu}_k$ is the kth moment of the $F_{1,\nu}$ distribution, $\hat{\sigma}_{1,j} = \sqrt{2\hat{\sigma}^2((2\Sigma - D)^{-1})_{jj}}$, $\hat{\sigma}_{2,j} = \sqrt{2\hat{\sigma}^2 D_{jj}^{-1}}$ and $\hat{\sigma}^2$ is the usual LS variance parameter estimate.

Proposition 4.2. *If $k < \nu/2$, then E_j is a valid e-value for each $j = 1, ..., d$.*

Proof. If $k < \nu/2$, then $\tilde{\mu}_k$ exists. Under H_0, we have

$$\begin{aligned} \left(\frac{\hat{\beta}_{1,j}}{\hat{\sigma}_{1,j}}\right)^{2k} &= \left(\left(\frac{\hat{\beta}_{1,j}}{\sigma_{1,j}}\right)^2 \cdot \left(\frac{\sigma_{1,j}}{\hat{\sigma}_{1,j}}\right)^2\right)^k \\ &\sim \left(N(0,1)^2 \cdot \frac{1}{\chi^2_\nu/\nu}\right)^k \\ &\sim \left(\frac{\chi^2_1/1}{\chi^2_\nu/\nu}\right)^k \\ &\sim F^k_{1,\nu}, \end{aligned}$$

so that $\mathbb{E}[E_j] = 1$. Here we used the independence of $\hat{\beta}_1$ and $\hat{\sigma}$. □

It is clear that the P_j are valid p values. Unlike the known variance section, we do not conclude that the method controls the FDR; the p values share dependence through $\hat{\sigma}$, and so do the set of p values and set of e values. However, in simulations, the method obtains FDR $\leq \alpha$ because $\hat{\sigma} \approx \sigma$. That is, when n is large enough, the p values and e values behave as if you had known σ, which is what we would expect. Section 4.8 provides more detail.

4.3. **An Alternative Method for Unknown Variance.** When the variance is estimated in the previous method, we use the fact that $\hat{\sigma}$ is close to σ. An alternative, then, is to plug in $\hat{\sigma}$ in place of σ for the known variance method, essentially pretending that σ is known. That is, we will have

$$P_j = 2\left(1 - \Phi\left(\frac{|\hat{\beta}_{2,j}|}{\hat{\sigma}_{2,j}}\right)\right), \tag{4.3}$$

$$E_j = \frac{\left(\dfrac{\hat{\beta}_{1,j}}{\hat{\sigma}_{1,j}}\right)^{2k}}{\mu_{2k}}, \tag{4.4}$$

where μ_{2k} is the $2k$th moment of the $N(0,1)$ distribution.

The issue is that these are not valid p values or e values. The point is that, as before, they are approximately correct. This alternative method and the previous method usually perform the same in simulations. Since the first method provided

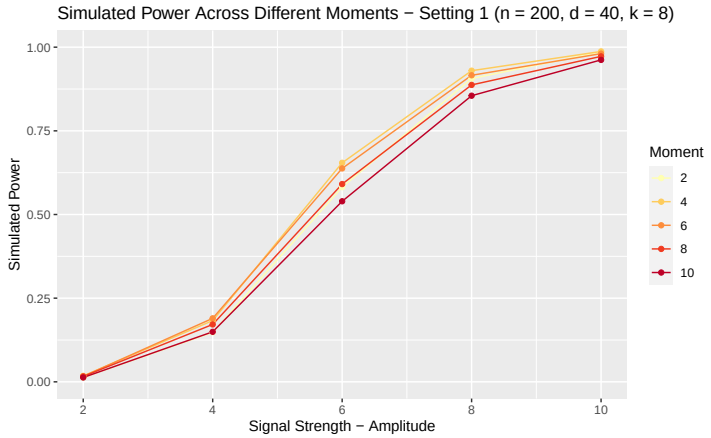

FIGURE 4. Different powers of the test statistic in the e value calculation of the ep-moment method. Fourth power seems to be favored.

provably valid p values and e values, we may as well use the first one instead of this one.

4.4. **Comparison of Different Powers of the Test Statistic.** We can compare the use of different values of k in the calculation of the e value. There is some evidence that using higher powers will increase power of the test. However, in the unknown variance setting, trying to use higher powers in general will often lead to NaN's being produced in R because the moment calculation relies on the Gamma function, which becomes large when n is large compared to $2d$. One can use a simplified formula for the first two moments since mean and variance have simpler formulas (I've only implemented the mean function at the moment to sidestep the Gamma function calculations), but general moment calculations would use the Gamma function.

4.5. **"Dampening" e values.** One can redefine the e values by $E'_j = \frac{1}{2} + \frac{1}{2}E_j$ where E_j are the e values defined in the methods above. This seemed to slightly increase power for higher moments and decreased power for the second moment. Results are shown in Figure 5.

4.6. **Simulation Performance of the ep-Moment method (known variance) vs. Sarkar and Tang (known variance).** To get a proper comparison for the ep-Moment method with known σ, we can adapt [4]'s method to use a z-test instead of the t distribution. In these simulations, we used an autoregressive design matrix, as described in [4]. Results are shown in Figure 6.

4.7. **Simulation Performance of the ep-Moment Method vs. all Other Methods.** The new ep-Moment method performs slightly better than Sarkar and Tang's method. However, the method is usually overshadowed by the newer methods such as dependence adjusted BH (dBH, from [1]) and the calibrated Knockoffs (cKnockoffs, from [3]). Even knockoffs will beat this method in some settings. Furthermore, this method has the same drawbacks as the other knockoff-based methods

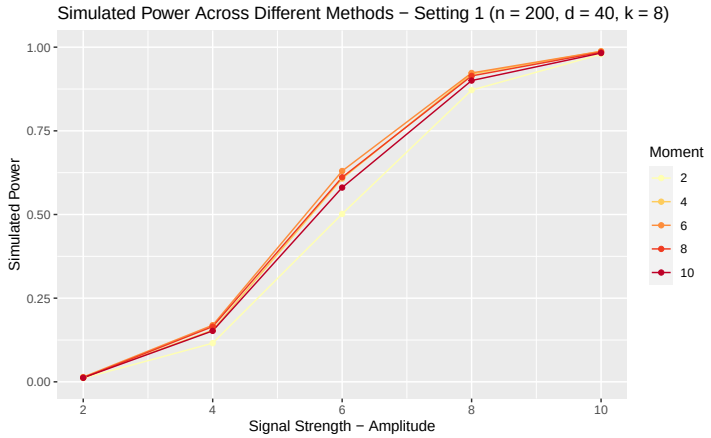

FIGURE 5. E values with dampening can modify the power slightly.

in the MCC setting: it is rendered powerless here. Interestingly, it performed the best in the IID normal setting.

The first 3 simulation settings were those from [4]. The others were based off those in [3], where we vary α and fix the signal strength size. In the settings based off [3], we consider an MCC design setting and an IID normal design setting with $n = 1000, d = 100$, and $k = 10$. The strength size was fixed such that base BH will have power of about 50% when $\alpha = .20$ for the given setting; we use $\beta = 3.66$ for the MCC setting and $\beta = .083$ for the IID normal setting. An advantage of dBH not shown here is that computation time was by far the fastest of the methods; cKnockoff had the slowest by far.

4.8. **Sanity Checks for the Unknown Variance Method.** Empirically, the test controls the FDR because the p values from the F distribution will approximate the p values obtained from the normal distribution when σ is known. Likewise for the e values.

Proposition 4.5. *If $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$ and f_n are continuous functions with $f_n \to f$ uniformly, with f uniformly continuous, then $|f_n(X_n) - f(Y_n)| \xrightarrow{\mathbb{P}} 0$.*

Proof. By uniform convergence, we have $|f_n(X_n) - f(X_n)| \to 0$. Second, let $\varepsilon > 0$; let $\delta > 0$ such that $|x - y| \leq \delta$ implies $|f(x) - f(y)| \leq \varepsilon$. Then

$$\begin{aligned}\mathbb{P}(|f(X_n) - f(Y_n)| > \varepsilon) &\leq \mathbb{P}(|X_n - Y_n| > \delta) \\ &\to 0.\end{aligned}$$

Then,

$$|f_n(X_n) - f(Y_n)| \leq \underbrace{|f_n(X_n) - f(X_n)|}_{\xrightarrow{\mathbb{P}}0} + \underbrace{|f(X_n) - f(Y_n)|}_{\xrightarrow{\mathbb{P}}0},$$

from which the statement follows. □

In the setting below, consider d to be fixed, suppose that H_0 holds for β_j where $j \in \{1, ..., d\}$, and $n \to \infty$.

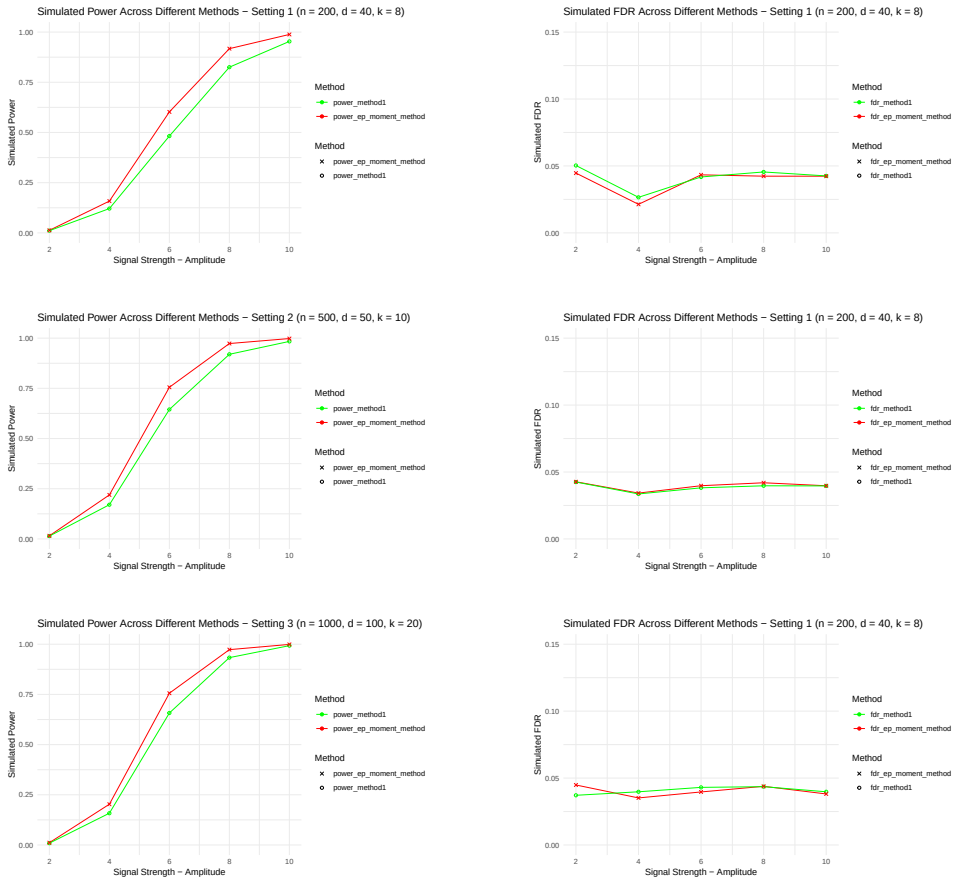

FIGURE 6. In this setting, the ep-Moment method is more powerful than Method 1 of [4], i.e. Bonferroni-BH, and with nearly identical FDR.

Proposition 4.6.

$$\left|\frac{\hat{\beta}_{2,j}^2}{\hat{\sigma}_{2,j}^2} - \frac{\hat{\beta}_{2,j}^2}{\sigma_{2,j}^2}\right| \xrightarrow{\mathbb{P}} 0.$$

Proof. We have $\hat{\sigma}^2 \sim \sigma^2 \chi_\nu^2/\nu \xrightarrow{a.s.} \sigma^2$ by the strong law of large numbers. Then,

$$\left|\frac{\hat{\beta}_{2,j}^2}{\hat{\sigma}_{2,j}^2} - \frac{\hat{\beta}_{2,j}^2}{\sigma_{2,j}^2}\right| = \left|\underbrace{\frac{\hat{\beta}_{2,j}^2}{\sigma_{2,j}^2}}_{\xrightarrow{d}\chi_1^2}\underbrace{\left(\frac{\sigma_{2,j}^2}{\hat{\sigma}_{2,j}^2} - 1\right)}_{\xrightarrow{a.s.}0}\right| \xrightarrow{\mathbb{P}} 0,$$

by Slutsky's theorem. □

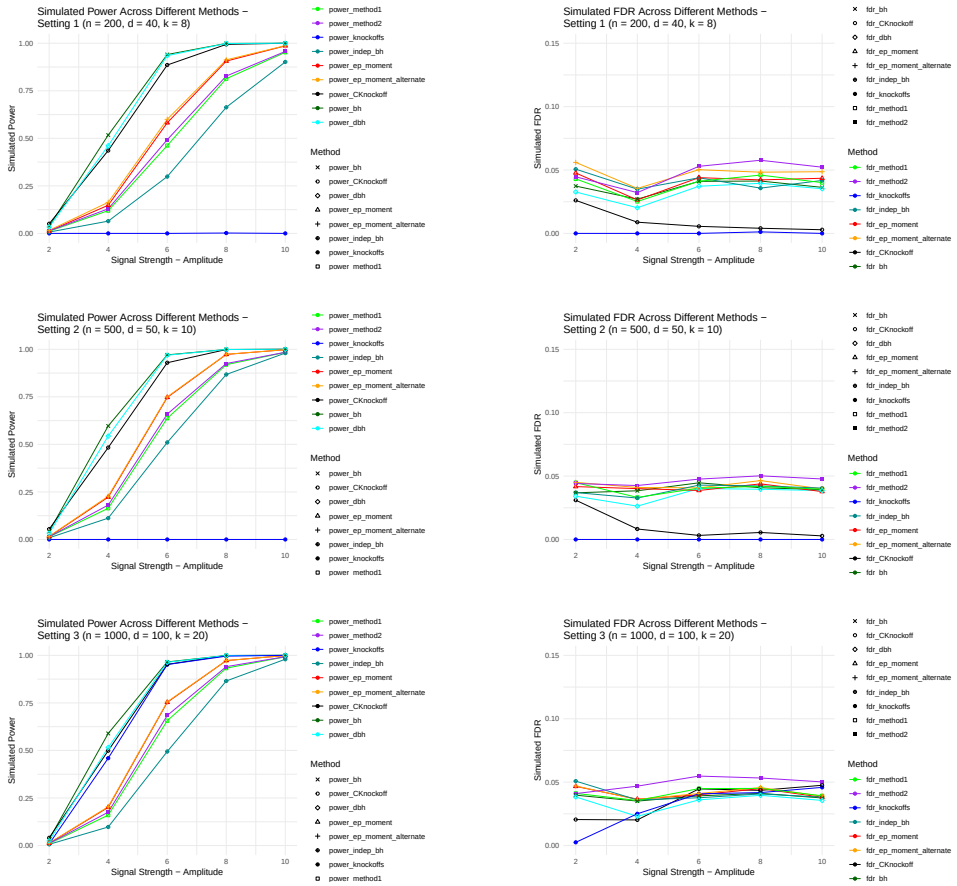

FIGURE 7. In the autoregressive design setting, the ep-Moment method has moderate power compared to the more advanced alternatives dBH and cKnockoffs. Base knockoffs are more powerful if n is large. The ep_moment_alternate method is the method referred to in Section 4.3.

Corollary 4.7. *Let* $\mathbf{F}_\nu$ *denote the CDF of an* $F_{1,\nu}$ *random variable, and consider* $\mathbf{F}_{\chi_1^2}$ *to be the CDF of a* χ_1^2 *random variable. Then*

$$\left| \mathbf{F}_\nu \left(\frac{\hat{\beta}_{2,j}^2}{\hat{\sigma}_{2,j}^2} \right) - \mathbf{F}_{\chi_1^2} \left(\frac{\hat{\beta}_{2,j}^2}{\sigma_{2,j}^2} \right) \right| \xrightarrow{\mathbb{P}} 0.$$

Proof. Slutsky's theorem and the law of large numbers implies that $F_{1,\nu} \xrightarrow{d} \chi_1^2$. Since χ_1^2 is continuous, we therefore have that $\mathbf{F}_\nu \to \mathbf{F}_{\chi_1^2}$ uniformly and $\mathbf{F}_{\chi_1^2}$ is uniformly continuous. Now, Proposition 4.6 implies we may apply Proposition 4.5. □

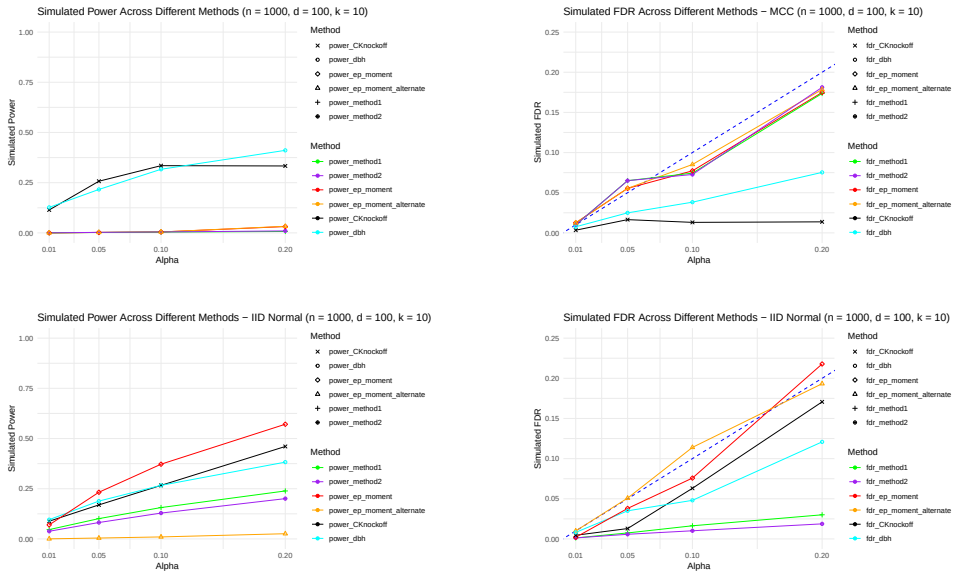

FIGURE 8. The ep-Moment method is powerless in the MCC setting, where only cKnockoffs and dBH are viable. However, it is the most powerful in the IID Normal setting. The blue dashed line is $y = x$.

Now, since $1 - \mathbf{F}_{\chi_1^2}(t) = 2(1 - \Phi(\sqrt{t})$ for $t \geq 0$, we have

$$1 - \mathbf{F}_{\chi_1^2}\left(\frac{\hat{\beta}_{2,j}^2}{\sigma_{2,j}^2}\right) = 2\left(1 - \Phi\left(\frac{|\hat{\beta}_{2,j}|}{\sigma_{2,j}}\right)\right),$$

from which it follows that

$$\left|1 - \mathbf{F}_{\nu}\left(\frac{\hat{\beta}_{2,j}^2}{\hat{\sigma}_{2,j}^2}\right) - 2\left(1 - \Phi\left(\frac{|\hat{\beta}_{2,j}|}{\sigma_{2,j}}\right)\right)\right| \xrightarrow{\mathbb{P}} 0.$$

That is, the p values obtained from the method where σ is estimated will approximate those obtained from the method where σ is known when n is large.

Next, we see that the e values in the estimated σ method will also approximate the e values for the known σ method.

Proposition 4.8.

$$\left|\left(\frac{\hat{\beta}_{1,j}}{\hat{\sigma}_{1,j}}\right)^{2k} - \left(\frac{\hat{\beta}_{1,j}}{\sigma_{1,j}}\right)^{2k}\right| \xrightarrow{\mathbb{P}} 0.$$

Proof. Similar argument as before, along with the continuous mapping theorem. □

Proposition 4.9.

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\frac{\hat{\beta}_{1,j}}{\hat{\sigma}_{1,j}}\right)^{2k}\right] = \mathbb{E}[N(0,1)^{2k}]. \tag{4.10}$$

Proof. If $k = 1$, then the LHS of (4.10) is $\mathbb{E}[F_{1,\nu}] = \frac{\nu}{\nu-2} \to 1 = \mathbb{E}[N(0,1)^2]$. For general k, we have

$$\begin{aligned}
\mathbb{E}[F_{1,\nu}^k] &= \nu^k \frac{\Gamma(\frac{1}{2}+k)\Gamma(\frac{\nu}{2}-k)}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \\
&\to 2^k \frac{\Gamma(\frac{1}{2}+k)}{\sqrt{\pi}} \\
&= (2k-1)!! \\
&= \mathbb{E}[N(0,1)^{2k}].
\end{aligned}$$

The first equation is a general fact about the F distribution; the second line is a limit calculation that follows from Stirling's formula. The third is a property of the double factorial, and the fourth is a well known property of the normal distribution. □

Hence, we see that the e values have the desired limiting behavior.

References

[1] Fithian, W., & Lei, L. (2022). Conditional calibration for false discovery rate control under dependence. The Annals of Statistics, 50(6). https://doi.org/10.1214/21-aos2137

[2] Ignatiadis, N., Wang, R., & Ramdas, A. (2023). e-values as unnormalized weights in multiple testing. Biometrika, 111(2), 417–439. https://doi.org/10.1093/biomet/asad057

[3] Luo, Y., Fithian, W., & Lei, L. (2023). Improving Knockoffs with Conditional Calibration. Arxiv Preprint.

[4] Sarkar, S. K., & Tang, C. Y. (2021). Adjusting the benjamini–hochberg method for controlling the false discovery rate in knockoff-assisted variable selection. Biometrika, 109(4), 1149–1155. https://doi.org/10.1093/biomet/asab066