JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL 118, NO. 543, 1586–1589: Theory and Methods Discussion
https://doi.org/10.1080/01621459.2023.2224412

Comments on "A Scale-Free Approach for False Discovery Rate Control in Generalized Linear Models"

Sai Li[a], Yisha Yao[b], and Cun-Hui Zhang[c]

[a] Institute of Statistics and Big Data, Renmin University of China, Beijing, China; [b] Department of Biostatistics, Yale University, New Haven, CT; [c] Department of Statistics, Rutgers University, Piscataway, NJ

ARTICLE HISTORY
Received May 2023
Accepted June 2023

We congratulate Chenguang Dai, Buyu Lin, Xin Xing and Jun Liu on their development of interesting methods for asymptotic FDR control and related theory in generalized linear models. We appreciate this opportunity to comment on their thought-provoking paper.

1. Introduction

Consider testing null hypotheses $H_j : \theta_j = 0$ with statistics Z_j, $j \in [p]$. Let $S_0 = \{j \in [p] : \theta_j = 0\}$ be the index set for the true nulls. Suppose it is sensible to reject H_j when Z_j is large. Let $\phi_t(x) = I\{x \geq t\}$ denote the threshold test. For tests with a common threshold level t, $R_t(Z_{[p]}) = \sum_{j=1}^{p} \phi_t(Z_j)$ is the total number of rejected hypotheses and $R_{0,t}(Z_{[p]}) = \sum_{j \in S_0} \phi_t(Z_j)$ the total number of false rejections. Benjamini and Hochberg (1995) advocated to measure the Type-I error by FDR in multiple testing and proved that for independent test statistics Z_j with a known common continuous null survival function G_0, the adaptive threshold level

$$T = \inf\{t : R_t(Z_{[p]}) \geq G_0(t)p/q\} \tag{1.1}$$

of the global test for the intersection null $\cap_{j=1}^{p} H_j$ in Simes (1986) also controls the FDR:

$$\mathrm{FDR} = \mathbb{E}\left[\frac{R_{0,T}(Z_{[p]})}{1 \vee R_T(Z_{[p]})}\right] \leq \frac{p_0 q}{p} \tag{1.2}$$

where $p_0 = |S_0|$. Benjamini and Yekutieli (2001) relaxed the mutual independence assumption on the test statistics to the positive regression dependency on each one from the subset S_0 (PRDS), and the strict null assumption to $\mathbb{P}\{Z_k \geq t\} \leq G_0(t)$ with $G_0(t) = q_k/p$ for $k \in [p], j \in S_0$.

Barber and Candes (2015) introduced knockoffs to expand the realm of non-asymptotic FDR control. Statistics $\tilde{Z}_1, \ldots, \tilde{Z}_p$ are knockoffs of $Z_1, \ldots, Z_p$ if $(Z_j, \tilde{Z}_j, j \in B, Z_k, \tilde{Z}_k, k \notin B)$ has the same joint distribution as its B-swapped version $(\tilde{Z}_j, Z_j, j \in B, Z_k, \tilde{Z}_k, k \notin B)$ for every null subset $B \subseteq S_0$. Let $\tilde{Z}_1, \ldots, \tilde{Z}_p$ be knockoffs of $Z_1, \ldots, Z_p$,

$$s_j = \mathrm{sgn}(Z_j - \tilde{Z}_j), \quad W_j = s_j w(Z_j, \tilde{Z}_j), \tag{1.3}$$

with a function satisfying $w(x, \tilde{x}) = w(\tilde{x}, x) \geq 0$. They proposed to use the threshold test $\phi_T(W_j)$ at the adaptive threshold level

$$T(W_{[p]}) = \inf\left\{t : \frac{\delta + R_{-t}(W_{[p]})}{1 \vee R_t(W_{[p]})} \leq q\right\}, \tag{1.4}$$

and proved FDR $\leq q$ for the rule with $\delta = 1$ and a related error bound for $\delta = 0$. A crucial feature of the knockoffs is the sign symmetry $\{s_j, j \in S_0\} \sim$ uniform$\{-1, 1\}^{S_0}$ in the null set. The knockoff approach has been further developed in Candes et al. (2018) and Huang and Janson (2020).

The beauty of the above results is the non-asymptotic nature of the FDR guarantee. Still, the PRDS assumption and the knowledge of the null survival function, or the knockoffs, may not be available in practice.

Following Xing, Zhao, and Liu (2021) and Dai et al. (2022), Dai et al. (2023) advocates the use of the thresholding level (1.4) on mirror statistics. Sufficient conditions are developed for approximate FDR control in moderate- and high-dimensional generalized linear models (GLM). Mirror statistics can be constructed by data splitting (DS) when asymptotically normal estimates of θ_j are available. Let $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ be two independent estimators of θ. The mirror statistics are defined by

$$s_j = \mathrm{sgn}(\hat{\theta}_j^{(1)})\,\mathrm{sgn}(\hat{\theta}_j^{(2)}), \quad M_j = s_j M(|\hat{\theta}_j^{(1)}|, |\hat{\theta}_j^{(2)}|), \tag{1.5}$$

with a function satisfying $M(x, \tilde{x}) = M(\tilde{x}, x) \geq 0$ as the $w(\cdot, \cdot)$ in (1.3). Suppose

$$\hat{\theta}_j^{(k)}/a_n \xrightarrow{D} N(\mu_{n,j}, 1), \quad k = 1, 2, \tag{1.6}$$

for some unknown parameters a_n and $\mu_{n,j}$ satisfying $\mathrm{sgn}(\mu_{n,j}) = \mathrm{sgn}(\theta_j)$. For example, in Proposition 3.1 of Dai et al. (2023), $\theta_j = \beta_j/\sigma, a_n = a^*/\sqrt{n}$ and $\mu_{n,j} = a^*\theta_j/a_n$. Similar to the test

CONTACT Cun-Hui Zhang czhang@stat.rutgers.edu Department of Statistics, Rutgers University, Piscataway, NJ.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

statistic W_j in (1.3), M_j is approximately marginally symmetric when $\theta_j = 0$ and likely to be positive and large when $\theta_j \neq 0$. This motivates the use of the threshold rule $\langle f\rangle_t$ on the mirror statistics at $t = T_\delta(M[p])$ in (1.4). In an asymptotic analysis, a condition on the level of dependence can be imposed instead of the independence of the signs $\{s_j, j \in S_0\}$ as required in the non-asymptotic FDR control with knockoffs.

Compared with the knockoff approach, the mirror statistics are more readily available with sufficiently large samples. Compared with asymptotic FDR control with the Benjamini-Hochberg rule (1.1) based on (1.6), the mirror statistics approach does not require a specification of the scale σ_n. Moreover, as mentioned in Dai et al. (2023), the asymptotic symmetry $\mathbb{P}\{M_j > t\} \approx \mathbb{P}\{M_j < -t\}$ for $j \in S_0$, sufficient for the mirror statistic to make sense, may kick in with a smaller sample size requirement than the asymptotic normality (1.6).

2. Relationship of Mirror Statistics to Knockoffs

As functions of data, mirror statistics are equivalent to knockoff test statistics through an algebraic transformation. Consider two-sided tests for simplicity. Given estimators $\widehat{\theta}^{(1)}$ and $\widehat{\theta}^{(2)}$, let

$$z_j = |\widehat{\theta}_j^{(1)} + \widehat{\theta}_j^{(2)}|/2, \quad z_j' = |\widehat{\theta}_j^{(1)} - \widehat{\theta}_j^{(2)}|/2, \; i \in [p]$$

We have $|\widehat{\theta}_j^{(1)}| \vee |\widehat{\theta}_j^{(2)}| = z_j + z_j'$, $|\widehat{\theta}_j^{(1)}| \wedge |\widehat{\theta}_j^{(2)}| = |z_j - z_j'|$, and $\mathrm{sgn}(\widehat{\theta}_j^{(1)})\mathrm{sgn}(\widehat{\theta}_j^{(2)}) = \mathrm{sgn}(z_j - z_j')$ with the convention $\mathrm{sgn}(0) = 0$. Thus, the mirror statistics can be written as

$$\begin{aligned} M_j &= \mathrm{sgn}(\widehat{\theta}_j^{(1)})\mathrm{sgn}(\widehat{\theta}_j^{(2)}) f_M(|\widehat{\theta}_j^{(1)}|, |\widehat{\theta}_j^{(2)}|) \\ &= \mathrm{sgn}(z_j - z_j') f_W(z_j, z_j') \\ &= W_j \end{aligned} \tag{2.1}$$

with $f_W(x, x') = f_M(x + x', |x - x'|)$. Conversely, $f_M(x, x') = f_W((x + x')/2, |x - x'|/2)$ given a choice of $f_W(\cdot, \cdot)$. For example, M_j with $f_M(x, x') = (|x| + |x'|)/2$ matches W_j with $f_W(x, x') = |x| \vee |x'|$ and M_j with $f_M(x, x') = |xx'|$ matches W_j with $f_W(x, x') = |x^2 - x'^2|$;

$$M_j = \begin{cases} \mathrm{sgn}(z_j - z_j')(z_j \vee z_j'), & f_M(x, x') = (|x| + |x'|)/2, \\ \mathrm{sgn}(z_j - z_j')|z_j^2 - z_j'^2|, & f_M(x, x') = |xx'|. \end{cases}$$

It can be seen that the natural choice is $f_M(x, x') = (|x| + |x'|)/2$ in the first example.

The algebraic transformation of W_j to M_j is not guaranteed to offer new statistical insight, and vice versa. For example, when W_j are defined through the Lasso path in linear regression with non-orthogonal designs, the interpretation of the corresponding M_j is unclear.

When $\widehat{\theta}^{(1)}$ and $\widehat{\theta}^{(2)}$ are two independent copies of a Gaussian vector $N(\theta_{n,[p]}, \Sigma_n)$,

$$z_{[p]} \sim |N(\theta_{n,[p]}, \Sigma_n/2)|, \quad z_{[p]}' \sim |N(0, \Sigma_n/2)|,$$

and $z_{[p]}$ and $z_{[p]}'$ are independent. Thus, the use of mirror statistics is equivalent to generating a copy of noise vector and treating it as a knockoff of $z_{[p]}$, with the correspondence between $f_M(\cdot)$ and $f_W(\cdot)$ in (2.1). For example, in linear regression with a design matrix X of rank p and Gaussian noise with a known noise level σ, Gaussian mirror (Xing, Zhao, and Liu 2021) can be constructed in one shot by generating $z_{[p]}'$ with $\Sigma_n/2 = \sigma^2(X^TX)^{-1}$. Mirror statistics $M_{[p]}$ are equivalent to test statistics $W_{[p]}$ based on knockoffs if and only if Σ_n is diagonal.

In an asymptotic analysis based on (1.6), the mirror statistic and knockoff methods would have asymptotically the same FDR and power due to their algebraic equivalence (2.1) when the dependence between the test statistics has only an infinitesimal impact on the operating characteristics of the tests. This condition on dependence is typically imposed in the literature through the sparsity of the correlation of the estimates in (1.6). In this asymptopia, an advantage of mirror statistics is their availability through DS.

Asymptotic FDR control in moderately high- and high-dimensional settings has been considered in Liu (2013), Xia, Cai, and Li (2018), Javanmard and Javadi (2019), and Ma, Tony Cai, and Li (2021) with the Benjamini-Hochberg (BH) rule, and in Xing, Zhao, and Liu (2021) and Dai et al. (2022) with mirror statistics. The theoretical results in Dai et al. (2023) on data-splitting-based mirror statistics in GLM further develop this direction.

3. Scale-Free FDR Control

For asymptotic FDR control, the unknown scaling factor σ_n^* in Sur and Candes (2019) and Proposition 3.1 of Dai et al. (2023) is no longer a sticking issue with DS. Under (1.6), scale-free asymptotic FDR control can be achieved by the BH rule with familiar Student's t-statistics. Still, mirror statistics may rely less heavily on the asymptotic normality.

When the asymptotic normality holds with a common asymptotic variance as in (1.6) and p is large, the BH rule (1.1) can be used with the two-sided test statistics

$$Z_j = \frac{|\widehat{\theta}_j^{(1)} + \widehat{\theta}_j^{(2)}|}{\left(\sum_{i=1}^{p}\left(\widehat{\theta}_i^{(1)} - \widehat{\theta}_i^{(2)}\right)^2/(p-1)\right)^{1/2}} \tag{3.1}$$

and the standard absolute Gaussian survival function $G_0(t) = 2\Phi(-t)$.

We compare this procedure (3.1), denoted as "θn-BH", with DS and Gaussian mirror (GM) in Dai et al. (2023), the BH rule on the MLE (BHq), BH rule with adjusted p-values (ABHq) in Sur and Candes (2019), and model-X knockoff (KN) in Candes et al. (2018). Specifically, Figure 1 reports the simulation comparison among the above six procedures for feature selection in logistic regression in the same setting as in Figure 3 of Dai et al. (2023). Each point in Figure 1 represents the average of 50 independent replications with $N(0, \Sigma)$ iid designs, where $\Sigma_{ij} = r^{|i-j|}$, $q = 0.1$, $n = 3000$, $p = 500$, $\|\beta^*\|_0 = 50$, and $|\beta_j| =$ signal for $j \in S_0^c$. It can be seen from these results that θn-BH quite consistently exhibits FDR $\approx qp_0/p = 0.09$, slightly more power than DS, and less power than ABHq, GM and KN. We note that ABHq, GM and KN rely more heavily on Gaussian design.

In the case of $p > n$, debiased Lasso exhibits heteroscedasticity (Candes et al. 2018; Dai et al. 2023). In the presence of heteroscedasticity, $\widehat{\theta}_j^{(k)}/\sigma_{n,j} \approx N(\mu_{n,j}, 1)$, scale-free asymptotic FDR control can be achieved with the absolute Cauchy

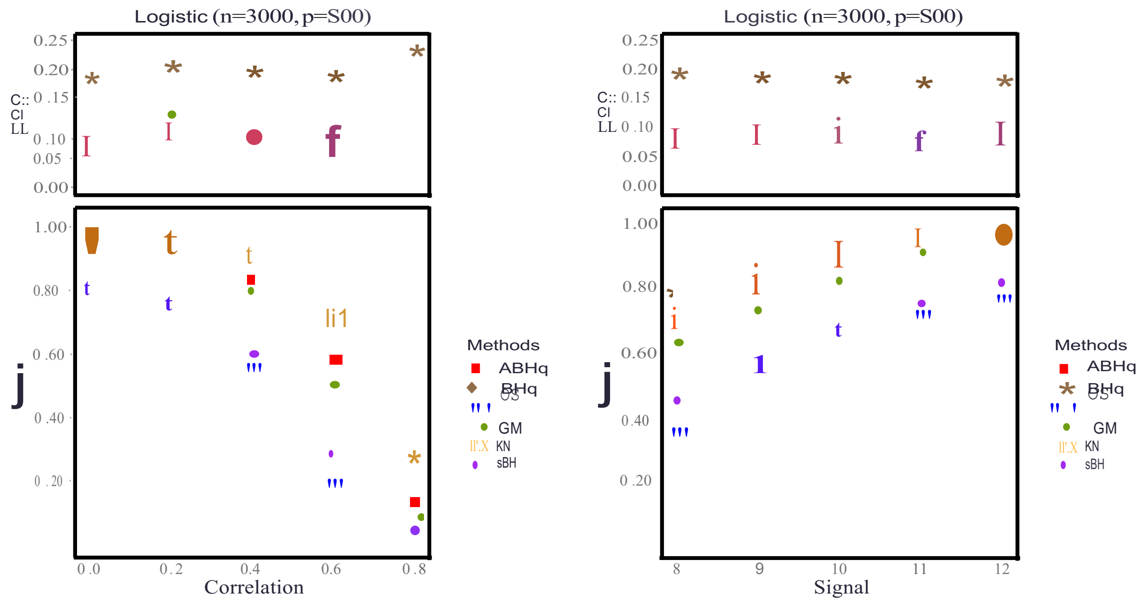

Figure 1. Left panel: the signal strength is fixed at $|fJt| = 11$ for $j \in S8$, and the correlations among the covariates vary. Right panel: the correlation is fixed at , = 0.2, and the signal strength varies. Here "sBH" stands for an-BH.

statistic Zj = |oJ| + 0j² || / |oJ| 0j² | and the corresponding $G_0(t) = 1 - (2/n)\arctan(t)$. However, our simulation experiment demonstrates that a cost of this robustness against heteroscedasticity is a significant loss of power.

4. Multiple Data Splitting

As knockoff and mirror statistics are randomized, it would be interesting to find methods to aggregate multiple randomized tests based on them for further improvement. See for example the discussion in section 7.2.3 of Candes et al. (2018). In Dai et al. (2022), the authors proposed a multiple data splitting (MDS) strategy to achieve this goal.

Similar to Dai et al. (2022), Proposition 3.3 seems to focus on special regimes aimed at the nearly complete FDR control in the form of FDR = o(1). This can be seen as follows.

Let I j = E*[JU E S}/(IS i v 1)] for a randomized selector S c J;p] where E* denotes avera&ng over many copies 0! random d Sgiven data. If the selector S controls the FDP = IS nS₀1/ (1S1 v 1), I:tSo I; :S q + op(1). Let $I(1) \leq \cdots \leq I(p)$ be the ordered entries of $I[p]$. Define $I(m)$ by

$$J(1) + \cdots + I(m) \leq q < I(1) + \cdots + I(m+1). \quad (4.1)$$

Dai et al. (2023) proposes to test H_j by thresholding I_j at level $I(m)$ and provides sufficient conditions for the asymptotic FDP control by such MDS schemes. The following lemma provides a sketch of a more explicit version of the proofs in Dai et al. (2022, 2023).

Lemma 1. Let $S_1 = [p] \setminus S_0$, $P_1 = P - P_0$, and I[p be any vector of nonnegative statistics satisfying I:fESo I; :S $q + E_1$. Let m be as in (4.1). In the event where $p - m$ κ1p1 and L ESoj,E SI{Ij :'S I;} :S Ei PoP1,

$$FDP = \frac{L\ iES_0\ I\{I_j > I_{(m)}\}}{p - m}$$

< max { 1 + E Po + E2./popl E2 Po/P1 , 1 + E1Po + Pi KJ } , (4.2)

where $E_j = E_i/(q + E_1)$.

In the nonsparse case, Proposition 3.3 of Dai et al. (2023) provides FDP = op(1) because the assumptions imply E_1 o(1), E_2 = o(1) and κ 1 is bounded away from zero in (4.2).

5. Fast Data Splitting Methods

In sparse GLM where Data = $(X,y) \in \mathbb{R}^{n \times (p+1)}$, a debiased estimator of $f3j^*$ can be written as ''''''''(7J, uj, Data) where $7J$ is an initial estimate of $/3^*$ and u_j is an estimate of the direction u_j for the least favorable submodel for the estimation of f3/- !!'/ and u are estimated by the Lasso, the computational cost of f3/'t;,. given Data is

$$\text{Cost}\left(-''(.Bww,'\ \text{Data}), j \in [p]\right) \times p\text{Cost(Lasso)}.$$

We ignore the computational cost of estimating scaling factors such as the < and On in the discussion below (1.6) as they could be viewed as byproducts of the debiased Lasso program. Let Data(1,l , Data ,(2,l , e E [11 , be data generated in MDS. If jk.l)= ,B......,(Data<k,l)) and rfk .l) = u':'''°(DataCk.l)) are used, the computational cost of MDS is $T \times p \times$ Cost(Lasso).

Under proper sparsity and regularity conditions, the debiased Lasso provides

$$''''''(,B'''''',ut'', \text{Data}) - .ie;a'(f3^*, u_j, \text{Data}) = op(n^{-1/2}),$$

so that the asymptotic normality is valid with a single uj'°(Data) for each j based on the entire data in DS and MDS. We propose such procedures as fast DS and MDS algorithms in Table 1.

OS.fast needs to run Lasso $p + 2$ times while DS $2(p+1)$ times. The computational cost of MOS.fast is of the order (T +

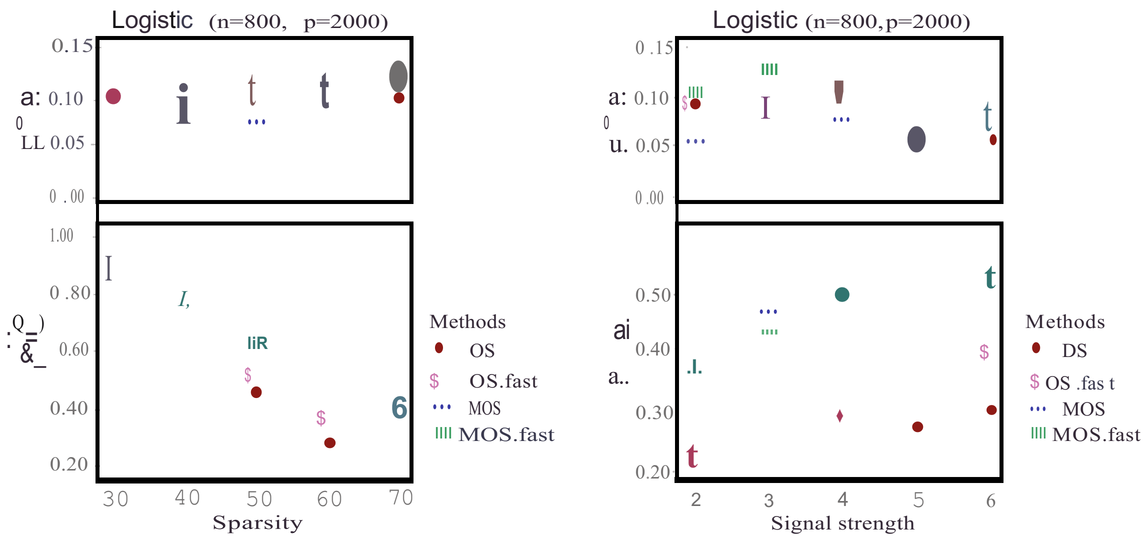

Figure 2. Left panel: the signal strength is 1/3/ l = 4 for j E 5 , and the sparsity P1 = 15 l varies. Right panel: the sparsity is fixed at 60 and the signal strength varies.

Table 1. Fast DS and MDS algorithms.

DS.fast(Data)

1. Compute uj = u\-"""(Data), j E [p].
2. Generate Data<f>, Data <2> by splitting data at random.
3. For j E [p], compute M' (pu " (Data <k>), uj, Data <k>) and respective scaling factors, $k = 1, 2$.
4. Compute 8 1), ¥ (2) an dMj = M j(0(1), 0 2>), e n d(For j).
5. Compute and output <Þ1 (M[p]),

MDS.fast(Data, U[p]) with input U[p] as in Step 1 of DS.fast

1. For e E en, generate Data <l, l), Data <2,l> by splitting data at random.
2. For j E [p], compute " ("fi'--"" (Data <k,l)) ,Uj, Data<k,l)) and respective scaling factors, $k = 1, 2$
3. Compute eJ 1,l) if.2 l) and My > = M j(eJ1,l), 8j2 ,l))(nd(For l
4. Compute <Þk (M 1), s(l) and 11) =lU E s(l >J; jW >l, j E [p], end(For e)
5. Compute /j = l:t,lt / T, j E [p], and /(m) as in (4.1).
6. Compute and output /{/ > /(m)}, j E [p].

p) x Cost(Lasso) while the cost of MDS is of the order $T \times p \times$ Cost(Lasso).

Figure 2 reports simulation performance of DS, DS.fast, MDS and MDS.fast in the settings of Figure 5 in Dai et al. (2023). Each point is averaged based on 50 independent replications. The results demonstrate comparable performance between DS and DS.fast and that between MDS and MDS.fast.

Supplementary Materials

In the supplement, we provide a proof of Lemma 1 and the R code for the simulation results reported in Figures 1 and 2.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Li's research is supported in part by the National Natural Science Foundation of China (grant no. 12201630.) Yao's research is supported in part by U.S. National Institutes of Health grants ROlHGO10171 and R01MH116527 and National Science Foundation grant DMS-2112711. Zhang's research is supported in part by National Science Foundation grants CCF-1934924, DMS-2052949 and DMS-2210850

References

Barber, R. F., and Candes, E. J. (2015), "Controlling the False Discovery Rate via Knockoffs," *The Annals of statistics,* 43, 2055 - 2085. (1586]

Benjamini, Y., and Hochberg, Y. (1995) , "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society,* Series B, 57, 289- 300. [1586]

Benjamini, Y., and Yekutieli, D. (2001), "The Control of the False Discovery Rate in Multiple Testing Under Dependency;' *Annals of statistics,* 29 , 1165 - 1188. (1586)

Candes, E., Fan, Y., Janson, L. and Lv, J. (2018), "Panning for Gold: Model-x Knockoffs for High-Dimensional Controlled Variable Selection," *Journal of the Royal Statistical Society,* Series B, 80, 551- 577. [1586, 1587, 1588)

Dai, C., Lin, B., Xing, X., and Liu, J. S. (2022), "False Discovery Rate Control via Data Splitting," *Journal of the American Statistical Association,* 1- 18, DOI : 10.1080/01621459 .2022.2060113 . [1586, 1587, 1588)

- - - (2023), "A Scale-Free Approach for False Discovery Rate Control in Generalized Linear Models;' *Journal of the American Statistical Association,* 118, 1551- 1565. [1586, 1587, 1588, 1589)

Huang, D., and Janson, L. (2020), "Relaxing the Assumptions of Knockoffs by Conditioning;' *The Annals of Statistics,* 48, 3021- 3042. (1586)

Javanmard, A., and Javadi, H. (2019), "False Discovery Rate Control via Debiased Lasso," *Electronic Journal of Statistics,* 13, 1212 - 1253 . (1587)

Liu, W. (2013), "Gaussian Graphical Model Estimation with False Discovery Rate Control," *The Annals of Statistics,* **41,** 2948- 2978. [1587)

Ma, R., Tony Cai, T., and Li, H. (2021), "Global and Simultaneous Hypothesis Testing for High-Dimensional Logistic Regression Models," *Journal of the American Statistical Association,* 116, 984- 998. [1587)

Simes, R. J. (1986), "An Improved Bonferroni Procedure for Multiple Tests of Significance; *Biometrika,* 73, 751- 754. [1586)

Sur, P., and Candes, E. J. (2019), "A Modern Maximum-Likelihood Theory for High-Dimensional Logistic Regression; *Proceedings of the National Academy of Sciences,* 116, 14516- 14525. [1587)

Xia, Y., Cai, T. T., and Li, H. (2018), "Joint Testing and False Discovery Rate Control in High-Dimensional Multivariate Regression: *Biometrika,* 105, 249- 269. I1s 87)

Xing, X., Zhao, Z., and Liu, J.S. (2021), "Controlling False Discovery Rate Using Gaussian Mirrors: *Journal of the American Statistical Association,* 118, 222- 241. (1586 ,1587)