# Rejoinder: A Scale-free Approach for False Discovery Rate Control in Generalized Linear Models

Chenguang Dai,[*1], Buyu Lin,[*1], Xin Xing[2], and Jun S. Liu[1]

[1]*Department of Statistics, Harvard University*
[2]*Department of Statistics, Virginia Tech*

[*]These authors contribute equally to this work.

# 1 Introduction

We thank all the discussants for their critical observations and inspiring comments, which not only help improve our own understanding of the area but also stimulate many new thoughts for future research. We are grateful to the editors for organizing this thought-provoking event and to all the discussants for their invaluable contributions.

There are many exciting new ideas proposed by the discussants that are worthy of further investigations. Claeskens, Jansen, and Zhou (Claeskens&Jansen&Zhou) raised important questions regarding the ultra-sparse setting and the impact of tuning parameter choices; Janson pointed out the two-stage formulation in Li and Fithian (2021) and possible extensions of DS; Law and Bühlmann (Law&Bühlmann) provided detailed power calculations for GM and DS; Li, Yao and Zhang (Li&Yao&Zhang) outlined numerous novel findings (e.g., scale-free FDR control, fastMDS etc.); Xia and Cai (Xia&Cai) illustrated better ways of adjusting the BHq procedure and the bias-variance tradeoff; Zhang and Ma (Zhang&Ma) pointed out a possible way of improving DS by combining with factor-adjusted techniques and suggested future researches for regression with hierarchically-structured predictors; and both Zhang&Ma and Xia&Cai pointed out a potential connection between MDS and e-value based approaches.

In this rejoinder, due to the space limit, we are only able to highlight some comments raised by the discussants and provide some analyses in response to a few of their suggestions. In Section 2, we reformulate DS into a two-stage algorithm and discuss new potentials to construct mirror statistics. In Section 3, we connect the MDS procedure with e-value and propose a new algorithm to derandomize DS. In Section 4, we compare the ranking qualities of different FDR control procedures and thus provide a way to theoretically compare their powers. In Section 5, we discuss possible improvements and future research directions.

# 2 Reformulation to Facilitate Use of Prior Information

## 2.1 A unified framework for Knockoff and DS

A common feature of Knockoff, DS, and GM is that they construct test statistics by contrasting two important measures of each feature. We here revisit the contrasting formulation from a conditional

inference point of view to gain some new insights.

Li and Fithian (2021) showed that Knockoff can be recast as a conditional inference procedure on a "whitened" estimator. Inspired by Janson's comment, we present a similar reformulation. This formulation not only brings together DS and Knockoff within a unified framework but also sheds light on how to construct more powerful mirror statistics. It further complements the algebraic equivalence between Knockoff and DS as pointed out by Li&Yao&Zhang. 1 Under the regression setting with $p$ features, and more generally, we consider $p$ hypotheses $H_j$, $j \in [p]$, each associated with the null $\beta_j = 0$. Suppose we have two independent estimates $\widehat{\beta}_j^{(1)}$ and $\widehat{\beta}_j^{(2)}$ for coefficient $\beta_j$, of which $\widehat{\beta}_j^{(2)}$ follows a normal distribution centered at $\beta_j$. DS can be viewed as a special case of the following two-stage algorithm.

Stage 1. We accomplish the following two tasks with only access to $\widehat{\beta}^{(1)}$, $|\widehat{\beta}^{(2)}|$, and domain knowledge or information from independent data sources: (a) order the $p$ hypotheses, $H_{(1)}, \ldots, H_{(p)}$, so that the front ones are more likely rejected; and (b) make a guess $\psi_j$ on the sign of $\beta_j$, where $\psi_j = 1$ corresponds to $\beta_j > 0$ and $-1$, otherwise.

Stage 2. We verify the consistency of our guesses by comparing $\psi_j$ with $\mathrm{sign}(\widehat{\beta}_j^{(2)})$. For each $k$, we reject those hypotheses among $H_{(1)}, \ldots, H_{(k-1)}$ with $\mathrm{sign}(\widehat{\beta}_{(j)}^{(2)}) = \psi_{(j)}$. We estimate the FDP of this rejection set as:

$$\mathrm{FDP}_k = \frac{1 + \sum_{i<k} \mathbb{1}(\mathrm{sign}(\widehat{\beta}_{(i)}^{(2)}) \neq \psi_{(i)})}{\sum_{i<k} \mathbb{1}(\mathrm{sign}(\widehat{\beta}_{(i)}^{(2)}) = \psi_{(i)}) \vee 1}. \tag{1}$$

The procedure ends by finding $\hat{k} = \arg\max_k \{\mathrm{FDP}_k \leq q\}$.

To show that DS is a special case of the above two-stage algorithm, we set $\psi_j = \mathrm{sign}(\widehat{\beta}_j^{(1)})$ and sort the hypotheses based on $|M_j|$ in the decreasing order. Recall the mirror statistics:

$$M_j = \mathrm{sign}(\widehat{\beta}_j^{(1)})\mathrm{sign}(\widehat{\beta}_j^{(2)}) f(|\widehat{\beta}_j^{(1)}|, |\widehat{\beta}_j^{(2)}|), \tag{2}$$

in which $f(u, v)$ is non-negative, symmetric about $u$ and $v$, and monotonically increasing in both $u$ and $v$. For each $k$, we have:

$$\mathrm{FDP}_k = \frac{1 + \sum_{i<k} \mathbb{1}(\mathrm{sign}(\widehat{\beta}_{(i)}^{(2)}) \neq \psi_{(i)})}{\sum_{i<k} \mathbb{1}(\mathrm{sign}(\widehat{\beta}_{(i)}^{(2)}) = \psi_{(i)}) \vee 1} = \frac{1 + \sum_{i<k} \mathbb{1}(M_{(i)} < 0)}{\sum_{i<k} \mathbb{1}(M_{(i)} > 0) \vee 1} = \frac{1 + \sum_j \mathbb{1}(M_j < -|M_{(k)}|)}{\sum_j \mathbb{1}(M_j > |M_{(k)}|) \vee 1}, \tag{3}$$

where the last equality holds because the order of the hypotheses is given by $|M_j|$. Note that the key step in Algorithm 1 in Dai et al. (2023) is to find the smallest $t > 0$, denoted as $\tau_q$, such that

$$\widehat{\mathrm{FDP}}(t) = \frac{1 + \sum_j \mathbb{1}(M_j < -t)}{\sum_j \mathbb{1}(M_j > t) \vee 1} \leq q. \tag{4}$$

We then reject all the hypotheses with $M_j > \tau_q$. This is equivalent to Stage 2 since the value of $\widehat{\mathrm{FDP}}(t)$ only changes when $t$ crosses some $|M_j|$.

We should emphasize that $\mathrm{sign}(\widehat{\beta}_j^{(2)})$ can not be used at Stage 1. Specifically, the validity of the two-stage algorithm relies on the following approximation:

$$\sum_{i<k} \mathbb{1}(\mathrm{sign}(\widehat{\beta}_{(i)}^{(2)}) \neq \psi_{(i)}, H_{(i)} \text{ is null}) \approx \sum_{i<k} \mathbb{1}(\mathrm{sign}(\widehat{\beta}_{(i)}^{(2)}) = \psi_{(i)}, H_{(i)} \text{ is null}). \tag{5}$$

The above equation holds since $\mathrm{sign}(\widehat{\beta}_{(j)}^{(2)})$ has equal probability of being $\pm 1$ under the null and is independent of the ordering of the hypotheses and the choice of $\psi_j$.

## 2.2 Prior-assisted DS

The flexible two-stage formulation of DS enables us to incorporate prior information into the procedure. For example, in genetic studies, we may have domain knowledge and other independent data showing that certain genetic variations are more likely associated with a given disease. To illustrate this, we consider the following toy example.

Assume that $\widehat{\beta}_j^{(1)}$ and $\widehat{\beta}_j^{(2)}$ independently follow $N(\beta_j, 1)$, where $\beta_j = \sqrt{\delta \log p}$ for $j \in S_1$, with $\delta$ controlling the signal-to-noise ratio. Suppose we know that all the $\beta_j$'s are all non-negative. We consider the following three mirror statistics:

$$\begin{aligned} M_j &= \mathrm{sign}(\widehat{\beta}_j^{(1)})\mathrm{sign}(\widehat{\beta}_j^{(2)})(|\widehat{\beta}_j^{(1)}| + |\widehat{\beta}_j^{(2)}|), \\ \widetilde{M}_j &= \mathrm{sign}(\widehat{\beta}_j^{(2)})(|\widehat{\beta}_j^{(1)}| + |\widehat{\beta}_j^{(2)}|), \\ \widehat{M}_j &= \mathrm{sign}(\widehat{\beta}_j^{(2)})(\widehat{\beta}_j^{(1)} + |\widehat{\beta}_j^{(2)}|). \end{aligned} \tag{6}$$

From the two-stage perspective, we can carefully specify $\psi_j$ and order the hypotheses using the prior information that $\beta_j$'s are non-negative. While $M_j$ sets $\psi_j = \mathrm{sign}(\widehat{\beta}_j^{(1)})$ without considering any prior information, $\widetilde{M}_j$ and $\widehat{M}_j$ set their $\psi_j = 1$ to be consistent with the prior. As for ordering the hypotheses, $|\widehat{\beta}_j^{(1)} + |\widehat{\beta}_j^{(2)}||$ used in $\widehat{M}_j$ is a better ranking than $|\widehat{\beta}_j^{(1)}| + |\widehat{\beta}_j^{(2)}|$ used in $M_j$ and $\widetilde{M}_j$. Indeed, if we know $\beta_j \geq 0$ and see that $\widehat{\beta}_j^{(1)} < 0$, we should downgrade the priority of $H_j$.

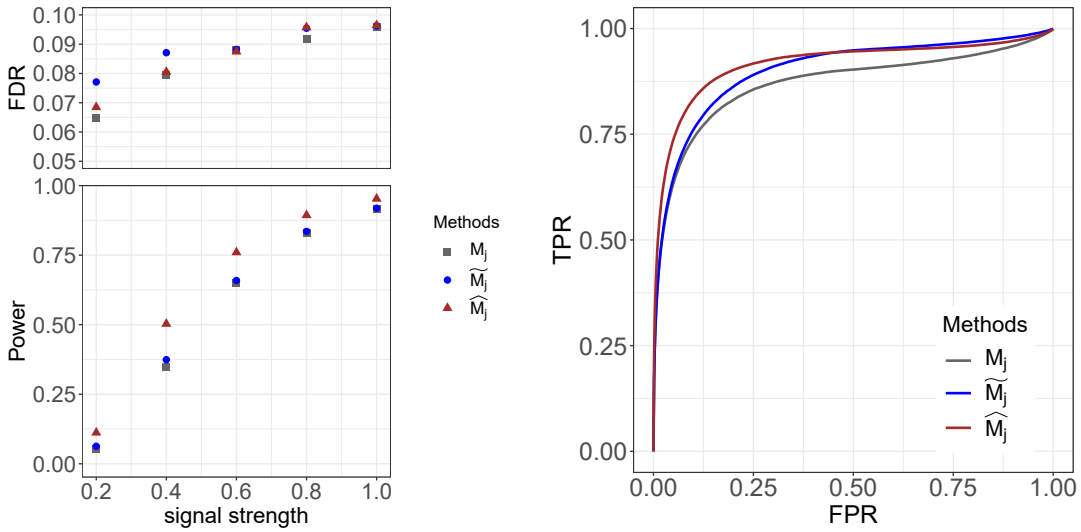


Figure 1: Empirical FDRs, powers (left figure) and ROC curve (right figure) for the Normal mean model. We set the number of features $p = 800$ and the number of relevant features $p_1 = 160$. We vary the signal strength from 0.2 to 1.0. Each dot and each line represent the average from 500 independent runs.

We set the number of features as $p = 800$, of which $p_1 = 160$ are relevant features. We replicate DS 500 times, and the results of the FDR, power, and ROC curve are shown in Figure 1. Empirically, all three mirror statistics control FDR across all settings, with $\widehat{M}_j$ enjoying the best ROC curve and the highest power, indicating that a better choice of $\psi_j$ and a carefully-designed ordering strategy can both boost the power. This example demonstrates that given prior information, there could be better choices of mirror statistics than the form of $M_j$ in (6), which was proven to be optimal in a general prior-free setting (see Proposition 2.1 in Dai et al. (2022)).

Note that $\widehat{M}_j$ may still be sub-optimal mirror statistics in this case. It is an interesting challenge for the researchers to figure out how to systematically and optimally incorporate prior information in DS and other FDR control methods to improve their powers.

# 3 Connections to E-value

## 3.1 Derandomize DS via e-value

As pointed out in the discussions of Xia&Cai and Zhang&Ma, the e-BH procedure (Wang and Ramdas, 2022) has received much attention recently as an alternative to p-value based FDR control

methods. Ren and Barber (2023) and Xia&Cai show that both Knockoff and DS can be formulated as an e-BH procedure. This leads to a different way of derandomizing DS using e-values than MDS.

Let $M_j^{(k)}$ denote the mirror statistic of feature $X_j$ in the $k$-th data split, $k \in [m]$. Define

$$e_j^{(k)} = p \frac{\mathbb{1}(M_j^{(k)} > \tau_\alpha^{(k)})}{1 + \sum_{s=1}^{p} \mathbb{1}(M_s^{(k)} < -\tau_\alpha^{(k)})}, \tag{7}$$

in which $\tau_\alpha^{(k)}$ is the threshold defined in Algorithm 1 in Dai et al. (2023). Note that FDR control of the e-BH procedure outlined in the Algorithm 1 below holds for any $\alpha > 0$, although this parameter can potentially affect power. Ren and Barber (2023) suggested $\alpha = q/2$, and pointed out that $\alpha$ should be smaller than $q$ when $m > 1$; otherwise, the e-BH procedure may have zero power. In practice, without losing asymptotic FDR control, we may try multiple choices of $\alpha < 1$ and choose the one that yields the largest selection set.

**Algorithm 1** Derandomize DS: an e-BH procedure.

1. Calculate the average e-values: $\bar{e}_j = \sum_{k=1}^{m} e_j^{(k)}/m,\ j \in [p]$.
2. Sort the average e-values: $\bar{e}_{(1)} \geq \bar{e}_{(2)} \geq \ldots \geq \bar{e}_{(p)}$.
3. Given a designated FDR level $q \in (0, 1)$, find the largest $\ell \in [p]$ such that $\bar{e}_{(\ell)} \geq p/(\ell q)$.
4. Select the features $\widehat{S} = \{j : \bar{e}_j \geq p/(\ell q)\}$.

**Theorem 3.1.** *Under Assumptions 3.1 and 4.1 in Dai et al. (2023) for the moderate-dimensional and the high-dimensional regime, respectively, the e-BH procedure in Algorithm 1 asymptotically controls FDR at any designated level $q \in (0, 1)$.*

The proof of Theorem 3.1 is very similar to that of Theorem 3 in Ren and Barber (2023). The key is to establish

$$\mathbb{E}\left[\frac{\sum_{j \in S_0} \mathbb{1}(M_j^{(k)} > \tau_\alpha^{(k)})}{\sum_{j \in S_0} \mathbb{1}(M_j^{(k)} < -\tau_\alpha^{(k)})}\right] \leq 1 + o(1)$$

for $\forall \alpha > 0$, which, as shown in Remark 3.4 in Dai et al. (2023), is also crucial for justifying the DS procedure, and relies on certain assumptions on the covariance matrix of the features.

## 3.2 Comparison with MDS

The inclusion rate and the e-value has the following relationship.

**Theorem 3.2.** *$\forall \alpha > 0$, assume MDS targets at the FDR level $\alpha$ and the e-value is defined as in Equation (7), then the inclusion rate $I_j$ and the average e-value $\bar{e}_j$ satisfy*

$$\frac{\alpha \bar{e}_j}{p I_j} \to 1$$

*under Assumptions 3.1(2) or 4.1(3)(b) of Dai et al. (2023) for moderate or high dimensions, respectively.*

**Remark 3.1.** *Note that both the inclusion rate $I_j$ and the average e-value $\bar{e}_j$ can be written as*

$$\frac{1}{m} \sum_{k=1}^{m} w_j^{(k)} \mathbb{1}(M_j^{(k)} > \tau_\alpha^{(k)}),$$

*in which the weight $w_j^{(k)}$ is*

$$I_j : \ 1/w_j^{(k)} = \sum_{s=1}^{p} \mathbb{1}(M_s^{(k)} > \tau_\alpha^{(k)}); \quad \bar{e}_j : \ p/w_j^{(k)} = 1 + \sum_{s=1}^{p} \mathbb{1}(M_s^{(k)} < -\tau_\alpha^{(k)}).$$

*By the definition of $\tau_\alpha$, we have $\alpha \bar{e}_j \geq p I_j$. Theorem 3.2 shows that this inequality in fact becomes an equality as $p \to \infty$. In fact, the above theorem holds true only if $\min_{k \in [m]} |\widehat{S}^{(k)}| \to \infty$, which can be guaranteed by Assumption 3.1(2) and Assumption 4.1(3)(b) of Dai et al. (2023).*

We shall emphasize that MDS and the e-BH procedure are still very different despite of the nice coincidence in Theorem 3.2. For example, the connection between their selection rules remains unclear. We empirically compare the two procedures on a simple Normal means model. We set $n = 500$, $p = 800$, and sample $X_{ij}$ from $N(\mu_j, 1)$ for $i \in [n], j \in [p]$. We set 20% of $\mu_j$'s to be nonzero, and generate them from $N(0, 2r \log p/n)$. $r$ is referred to as the signal strength, and we test out scenarios with $r$ varing from 0.2 to 1.8. The mirror statistic for each $\mu_j$ is

$$\text{sign}(\bar{X}_j^{(1)} \bar{X}_j^{(2)})(|\bar{X}_j^{(1)}| + |\bar{X}_j^{(2)}|),$$

where $\bar{X}_j^{(1)}$ and $\bar{X}_j^{(2)}$ are the sample means on each half of the data. We also vary the designated FDR control level $q$ from 0.05 to 0.3. The simulation results are summarized in Figure 2. Both methods control FDR well across most of the settings. MDS tends to have a higher power when the signal is weak or the designated level $q$ is small.

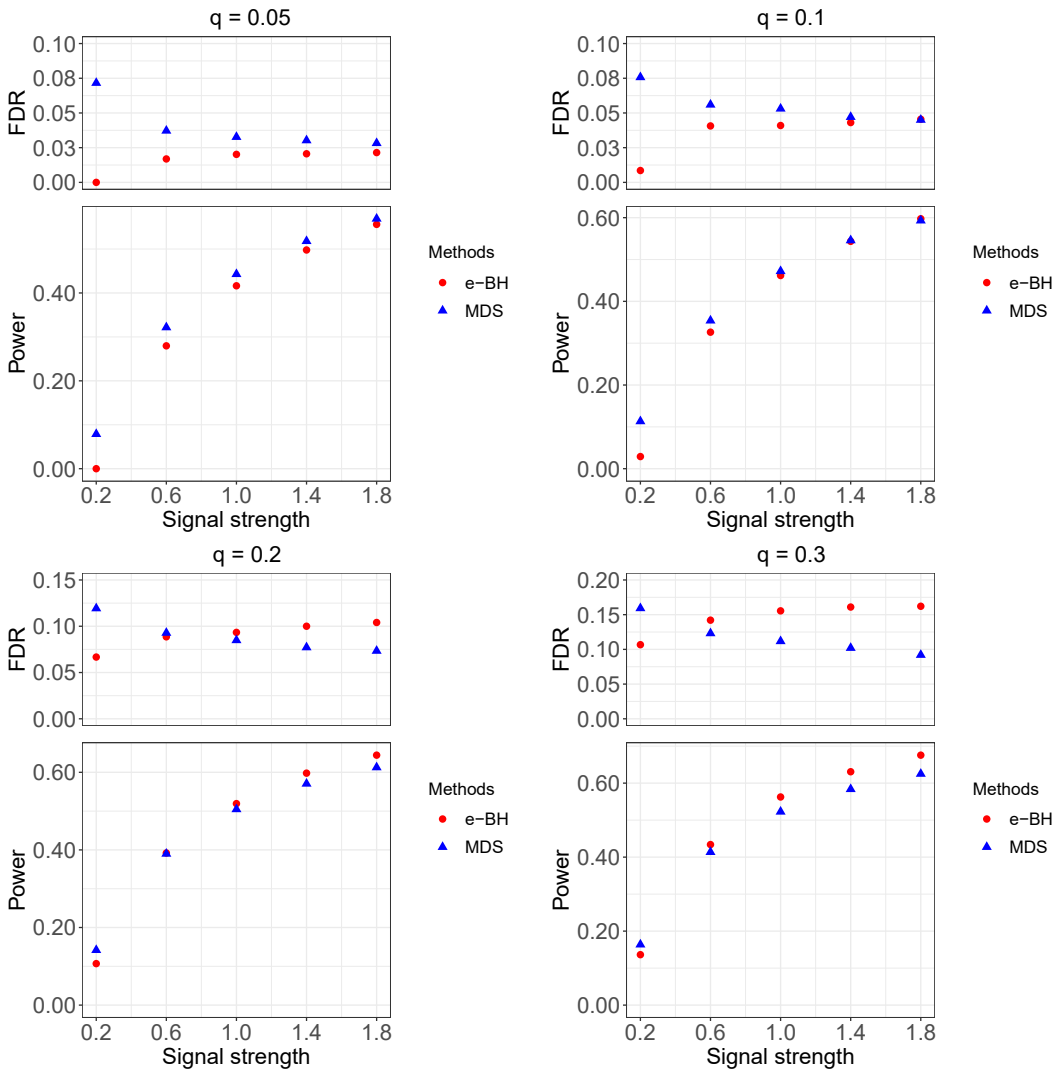


Figure 2: Empirical FDRs and powers for the Normal mean model. We set the number of samples $n = 500$ and the number of features $p = 800$. We vary the signal strength from 0.2 to 1.8, and vary the designated FDR control level $q$ from 0.05 to 0.3. Each dot represents the average from 100 independent runs.

# 4 Power Analysis

Law&Bühlmann compared GM and DS for linear models in the moderate-dimensional regime where $p/n \to \kappa \in [0, 1/2)$. We here extend their results to Knockoffs. Consider the linear model:

$$y = X\beta^\star + \epsilon, \tag{8}$$

in which $X_{n\times p}$ is a random design matrix and $\epsilon_{n\times 1} \sim N(0, \sigma^2 I_n)$. Let $\widetilde{X}$ be the Knockoff features. Let $\Sigma$ and $\Sigma^\star$ denote the covariance matrices for $X$ and $[X, \widetilde{X}]$, respectively. Theorem 4.1 below summaries the variance calculations for the regression coefficient estimator $\widehat{\beta}$ of different methods. Similar calculations also apply to generalized linear models (GLMs) in the moderate-dimensional regime.

**Theorem 4.1.** *Consider the linear model in (8) where the rows of $X$ are i.i.d. samples from $N(0, \Sigma)$. Denote $\Theta = \Sigma^{-1}$, $\Theta^\star = \Sigma^{\star -1}$ and $\tau_j^2 = 1/\Theta_{jj}$, $\omega_j^2 = 1/\Theta^\star_{jj}$, $j \in [p]$. When $n > 2p + 2$, we have*

| Estimator | $\widehat{\beta}_j^{\mathrm{OLS}}$ | $\widehat{\beta}_j^{\mathrm{KN}}$ | $(\widehat{\beta}_{j,1}^{\mathrm{GM}} + \widehat{\beta}_{j,2}^{\mathrm{GM}})/2$ | $(\widehat{\beta}_{j,1}^{\mathrm{DS}} + \widehat{\beta}_{j,2}^{\mathrm{DS}})/2$ |
|---|---|---|---|---|
| Variance | $\dfrac{\sigma^2}{\tau_j^2(n-p-1)}$ | $\dfrac{\sigma^2}{\omega_j^2(n-2p-1)}$ | $\dfrac{\sigma^2}{\tau_j^2(n-p-2)}$ | $\dfrac{\sigma^2}{\tau_j^2(n-2p-2)}$ |

OLS yields the lowest variance of $\widehat{\beta}$, which is free of additional noise (GM, Knockoff) or sample splitting (DS). Besides, when $p/n \to \kappa$, the relative efficiency of GM, DS and Knockoff against OLS are 1, $(1-2\kappa)/(1-\kappa)$, and $((1-2\kappa)\omega_j^2)/((1-\kappa)\tau_j^2)$, respectively, i.e., OLS = GM > DS > Knockoff. The gap between GM and DS vanishes as $\kappa \to 0$, and the gap between DS and Knockoff is large if $\omega_j^2/\tau_j^2$ is small. Since $\Sigma$ is a principal sub-matrix of $\Sigma^\star$, we have $\omega_j^2 \le \tau_j^2$, where the inequality is often strict. Specifically, $\tau_j^2$ and $\omega_j^2$ are the conditional variances of $X_j$ given $X_{-j}$ and $[X_{-j}, \widetilde{X}]$, respectively. Thus, $\omega_j^2 = \tau_j^2$ only if $X_j \perp \widetilde{X} \mid X_j$. If we use OLS to rank features, the optimal construction of $\widetilde{X}$ should be the one that maximizes $\omega_j^2$, i.e., the MVR-Knockoff procedure proposed in Spector and Janson (2020).

We consider a special case where $\Sigma$ is block-wise diagonal. Each block is a $2\times 2$ matrix, in which the diagonal and the off-diagonal elements are 1 and $\rho$, respectively. In this case, $\tau_j^2 = 1 - \rho^2$ and all $\omega_j^2$'s are equal for $j \in [p]$. We calculate the relative efficiency of the MVR-Knockoff procedure

against DS, i.e., $\omega_j^2/\tau_j^2$, for $\rho$ varying from 0 to 0.9. Table 1 shows that the relative efficiency of MVR-Knockoffs decreases as the correlation $\rho$ increases. This is consistent with the patterns observed in Figures 2, 3, and 4 of Dai et al. (2023).

| $\rho$ | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\omega_j^2/\tau_j^2$ | 1.00 | 0.99 | 0.97 | 0.93 | 0.89 | 0.85 | 0.82 | 0.78 | 0.75 | 0.72 |

Table 1: Relative efficiency of the MVR-Knockoff procedure against DS.

We end this section by clarifying the question raised by Claeskens&Jansen&Zhou regarding the power of DS and MDS in ultra-sparse models. Specifically, they see a power decrease of general contrast-based methods outlined in Section 2 (e.g., DS/MDS/Knockoff) when the model is ultra-sparse. For example, when the number of relevant features is only 5, if we set $q = 0.1$, the selection rule in (4) will make no rejections even when the signal-to-noise ratio is arbitrarily high (because of "+1" in its numerator). This also explains why the power gap between BHq and the contrast-based methods disappears when we set a larger $q$.

# 5 Future Directions

Motivated by the fascinating comments and analyses from our discussants, in this section, we list some future directions that may enhance the performance and broaden the applicability of DS and other FDR-control methods.

A number of discussants have suggested alternative approaches to DS. Janson highlighted the post-selection literature, which provides a novel approach for obtaining two independent estimators by introducing additional randomness to the response variable. Rasines and Young (2022) showed that this approach allows for a higher selection and inferential power than DS. Li&Yao&Zhang proposed a "mirror-statistic-free" method, which relies on well-designed $t$-statistics and is also free of estimating the variance factor. Their simulation showed that the proposed $\sigma_n$-BH method achieved a comparable or even higher power compared to DS. However, their approach still faces challenges of instability and potential power loss associated with data splitting, and the MDS framework may help resolve these issues. Xia&Cai corrected the BHq procedure proposed in Ma

et al. (2020) by adopting the variance estimator adjustment from Liu et al. (2021), which can be a subject of independent investigation.

Law&Bühlmann proposed a high dimensional GM approach for linear models and demonstrated encouraging results in simulation studies. However, their method is computationally demanding, which requires a total of $p^2$ Lasso fittings. The fast data-splitting technique proposed by Li&Yao& Zhang offers potential solutions to speed up GM in high dimensions. Furthermore, extending the methodology to high-dimensional GLMs presents an intriguing avenue for future explorations.

Claeskens&Jansen& Zhou raised an important point regarding the choice and the impact of the hyper-parameters in our and other related FDR control methods. Theoretically speaking, the hyper-parameters should be specified so that the *sure screening* property (Dai et al., 2022) or the asymptotic normality (Dai et al., 2023) holds. It is important to investigate the robustness of our method with respect to different choices of the hyper-parameters and examine how they affect the performance of the approach.

It is of great value to generalize the DS framework to more structured and non-linear problems. Zhang&Ma pointed out some promising applications of our method, e.g., high dimensional interaction analysis and collective analysis of data from multiple resources. For non-linear problems, Zhao and Xing (2023) combines the DS procedure with sliced inverse regression to control the FDR without assuming any conditional distribution of the response. Exploring group feature selection and investigating additional applications in additive models could also yield fruitful insights.

Our rejoinder aims at stimulating further explorations in combining classic data-splitting approaches with FDR control. The discussions so far have highlighted several promising directions for future research. We would like to express our sincere gratitude once again to the discussants and editors for providing us with this invaluable opportunity to engage in this meaningful and insightful intellectual exchange.

# References

Dai, C., B. Lin, X. Xing, and J. S. Liu (2022). False discovery rate control via data splitting. *Journal of the American Statistical Association*.

Dai, C., B. Lin, X. Xing, and J. S. Liu (2023). A scale-free approach for false discovery rate control in generalized linear models. *Journal of the American Statistical Association 0*(0), 1–15.

Li, X. and W. Fithian (2021). Whiteout: when do fixed-x knockoffs fail?

Liu, M., Y. Xia, K. Cho, and T. Cai (2021). Integrative high dimensional multiple testing with heterogeneity under data sharing constraints. *The Journal of Machine Learning Research 22*(1), 5607–5632.

Ma, R., T. Tony Cai, and H. Li (2020). Global and simultaneous hypothesis testing for high-dimensional logistic regression models. *Journal of the American Statistical Association*, 1–15.

Rasines, D. G. and G. A. Young (2022, 12). Splitting strategies for post-selection inference. *Biometrika*, asac070.

Ren, Z. and R. F. Barber (2023). Derandomized knockoffs: leveraging e-values for false discovery rate control.

Spector, A. and L. Janson (2020). Powerful knockoffs via minimizing reconstructability. *arXiv preprint: 2011.14625*.

Wang, R. and A. Ramdas (2022, 01). False Discovery Rate Control with E-values. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(3), 822–852.

Zhao, Z. and X. Xing (2023). On the testing of multiple hypothesis in sliced inverse regression.