On p-value combination of independent and non-sparse signals: asymptotic efficiency and Fisher ensemble

Yusi Fang, Chung Chang, and George C. Tseng*

University of Pittsburgh and National Sun Yat-sen University

Supplementary Material

*Corresponding author: George Tseng; University of Pittsburgh; Email: ctseng@pitt.edu

S1 Supplementary theoretical results

S1.1 Asymptotic efficiencies of p-value combination methods

In this subsection, we outline the asymptotic efficiencies of multiple p-value combination methods mentioned in Section 2: Fisher, Stouffer, Pareto, Cauchy (CA), Berk-Jones (BJ) and higher criticism (HC). All technical proofs are left to Section S2. For Fisher test, combined with Lemma 1 and by almost the same argument in Littell and Folks (1973), one can show that Fisher test attains ABO in the modified partial signal setting. Similarly, by similar argument as above, the exact slope of Stouffer is

$$C_{\text{Stouffer}}(\vec{\theta}) = \frac{1}{K}\Big[\sum_{i=1}^{\ell}(\lambda_i c_i(\theta_i))^{\frac{1}{2}}\Big]^2.$$

Although generally $C_{\text{Stouffer}} \leqslant \sum_{i=1}^{\ell} \lambda_i c_i(\theta_i)$ and Stouffer is not ABO, Stouffer becomes ABO when all the p-values contain true signals with equal effects $\lambda_1 c_1(\theta_1) = \ldots = \lambda_K c_K(\theta_K) > 0$. Theorem S1 below describes asymptotic efficiency property of Fisher and Stouffer.

Theorem S1 (extended from Littell and Folks (1973); Fisher is ABO and Stouffer is generally not ABO)**.** *Under the setup in Section 2.1, Fisher is ABO with exact slope* $C_{Fisher}(\vec{\theta}) = \sum_{i=1}^{\ell} \lambda_i c_i(\theta_i)$. *Stouffer is generally not ABO with exact slope* $C_{Stouffer}(\vec{\theta}) = \frac{1}{K}\Big[\sum_{i=1}^{\ell}(\lambda_i c_i(\theta_i))^{\frac{1}{2}}\Big]^2$. *Stouffer is ABO when all signals combined have equal sample-size adjusted exact slope:* $\lambda_1 c_1(\theta_1) = \ldots = \lambda_K c_K(\theta_K) > 0$.

We next study two methods by heavy-tailed distribution transformation, Pareto and CA, as follows

$$T_{\text{Pareto}}(\eta) = \sum_{i=1}^{K} \frac{1}{p_i^{\eta}} \text{ with some } \eta > 0, \quad T_{\text{CA}} = \frac{1}{K}\sum_{i=1}^{K} \tan(\pi(\frac{1}{2} - p_i)).$$

Methods of this category are in the form of statistics $T(\vec{p}) = \sum_{i=1}^{n} g(p_i) = \sum_{i=1}^{n} F_U^{-1}(1-p_i)$ to sum up transformed p-values, where the transformation $g(p)$ is the inverse CDF of U. Indeed, for Cauchy, $U \overset{D}{\sim} \text{CAU}(0,1)$ (standard Cauchy), and $U \overset{D}{\sim} \text{Pareto}(\frac{1}{\eta}, 1)$ for Pareto. Intuitively, methods by light-tailed distribution transformations (e.g., Stouffer and Fisher) achieve better asymptotic efficiency as a thin-tailed distribution generates more comparable contributions from marginally significant p-values with frequent signals, while methods by heavy-tailed distribution focus more on the extreme effects and downweigh the frequent small effects. For example, Stouffer test transforms p-values 10^{-2} and 10^{-6} to 2.32 and 4.75 while $\tan(\pi(1/2-p))$ for Cauchy test transforms the same p-values to 31.82 and 3.82×10^5, which makes the contribution from very small p-value (10^{-6}) dominate that from the moderate one (10^{-2}). The following two theorems show that CA and Pareto are generally not ABO and they are ABO if and only if there is only one true signal among p-values.

Theorem S2. *Under the setup in Section 2.1,* $T_{Pareto}(\eta)$ *is generally not ABO with exact slope* $C_{Pareto}^{(\eta)}(\vec{\theta}) = \max_{1 \leqslant i \leqslant K} \lambda_i c_i(\theta_i)$.

Theorem S3. *Under the setup in Section 2.1,* T_{CA} *is generally not ABO with exact slope* $C_{CA}(\vec{\theta}) = \max_{1 \leqslant i \leqslant K} \lambda_i c_i(\theta_i)$.

Both exact slopes of CA and Pareto are $\max_{1 \leqslant i \leqslant K} \lambda_i c_i(\theta_i)$, which is also the exact slope of minP, shown in Littell and Folks (1973). This suggests that CA and Pareto are more powerful for detecting sparse signals as minP.

We continue investigating the asymptotic efficiencies of BJ and HC, which can be viewed as goodness-of-fit tests:

$$T_{\mathrm{HC}} = \max_{1 \leqslant i \leqslant K} \sqrt{K} \frac{i/n - p_{(i)}}{\sqrt{p_{(i)}\left(1 - p_{(i)}\right)}}$$

$$T_{\mathrm{BJ}} = \max_{1 \leqslant i \leqslant K} \mathrm{I}_{\{p_{(i)} < \frac{i}{K}\}} \left[\frac{i}{K} \log\left(\frac{i/K}{p_{(i)}}\right) + \left(1 - \frac{i}{K}\right) \log\left(\frac{1 - i/K}{1 - p_{(i)}}\right) \right].$$

As goodness-of-fit tests, both test statistics can test whether the underlying distribution is Unif$(0,1)$ given K independent observed p-values $p_1, \ldots, p_K$. Both BJ and HC are mainly applied to the scenarios of detecting weak and sparse signals (Donoho and Jin, 2004; Berk and Jones, 1979; Li and Siegmund, 2015). The following theorem shows that BJ is generally not ABO.

Theorem S4. *Under setup in Section 2.1, T_{BJ} is not ABO with exact slope*

$$C_{BJ}(\vec{\theta}) = \max_{1 \leqslant i \leqslant K} i \lambda_i c_i(\theta_i).$$

The following proposition shows that HC generally is not ABO even for combining two p-values with equal effects.

Proposition S1. *Suppose p_1 and p_2 are two independent p-values such that*

$$-\frac{2}{n} \log(p_i) \to c_i(\theta_i) \quad \text{as } n \to \infty \text{ for } i = 1, 2,$$

with probability one. Then for $c_1(\theta_1) = c_2(\theta_2) = c_0 > 0$, T_{HC} is not ABO with exact slope $C_{HC}(\vec{\theta}) = c_0$.

S1.2 Type I error control of FE and FE$_{CS}$

In this subsection, we provide more details on the type I error control of FE and FE$_{\text{CS}}$ computation using the Pareto$(1,1)$ distribution. Assume X follows Pareto$(1,1)$. As suggested by Theorems 1 and 2 in Fang et al. (2023), under the null, the upper tail of distribution of the average of $1/p_1,\ldots,1/p_L$ with unknown dependence structure can be approximated by that of Pareto$(1,1)$, in a sense that for sufficiently large $t>0$ (corresponding to sufficiently small significant level α),

$$\frac{\mathbb{P}\big(\frac{1}{L}\sum_{i=1}^{L} 1/p_i > t\big)}{\mathbb{P}\big(X>t\big)} \approx 1. \tag{S1.1}$$

Hence for FE and FE$_{\text{CS}}$ respectively, one can show that for sufficiently large $t>0$ (corresponding to sufficiently small α),

$$\frac{1-F_{T_{\text{FE}}(t)}}{1-F_{\text{Pareto}(1,1)}(t)} \approx 1$$
$$\frac{1-F_{T_{\text{FE}_{\text{CS}}}(t)}}{1-F_{\text{Pareto}(1,1)}(t)} \approx 1,$$

which justifies the type I error control procedures for FE and FE$_{\text{CS}}$ using Pareto$(1,1)$, respectively. Table S1 in Section S3.1 numerically justifies accuracy of the above fast-computing procedure, where we show that type I error control procedures for FE and FE$_{\text{CS}}$ are accurate for $\alpha = 0.0001 \sim 0.05$ across a broad range of K (5 to 100). Note $1-F_{\text{Pareto}(1,1)}(t) = 1/t$, combined with equation (S1.1), one can show that the above procedures are equivalent to directly using

$$\frac{L}{\sum_{i=1}^{L} 1/p_i}$$

as p-value for statistic $\frac{1}{L}\sum_{i=1}^{L} 1/p_i$, which is suggested by Wilson (2019).

S2 Technical arguments

In this section, we present the technical arguments for proving the theoretical results. For any random variable X with CDF F, the corresponding survival function is denoted by $\bar{F} = 1 - F(t)$. For two positive functions $u(\cdot)$ and $v(\cdot)$, we denote by $u(t) \sim v(t)$ if $\lim_{t\to\infty} \frac{u(t)}{v(t)} = 1$. Also, $u(t) \gtrsim v(t)$ if $\lim_{t\to\infty} \frac{u(t)}{v(t)} > 1$ and $u(t) \lesssim v(t)$ if $\lim_{t\to\infty} \frac{u(t)}{v(t)} < 1$.

S2.1 Proof of results of modified Fisher's methods: Lemma 1 and Theorems 1-6

In this subsection, we prove Theorems 1-6. Before proceeding to the proofs, we first prove Lemma 1 and introduce Lemma S1-S3.

Proof of Lemma 1. For $\theta \in \Theta_0$, note that $-\log p^{(n)}$ follows exponential distribution with rate parameter 1 (denoted by EXP(1)) since the p-value p_n is distributed uniformly in $(0, 1)$. Consider the sequence of random variables $Y_1, Y_2, \ldots$, where $Y_1, Y_2 \ldots$, identically follow EXP(1). Define event $A_n = \{\frac{Y_n}{n} > \frac{\log(n(n+1))}{n}\}$. Then since $\sum_{i=1}^{+\infty} \mathbb{P}(A_n) < \infty$, by the Borel–Cantelli lemma, we have $\mathbb{P}\left(\limsup_{n\to+\infty} A_n\right)$=0. Hence $\frac{Y_n}{n}$ converges to zero with probability one. □

Lemma S1 (Bahadur et al. (1960)). *Let* $F_{\chi_k}(x) = \mathbb{P}(\chi_k \leqslant x)$, *where* $\chi_k = \sqrt{\chi_k^2}$ *and* χ_k^2 *follows chi-squared distribution with degrees of freedom* k. *Then* $\log(\bar{F}_{\chi_k}(x)) \to -\frac{1}{2}x^2(1 + o(1))$ *as* $x \to \infty$.

Lemma S2 (Savage (1969)). *Suppose* $\left\{T^{(n)}\right\}$ *is a sequence of test statistics which satisfies the following two properties:*

1. *There exists a function* $b(\theta), 0 < b(\theta) < \infty$, *such that* $T^{(n)}/\sqrt{n} \to b(\theta)$ *with probability one.*

2. *There exists a function* $f(t), 0 < f(t) < \infty$, *which is continuous in some open set containing the range of* $b(\theta)$ *such that for each* t *in the open set:*

$$-\frac{1}{n}\log\left[1 - F_n(\sqrt{n}t)\right] \to f(t),$$

where F_n *is the continuous CDF function of some random variable* X_n.

Then

$$-\frac{2}{n}\log\left[1 - F_n(T^{(n)})\right] \to 2f(b(\theta))$$

with probability one.

Remark S1. The condition $0 < f(t) < \infty$ implicitly puts restrictions on the choice of X_n (corresponding to F_n). For example, the rate of the upper tail of X_n should not be too fast. Indeed, $X_n \overset{D}{\sim} \text{Unif}(0,1)$ leads to a too fast rate ($F_{\text{Unif}(0,1)}(\sqrt{n}t) = 0$ for any $\sqrt{n}t > 1$), resulting in $f(t) = +\infty$ that clearly does not satisfy the conditions of Lemma S2.

Remark S2. When F_n is the CDF of $T^{(n)}$, Lemma S2 becomes Theorem 1 in Littell and Folks (1973), which will be used in the proof of Theorem S2; We will use Lemma S2 in the proofs of Theorems 1 and 3 to 6, where $F_n = F$ for some F and all n.

Lemma S3. *Under the setup in Section 2.1, define the following two index sets:*

$$\mathcal{D}^* = \{i : c_i(\theta_i) > 0\}\,;\ \hat{\mathcal{D}} = \left\{i : p_i \leqslant p_{(\ell)}\right\}.$$

Then we have, as $n \to \infty$, $\hat{\mathcal{D}} \to \mathcal{D}^*$ *with probability one. And if* $\lambda_i c_i(\theta) > \lambda_{i'} c_{i'}(\theta) > 0$, $\frac{p_{i'}}{p_i} \to +\infty$ *with probability one.*

Proof. To prove the first claim, first denote $\mathcal{D}^{*c} = \{i : i \notin \mathcal{D}^*\}$. For any $i' \in \mathcal{D}^{*c}$, and any $i \in \mathcal{D}^*$, by Lemma 1, we have $\log(p_{i'}/p_i)/n \to \lambda_i c_i(\theta_i) - 0 > 0$ with probability one. Hence p_i is smaller order of $p_{i'}$ as $n \to \infty$, which completes the proof. For the second claim, simply note for any $\lambda_i c_i(\theta) > \lambda_{i'} c_{i'}(\theta) > 0$, $\log(p_{i'}/p_i)/n \to \lambda_i c_i(\theta) - \lambda_{i'} c_{i'}(\theta) > 0$ with probability one. Then the result follows. □

Corollary S1. *Under the alternative in Section 2.1, with probability one, we have*

$$-\frac{2}{n}\sum_{i=1}^{j} \log p_{(i)} \to \begin{cases} \sum_{i=1}^{j} \lambda_i c_i(\theta_i) & 1 \leqslant j \leqslant \ell \\ \\ \sum_{i=1}^{\ell} \lambda_i c_i(\theta_i) & \ell < j \leqslant K. \end{cases}$$

Proof. Combine the results of Lemmas 1 and S3, the results follow. □

The proof of Theorem 1 below will use the second equivalent form of AFs to derive the exact slope:

$$T'_{\mathrm{AFs}} = \min_{\vec{w}} \bar{F}_{\chi^2_{2(\sum_{i=1}^{K} w_i)}}(-2\sum_{i=1}^{K} w_i \log p_i),$$

where $\vec{w} = (w_1, \ldots, w_K) \in \{0,1\}^K$ is the vector of binary weights that identify the candidate subset of p-values with true signals. In addition, denote

$$\hat{\vec{w}} = \operatorname*{argmin}_{\vec{w}} \bar{F}_{\chi^2_{2(\sum_{i=1}^{K} w_i)}}(-2\sum_{i=1}^{K} w_i \log p_i),$$

and $\vec{w}^* = \{\vec{w} : w_k = 1 \text{ if } \theta_i \in \Theta_0 \text{ or } 0 \text{ if } \theta_i \in \Theta_0\}$ as the weight vector identifying the true signals. Also denote $\hat{\vec{w}} = (\hat{w}_1, \ldots, \hat{w}_K)$. For the original form

$$T_{\mathrm{AFs}} = \max_{1 \leqslant j \leqslant K} -\log(\bar{F}_{\chi^2_{2j}}(-2\sum_{i=1}^{j} \log p_{(i)})),$$

we denote correspondingly $\hat{j} = \text{argmax}_j - \log \bar{F}_{\chi^2_{2j}}\big(- 2\sum_{i=1}^{j} \log p_{(i)}\big)$. Since $p_1, \ldots, p_K$ are independent with each other, we have

$$-2\sum_{i=1}^{\hat{j}} \log p_{(i)} = -2\sum_{i=1}^{K} \hat{w}_i \log p_i.$$

Proof of Theorem 1. Denote

$$T(\vec{w};\vec{p}) = -2\sum_{i=1}^{K} w_i \log p_i$$

$$L(T(\vec{w};\vec{p})) = \bar{F}_{\chi^2_{2d(\vec{w})}}(T(\vec{w};\vec{p})),$$

where $d(\vec{w}) = \sum_{i=1}^{K} w_i$ and $\vec{p} = (p_1, \ldots, p_K)$. Essentially, $T'_{\text{AFs}} = \min_{\vec{w}} L(T(\vec{w};\vec{p}))$. Further denote by $L_{\text{obs}} = \min_{\vec{w}} L(T(\vec{w};\vec{p}_{\text{obs}}))$ the observed value of T'_{AFs}.

Let $\mathbb{P}_0$ be the probability measure of $\vec{p} = (p_1, \ldots, p_K)$ under the null and U_{AFs} be the random variable that follows the same distribution of T'_{AFs} under the null. For any fixed $\vec{w}$, denote by $U(\vec{w},\vec{p})$ the random variable follows the same distribution of $\bar{F}_{\chi^2_{2d(\vec{w})}}(T(\vec{w};\vec{p}))$ under the null. Further denote $\Omega_j = \{\vec{w} : d(\vec{w}) = j\}$ for $j = 1, \ldots, K$. Then we have:

$$\begin{aligned} p_{\text{AFs}} = F_{U_{\text{AFs}}}(L_{\text{obs}}) &= 1 - \bar{F}_{U_{\text{AFs}}}(L_{\text{obs}}) \\ &= 1 - \mathbb{P}_0\Big(\bigcap_{j=1}^{K} \bigcap_{\vec{w}\in\Omega_j} U(\vec{w},\vec{p}) \geqslant L_{\text{obs}}\Big). \end{aligned} \tag{S2.2}$$

By Bonferroni's inequality,

$$\begin{aligned} (S2.2) &\leqslant 1 - \Big[1 - \sum_{j=1}^{K} \mathbb{P}_0\Big(\bigcup_{\vec{w}\in\Omega_j} U(\vec{w},\vec{p}) \leqslant L_{\text{obs}}\Big)\Big] \\ &\leqslant \sum_{j=1}^{K} \sum_{\vec{w}\in\Omega_j} \bar{F}_{\chi^2_{2j}}\big(\bar{F}^{-1}_{\chi^2_{2j}}(L_{\text{obs}})\big) = \big(2^K - 1\big)\, L_{\text{obs}}, \end{aligned} \tag{S2.3}$$

where $\bar{F}^{-1}_{\chi^2_{2j}}(\alpha)$ denotes the $1-\alpha$ quantile of χ^2_{2j}. Further note

$$\begin{aligned}-\frac{2}{n}\log\big(L_{\text{obs}}\big) &= -\frac{2}{n}\log\big(1 - F_{\chi^2_{2d(\hat{\vec{w}})}}\big(-2\sum_{i=1}^{K}\hat{w}_i\log p_i\big)\big)\\ &= -\frac{2}{n}\log\bar{F}_{\chi_{2d(\hat{\vec{w}})}}\big((-2\sum_{i=1}^{K}\hat{w}_i\log p_i)^{\frac{1}{2}}\big).\end{aligned}$$

Since $1 \leqslant d(\hat{\vec{w}}) \leqslant K$,

$$\begin{aligned}-\frac{2}{n}\log\bar{F}_{\chi_{2K}}\big((-2\sum_{i=1}^{K}\hat{w}_i\log p_i)^{\frac{1}{2}}\big) &\leqslant -\frac{2}{n}\log\big(L_{\text{obs}}\big)\\ &\leqslant -\frac{2}{n}\log\bar{F}_{\chi_2}\big((-2\sum_{i=1}^{K}\hat{w}_i\log p_i)^{\frac{1}{2}}\big).\end{aligned} \tag{S2.4}$$

Denote $\hat{j} = \text{argmax}_j - \log\bar{F}_{\chi^2_{2j}}\big(-2\sum_{i=1}^{j}\log p_{(i)}\big)$, as $p_1,\ldots,p_K$ are independent with each other, we have

$$-2\sum_{i=1}^{K}\hat{w}_i\log p_i = -2\sum_{i=1}^{\hat{j}}\log p_{(i)}.$$

In the proof of Theorem 4 latter, we will show that $\hat{j} \geqslant \ell$ with probability one (equation (S2.11) in the proof of Theorem 4). Hence by Corollary S1, under the alternative, we have $\frac{\sqrt{-2\sum_{i=1}^{K}\hat{w}_i\log p_i}}{\sqrt{n}} = \frac{\sqrt{-2\sum_{i=1}^{\hat{j}}\log p_{(i)}}}{\sqrt{n}} \to (\sum_{i=1}^{\ell}\lambda_i c_i(\theta_i))^{\frac{1}{2}}$ with probability one. Further combined with and Lemmas S1 and S2 and equation (S2.4), we have

$$-\frac{2}{n}\log\big(L_{\text{obs}}\big) \to \sum_{i=1}^{\ell}\lambda_i c_i(\theta_i).$$

Combined with (S2.3), we have under the alternative

$$-\frac{2}{n}\log p_{\text{AFs}} \geqslant \sum_{i=1}^{\ell}\lambda_i c_i(\theta_i)$$

with probability one. Then the result follows.

Proof of Theorem 2. Note that by Theorem S1, we have that Fisher is ABO with exact slope $C_{\text{Fisher}}(\vec{\theta}) = \sum_{i=1}^{\ell} \lambda_i c_i(\theta_i)$, then combine with Theorem 2.6 in Berk and Jones (1978), the result follows. □

□

Proof of Theorem 3. Case when $\ell \geqslant 2$:

Assume $j^* = \text{argmax}_j \frac{\sum_{i=1}^{j} \lambda_i c_i(\theta_i)}{B_j}$. We first prove that under alternative,

$$\underset{j}{\text{argmax}}\, T_A \to j^* \tag{S2.5}$$

with probability one as $n \to \infty$. Indeed, for $\forall j' \neq j^*$, suppose the following event holds:

$$\begin{aligned} &\frac{-2\sum_{i=1}^{j'} \log p_{(i)} - A_{j'}}{B_{j'}} > \frac{-2\sum_{i=1}^{j^*} \log p_{(i)} - A_{j^*}}{B_{j^*}} \\ &\Leftrightarrow -2B_{j^*} \sum_{i=1}^{j'} \log p_{(i)} + A_{j^*} B_{j'} > -2B_{j'} \sum_{i=1}^{j^*} \log p_{(i)} + A_{j'} B_{j^*} \\ &\Leftrightarrow \frac{-2B_{j^*} \sum_{i=1}^{j'} \log p_{(i)}/n + A_{j^*} B_{j'}/n}{-2B_{j'} \sum_{i=1}^{j^*} \log p_{(i)}/n + A_{j'} B_{j^*}/n} > 1. \end{aligned} \tag{S2.6}$$

Here without loss of generality we assume both $A_{j^*}B_{j'}$ and $A_{j'}B_{j^*}$ in the second inequality are positive. Otherwise one can always move the smaller term to the other side of the inequality and still use almost the same arguments as follows. However, under the setup in Section 2.1, note that by Lemmas 1 and S3 and $j^* = \text{argmax}_j \frac{\sum_{i=1}^{j} \lambda_i c_i(\theta_i)}{B_j}$, under the alternative we have

$$\frac{-2B_{j^*} \sum_{i=1}^{j'} \log p_{(i)}/n + A_{j^*} B_{j'}/n}{-2B_{j'} \sum_{i=1}^{j^*} \log p_{(i)}/n + A_{j'} B_{j^*}/n} \to \frac{B_{j^*}}{B_{j'}} \cdot \frac{\sum_{i=1}^{j'} \lambda_i c_i(\theta_i)}{\sum_{i=1}^{j^*} \lambda_i c_i(\theta_i)} < 1$$

as $n \to \infty$ with probability one, which contradicts to equation (S2.6). Hence (S2.5) holds. Let U_{A} be the random variable that follows the same distribution of T_A under the null.

Denote by F_{U_A} the CDF of the U_{A} and $\bar{F}_{U_A} = 1 - F_{U_A}$ as the corresponding survival function, respectively. Similarly, for the following test statistic

$$T_{A_j} = \frac{-2\sum_{i=1}^{j} \log p_{(i)} - A_j}{B_j},$$

let U_{A_j} be the random variable that follows the same distribution of T_{A_j} under the null. And define $F_{U_{A_j}}$ and $\bar{F}_{U_{A_j}}$ as the CDF and survival function of U_{A_j}, respectively. Furthermore, define the test statistic

$$T_j = -2\sum_{i=1}^{j} \log p_{(i)}$$

and U_j as the random variable that follows the same distribution of T_j under the null and let F_{U_j} and $\bar{F}_{U_j}$ be the CDF and survival function of U_j, respectively. Pick $j = 1$, then we have:

$$\bar{F}_{U_A}(T_A) \geqslant \bar{F}_{U_{A_1}}(T_A) = \bar{F}_{U_1}(B_1 T_A + A_1),$$

Denote $T^{(n)} = \sqrt{B_1 T_A + A_1}$, with (S2.5) holds, by Lemmas 1 and S3, under the alternative, we have,

$$\frac{T^{(n)}}{\sqrt{n}} = n^{-\frac{1}{2}}(B_1 T_A + A_1)^{\frac{1}{2}} \to \left[(B_1/B_{j^*})\sum_{i=1}^{j^*} \lambda_i c_i(\theta_i)\right]^{\frac{1}{2}}$$

with probability one. Note for $t > 0$ we have

$$\bar{F}_{\chi_2}(\sqrt{n}t) = \bar{F}_{\chi_2^2}(nt^2) \leqslant \bar{F}_{U_1}(nt^2) \leqslant \bar{F}_{\chi_{2K}^2}(nt^2) = \bar{F}_{\chi_{2K}}(\sqrt{n}t).$$

Hence by Lemma S1,

$$-\frac{1}{n}\log \bar{F}_{U_1}(nt^2) = -\frac{1}{n}\log \bar{F}_{\sqrt{U_1}}(\sqrt{n}t) \to \frac{t^2}{2}.$$

Hence by Lemma S2, under the alternative, we have

$$-\frac{2}{n}\log\big(\bar{F}_{U_A}(T_A)\big) \leqslant -\frac{2}{n}\log \bar{F}_{U_1}\big((T^{(n)})^2\big) = -\frac{2}{n}\log \bar{F}_{\sqrt{U_1}}(T^{(n)}) \\ \to \frac{B_1}{B_{j^*}}\sum_{i=1}^{j^*}\lambda_i c_i(\theta_i) < \sum_{i=1}^{\ell}\lambda_i c_i(\theta_i) \tag{S2.7}$$

with probability one. Here the last inequality is due to $\ell \geqslant 2$ and B_j is a strictly increasing function. Hence T_A is still not ABO.

Case when $\ell = 1$:

First we prove that $j^* \to 1$ with probability one under the alternative. Note here we assume $\ell = 1$ and B_j increases as j increases. Note that $c_i(\theta_i) = 0$ with probability one for all $i > 1$, hence

$$\max_j \frac{\sum_{i=1}^{j}\lambda_i c_i(\theta_i)}{B_j} \to \frac{\lambda_1 c_1(\theta)}{B_1} \tag{S2.8}$$

with probability one. Hence $j^* \to 1$ with probability one. Then we have:

$$\bar{F}_{U_A}(T_A) \leqslant \sum_{j=1}^{K}\bar{F}_{U_{A_j}}(T_A) = \sum_{j=1}^{K}\bar{F}_{T_j}\left(B_j T_A + A_j\right) \leqslant K\cdot \bar{F}_{\chi^2_{2K}}\big(B_1 T_A + \min_j A_j\big).$$

By combining (S2.8) and Lemmas 1 and S3, under the alternative, we have

$$\frac{\sqrt{B_1 T_A + \min_j A_j}}{\sqrt{n}} \to \sqrt{\lambda_1 c_1(\theta)}$$

with probability one. And by Lemma S1

$$-\frac{1}{n}\log\big(1 - F_{\chi_{2K}}(\sqrt{n}t)\big) \to \frac{1}{2}t^2.$$

In addition,

$$-\frac{2}{n}\log \bar{F}_{U_A}(T_A) \geqslant -\frac{2}{n}\Big[\log \bar{F}_{\chi^2_{2K}}\big(B_1 T_A + \min_j A_j\big) + \log K\Big]. \tag{S2.9}$$

Hence by Lemma S2, under alternative, $(S2.9) \to \lambda_1 c_1(\theta)$ with probability one. Then we conclude that when $\ell = 1$, T_A is ABO. □

Remark S3. It can be shown that T_A generally does not has signal selection consistency. Recall that T_A picks $j^* = \text{argmax}_j \frac{\sum_{i=1}^{j} \lambda_i c_i(\theta_i)}{B_j}$ with probability one as shown in the proof. To give a counter example, we consider $B_j = \sqrt{\sum_{i=1}^{K} w(i,j)}$ (corresponding to T_{AFz}), where $K = 2$. We assume there is only two signals, with $\lambda_1 c_1(\theta_1) = 9$ and $\lambda_2 c_2(\theta_2) = 1$. Then one can show $j^* = 1$ here, i.e., T_A picks the wrong subset of p-values with probability one. Since B_j is a strictly increasing function, we can easily show that $j^* \leq \ell$ always holds and $j^* < \ell$ in general.

The proof of Theorem 4 will use the first equivalent form of AFs,

$$T_{\text{AFs}} = \max_{1 \leqslant j \leqslant K} -\log \bar{F}_{\chi^2_{2j}}(-2\sum_{i=1}^{j} \log p_{(i)}).$$

Proof of Theorem 4. The goal is to prove $\hat{\vec{w}} \to \vec{w}^*$ in probability as $n \to \infty$ under the alternative. Recall by Corollary S1, we have, under the alternative,

$$-\frac{2}{n}\sum_{i=1}^{j} \log p_{(i)} \to \begin{cases} \sum_{i=1}^{j} \lambda_i c_i(\theta_i) & 1 \leqslant j \leqslant \ell \\ \sum_{i=1}^{\ell} \lambda_i c_i(\theta_i) & \ell < j \leqslant K \end{cases} \tag{S2.10}$$

with probability one as $n \to +\infty$. Define index sets

$$\mathcal{S}_1 = \{i: \ w_i^* = 1 \text{ and } \hat{w}_i = 0\}\,; \ \mathcal{S}_2 = \{i: \ w_i^* = 0 \text{ and } \hat{w}_i = 1\}\,.$$

Recall that we assume the first $\ell \leqslant K$ studies are with exact slopes $c_i(\theta) > 0$. The following arguments are based on the first equivalent form of AFs, denoted by T_{AFs}.

We first prove $\mathcal{S}_1 \to \varnothing$ in probability. Indeed, we claim a stronger result that $\mathcal{S}_1 \to \varnothing$ with probability one under the alternative. By Lemmas 1 and S3, as $n \to +\infty$, the first smallest ℓ p-values converge to the first ℓ p-values with exact slopes strictly greater than 0. Hence it suffices to prove that for

$$\hat{j} = \underset{j}{\operatorname{argmax}} - \log \bar{F}_{\chi^2_{2j}}\big(- 2 \sum_{i=1}^{j} \log p_{(i)}\big),$$

as $n \to +\infty$, we have $\hat{j} \geqslant \ell$ with probability one. Indeed, for any $j' < \ell$, by Lemmas S2 and S3 and equation (S2.10),

$$\begin{aligned}
\frac{- \log \bar{F}_{\chi^2_{2j'}}\big(- 2 \sum_{i=1}^{j'} \log p_{(i)}\big)}{- \log \bar{F}_{\chi^2_{2\ell}}\big(- 2 \sum_{i=1}^{\ell} \log p_{(i)}\big)} &= \frac{-(1/n) \log \bar{F}_{\chi^2_{2j'}}\big(- 2 \sum_{i=1}^{j'} \log p_{(i)}\big)}{-(1/n) \log \bar{F}_{\chi^2_{2\ell}}\big(- 2 \sum_{i=1}^{\ell} \log p_{(i)}\big)} \\
&= \frac{-(1/n) \log \bar{F}_{\chi_{2j'}}\big((-2 \sum_{i=1}^{j'} \log p_{(i)})^{\frac{1}{2}}\big)}{-(1/n) \log \bar{F}_{\chi_{2\ell}}\big((-2 \sum_{i=1}^{\ell} \log p_{(i)})^{\frac{1}{2}}\big)} \\
&\to \frac{\sum_{i=1}^{j'} \lambda_i c_i(\theta)}{\sum_{i=1}^{\ell} \lambda_i c_i(\theta)} < 1
\end{aligned}$$

with probability one. Hence as $n \to +\infty$,

$$\hat{j} \geqslant \ell \qquad \text{(S2.11)}$$

with probability one, i.e., $\mathcal{S}_1 \to \varnothing$ with probability one.

We then prove $\mathcal{S}_2 \to \varnothing$ in probability under the alternative, which is essentially to prove $\hat{j} \leqslant \ell$ in probability. To prove this, pick arbitrary $j > \ell$, and note event $\hat{j} = j$ is equivalent to event

$$\frac{\bar{F}_{\chi^2_{2j}}\big(- 2 \sum_{i=1}^{j} \log p_{(i)}\big)}{\bar{F}_{\chi^2_{2\ell}}\big(- 2 \sum_{i=1}^{\ell} \log p_{(i)}\big)} \leqslant 1. \qquad \text{(S2.12)}$$

Then we have

$$
\begin{aligned}
(S2.12) \Leftrightarrow & \sum_{i=0}^{j-1} \frac{1}{i!}\big(-\sum_{k=1}^{j} \log p_{(k)}\big)^i \exp\big(\sum_{k=1}^{j} \log p_{(k)}\big) \\
& \leqslant \sum_{i=0}^{\ell-1} \frac{1}{i!}\big(-\sum_{k=1}^{\ell} \log p_{(k)}\big)^i \exp\big(\sum_{k=1}^{\ell} \log p_{(k)}\big) \qquad \text{(S2.13)} \\
& \Leftrightarrow \exp\big\{\sum_{k=\ell+1}^{j} \log p_{(k)}\big\} \leqslant \frac{\sum_{i=0}^{\ell-1} \frac{1}{i!}\big(-\sum_{k=1}^{\ell} \log p_{(k)}\big)^i}{\sum_{i=0}^{j-1} \frac{1}{i!}\big(-\sum_{k=1}^{j} \log p_{(k)}\big)^i} \\
& \Leftrightarrow \underbrace{\prod_{k=\ell+1}^{j} p_{(k)}}_{I} \leqslant \underbrace{\frac{\sum_{i=0}^{\ell-1} \frac{1}{i!}\big(-\sum_{k=1}^{\ell} \log p_{(k)}\big)^i}{\sum_{i=0}^{j-1} \frac{1}{i!}\big(-\sum_{k=1}^{j} \log p_{(k)}\big)^i}}_{II}. \qquad \text{(S2.14)}
\end{aligned}
$$

(S2.13) is due to relationship between Poisson distribution and chi-squared distribution. Note

$$
II = \underbrace{\frac{\sum_{i=0}^{\ell-1} \frac{1}{i!}\big(-\sum_{k=1}^{\ell} \log p_{(k)}\big)^i / \big(\frac{n}{2}\big)^{\ell-1}}{\sum_{i=0}^{j-1} \frac{1}{i!}\big(-\sum_{k=1}^{j} \log p_{(k)}\big)^i / \big(\frac{n}{2}\big)^{j-1}}}_{III} \cdot \frac{1}{\big(\frac{n}{2}\big)^{j-\ell}},
$$

and

$$
III \to \frac{(j-1)!}{(\ell-1)!} \cdot \frac{1}{\big(\sum_{k=1}^{\ell} \lambda_i c_i(\theta_i)\big)^{j-\ell}}
$$

with probability one. Hence $II = O\big(\frac{1}{n^{j-\ell}}\big)$ with probability one. While for I, with probability one, it is the product of the first $(j-\ell)$-th smallest p-values of $K-\ell$ i.i.d. p-values following Unif$(0,1)$ as $n \to +\infty$. Hence $I = O_p(1)$ under the alternative. Hence the probability that event (S2.14) holds converges to zero as $n \to +\infty$. Then the result follows. □

Let $R_j = -\sum_{i=1}^{j} \log p_{(i)} = \frac{T_j}{2}$, to prove Theorem 5, we need the following Lemma to carefully quantify the upper tails of R_j when $1 < j < K$ and under the null:

Lemma S4 (Nagaraja (2006))**.** *Let $F_{R_j}(t)$ and $\bar{F}_{R_j}(t)$ be the CDF and survival function of*

R_j under the null, separately. For $1 < j < K$, we have:

$$\bar{F}_{R_j}(t) = \sum_{i=1}^{K-j} w_i \exp\{-c_i t / c_{K-j+1}\} \frac{1}{(j-1)!} \int_0^t \exp(d_i y)\, y^{j-1} dy + \sum_{k=0}^{j-1} e^{-t} \frac{t^k}{k!},$$

where $c_i = K - i + 1$, $d_i = \frac{c_i}{c_{K-j+1}} - 1$. And

$$w_i = \prod_{k=1;k\neq i}^{K-j} \frac{K-k+1}{i-k}.$$

Further calculation leads to

$$\bar{F}_{R_j}(t) = \sum_{i=1}^{K-j} w_i \exp\{-t\} \frac{1}{(j-1)!} \left\{ \sum_{m=0}^{j-1} (-1)^m t^{j-1-m} \frac{1}{d_i^{m+1}} \frac{(j-1)!}{(j-1-m)!} \right\} + \sum_{k=0}^{j-1} e^{-t} \frac{t^k}{k!}. \tag{S2.15}$$

Proof of Theorem 5. We consider the $T_{\mathrm{AFp}} = \max_{j\in\mathcal{S}} -\log \bar{G}_j(-2\sum_{i=1}^{j} \log p_{(i)})$, where $\bar{G}_j = 1 - G_j(t)$ and $G_j(t)$ denotes the CDF function of $T_j = -2\sum_{i=1}^{j} \log p_{(i)}$ under the null. Let

$$\hat{j} = \operatorname*{argmax}_j -\log \bar{G}_j\big(-2\sum_{i=1}^{j} \log p_{(i)}\big).$$

By Lemma S3, it suffices to show $\hat{j} \to \ell$ in probability. We First show that the choice of $\hat{j} \geqslant \ell$ with probability one as n diverges under the alternative. Indeed, by the following inequality

$$\mathbb{P}(\chi^2_{2j} > t) \leqslant \bar{G}_j(t) \leqslant \mathbb{P}(\chi^2_{2K} > t), \tag{S2.16}$$

we have

$$-2\log \bar{G}_j\big(-2\sum_{i=1}^{j} \log p_{(i)}\big) \leqslant -2\log \bar{F}_{\chi^2_{2j}}\big(-2\sum_{i=1}^{j} \log p_{(i)}\big)$$
$$-2\log \bar{G}_j\big(-2\sum_{i=1}^{j} \log p_{(i)}\big) \geqslant -2\log \bar{F}_{\chi^2_{2K}}\big(-2\sum_{i=1}^{j} \log p_{(i)}\big).$$

Then by Lemma 1, Lemmas S1-S3, and Corollary S1, we have:

$$-\frac{2}{n}\log\bar{G}_j\big(-2\sum_{i=1}^{j}\log p_{(i)}\big)\to\begin{cases}\sum_{i=1}^{j}\lambda_i c_i(\theta_i) & j<\ell\\ \sum_{i=1}^{\ell}\lambda_i c_i(\theta_i) & j\geqslant\ell\end{cases}$$

with probability one. Hence $\hat{j}\geqslant\ell$ with probability one as n goes to infinity under the alternative. Now we show $\hat{j}\leqslant\ell$ in probability as n goes to infinity. Indeed, for any $j>\ell$, consider the following event:

$$\frac{A}{B}=\frac{\bar{G}_j\left(\sum_{i=1}^{j}-2\log p_{(i)}\right)}{\bar{G}_\ell\left(\sum_{i=1}^{\ell}-2\log p_{(i)}\right)}\leqslant 1. \tag{S2.17}$$

It suffices to show probability of the above event goes to zero under the alternative.

For the case $1<\ell<j<K$, by Lemma S4,

$$\begin{aligned}&\frac{A}{B}\leqslant 1\\ \Leftrightarrow\ &\frac{\bar{F}_{R_j}\left(\sum_{i=1}^{j}-\log p_{(i)}\right)}{\bar{F}_{R_\ell}\left(\sum_{i=1}^{\ell}-\log p_{(i)}\right)}\leqslant 1\\ \Leftrightarrow\ &\underbrace{\prod_{i=\ell+1}^{j}p_{(i)}}_{I}\leqslant\underbrace{\frac{\sum_{i=1}^{K-\ell}w_i\frac{1}{(\ell-1)!}\left\{\sum_{m=0}^{\ell-1}(-1)^m R_\ell^{\ell-1-m}\frac{1}{d_i^{m+1}}\frac{(\ell-1)!}{(\ell-1-m)!}\right\}+\sum_{k=0}^{\ell-1}\frac{R_\ell^k}{k!}}{\sum_{i=1}^{K-j}w_i\frac{1}{(j-1)!}\left\{\sum_{m=0}^{j-1}(-1)^m R_j^{j-1-m}\frac{1}{d_i^{m+1}}\frac{(j-1)!}{(j-1-m)!}\right\}+\sum_{k=0}^{j-1}\frac{R_j^k}{k!}}}_{II}.\end{aligned}$$

Note that

$$II\cdot\frac{\left(\frac{n}{2}\right)^{j-1}}{\left(\frac{n}{2}\right)^{\ell-1}}=II\cdot\big(\frac{n}{2}\big)^{j-\ell}\to C_{K,j,\ell}\big(\sum_{i=1}^{\ell}\lambda_i c_i(\theta_i)\big)^{j-\ell}$$

with probability one, where $C_{K,i,\ell}$ is some constant that depends on K, i and ℓ. Hence $II\to 0$ with probability one as n diverges. By Lemma S3, we note I is the product of the first $(j-\ell)$-th smallest p-values of $K-\ell$ i.i.d. p-values following Unif$(0,1)$ as $n\to+\infty$. Hence $I=O_p(1)$. And the probability of event $A/B\leqslant 1$ goes to 0 in probability. For the case $1 < \ell < i = K$, we note that

$$
\begin{aligned}
&\frac{A}{B} \leqslant 1 \\
\Leftrightarrow\ & \frac{\bar{F}_{\chi^2_{2K}}\left(\sum_{i=1}^{K} -\log p_{(i)}\right)}{\bar{F}_{R_\ell}\left(\sum_{i=1}^{\ell} -\log p_{(i)}\right)} \leqslant 1 \\
\Leftrightarrow\ & \underbrace{\prod_{i=\ell+1}^{K} p_{(i)}}_{III} \leqslant \underbrace{\frac{\sum_{i=1}^{K-\ell} w_i \frac{1}{(\ell-1)!}\left\{\sum_{m=0}^{\ell-1}(-1)^m R_\ell^{\ell-1-m}\frac{1}{d_i^{m+1}}\frac{(\ell-1)!}{(\ell-1-m)!}\right\} + \sum_{k=0}^{\ell-1}\frac{R_\ell^k}{k!}}{\sum_{i=0}^{K-1}\frac{1}{i!}\left(R_K\right)^i}}_{IV}.
\end{aligned}
$$

Note that

$$
IV \cdot \frac{\left(\frac{n}{2}\right)^{K-1}}{\left(\frac{n}{2}\right)^{\ell-1}} = IV \cdot \left(\frac{n}{2}\right)^{K-\ell} \to C_{K,\ell}\Big(\sum_{i=1}^{\ell} \lambda_i c_i(\theta_i)\Big)^{K-\ell}
$$

with probability one, where $C_{K,\ell}$ is some constant that depends on K and ℓ. Hence $IV \to 0$ with probability one as n diverges. By Lemma S3, we note III is the product of $K-\ell$ i.i.d. p-values following $\mathrm{Unif}(0,1)$ as $n \to +\infty$. Hence $III = O_p(1)$. And the probability of event $A/B \leqslant 1$ goes to 0 in probability.

Now we consider the case $1 = \ell < j < K$. By inequality $(1+x)^K \geq 1 + Kx$ for $x > -1$, we have

$$
\begin{aligned}
\mathbb{P}(A/B \leqslant 1) &= \mathbb{P}\Big(\frac{\bar{F}_{R_j}(\sum_{i=1}^{j} -\log p_{(i)})}{\bar{F}_{R_1}(-\log p_{(1)})} \leqslant 1\Big) \\
&= \mathbb{P}\Big(\frac{\bar{F}_{R_j}(\sum_{i=1}^{j} -\log p_{(i)})}{1-(1-\exp(\log p_{(1)}))^K} \leqslant 1\Big) \\
&\leq \mathbb{P}\Big(\frac{\bar{F}_{R_j}(\sum_{i=1}^{j} -\log p_{(i)})}{K\exp(-R_1)} \leqslant 1\Big).
\end{aligned}
$$

Hence it suffices to show $\mathbb{P}\Big(\frac{\bar{F}_{R_j}(\sum_{i=1}^{j} -\log p_{(i)})}{K\exp(-R_1)} \leqslant 1\Big) \to 0$ as n diverges. Note that by Lemma

S4, we have

$$\frac{\bar{F}_{R_j}(\sum_{i=1}^{j} - \log p_{(i)})}{K \exp(-R_1)} \leqslant 1$$
$$\Leftrightarrow \underbrace{\prod_{i=2}^{j} p_{(i)}}_{V} \leqslant \underbrace{\frac{K}{\sum_{i=1}^{K-j} w_i \frac{1}{(j-1)!} \left\{ \sum_{m=0}^{j-1} (-1)^m R_j^{j-1-m} \frac{1}{d_i^{m+1}} \frac{(j-1)!}{(j-1-m)!} \right\} + \sum_{k=0}^{j-1} \frac{R_j^k}{k!}}}_{VI}$$

Note that

$$VI \cdot \left(\frac{n}{2}\right)^{K-1} \to C_K \big(\lambda_1 c_1(\theta_1)\big)^{K-1}$$

with probability one, where C_K is some constant that depends on K. Hence $VI \to 0$ with probability one as n diverges. By Lemma S3, we note V is the product of $j-1$ smallest p-values of $K-1$ i.i.d. p-values following Unif$(0,1)$ as $n \to +\infty$. Hence $V = O_p(1)$. And the probability of event $A/B \leqslant 1$ goes to 0 in probability. The arguments for the case $1 = \ell < j = K$ is quite similar, hence we omit the details. Combine the above results, we have $\hat{j} \leq \ell$ in probability as n diverges. Then the conclusion follows. □

We prove Theorem 6 by proving the following test statistic in a more general form is ABO:

$$T(\tau_1, \tau_2) = \sum_{i=1}^{K} \left(-2\log(p_i) + 2\log(\tau_2)\right) \mathrm{I}_{\{p_i \leq \tau_1\}} \text{ with } 0 \leqslant \tau_1, \tau_2 \leqslant 1.$$

When $\tau_1 = \tau$ and $\tau_2 = 1$, $T(\tau_1, \tau_2) = T_{\text{TFhard}}(\tau)$; and when $\tau_1 = \tau_2 = \tau$, $T(\tau_1, \tau_2) = T_{\text{TFsoft}}(\tau)$. The proof of the Theorem 6 requires the following additional lemma:

Lemma S5 (Zhang et al. (2020))**.** *Assume $p_1, \ldots, p_K \sim$ Unif$(0,1)$ independently and identically. Denote by $U(\tau_1, \tau_2)$ the random variable that follows the same distribution of $T(\tau_1, \tau_2)$*

under the null. Then

$$\bar{F}_{U(\tau_1,\tau_2)}(t) = \big(1-\tau_1\big)^K I_{\{t\leqslant 0\}} + \sum_{i=1}^{K}\binom{K}{i}\tau_1^i\big(1-\tau_1\big)^{K-i}\bar{F}_{\chi^2_{2i}}\big(t+2i\log\big(\tau_1/\tau_2\big)\big) \tag{S2.18}$$

Proof of Theorem 6. We only prove the case of $\tau_2 \leqslant \tau_1$ as the case of $\tau_2 > \tau_1$ can be proved by similar arguments. Let $F_{U(\tau_1,\tau_2)}(t)$ and $\bar{F}_{U(\tau_1,\tau_2)}(t)$ be the CDF and survival function of $U(\tau_1,\tau_2)$. Consider test statistic $\sqrt{T(\tau_1,\tau_2)}$. Under the setup in Section 2.1 and the alternative, by Lemmas 1 and S3, we have

$$\frac{\sqrt{T(\tau_1,\tau_2)}}{\sqrt{n}} = \frac{\sqrt{\sum_{i=1}^{K}\big(-2\log p_i + 2\log\tau_2\big)\mathrm{I}_{\{p_i\leq\tau_1\}}}}{\sqrt{n}} \to \big(\sum_{i=1}^{\ell}\lambda_i c_i(\theta_i)\big)^{\frac{1}{2}} \tag{S2.19}$$

with probability one as $n\to\infty$. In addition, by Lemma S1, for each $i=1,\ldots,K$,

$$-\frac{1}{n}\log\bar{F}_{\chi^2_{2i}}\left(nt^2+2i\log\left(\tau_1/\tau_2\right)\right) = -\frac{1}{n}\log\bar{F}_{\chi_{2i}}\left(\sqrt{nt^2+2i\log\left(\tau_1/\tau_2\right)}\right) \to \frac{t^2}{2}$$

as $n\to\infty$. Note by Lemma S5, for $t>0$ we have

$$\begin{aligned}
\bar{F}_{\sqrt{U(\tau_1,\tau_2)}}(\sqrt{n}t) &= \bar{F}_{U(\tau_1,\tau_2)}(nt^2) \\
&\geqslant \bar{F}_{\chi^2_2}\left(nt^2+2K\log(\tau_1/\tau_2)\right)\sum_{i=1}^{K}\binom{K}{i}\tau_1^i\left(1-\tau_1\right)^{K-i} \\
\bar{F}_{\sqrt{U(\tau_1,\tau_2)}}(\sqrt{n}t) &= \bar{F}_{U(\tau_1,\tau_2)}(nt^2) \\
&\leqslant \bar{F}_{\chi^2_{2K}}\left(nt^2+2\log\left(\tau_1/\tau_2\right)\right)\sum_{i=1}^{K}\binom{K}{i}\tau_1^i\left(1-\tau_1\right)^{K-i}.
\end{aligned}$$

Hence

$$-\frac{1}{n}\log\bar{F}_{\sqrt{U(\tau_1,\tau_2)}}\left(\sqrt{n}t\right) \to \frac{t^2}{2} \tag{S2.20}$$

with probability one as $n\to\infty$. By combining (S2.19) and (S2.20) and applying Lemma

S2, we have for the exact slope of $T(\tau_1, \tau_2)$,

$$\begin{aligned} C_{T(\tau_1,\tau_2)} &= -\frac{2}{n}\log \bar{F}_{U(\tau_1,\tau_2)}\big(T(\tau_1,\tau_2)\big) \\ &= -\frac{2}{n}\log \bar{F}_{\sqrt{U(\tau_1,\tau_2)}}\big(\sqrt{T(\tau_1,\tau_2)}\big) \to \sum_{i=1}^{\ell} \lambda_i c_i(\theta_i). \end{aligned}$$

Hence $T(\tau_1, \tau_2)$ is ABO. □

S2.2 Proof of Theorem 7

Proof of Theorem 7. Let $U_{\mathrm{RV}}(\gamma)$ be the random variable that follows the same distribution of $T_{\mathrm{RV}}(\gamma)$ under the null. Denote by $F_{U_{\mathrm{RV}}(\gamma)}(t)$ and $\bar{F}_{U_{\mathrm{RV}}(\gamma)}(t)$ the CDF and survival function of $T_{\mathrm{RV}}(\gamma)$ under the null. Furthermore, under the null, let $U(\gamma)$ be the random variable such that $U(\gamma) \in R_{-\gamma}$. Hence $g_\gamma(p_{T_i}) = F^{-1}_{U(\gamma)}(1-p_{T_i})$ follows the same distribution of $U(\gamma)$ under the null. Let $t_i = F^{-1}_{U(\gamma)}(1-p_{T_i})$. Consequently, under the alternative, for i such that $C_i(\vec{\theta}) > 0$, $p_{T_i} = \bar{F}_{U(\gamma)}(t_i)$ and $\bar{F}_{U(\gamma)}(t_i)/(L(t_i)t_i^{-\gamma}) \to 1$ with probability one. We have, under the alternative, as $n \to +\infty$,

$$-\frac{2}{n}\log\left(\bar{F}_{U(\gamma)}(t_i)/(L(t_i)t_i^{-\gamma})\right) - \frac{2}{n}\log\left(L(t_i)t_i^{-\gamma}\right) = -\frac{2}{n}\log(p_{T_i}) \to C_i(\vec{\theta})$$

with probability one. Hence $-\frac{2}{n}\log\left(L(t_i)t_i^{-\gamma}\right) \to C_i(\vec{\theta})$ with probability one. By the basic property of slowly varying function, we have $L(t_i) = o(t_i^{\gamma})$ with probability one for any γ. Hence for i such that $C_i(\vec{\theta}) > 0$,

$$-\frac{2}{n}\log(t_i^{-\gamma}) \to C_i(\vec{\theta}) \tag{S2.21}$$

with probability one. Let $t_0 = \sum_{i=1}^{L} F^{-1}_{U(\gamma)}(1-p_{T_i}) = \sum_{i=1}^{L} t_i$, then by Bonferroni's inequality, we have $\bar{F}_{U_{\mathrm{RV}}(\gamma)}(t_0) \leqslant L \cdot \bar{F}_{U(\gamma)}(\frac{t_0}{L})$ with probability one. then we have

$$
\begin{aligned}
& -\frac{2}{n}\log \bar{F}_{U_{\mathrm{RV}}(\gamma)}(T_{\mathrm{RV}}(\gamma)) \\
& \geqslant -\frac{2}{n}\log\big(L\bar{F}_{U(\gamma)}(\frac{t_0}{L})/(L(t_0)L^{\gamma+1}t_0^{-\gamma})\big) - \frac{2}{n}\log(L(t_0)L^{\gamma+1}t_0^{-\gamma}) \\
& = \underbrace{-\frac{2}{n}\log\big(\bar{F}_{U(\gamma)}(\frac{t_0}{L})/(L(t_0)L^{\gamma}t_0^{-\gamma})\big)}_{(A)} + \underbrace{\frac{2\gamma\log t_0 - 2\log L^{\gamma+1}L(t_0)}{n}}_{(B)}.
\end{aligned}
$$

Under the alternative, for (A), with $\max_{1\leqslant i\leqslant L} C_i(\vec{\theta}) > 0$ and either Conditions (C1) or (C2) holds, we have $t_0 \to +\infty$ with probability one. Then we have

$$
\bar{F}_{U(\gamma)}(\frac{t_0}{L})/(L(t_0)L^{\gamma}t_0^{-\gamma}) = \Big[\bar{F}_{U(\gamma)}(\frac{t_0}{L})/[L(\frac{t_0}{L})(\frac{t_0}{L})^{-\gamma}]\Big]\cdot\Big[L(\frac{t_0}{L})/L(t_0)\Big] \to 1
$$

with probability one, where the first term converges to 1 by the regularly varying tailed distribution definition and the second term converges to 1 by the definition of slow-varying distribution. Hence we have $(A) \to 0$ with probability one. For (B), we first assume Condition (C2) holds. Let $C_{i^*}(\vec{\theta}) = \max_{1\leqslant i\leqslant L} C_i(\vec{\theta})$, then under the alternative, by (S2.21) we have

$$
\begin{aligned}
& \frac{2\gamma}{n}\log t_0 = \frac{2\gamma}{n}\log\big(\sum_{i=1}^{L} t_i\big) \geqslant \frac{2\gamma}{n}\max_{1\leqslant i\leqslant L}\{\log(t_i)\} \to C_{i^*}(\vec{\theta}) \\
& \frac{2\gamma}{n}\log t_0 = \frac{2\gamma}{n}\log\big(\sum_{i=1}^{L} t_i\big) \leqslant \frac{2\gamma}{n}\max_{1\leqslant i\leqslant L}\{\log(t_i)\} + \frac{2\gamma\log L}{n} \to C_{i^*}(\vec{\theta})
\end{aligned}
$$

with probability one. Suppose Condition (C1) holds and Condition (C2) does not hold, it suffices to consider the worst case that $F^{-1}_{U(\gamma)}(1-p) \geqslant \nu$ for some $\nu < 0$ and $\forall p \in (0,1]$. Denote by index set $\mathcal{B} = \{i : C_i(\vec{\theta}) > 0\}$. Then under the alternative, with probability one we have

$$\frac{2\gamma}{n}\log t_0 = \frac{2\gamma}{n}\log\big(\sum_{i\in\mathcal{B}} t_i + \sum_{i\in\mathcal{B}^c} t_i\big) = \frac{2\gamma}{n}\log\big(\sum_{i\in\mathcal{B}} t_i\big) + \frac{2\gamma}{n}\log\big(1+\frac{\sum_{i\in\mathcal{B}^c} t_i}{\sum_{i\in\mathcal{B}} t_i}\big)$$
$$\geqslant \underbrace{\frac{2\gamma}{n}\log\big(\sum_{i\in\mathcal{B}} t_i\big)}_{(C)} + \underbrace{\frac{2\gamma}{n}\log\big(1+\frac{|\mathcal{B}^c|\nu}{\sum_{i\in\mathcal{B}} t_i}\big)}_{(D)},$$

where $|\mathcal{B}^c|$ denotes the cardinality of index set $\mathcal{B}^c$. For term (C), by (S2.21), under the alternative we have

$$\frac{2\gamma}{n}\log\big(\sum_{i\in\mathcal{B}} t_i\big) \geqslant \frac{2\gamma\max_{i\in\mathcal{B}}\{\log(t_i)\}}{n} = \frac{2\gamma\max_{1\leqslant i\leqslant L}\{\log(t_i)\}}{n} \to C_{i^*}(\vec{\theta})$$
$$\frac{2\gamma}{n}\log\big(\sum_{i\in\mathcal{B}} t_i\big) \leqslant \frac{2\gamma\max_{i\in\mathcal{B}}\{\log(t_i)\}}{n} + \frac{2\gamma\log|\mathcal{B}|}{n}$$
$$= \frac{2\gamma\max_{1\leqslant i\leqslant L}\{\log(t_i)\}}{n} + \frac{2\gamma\log|\mathcal{B}|}{n} \to C_{i^*}(\vec{\theta})$$

with probability one. Here we can also show that term (D) converges to zero with probability one as $n\to+\infty$. Hence $\frac{2\gamma}{n}\log t_0 = C_{i^*}(\vec{\theta})$ with probability one under the alternative. Further note $L(t_0) = o(t_0^{\gamma})$ with probability one, then we have $(B) = C_{i^*}(\vec{\theta})$ with probability one. Hence under the alternative

$$-\frac{2}{n}\log\bar{F}_{U_{\mathrm{RV}(\gamma)}}(T_{\mathrm{RV}}(\gamma)) = C_{i^*}(\vec{\theta})$$

as $n\to+\infty$ with probability one. □

Remark S4. The result of Theorem 7 also holds for the weighted version of $T_{\mathrm{RV}}(\gamma)$ by the similar arguments in the above proof:

$$T^{\epsilon}_{\mathrm{RV}}(\gamma) = \sum_{i=1}^{L}\epsilon_i g_\gamma(p_{T_i}) = \sum_{i=1}^{L}\epsilon_i F^{-1}_{U(\gamma)}(1-p_{T_i})$$

with $\sum_{i=1}^{L}\epsilon_i = 1$ and $\epsilon_i > 0$ for each $i = 1,\ldots,L$.

S2.3 Proofs of Theorems S2-S4 and Proposition S1

Lemma S6 (Mikosch (1999))**.** *Assume* $U_1(\gamma), \ldots, U_K(\gamma)$ *are i.i.d. random variables with distribution function* $F \in R_{-\gamma}$*. Then as* $t \to \infty$*, we have*

$$\mathbb{P}\left(U_1(\gamma) + \ldots + U_K(\gamma) > t\right) / \left(K\mathbb{P}\left(U_1(\gamma) > t\right)\right) \to 1. \tag{S2.22}$$

proof of Theorem S2. Denote $T_\eta = \sqrt{(1/\eta)\log\big(\sum_{i=1}^K 1/p_i^\eta\big)}$. Let $U(\eta)$ be the random variable that follows the same distribution of T_η under the null. Denote by $F_{U(\eta)}(t)$ and $\bar{F}_{U(\eta)}(t)$ the CDF and the survival function of T_η under the null. Further denote by $\mathbb{P}_0$ the probability measure of $\vec{p} = (p_1, \ldots, p_K)$ under the null. First note that $\bar{F}_{U(\eta)}(\sqrt{n}t) = \mathbb{P}_0\big(\sum_{i=1}^K 1/p_i^\eta > \exp(\eta n t^2)\big)$. Further note that $\frac{1}{p_i^\eta} \overset{D}{\sim} \text{Pareto}(\frac{1}{\eta}, 1) \in R_{-\frac{1}{\eta}}$ under the null, where the explicit form of survival function of Pareto distribution is $\bar{F}_{\text{Pareto}(\frac{1}{\eta},1)}(t) = t^{-\frac{1}{\eta}}$. Hence by Lemma S6 we have

$$\begin{aligned} &\bar{F}_{U(\eta)}(\sqrt{n}t)/(K\bar{F}_{\text{Pareto}(\frac{1}{\eta},1)}(\exp(\eta n t^2)) \\ &= \mathbb{P}_0\big(\sum_{i=1}^K 1/p_i^\eta > \exp(\eta n t^2)\big)/(K(\exp(\eta n t^2))^{-\frac{1}{\eta}}) \to 1, \end{aligned}$$

as $n \to +\infty$. Then we have,

$$-\frac{1}{n}\log\big(1 - F_{U(\eta)}(\sqrt{n}t)\big) \to t^2, \tag{S2.23}$$

as $n \to \infty$. We further claim under the alternative,

$$\frac{T_\eta}{\sqrt{n}} \to \sqrt{\max_{1 \leqslant i \leqslant K}\left\{\lambda_i c_i(\theta_i)\right\}/2} \tag{S2.24}$$

with probability one. Indeed, note

$$\max_{1 \leqslant i \leqslant K}\left\{\log(1/p_i^\eta)\right\} \leqslant \log\big(\sum_{i=1}^K 1/p_i^\eta\big) \leqslant \log K + \max_{1 \leqslant i \leqslant K}\left\{\log(1/p_i^\eta)\right\}.$$

Hence under the alternative, we have

$$\frac{1}{n}\log\big(\sum_{i=1}^{K} 1/p_i^{\eta}\big) \to \eta \max_{1\leqslant i\leqslant K} \lambda_i c_i(\theta)/2 \tag{S2.25}$$

with probability one. Then we have

$$\frac{T_\eta}{\sqrt{n}} = \frac{\sqrt{(1/\eta)\log\big(\sum_{i=1}^{K} 1/p_i^{\eta}\big)}}{\sqrt{n}} \to \sqrt{\max_{1\leqslant i\leqslant K} \lambda_i c_i(\theta)/2}$$

with probability one. Hence (S2.24) holds. Combining (S2.23) and (S2.24) and by Lemma S2, the result follows. □

Proof of Theorem S3. Note $T_{\text{CA}} = \frac{1}{K}\sum_{i=1}^{K}\cot(\pi p_i)$. Under the alternative, recall without loss of generality we assume that the first ℓ p-values correspond to non-zero exact slopes $c_i(\theta_i) > 0$ $(1 \leqslant i \leqslant \ell)$, while the remaining p-values correspond to the zero exact slopes ($p_i \sim \text{Unif}(0,1)$ for $\ell+1 \leqslant i \leqslant K$). For the p-values with non-zero exact slopes, by the Taylor's expansion $x\cot x - 1 = -\frac{x^2}{3} + o(x^2)$, under the alternative we have,

$$\frac{1}{K}\sum_{i=1}^{\ell}\big[\frac{1}{\pi p_i} - \frac{2\pi p_i}{3}\big] \leqslant \frac{1}{K}\sum_{i=1}^{\ell}\cot(\pi p_i) \leqslant \frac{1}{K}\sum_{i=1}^{\ell}\frac{1}{\pi p_i}$$

with probability one. Note $\frac{1}{K}\sum_{i=1}^{\ell}\left[\frac{1}{\pi p_i} - \frac{2\pi p_i}{3}\right] = \frac{1}{K}\left(1 - \frac{\sum_{i=1}^{K} 2\pi p_i/3}{\sum_{i=1}^{K} 1/\pi p_i}\right)\sum_{i=1}^{\ell}\frac{1}{\pi p_i}$ and under the alternative, with probability one,

$$\Big(1 - \frac{\sum_{i=1}^{\ell} 2\pi p_i/3}{\sum_{i=1}^{\ell} 1/\pi p_i}\Big) \to 1$$

$$\frac{1}{n}\log\big(\frac{1}{K}\sum_{i=1}^{\ell} 1/\pi p_i\big) \to \frac{1}{2}\max_{1\leqslant i\leqslant \ell}\lambda_i c_i(\theta), \tag{S2.26}$$

where (S2.26) is due to similar arguments for (S2.25) in the proof of Theorem S2 for $\eta = 1$. Hence we have

$$\frac{1}{n}\log\big(\frac{1}{K}\sum_{i=1}^{\ell}\cot(\pi p_i)\big) \to \frac{1}{2}\max_{1\leqslant i\leqslant \ell}\lambda_i c_i(\theta) \tag{S2.27}$$

with probability one.

Note that $\cot(\pi p_{\ell+1}), \ldots, \cot(\pi p_K) \overset{\text{i.i.d.}}{\sim} \mathrm{CAU}(0,1)$. Hence we have

$$\frac{1}{K}\sum_{i=\ell+1}^{K} \cot\left(\pi p_i\right) \overset{D}{\sim} \frac{K-\ell}{K} U_{\mathrm{CAU}(0,1)},$$

where $U_{\mathrm{CAU}(0,1)}$ denotes standard Cauchy random variable. Note that under the null, $T_{\mathrm{CA}} \overset{D}{\sim} \mathrm{CAU}(0,1)$. Hence $F_{\mathrm{CAU}(0,1)}(t)$ and $\bar{F}_{\mathrm{CAU}(0,1)}(t)$ are the CDF and survival function of T_{CA} under the null. Hence under the alternative, we have

$$\begin{aligned}
\bar{F}_{\mathrm{CAU}(0,1)}(T_{\mathrm{CA}}) &= \mathbb{P}(U_{\mathrm{CAU}(0,1)} > \frac{1}{K}\sum_{i=1}^{\ell}\cot(\pi p_i) + \frac{1}{K}\sum_{i=\ell+1}^{K}\cot(\pi p_i)) \\
&= \mathbb{P}\Big(\Big(1+\frac{K-\ell}{K}\Big)U_{\mathrm{CAU}(0,1)} > \frac{1}{K}\sum_{i=1}^{\ell}\cot(\pi p_i)\Big) \\
&= \mathbb{P}\Big(U_{\mathrm{CAU}(0,1)} > \frac{K}{2K-\ell}\cdot\frac{1}{K}\sum_{i=1}^{\ell}\cot(\pi p_i)\Big). \qquad \text{(S2.28)}
\end{aligned}$$

In addition, for $t > 1$,

$$\begin{aligned}
\bar{F}_{\mathrm{CAU}(0,1)}\left(t\right) &= \frac{1}{2} - \frac{1}{\pi}\arctan t = \frac{1}{\pi}\cdot\arctan\left(1/t\right) \leqslant \frac{1}{\pi t} \\
\bar{F}_{\mathrm{CAU}(0,1)}\left(t\right) &= \frac{1}{\pi}\arctan\frac{1}{t} \geqslant \frac{1}{\pi t}\cdot\frac{t^2}{1+t^2}.
\end{aligned}$$

By combining the above two inequalities with (S2.27) and (S2.28), under the alternative we have

$$-\frac{2}{n}\log\left(\bar{F}_{\mathrm{CAU}(0,1)}(T_{\mathrm{CA}})\right) \to \max_{1\leqslant i\leqslant \ell}\lambda_i c_i(\theta) = \max_{1\leqslant i\leqslant K}\lambda_i c_i(\theta)$$

with probability one. □

To prove Theorem S4, we introduce the following notations adopted from Zhang et al. (2020). Define

$$f_1^{\phi}(x,y) = x\log(\frac{x}{y}) + (1-x)\log(\frac{1-x}{1-y}).$$

Further define

$$f(x,y) = \sqrt{2Kf_1^{\phi}(x,y)} \qquad \text{if } y \leqslant x$$
$$= -\sqrt{2Kf_1^{\phi}(x,y)} \quad \text{if } y > x.$$

Note that $f(x,y)$ is strictly decreasing in y. When $T_{\mathrm{BJ}} > 0$, we have

$$\sqrt{2KT_{\mathrm{BJ}}} = \max_{1\leqslant i\leqslant K} f(\frac{i}{K}, p_{(i)}).$$

For each fixed x, define the inverse function of $f(x,\cdot)$ as $g(x,\cdot)$, i.e.,

$$g(x,\cdot) = f^{-1}(x,\cdot).$$

Proof of Theorem S4 . Let $i^* = \operatorname{argmax}_i i\lambda_i c_i(\theta)$ and note that by Lemma 1, $i^* \leqslant \ell$. We first show that under the alternative,

$$2KT_{\mathrm{BJ}}/n \to i^*\lambda_{i^*}c_{i^*}(\theta) \tag{S2.29}$$

with probability one. Denote

$$\hat{i} = \operatorname*{argmax}_i \Big\{\frac{i}{K}\log\Big(\frac{i/K}{p_{(i)}}\Big) + \Big(1-\frac{i}{K}\Big)\log\Big(\frac{1-i/K}{1-p_{(i)}}\Big)\Big\}\mathrm{I}_{\{p_{(i)}<\frac{i}{K}\}}.$$

We show that under the alternative $\hat{i} \to i^*$ with probability one. Indeed, for any $i \neq i^*$ and $i \leqslant \ell$, by Lemma S3, we have

$$\frac{(1/n)\Big[\frac{i}{K}\log\Big(\frac{i/K}{p_{(i)}}\Big) + \Big(1-\frac{i}{K}\Big)\log\Big(\frac{1-i/K}{1-p_{(i)}}\Big)\Big]\mathrm{I}_{\{p_{(i)}<\frac{i}{K}\}}}{(1/n)\Big[\frac{i^*}{K}\log\Big(\frac{i^*/K}{p_{(i^*)}}\Big) + \Big(1-\frac{i^*}{K}\Big)\log\Big(\frac{1-i^*/K}{1-p_{(i^*)}}\Big)\Big]\mathrm{I}_{\{p_{(i^*)}<\frac{i^*}{K}\}}} \to \frac{i\lambda_i c_i(\theta)}{i^*\lambda_{i^*}c_{i^*}(\theta)} < 1$$

with probability one. For any $i > \ell$, note that $1 - p_i$ still follows Unif$(0,1)$. Hence for any $i' > \ell$, by Lemmas 1 and S3, we have $-\frac{1}{n}\log p_{(i')} \to 0$ and $-\frac{1}{n}\log(1-p_{(i')}) \to 0$ with probability one. Hence

$$\frac{(1/n)\left[\frac{i'}{K}\log\left(\frac{i'/K}{p_{(i')}}\right)+\left(1-\frac{i'}{K}\right)\log\left(\frac{1-i'/K}{1-p_{(i')}}\right)\right]\mathrm{I}_{\{p_{(i')}<\frac{i'}{K}\}}}{(1/n)\left[\frac{i^*}{K}\log\left(\frac{i^*/K}{p_{(i^*)}}\right)+\left(1-\frac{i^*}{K}\right)\log\left(\frac{1-i^*/K}{1-p_{(i^*)}}\right)\right]\mathrm{I}_{\{p_{(i^*)}<\frac{i^*}{K}\}}}$$
$$\leqslant \frac{(1/n)\left[\frac{i'}{K}\log\left(\frac{i'/K}{p_{(i')}}\right)+\left(1-\frac{i'}{K}\right)\log\left(\frac{1-i'/K}{1-p_{(i')}}\right)\right]}{(1/n)\left[\frac{i^*}{K}\log\left(\frac{i^*/K}{p_{(i^*)}}\right)+\left(1-\frac{i^*}{K}\right)\log\left(\frac{1-i^*/K}{1-p_{(i^*)}}\right)\right]\mathrm{I}_{\{p_{(i^*)}<\frac{i^*}{K}\}}} \to 0$$

with probability one. Hence under the alternative $2KT_{\mathrm{BJ}}/n \to i^*\lambda_{i^*}c_{i^*}(\theta_{i^*})$ with probability one.

Denote by U_{BJ} the random variable follows the same distribution of $\sqrt{2KT_{\mathrm{BJ}}}$ under the null, and let

$$\mu_i = g(\frac{i}{K}, b) = f^{-1}(\frac{i}{K}, b),\ i = 1, 2, \ldots, K.$$

Let $F_{U_{\mathrm{BJ}}}$, $\bar{F}_{U_{\mathrm{BJ}}}$ be the CDF and survival function of U_{BJ}, respectively. Also let $F_{\mathrm{Beta}(\alpha,\beta)}$ and $\bar{F}_{\mathrm{Beta}(\alpha,\beta)}$ be the CDF and survival function of $\mathrm{Beta}(\alpha, \beta)$, respectively. By Theorem 5.1 in Zhang et al. (2020), we have,

$$F_{U_{\mathrm{BJ}}}(b) = \bar{F}_{\mathrm{Beta}(K,1)}(\mu_K) - \sum_{i=1}^{K-1} \frac{\mu_i^i}{i!} a_{i+1}, \tag{S2.30}$$

where

$$a_K = K!\bar{F}_{\mathrm{Beta}(1,1)}(\mu_K)$$
$$a_i = \frac{K!}{(K-i+1)!}\bar{F}_{\mathrm{Beta}(K-i+1,1)}(\mu_K) - \sum_{j=1}^{K-i} \frac{\mu_{i+j-1}^j}{j!} a_{i+j}$$

for $i = K-1, K-2, \ldots, 1$.

Since $\mu_i = g(\frac{i}{K}, b) = f^{-1}(\frac{i}{K}, b)$, for sufficiently large b, we have

$$b = \sqrt{2Kf_1^{\phi}(\frac{i}{K}, \mu_i)}.$$

Hence

$$\frac{b^2}{2K} = f_1^\phi(\frac{i}{K}, \mu_i) = \frac{i}{K}\log\frac{\frac{i}{K}}{\mu_i} + (1-\frac{i}{K})\log\frac{1-\frac{i}{K}}{1-\mu_i}.$$

Then

$$e^{-[\frac{b^2}{2K}-\frac{i}{K}\log\frac{i}{K}-(1-\frac{i}{K}\log(1-\frac{i}{K}))]} = \mu_i^{\frac{i}{K}}(1-\mu_i)^{1-\frac{i}{K}}. \tag{S2.31}$$

Note $f(x,y)$ is strictly decreasing in y, for $b\to\infty$, $\mu_i\to 0$. Denote

$$\mu_i = C_{i,b}e^{-\frac{b^2}{2i}},$$

where $C_{i,b}$ depends on i and b. We show that there exist $C_i > 0$ only depending on i, such that

$$\lim_{b\to\infty} C_{i,b} = C_i. \tag{S2.32}$$

Indeed, from equation (S2.31), we have

$$\lim_{b\to\infty}\frac{e^{-[\frac{b^2}{2K}-\frac{i}{K}\log\frac{i}{K}-(1-\frac{i}{K}\log(1-\frac{i}{K}))]}}{(C_{i,b}e^{-\frac{b^2}{2i}})^{\frac{i}{K}}(1-C_{i,b}e^{-\frac{b^2}{2i}})^{1-\frac{i}{K}}} = 1.$$

Hence

$$\lim_{b\to\infty}\frac{e^{\frac{i}{K}\log\frac{i}{K}+(1-\frac{i}{K}\log(1-\frac{i}{K}))}}{C_{i,b}^{\frac{i}{K}}} = 1.$$

Hence we have $\lim_{b\to\infty} C_{i,b} = C_i > 0$ for $i = 1,\ldots,K$. Hence for sufficiently large b, we have

$$\mu_i = (C_i + o(1))e^{-\frac{b^2}{2i}}.$$

As $\lim_{b\to\infty}\mu_i = 0$, for sufficiently large b, we have

$$a_k = K!\bar{F}_{\mathrm{Beta}(1,1)}(\mu_K) = K! + o(1).$$

Similarly, for $i = 1, \dots, K-1$ and sufficiently large b,

$$a_i = \frac{K!}{(K-i+1)!} + o(1).$$

For $\bar{F}_{U_{\mathrm{BJ}}} = 1 - F_{U_{\mathrm{BJ}}}$, we have

$$\bar{F}_{U_{\mathrm{BJ}}}(b) = \underbrace{F_{\mathrm{Beta}(K,1)}(\mu_K)}_{I} + \underbrace{\sum_{i=1}^{K-1} \frac{\mu_i^i}{i!} a_{i+1}}_{II}. \tag{S2.33}$$

As $\mu_i^i = C_{i,b}^i e^{-\frac{b^2}{2}} = (C_i^i + o(1))e^{-\frac{b^2}{2}}$ for sufficiently large b, we have

$$I = F_{\mathrm{Beta}(K,1)}(\mu_K) = \int_0^{\mu_K} K x^{K-1} dx = \mu_K^K = (C_K^K + o(1))e^{-\frac{b^2}{2}}.$$

Similarly,

$$II = \sum_{i=1}^{K-1} \frac{\mu_i^i}{i!} a_{i+1} = \sum_{i=1}^{K-1} \frac{(C_i^i + o(1))e^{-\frac{b^2}{2}}}{i!} \Big[\frac{K!}{(K-i+1)!} + o(1)\Big].$$

Hence for sufficiently large b,

$$\begin{aligned} (S2.33) = I + II &= \Big[C_K^K + \sum_{i=1}^{K-1} \frac{C_i^i}{i!} \frac{K!}{(K-i+1)!} + o(1)\Big] e^{-\frac{b^2}{2}} \\ &= (C(K) + o(1))e^{-\frac{b^2}{2}}, \end{aligned} \tag{S2.34}$$

where $C(K)$ only depends on K. Let $b = \sqrt{2KT_{\mathrm{BJ}}}$, combine equations (S2.29) and (S2.34), under the alternative, we have

$$-\frac{2\log \bar{F}_{U_{\mathrm{BJ}}}\big(\sqrt{2KT_{\mathrm{BJ}}}\big)}{n} \to i^* \lambda_{i^*} c_{i^*}(\theta_{i^*})$$

with probability one. □

Remark S5. It can be shown that T_{BJ} generally does not has signal selection consistency. Recall that T_{BJ} picks $i^* = \mathrm{argmax}_i \, i\lambda_i c_i(\theta_i)$ with probability one as shown in the proof.

Consider $K = 2$ and there is only two signals, with $\lambda_1 c_1(\theta_1) = 9$ and $\lambda_2 c_2(\theta_2) = 1$. Then one can show $i^* = 1$ here, i.e., T_{BJ} picks the wrong subset of p-values with probability one.

Below we use a counter example to show that higher criticism is generally not ABO. Let U_{HC} be the random variable that follows the same distribution of T_{HC} under the null. Denoted by $F_{U_{\mathrm{HC}}}$ and $\bar{F}_{U_{\mathrm{HC}}}$ the CDF and survival function of U_{HC}, respectively. To prove Proposition S1, we need the following Lemma to derive the survival function $\bar{F}_{U_{\mathrm{HC}}}$ under the finite-sample case.

Lemma S7 (Barnett and Lin (2014))**.** *For each $k = 1, \ldots, K$, let*

$$t_k = \Phi^{-1}\left[1 - \frac{2(K-k+1) + h^2 - h\left\{h^2 + 4(K-k+1) - 4(K-k+1)^2/K\right\}^{1/2}}{4\left(h^2 + K\right)}\right].$$

Denote $q_{1,a} = \mathbb{P}(S(t_1) = a)$ for $a = 0, 1, \ldots, K-1$. Here $S(t) = \sum_{j=1}^{K} I_{\{(|Z_j| \geqslant t)\}}$ is the binomial random variable with $Z_1, \ldots, Z_K \overset{i.i.d.}{\sim} N(0,1)$. Let

$$q_{k,a} = \sum_{m=0}^{K-k+1} I_{\{a \leqslant m\}} \binom{m}{a} \left\{\bar{\Phi}\left(t_k\right)/\bar{\Phi}\left(t_{k-1}\right)\right\}^a \times \left\{1 - \bar{\Phi}\left(t_k\right)/\bar{\Phi}\left(t_{k-1}\right)\right\}^{m-a} \frac{q_{k-1,m}}{\sum_{\ell=0}^{K-k+1} q_{k-1,\ell}}$$

for $k = 2, \ldots, K$ and $a = 0, 1, \ldots, K-k$. Then we have

$$\bar{F}_{U_{HC}}(h) = 1 - \prod_{k=1}^{K} \sum_{a=0}^{K-k} q_{k,a}.$$

Proof of Proposition S1. We first derive the exact form of $\bar{F}_{U_{\mathrm{HC}}}(h)$ for $K = 2$. By Lemma S7,

$$\bar{F}_{U_{\mathrm{HC}}}(h) = 1 - \prod_{k=1}^{2} \sum_{a=0}^{2-k} q_{k,a} = 1 - (q_{1,0} + q_{1,1}) q_{2,0}. \tag{S2.35}$$

We note

$$t_1 = \Phi^{-1}\left[1 - \frac{2(2-1+1)+h^2 - h\{h^2+8-4\cdot 4/2\}^{\frac{1}{2}}}{4(h^2+2)}\right] = \Phi^{-1}\left[1 - \frac{4}{4(h^2+2)}\right].$$

And

$$q_{1,0} = \mathbb{P}(S(t_1)=0) = [1-2(1-\Phi(t_1))]^2 = \left[1 - \frac{2}{h^2+2}\right]^2.$$

Also

$$q_{1,1} = \mathbb{P}(S(t_1)=1) = \frac{4}{h^2+2}(1 - \frac{2}{h^2+2}).$$

Hence $q_{1,1} + q_{1,0} = \left(1 - \frac{2}{h^2+2}\right)\left(1 + \frac{2}{h^2+2}\right)$. Further more,

$$t_2 = \Phi^{-1}\left[1 - \frac{2+h^2 - h\{h^2+4-4/2\}^{\frac{1}{2}}}{4(h^2+2)}\right] = \Phi^{-1}\left[\frac{3}{4} + \frac{h}{4\sqrt{h^2+2}}\right].$$

Then we have $\bar{\Phi}(t_2) = \frac{1}{4} - \frac{h}{4\sqrt{h^2+2}}$, also $\bar{\Phi}(t_1) = \frac{1}{h^2+2}$. Hence

$$\begin{aligned} q_{2,0} &= \sum_{m=0}^{1} \mathrm{I}_{\{0\leqslant m\}} \binom{m}{0} \left[\frac{\bar{\Phi}(t_2)}{\bar{\Phi}(t_1)}\right]^0 \left[1 - \frac{\bar{\Phi}(t_2)}{\bar{\Phi}(t_1)}\right]^m \cdot \frac{q_{1,m}}{q_{1,0}+q_{1,1}} \\ &= I_1 + I_2, \end{aligned}$$

where

$$\begin{aligned} I_1 &= \frac{q_{1,0}}{q_{1,0}+q_{1,1}} = \frac{1 - \frac{2}{h^2+2}}{1 + \frac{2}{h^2+2}} \\ I_2 &= \left[1 - \frac{h^2+2-h\sqrt{h^2+2}}{4}\right] \frac{q_{1,1}}{q_{1,0}+q_{1,1}} = \left[\frac{1}{2} - \frac{h^2}{4} + \frac{h\sqrt{h^2+2}}{4}\right] \frac{\frac{4}{h^2+2}}{1 + \frac{2}{h^2+2}}. \end{aligned}$$

Hence

$$I_1 + I_2 = \frac{1 + \frac{2}{(h^2+2)(\sqrt{h^2+2}+h)}}{1 + \frac{2}{h^2+2}}.$$

By plugging in all the quantities into (S2.35), we have,

$$\bar{F}_{U_{\mathrm{HC}}}(h) = 1 - \left(1 - \frac{2}{h^2+2}\right)\left(1 + \frac{2}{h^2+2}\right)\frac{1 + \frac{2}{(h^2+2)(\sqrt{h^2+2}+h)}}{1 + \frac{2}{h^2+2}}$$
$$= \frac{2}{h^2+2} - \frac{2}{(h^2+2)(\sqrt{h^2+2}+h)} + \frac{4}{(h^2+2)^2(\sqrt{h^2+2}+h))}. \qquad \text{(S2.36)}$$

Recall

$$T_{\mathrm{HC}} = \max_{1\leqslant i\leqslant 2} \sqrt{2}\frac{\frac{i}{2} - p_{(i)}}{\sqrt{p_{(i)}(1-p_{(i)})}} = \max_{1\leqslant i\leqslant 2} \frac{i}{\sqrt{2p_{(i)}(1-p_{(i)})}} - \sqrt{\frac{p_{(i)}}{1-p_{(i)}}}.$$

Under the alternative, note $T_{\mathrm{HC}}/(\sqrt{2}\exp(nc_0/4)) \to 1$ with probability one given $c_1(\theta_1) = c_2(\theta_2) = c_0 > 0$. Plugging into (S2.36), we have under the alternative in Proposition S1,

$$-\frac{2}{n}\log \bar{F}_{U_{\mathrm{HC}}}(T_{\mathrm{HC}}) \to c_0$$

with probability one as $n \to \infty$. □

Remark S6. One can note that under the alternative of combining two p-values with $c_1(\theta_1) = 2c_2(\theta_2) = 2c_0 > 0$, $\hat{i} = \mathrm{argmax}_i \sqrt{2}\frac{\frac{i}{2} - p_{(i)}}{\sqrt{p_{(i)}(1-p_{(i)})}} \to 1$ with probability one. Hence, HC is not consistent for selecting the subset of p-values with true signals.

S3 Supplementary simulation results

S3.1 Type I error control of FE and $\mathrm{FE}_{\mathrm{CS}}$

In this subsection, we numerically evaluate accuracy of type I error control using fast algorithm of independent Cauchy for the two methods proposed in Sections 4 and 5, FE and $\mathrm{FE}_{\mathrm{CS}}$. We simulate K p-values $p_1, \ldots, p_K \overset{D}{\sim} \mathrm{Unif}(0,1)$, and calculate the test statistics for the two methods respectively, where $1 - p_1, \ldots, 1 - p_K$ with the previously generated

Table S1: Accuracy of type I error control for FE and FE_{CS}

Methods	K	0.05	0.01	5×10^{-3}	1×10^{-3}
FE	5	$5.02{\times}10^{-2}$	$1.02{\times}10^{-2}$	$5.23{\times}10^{-3}$	$1.07{\times}10^{-3}$
	10	$5.12{\times}10^{-2}$	$9.96{\times}10^{-3}$	$5.05{\times}10^{-3}$	$1.17{\times}10^{-3}$
	20	$5.12{\times}10^{-2}$	$1.02{\times}10^{-2}$	$4.90{\times}10^{-3}$	$9.40{\times}10^{-4}$
	40	$5.11{\times}10^{-2}$	$9.80{\times}10^{-3}$	$5.15{\times}10^{-3}$	$1.02{\times}10^{-3}$
	60	$5.15{\times}10^{-2}$	$1.01{\times}10^{-2}$	$5.13{\times}10^{-3}$	$1.17{\times}10^{-3}$
	80	$5.31{\times}10^{-2}$	$1.10{\times}10^{-2}$	$5.72{\times}10^{-3}$	$1.05{\times}10^{-3}$
	100	$5.36{\times}10^{-2}$	$1.06{\times}10^{-2}$	$5.37{\times}10^{-3}$	$1.04{\times}10^{-3}$
FE_{CS}	5	$5.39{\times}10^{-2}$	$1.03{\times}10^{-2}$	$5.16{\times}10^{-3}$	$1.02{\times}10^{-3}$
	10	$5.51{\times}10^{-2}$	$1.02{\times}10^{-2}$	$5.23{\times}10^{-3}$	$1.15{\times}10^{-3}$
	20	$5.50{\times}10^{-2}$	$1.01{\times}10^{-2}$	$4.95{\times}10^{-3}$	$9.30{\times}10^{-4}$
	40	$5.52{\times}10^{-2}$	$1.06{\times}10^{-2}$	$5.02{\times}10^{-3}$	$9.80{\times}10^{-4}$
	60	$5.35{\times}10^{-2}$	$1.04{\times}10^{-2}$	$5.55{\times}10^{-3}$	$1.13{\times}10^{-3}$
	80	$5.78{\times}10^{-2}$	$1.13{\times}10^{-2}$	$5.25{\times}10^{-3}$	$9.90{\times}10^{-4}$
	100	$5.70{\times}10^{-2}$	$1.15{\times}10^{-2}$	$5.77{\times}10^{-3}$	$1.17{\times}10^{-3}$

p-values are used as one-sided p-values for FE_{CS}. We vary $K = 5, 10, 20, 40, 60, 80, 100$ for a wide range of numbers of combined p-values. Table S1 shows type I error control for the two methods under different significance levels α =0.05, 0.01, 0.001, 0.005, 0.001 using 10^5 times of simulations. Across wide ranges of K and $\alpha \leqslant 0.01$, type I error by the fast computing has less than 10% inflation, with improved accuracy for smaller α. As the worst case, the type I error control of FE_{CS} when $\alpha = 0.05$ is slightly anti-conservative but acceptable (in the range of 0.0539∼0.0578 for different K).

S3.2 Statistical power comparison for modified Fisher methods in the case of combining a small group of strong signals

In this subsection, we demonstrate the statistical power of Stouffer, Fisher, and 5 modified Fisher methods for combining a small group of strong signals. We simulate the alternatives with fixed numbers of true signals $\ell = 1, 2, \ldots, 6$ for $K = 20, 40, 80$ following the same simulation scheme in Section 3.3. For a given K and ℓ, we choose the smallest μ_0 such that the best method has at least 0.9 statistical power at $\alpha = 0.05$. The results are shown in Figure S1.

S3.3 Statistical power comparison for 12 existing p-value combination methods

In this subsection, we demonstrate the statistical power of 12 p-value combination methods: Fisher, AFp, AFz, oTFsoft, oTFhard, HC, minP, HM, BJ, Cauchy (CA), and Stouffer. For Figure S2, the signal strength μ_0 is chosen the same as Figure 1 in Section 3.3 for a given proportion of signals ℓ/K and number of combined p-values K. As expected, 4 added methods (HC, minP, HM, CA) that are designed for sparse signals and have very weak power for frequent signals. Although BJ is also designed for sparse signal scenarios, it has relatively higher power, comparable to AFz but much lower than Fisher and AFp. For Figure S3, the signal strength μ_0 is chosen the same as Figure S1 for a given number of signals ℓ and number of combined p-values K. 4 added methods (HC, minP, HM, CA) that are designed for sparse signals outperform Fisher and Stouffer, but still are comparative with modified Fisher's methods such as AFp and AFz.

S3.4 Statistical power comparison for FE in the case of combining a small group of strong signals

In this subsection, we demonstrate the statistical power of Fisher, AFp, and FE for combining a small group of strong signals. We simulate the alternatives with fixed numbers of true signals $\ell = 1, 2, \ldots, 6$ for $K = 20, 40, 80$ following the same simulation scheme in Section

3.3. For a given K and ℓ, we choose the smallest μ_0 such that the best method has at least 0.9 statistical power at $\alpha = 0.05$. The results are shown in Figure S4.

S3.5 Statistical power comparison for FE and FE2

In this subsection, we evaluate the statistical power of Fisher, AFp, FE, and the following FE2 that integrates Fisher, AFp and minP:

$$T_{\text{FE2}} = [1/p^{\text{Fisher}} + 1/p^{\text{AFp}} + 1/p^{\text{minP}}]/3.$$

The following Figures S5 and S6 present the results in settings similar to that of Figures 2 and S4, respectively. For Figure S5, we choose the smallest μ_0 that allows the best method to have power larger than 0.5 for a given proportion of signals ℓ/K and a number of combined p-values K. For Figure S6, we choose the smallest μ_0 that allows the best method to have power larger than 0.5 for a given proportion of signals ℓ and a number of combined p-values K. Although FE2 improves power over FE when the signal is very sparse, its power is much reduced when the signal is frequent, which is an important scenario in most applications. As a result, FE combining Fisher and AFp but not minP is recommended for general applications.

S3.6 Statistical power comparison for FE_{CS} in the case of combining a small group of strong signals

In this subsection, we demonstrate the statistical power of Pearson, FE, and FE_{CS} for combining a small group of strong signals. We simulate the alternatives with fixed numbers of true signals $\ell = 1, 2, \ldots, 6$ for $K = 20, 40, 80$ following the same simulation scheme in

Section 3.3. For a given K and ℓ, we choose the smallest μ_0 such that the best method has at least 0.9 statistical power at $\alpha = 0.05$. The results are shown in Figure S7.

S3.7 Numeric examples that harmonic mean outperforms Cauchy for Fisher ensemble

This subsection provides numeric examples that using harmonic mean is better than Cauchy in the FE and FE_{CS} construction (Equation (2) in the manuscript). Below, we follow the simulation scheme in Section 5.2 to generate data and the combined p-values, where we evaluate the power of FE_{CS} (using the harmonic mean), Pearson, and $\text{FE}_{\text{CS}}^{\text{Cauchy}}$ (using Cauchy). Figures S8 and S9 show the empirical power of the three methods. For figure S8, we choose the smallest μ_0 that allows the best method to have power larger than 0.5 at significance level $\alpha = 0.01$ for a given proportion of signals ℓ/K and a number of combined p-values K. For figure S9, we choose the smallest μ_0 that allows the best method to have power larger than 0.9 at significance level $\alpha = 0.05$ for a given proportion of signals ℓ and a number of combined p-values K. The results show that FE_{CS} largely outperforms the latter for $\ell/K \geqslant 0.4$ in Figure S8, as a consequence of the "$-\infty$ score" issue when using Cauchy.

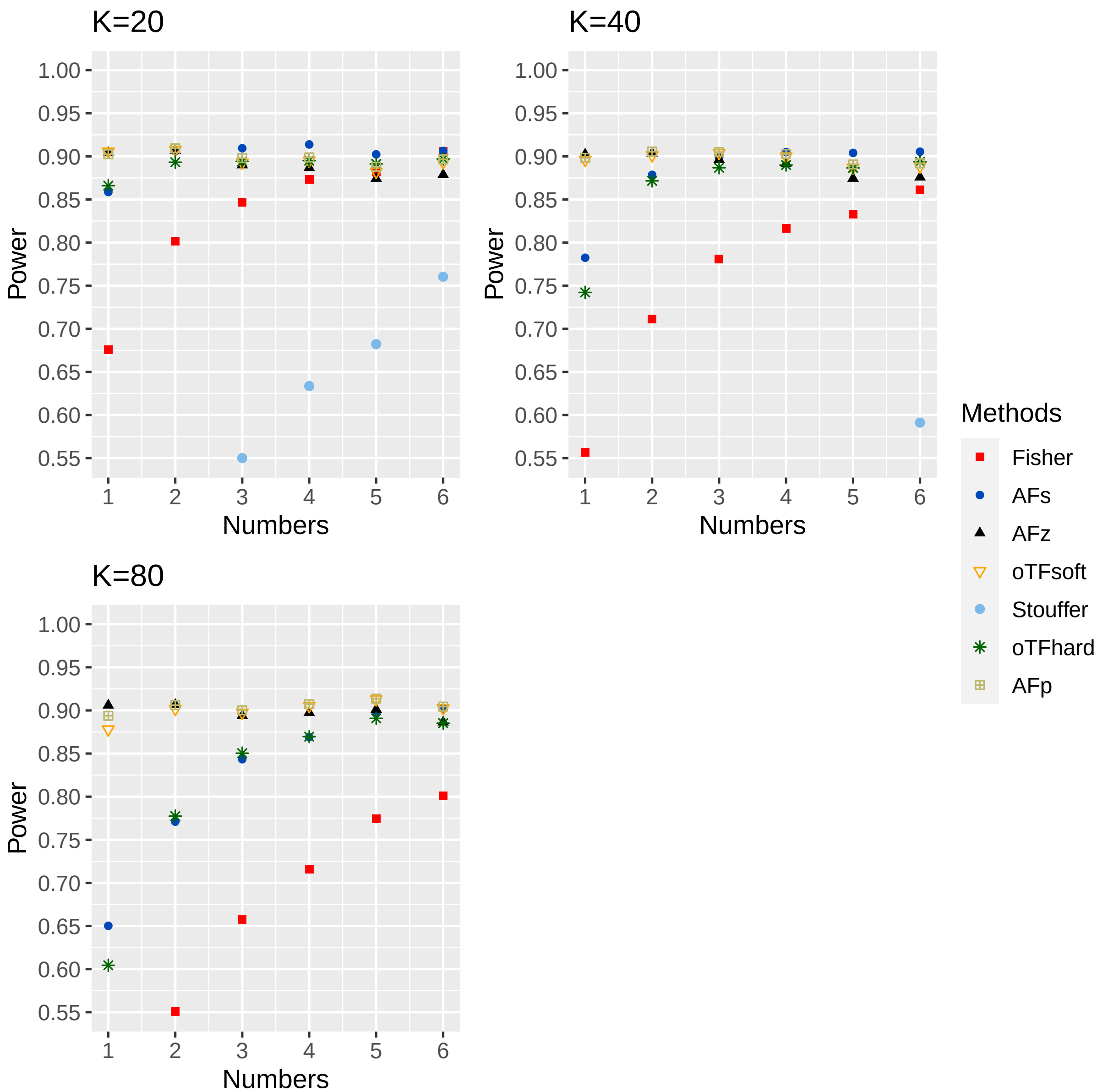

Figure S1: Statistical power of Fisher, Stouffer, and 5 modified Fisher's methods at significance level $\alpha = 0.05$ across varying numbers of true signals $\ell = 1, 2, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.9 statistical power. The standard errors are negligible compared to the scale of the mean power (smaller than 0.1% of the power) and hence omitted. The results of Stouffer and Fisher with a power smaller than 0.55 are omitted.

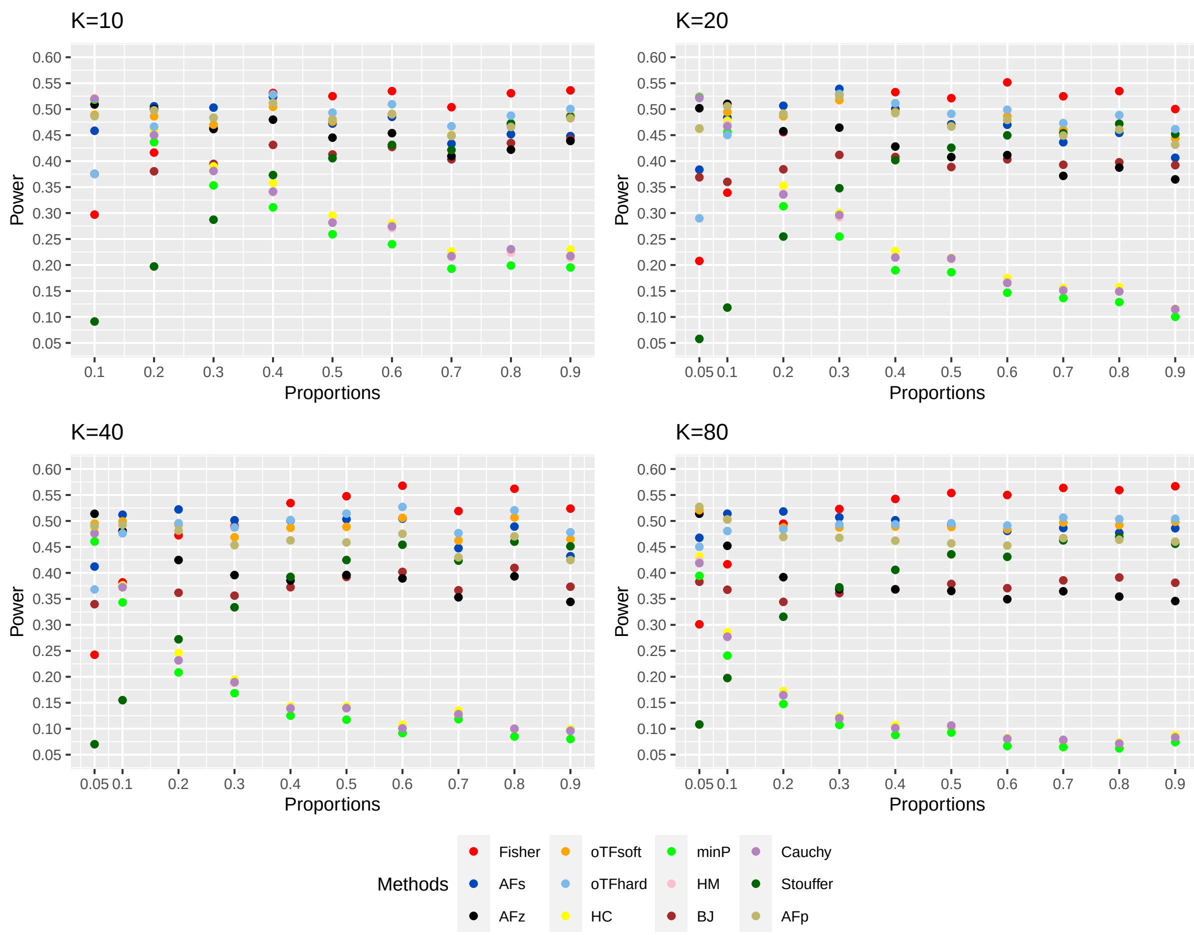

Figure S2: Statistical power of Fisher, AFs, AFp, AFz, oTFsoft, oTFhard, HC, minP, HM, BJ, Cauchy (CA), and Stouffer at significance level $\alpha = 0.01$ across varying proportions of signals $\ell/K = 0.05, 0.1, 0.2, \ldots, 0.9$ and varying numbers of combined p-values $K = 10, 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.5 statistical power. The standard errors are negligible and hence omitted.

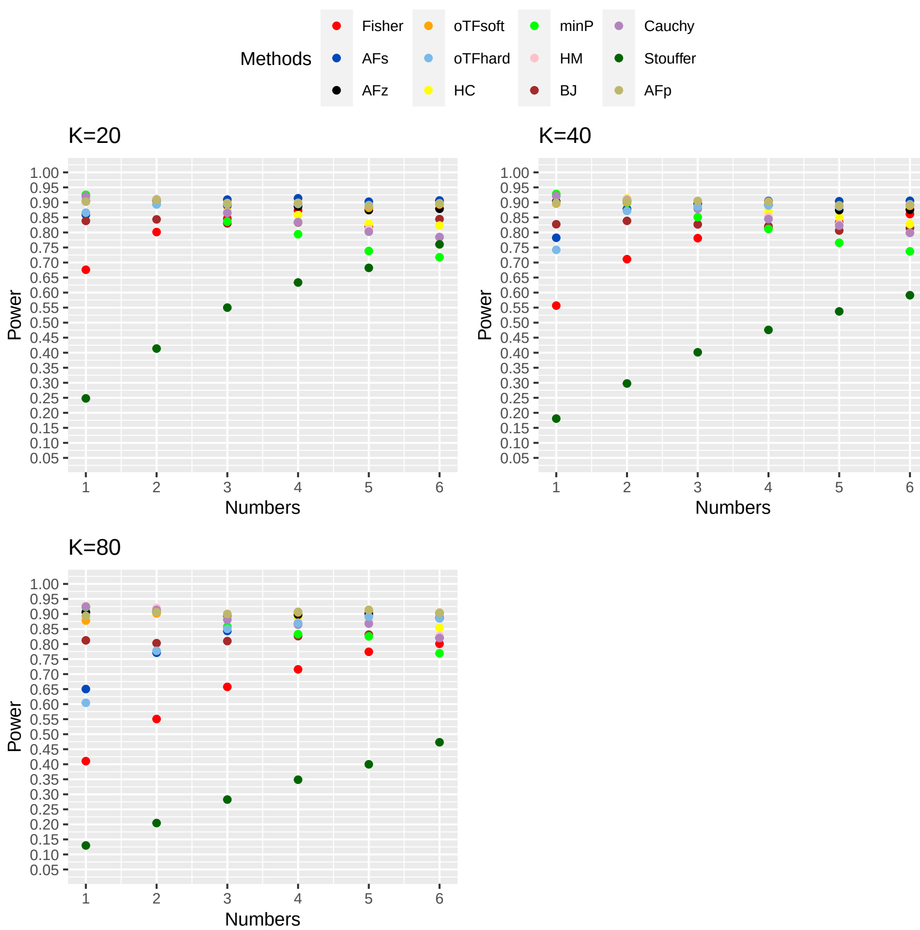

Figure S3: Statistical power of Fisher, AFs, AFp, AFz, oTFsoft, oTFhard, HC, minP, HM, BJ, Cauchy (CA), and Stouffer at significance level $\alpha = 0.05$ across varying numbers of signals $\ell = 1, 2, 3, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.9 statistical power. The standard errors are negligible and hence omitted.

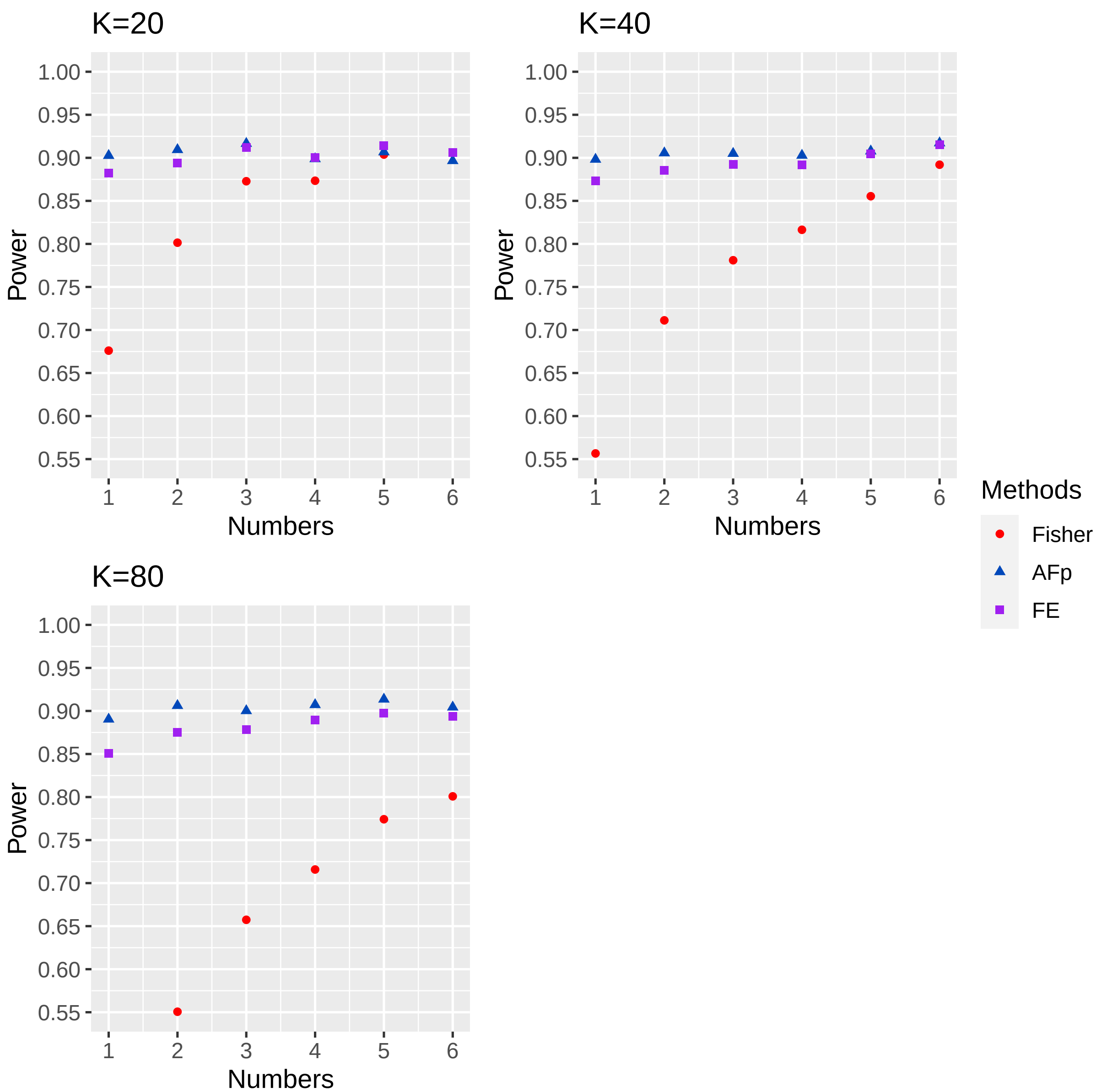

Figure S4: Statistical power of FE, Fisher, and AFp at significance level $\alpha = 0.05$ across varying numbers of signals $\ell = 1, 2, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.9 statistical power. The standard errors are negligible and hence omitted. Dots smaller than 0.55 are also omitted.

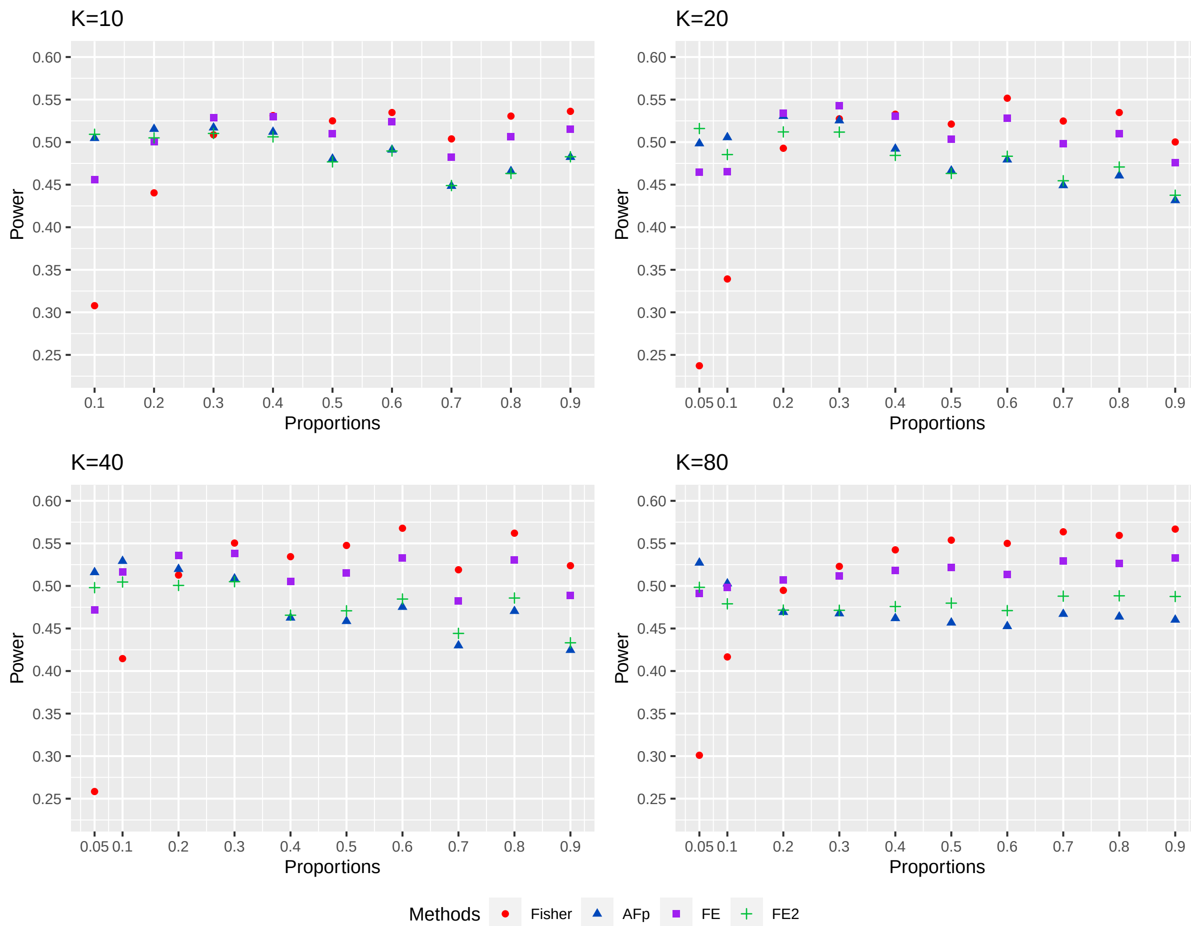

Figure S5: Statistical power of Fisher, AFp, FE, and FE2 at significance level $\alpha = 0.01$ across varying frequencies of signals $\ell/K = 0.05, 0.2, \ldots, 0.9$ and varying numbers of combined p-values $K = 10, 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.5 statistical power. The standard errors are negligible and hence omitted.

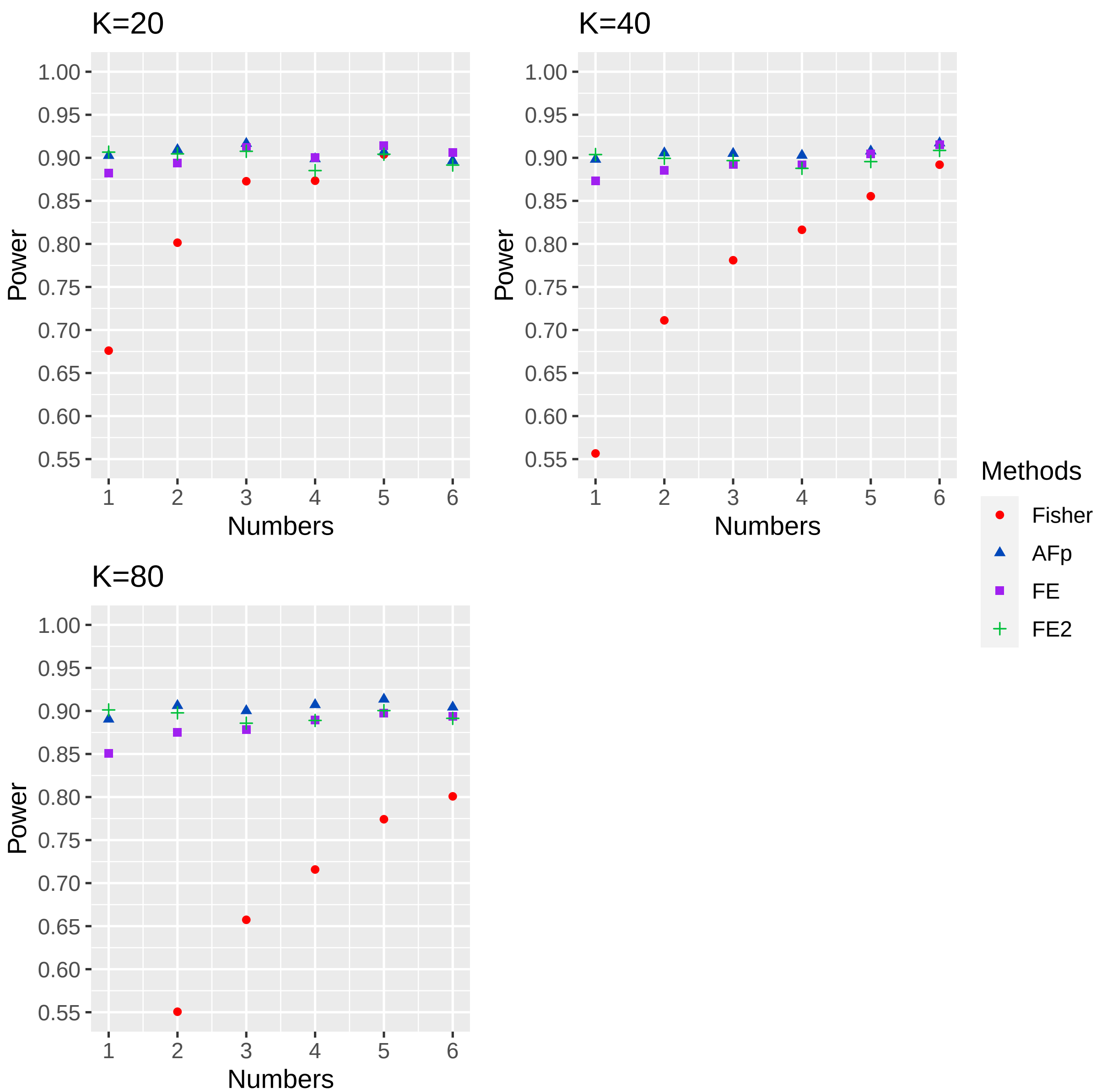

Figure S6: Statistical power of Fisher, AFp, FE, and FE2 at significance level $\alpha = 0.05$ across varying numbers of signals $\ell = 1, 2, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.9 statistical power. The standard errors are negligible and hence omitted. results of Fisher smaller than 0.55 are omitted.

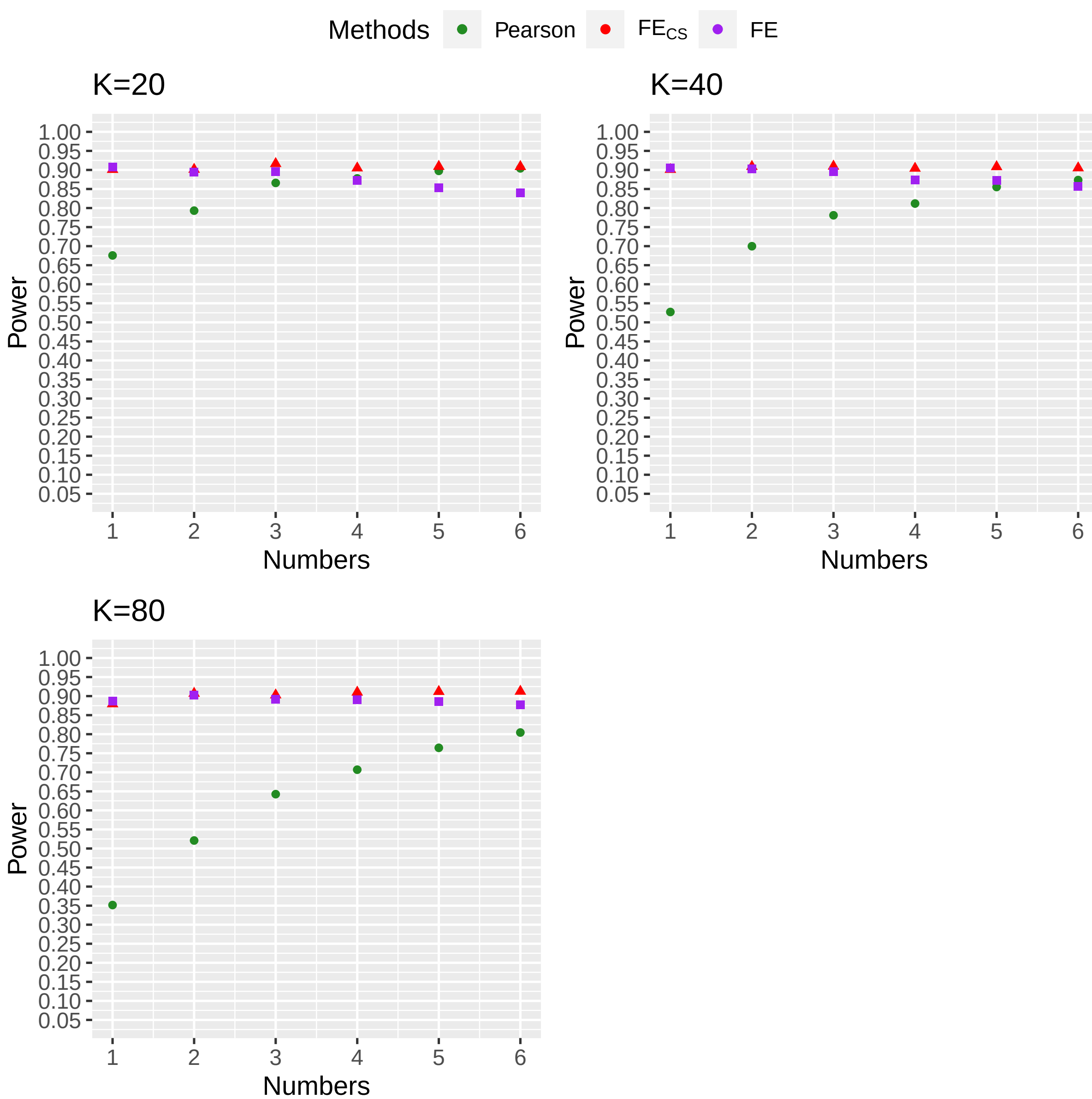

Figure S7: Statistical power of FE, FE_{CS}, and Pearson at significance level $\alpha = 0.05$ across varying numbers of signals $\ell = 1, 2, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. The standard errors are negligible and hence omitted.

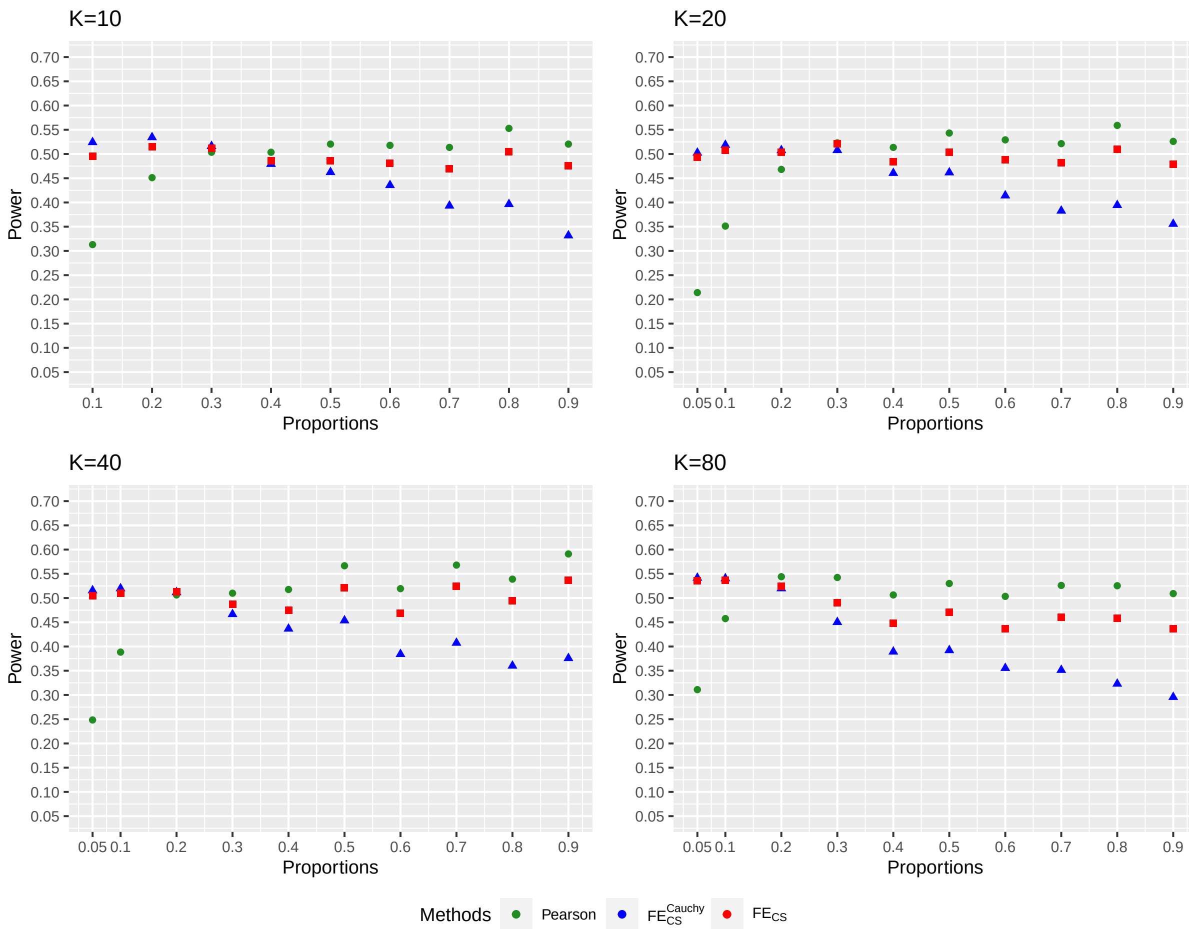

Figure S8: Statistical power of $\mathrm{FE}_{\mathrm{CS}}$, $\mathrm{FE}_{\mathrm{CS}}^{\mathrm{Cauchy}}$, and Pearson at significance level $\alpha = 0.01$ across varying frequencies of signals $\ell/K = 0.05, 0.1, 0.2, \ldots, 0.9$ and varying numbers of combined p-values $K = 10, 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.5 statistical power. The standard errors are negligible and hence omitted.

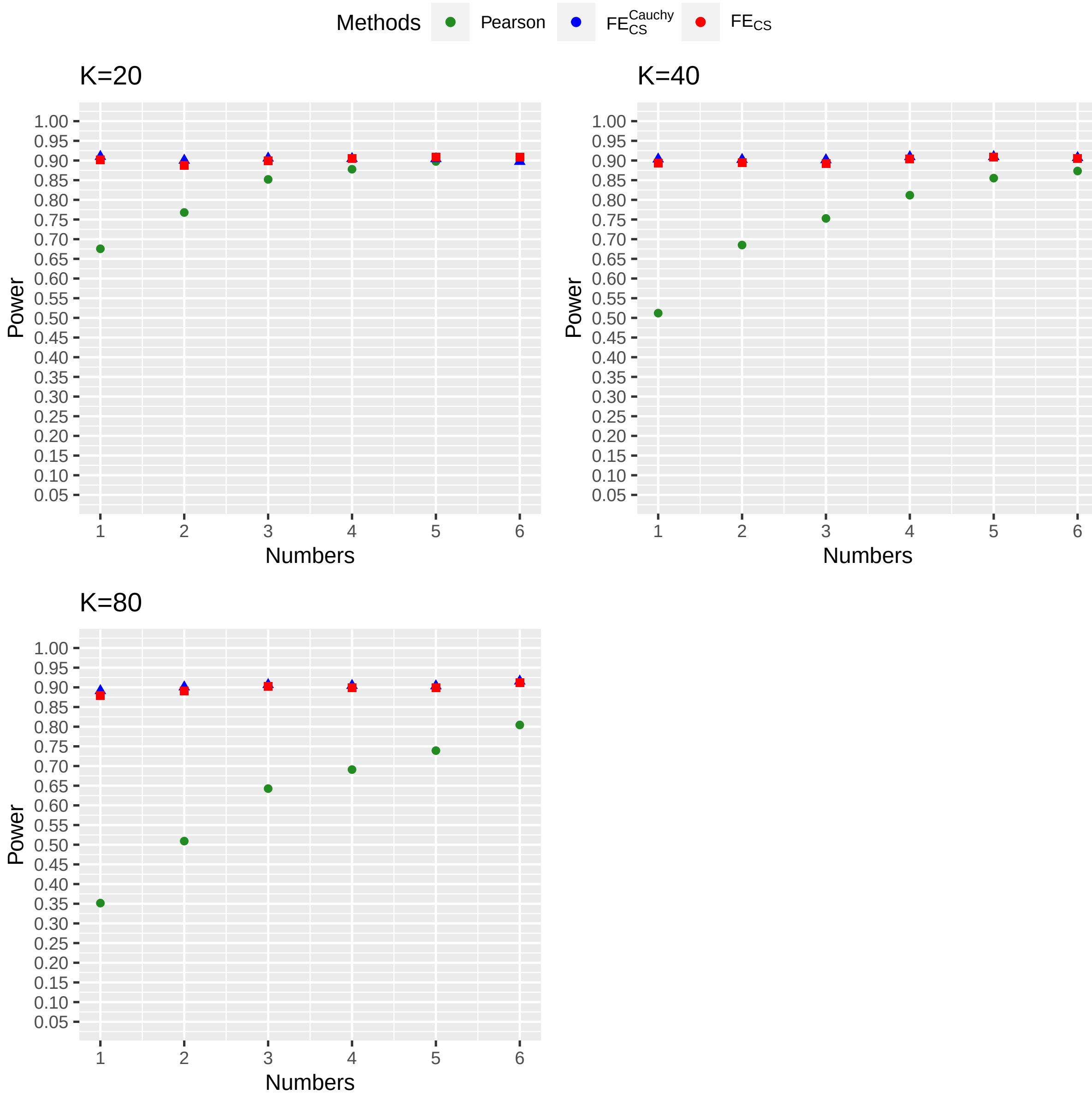

Figure S9: Statistical power of $\mathrm{FE}_{\mathrm{CS}}$, $\mathrm{FE}_{\mathrm{CS}}^{\mathrm{Cauchy}}$, and Pearson at significance level $\alpha = 0.05$ across varying numbers of signals $\ell = 1, 2, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.9 statistical power. The standard errors are negligible and hence omitted.

S4 Supplementary real application results

This section contains supplementary table and figures for the AGEMAP application result.

Table S2: Up-regulated/down-regulated age-related pathways detected in one-sided design by FE_{CS} with significance level $p \leqslant 0.01$. The reference columns of the 2 tables lists literature that supports the relationships between the pathways and aging/early development processes.

(a): Pathways by up-regulated genes

Pathways	p-values	References
Phagosome Maturation	0.0005	Vieira et al. (2002)
Glutathione Redox Reactions I	0.00085	Mandal et al. (2015); Erden-İnal et al. (2002)
Tryptophan Degradation III (Eukaryotic)	0.0006	Van der Goot and Nollen (2013)
FAT10 Cancer Signaling Pathway	0.0041	Canaan et al. (2014); Aichem and Groettrup (2016)
Isoleucine Degradation I	0.0058	Canfield and Bradshaw (2019); Salcedo et al. (2021)
Glutamine Biosynthesis I	0.0065	Meynial-Denis (2016); Canfield and Bradshaw (2019)
Histamine Biosynthesis	0.0065	Mazurkiewicz-Kwilecki and Nsonwah (1989); Terao et al. (2004)
Tumor Microenvironment Pathway	0.0060	Mori et al. (2018); Sandiford et al. (2018)
Glutaryl-CoA Degradation	0.0065	Porcellini et al. (2007)
Valine Degradation I	0.0079	Canfield and Bradshaw (2019); Salcedo et al. (2021)
Androgen Signaling	0.0047	He et al. (2018); Rey (2021); Zhou et al. (2015)

(b): Pathways by down-regulated genes

Pathways	p-values	References
EIF2 Signaling	0.00001	Ma et al. (2013)
Remodeling of Epithelial Adherens Junctions	0.0019	Parrish (2017)
Tight Junction Signaling	0.00028	Parrish (2017); Ren et al. (2014)
NER (Nucleotide Excision Repair, Enhanced Pathway)	0.0087	Maynard et al. (2009)

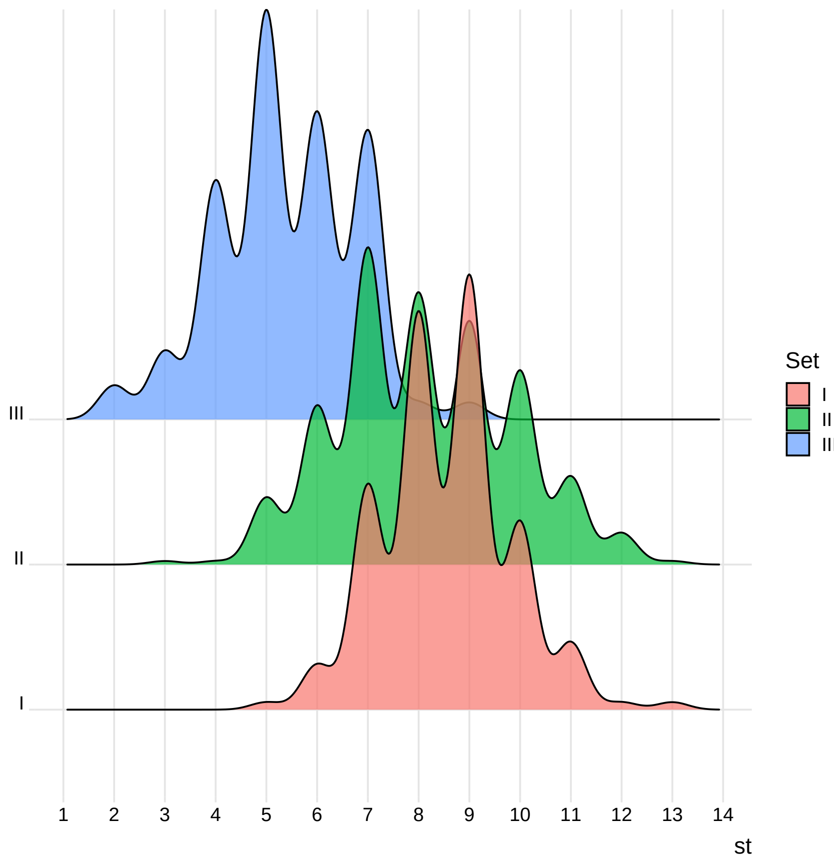

Figure S10: Distributions of numbers of p-values $p_{jk} \leqslant 0.05$ of each gene j in gene Categories I, II, and III in Figure 4.

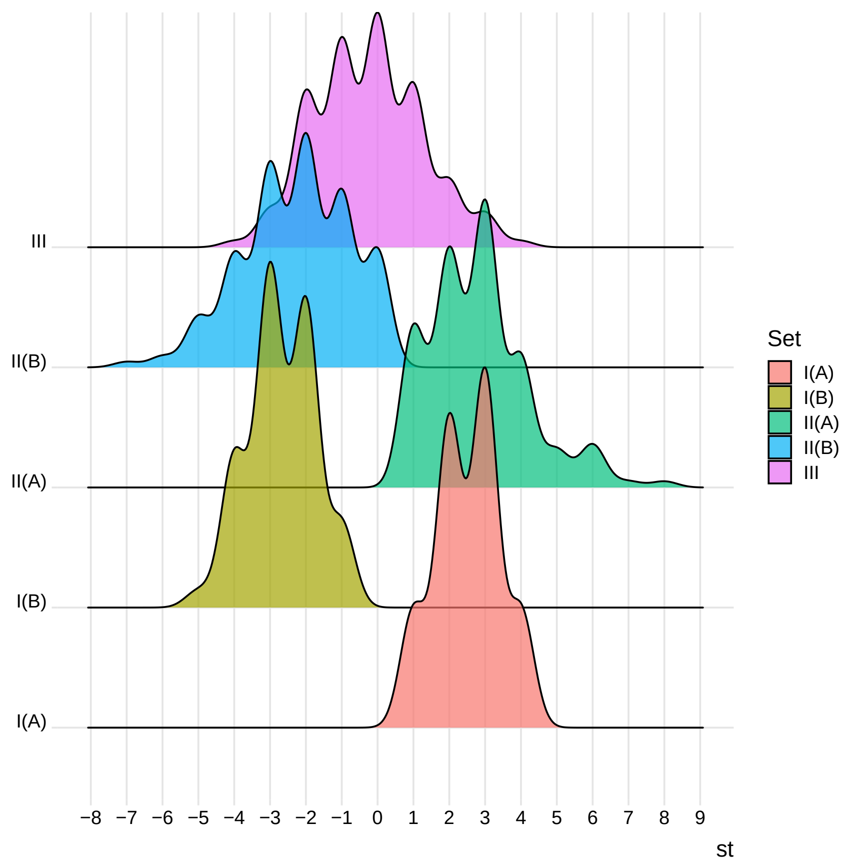

Figure S11: Distributions of quantities $S_{\text{sign},j} = \sum_{k=1}^{16} \text{sign}(\beta_{\text{age},jk}) \text{I}_{\{\min\{\tilde{p}_{jk}^{L}, \tilde{p}_{jk}^{R}\}\}}$ each gene j in gene Categories I(A), I(B), II(A), II(B), and III in Figure 5.

S5 Additional simulation results

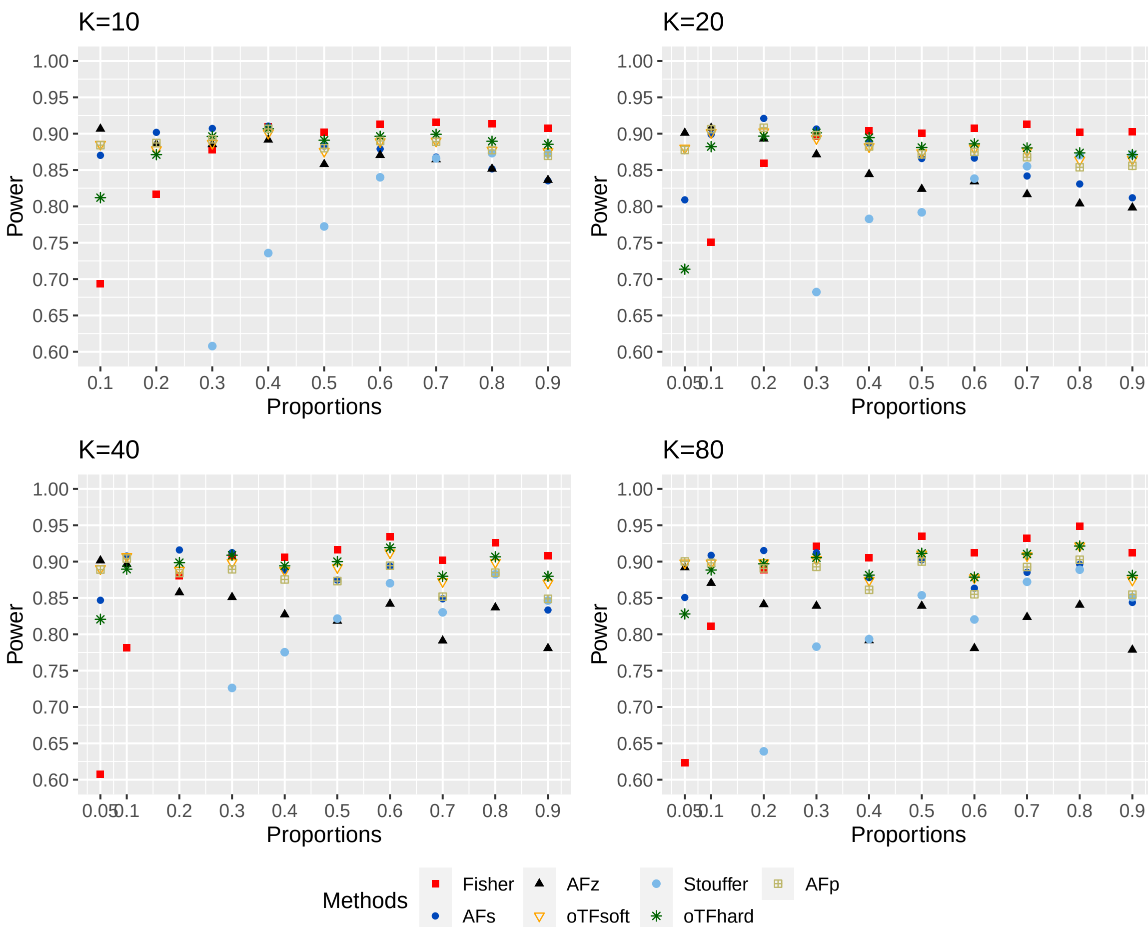

Figure S12: Statistical power of Fisher, AFs, AFp, AFz, oTFsoft, oTFhard, HC, minP, HM, BJ, Cauchy (CA), and Stouffer at significance level $\alpha = 0.01$ across varying proportions of signals $\ell/K = 0.05, 0.1, 0.2, \ldots, 0.9$ and varying numbers of combined p-values $K = 10, 20, 40, 80$. For each proportion ℓ/K and K, we choose the smallest μ_0 such that the best performer has at least 0.9 statistical power. The standard errors are negligible and hence omitted.

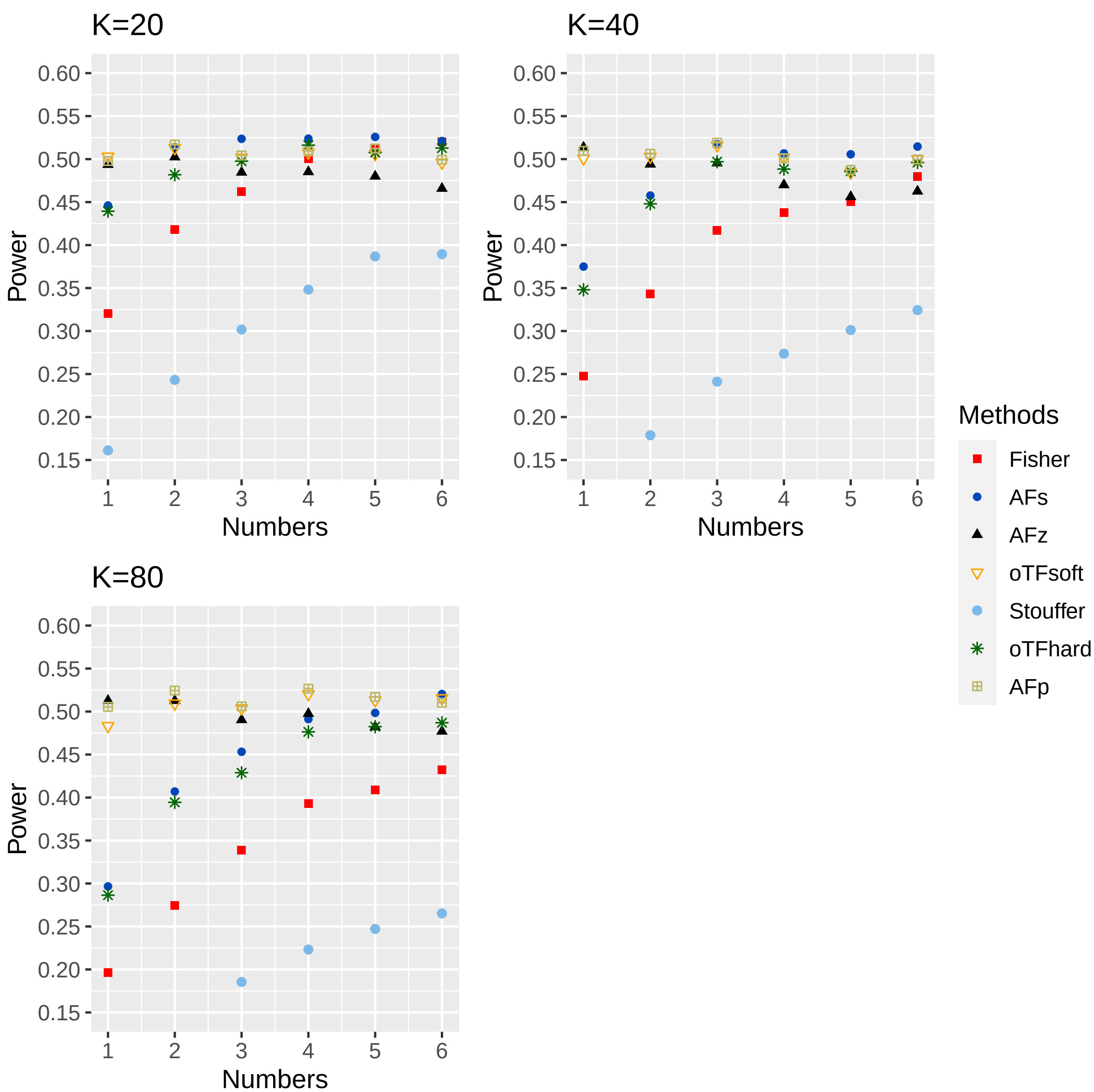

Figure S13: Statistical power of Fisher, Stouffer, and 5 modified Fisher's methods at significance level $\alpha = 0.05$ across varying numbers of true signals $\ell = 1, 2, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.5 statistical power. The standard errors are negligible compared to the scale of the mean power and hence omitted.

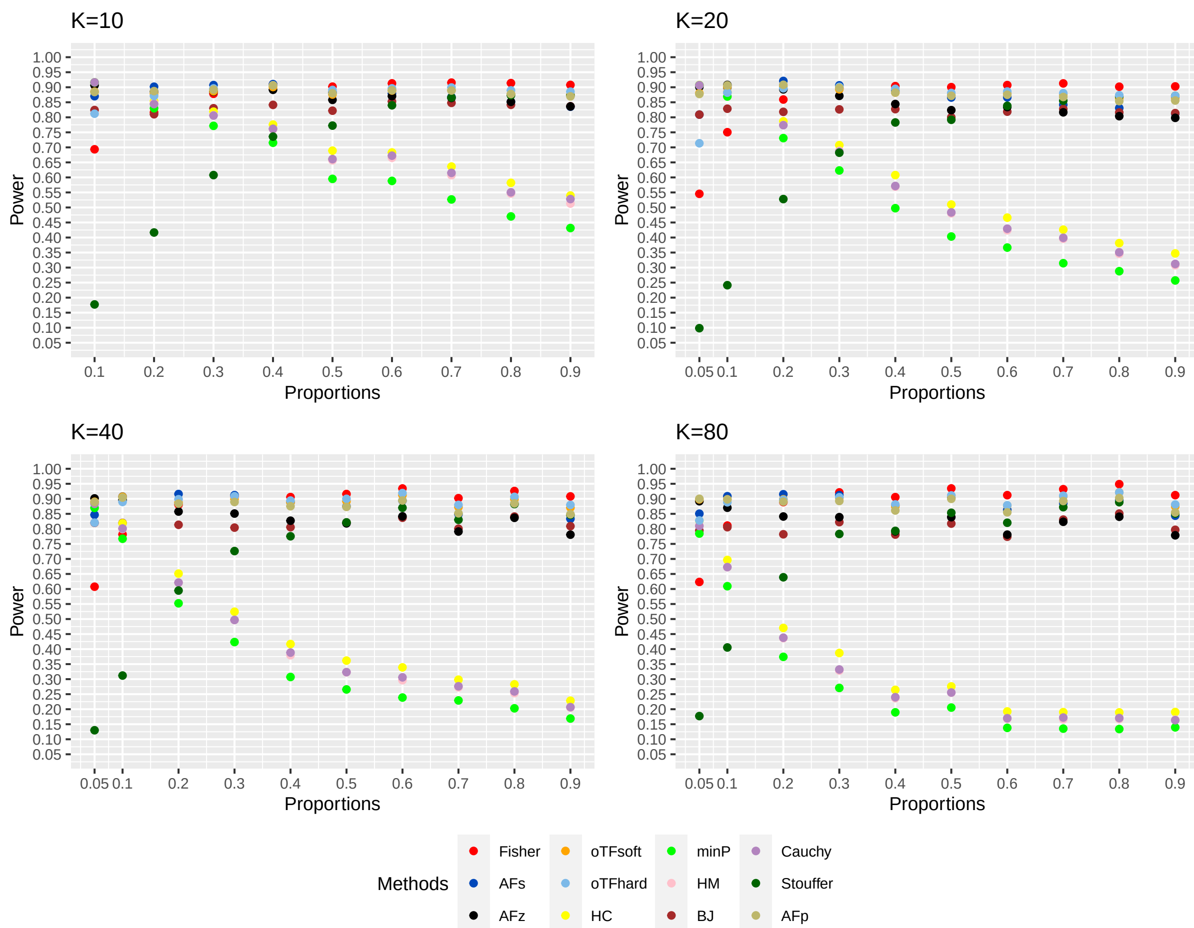

Figure S14: Statistical power of Fisher, AFs, AFp, AFz, oTFsoft, oTFhard, HC, minP, HM, BJ, Cauchy (CA), and Stouffer at significance level $\alpha = 0.01$ across varying proportions of signals $\ell/K = 0.05, 0.1, 0.2, \ldots, 0.9$ and varying numbers of combined p-values $K = 10, 20, 40, 80$. For each ℓ/K and K, we choose the smallest μ_0 such that the best performer has at least 0.9 statistical power. The standard errors are negligible and hence omitted.

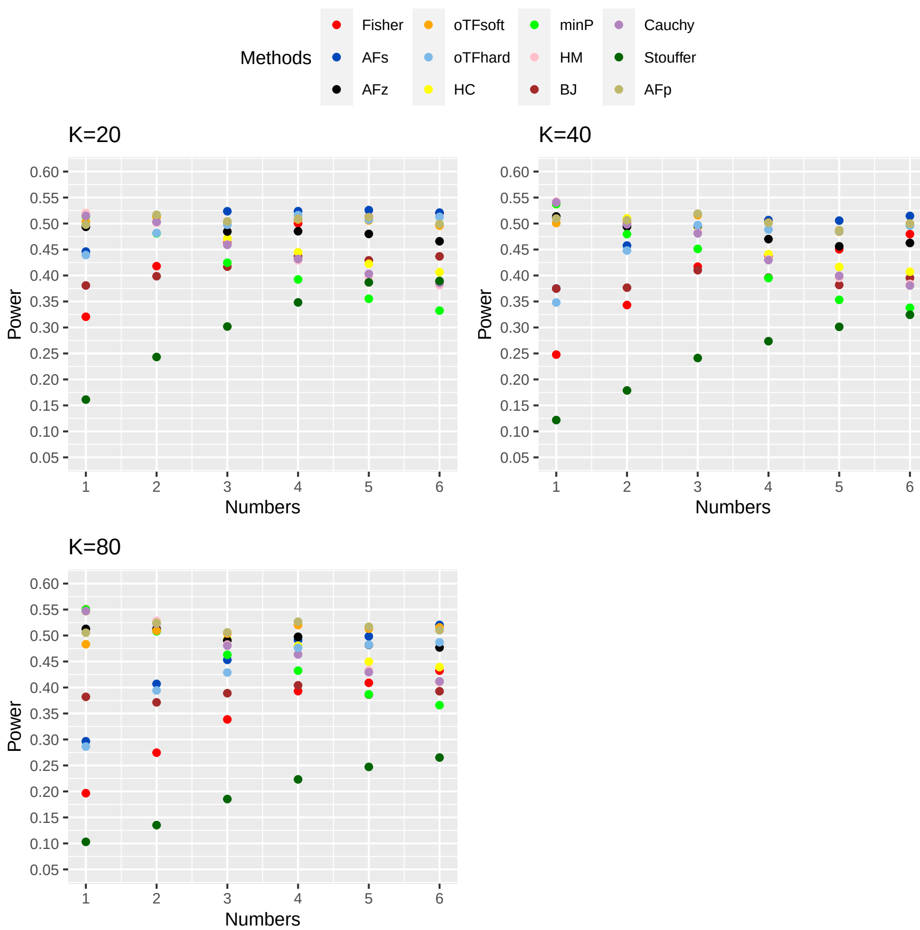

Figure S15: Statistical power of Fisher, AFs, AFp, AFz, oTFsoft, oTFhard, HC, minP, HM, BJ, Cauchy (CA), and Stouffer at significance level $\alpha = 0.05$ across varying numbers of signals $\ell = 1, 2, 3, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.5 statistical power. The standard errors are negligible and hence omitted.

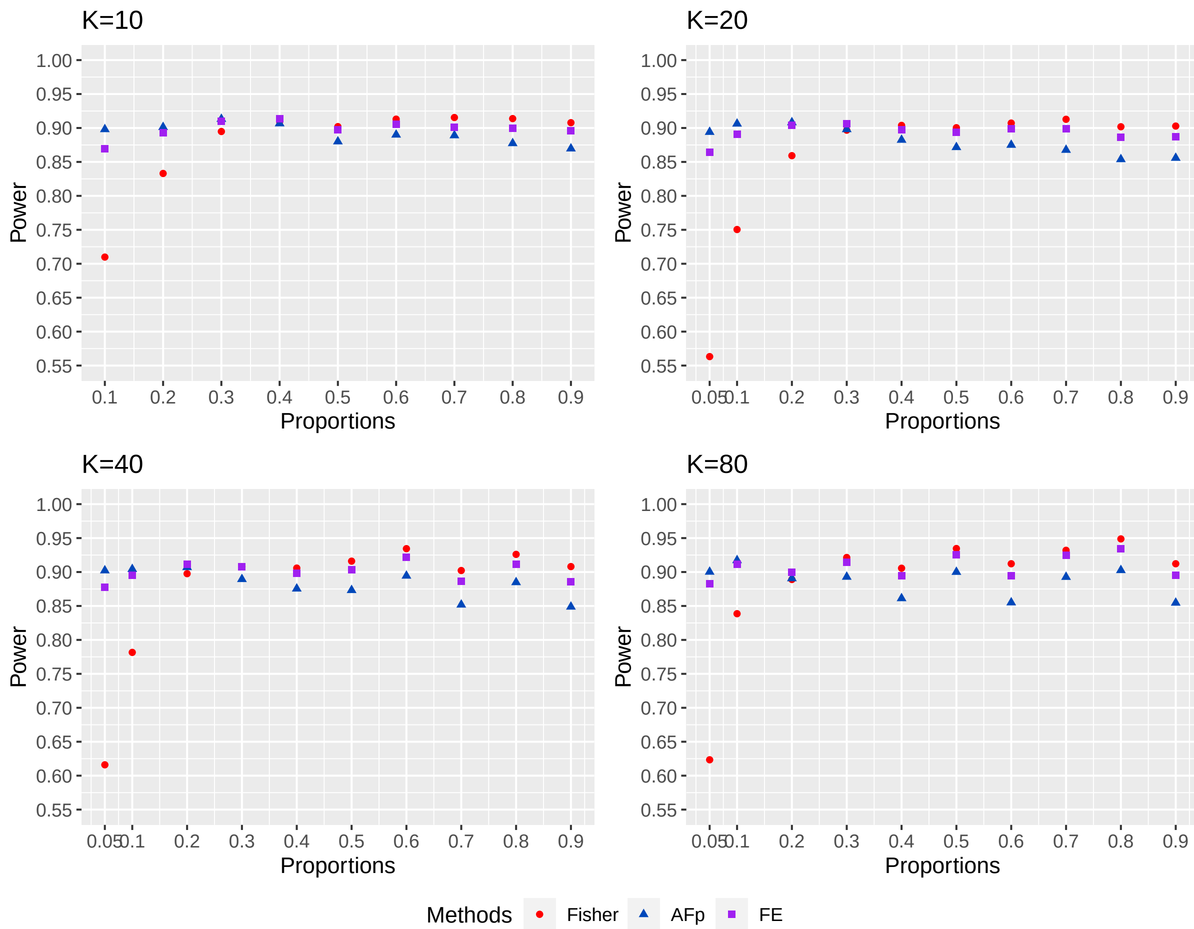

Figure S16: Statistical power of FE, Fisher, and AFp at significance level $\alpha = 0.01$ across varying proportions of signals $\ell/K = 0.05, 0.1\ldots, 0.9$ and varying numbers of combined p-values $K = 10, 20, 40, 80$. For each ℓ/K and K, we choose the smallest μ_0 such that the best performer has at least 0.9 statistical power. The standard errors are negligible and hence omitted.

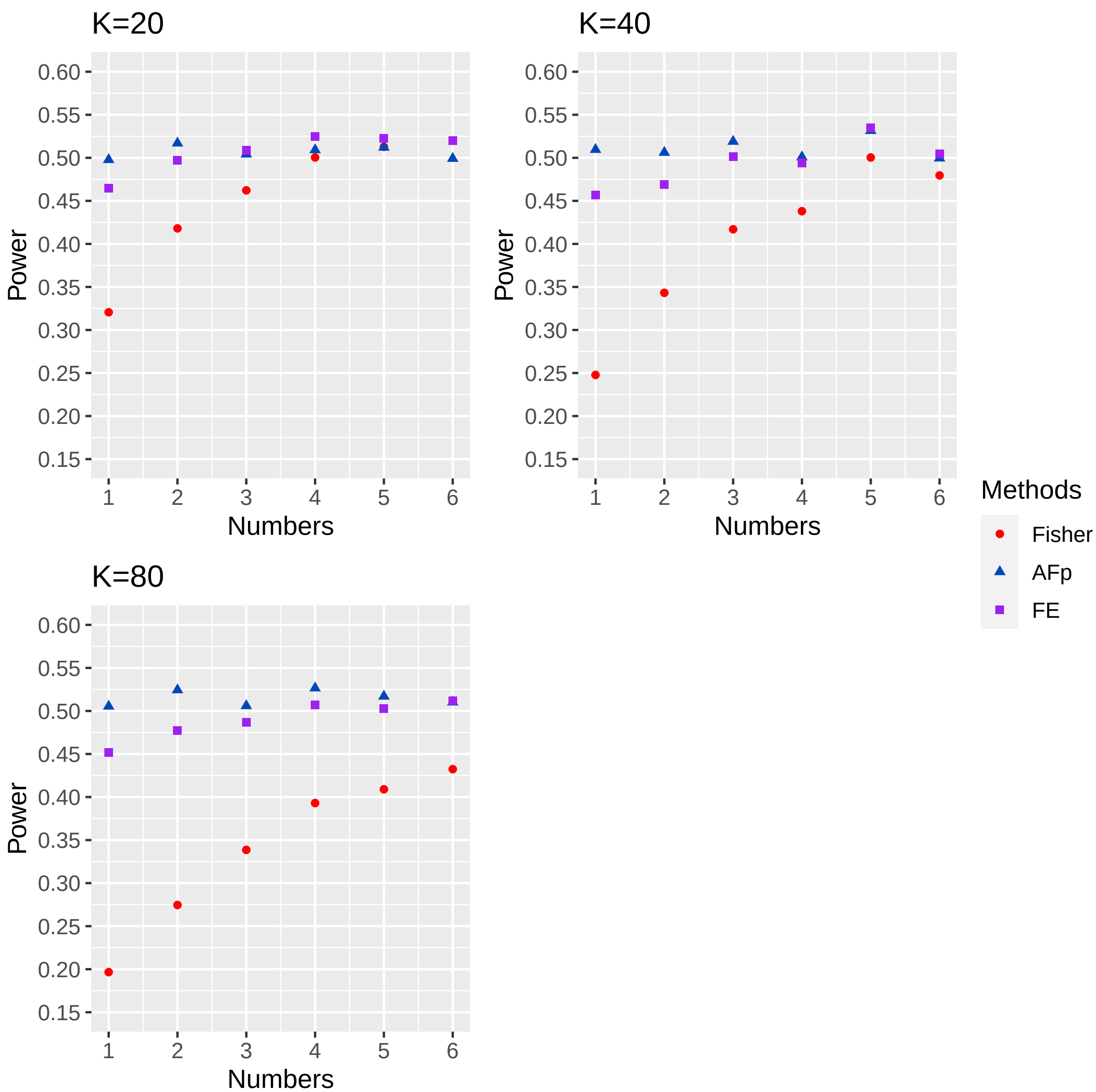

Figure S17: Statistical power of FE, Fisher, and AFp at significance level $\alpha = 0.05$ across varying numbers of signals $\ell = 1, 2, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.5 statistical power. The standard errors are negligible and hence omitted.

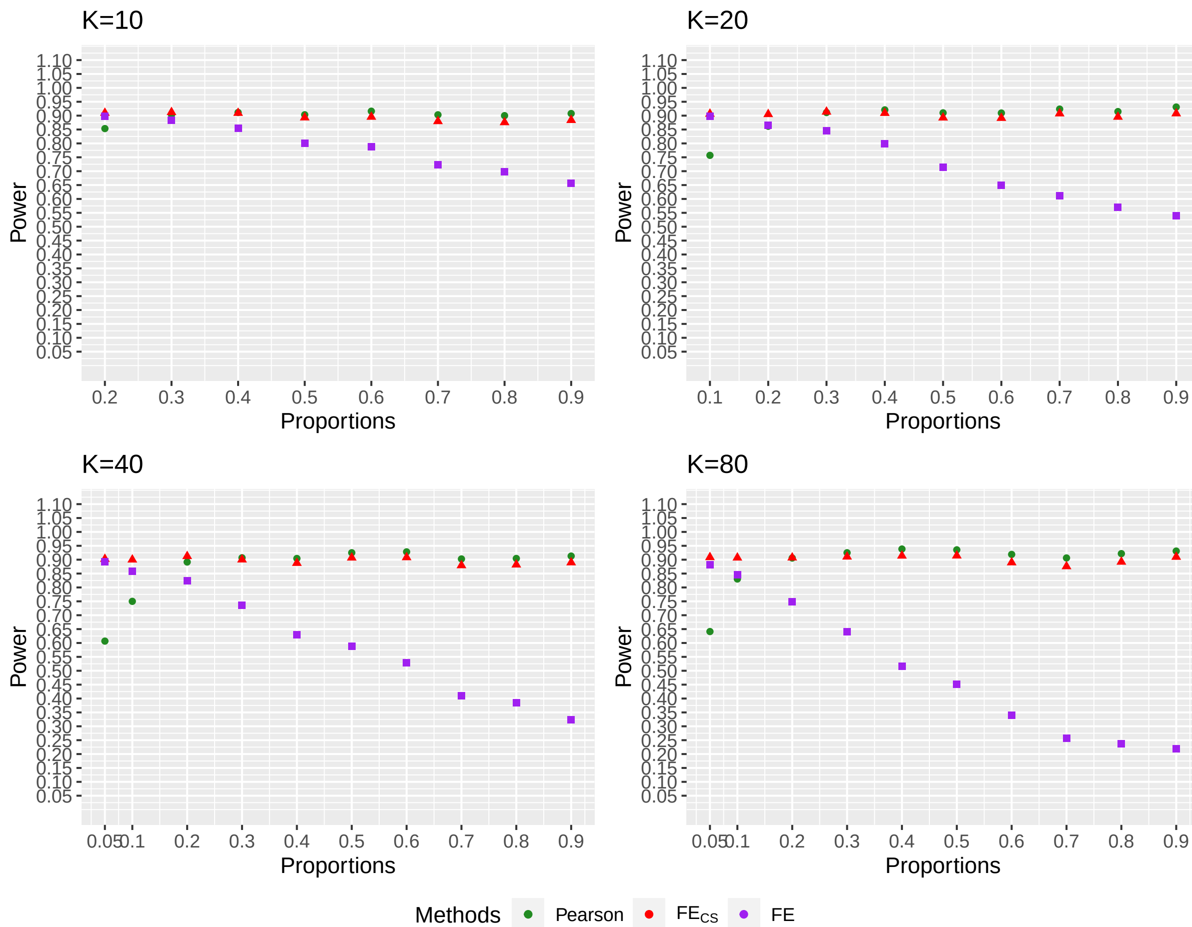

Figure S18: Statistical power of FE, FE_{CS}, and Pearson at significance level $\alpha = 0.01$ across varying proportions of signals $\ell/K = 0.05, 0.1, \ldots, 0.9$ and varying numbers of combined p-values $K = 10, 20, 40, 80$. For each ℓ/K and K, we choose the smallest μ_0 such that the best performer has at least 0.9 statistical power. The standard errors are negligible and hence omitted.

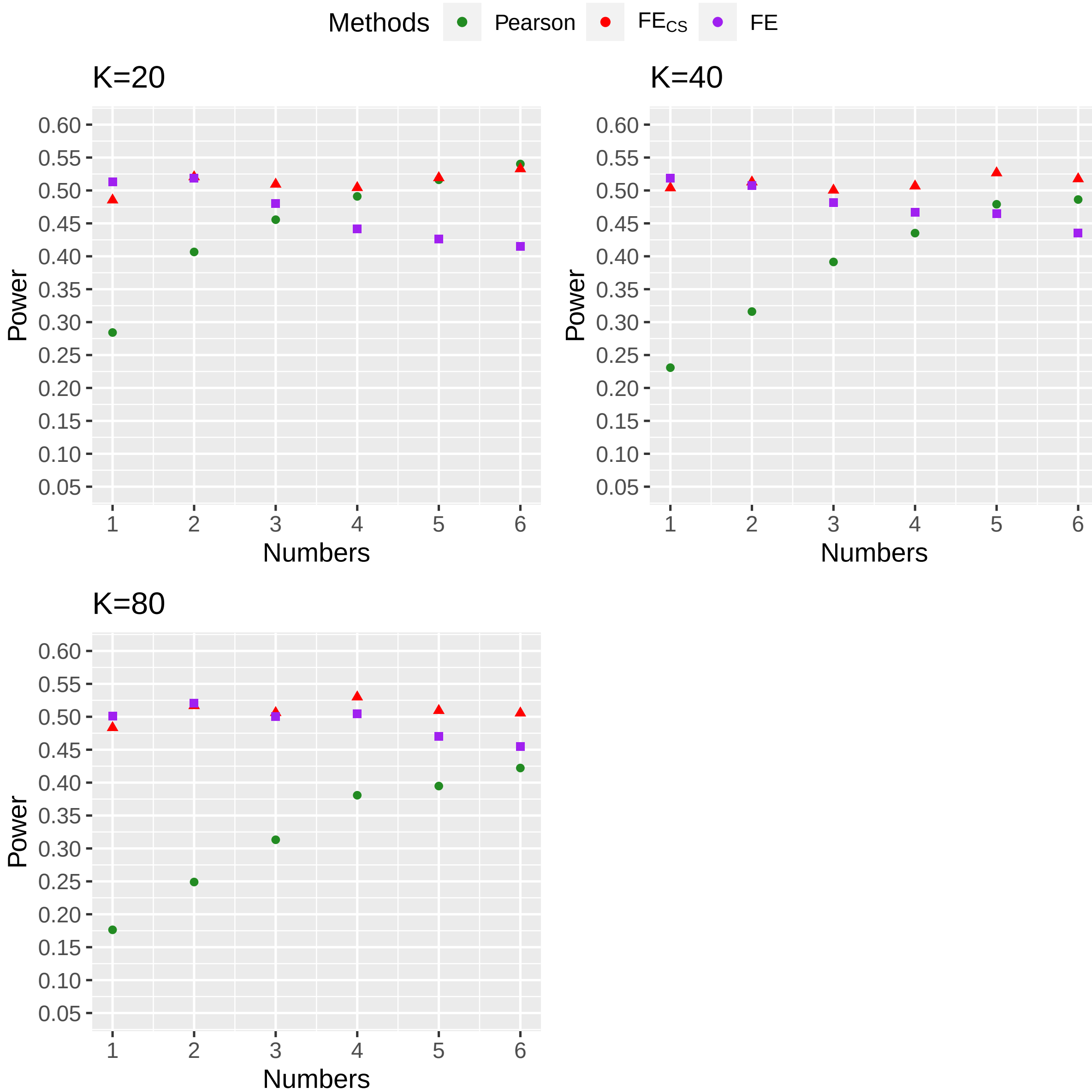

Figure S19: Statistical power of FE, FE_{CS}, and Pearson at significance level $\alpha = 0.05$ across varying numbers of signals $\ell = 1, 2, \ldots, 6$ and varying numbers of combined p-values $K = 20, 40, 80$. For each ℓ and K, we choose the smallest μ_0 such that the best performer has at least 0.5 statistical power. The standard errors are negligible and hence omitted.

References

Bahadur, R. R. et al. (1960). Stochastic comparison of tests. *The Annals of Mathematical Statistics 31*(2), 276–295.

Barnett, I. J. and X. Lin (2014). Analytical p-value calculation for the higher criticism test in finite-d problems. *Biometrika 101*(4), 964–970.

Berk, R. H. and D. H. Jones (1978). Relatively optimal combinations of test statistics. *Scandinavian Journal of Statistics*, 158–162.

Berk, R. H. and D. H. Jones (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Probability Theory and Related Fields 47*(1), 47–59.

Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics 32*(3), 962–994.

Fang, Y., G. C. Tseng, and C. Chang (2023+). Heavy-tailed distribution for combining dependent p-values with asymptotic robustness. *Statistica Sinica*. Also available as "`https://www3.stat.sinica.edu.tw/ss_newpaper/SS-2022-0046_na.pdf`".

Li, J. and D. Siegmund (2015). Higher criticism: p-values and criticism. *The Annals of Statistics 43*(3), 1323–1350.

Littell, R. C. and J. L. Folks (1973). Asymptotic optimality of Fisher's method of combining independent tests II. *Journal of the American Statistical Association 68*(341), 193–194.

Mikosch, T. (1999). *Regular variation, subexponentiality and their applications in probability theory*, Volume 99. Eindhoven University of Technology Eindhoven, The Netherlands.

Nagaraja, H. N. (2006). Order statistics from independent exponential random variables and the sum of the top order statistics. In *Advances in Distribution Theory, Order Statistics, and Inference*, pp. 173–185. Springer.

Savage, I. R. (1969). Nonparametric statistics: a personal review. *Sankhyā: The Indian Journal of Statistics, Series*

A, 107–144.

Wilson, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences 116*(4), 1195–1200.

Zhang, H., J. Jin, and Z. Wu (2020). Distributions and power of optimal signal-detection statistics in finite case. *IEEE Transactions on Signal Processing 68*, 1021–1033.

Zhang, H., T. Tong, J. Landers, and Z. Wu (2020). TFisher: A powerful truncation and weighting procedure for combining p-values. *The Annals of Applied Statistics 14*(1), 178–201.

References II: References for detected pathways

Aichem, A. and M. Groettrup (2016). The ubiquitin-like modifier FAT10 in cancer development. *The International Journal of Biochemistry and cell biology 79*, 451–461.

Canaan, A., J. DeFuria, E. Perelman, V. Schultz, M. Seay, D. Tuck, R. A. Flavell, M. P. Snyder, M. S. Obin, and S. M. Weissman (2014). Extended lifespan and reduced adiposity in mice lacking the FAT10 gene. *Proceedings of the National Academy of Sciences 111*(14), 5313–5318.

Canfield, C.-A. and P. C. Bradshaw (2019). Amino acids in the regulation of aging and aging-related diseases. *Translational Medicine of Aging 3*, 70–89.

Cotrina, M. L., Q. Gao, J. H.-C. Lin, and M. Nedergaard (2001). Expression and function of astrocytic gap junctions in aging. *Brain Research 901*(1-2), 55–61.

Daniel, C. W. and L. Young (1971). Influence of cell division on an aging process: life span of mouse mammary epithelium during serial propagation in vivo. *Experimental Cell Research 65*(1), 27–32.

Erden-İnal, M., E. Sunal, and G. Kanbak (2002). Age-related changes in the glutathione redox system. *Cell*

Biochemistry and Function 20(1), 61–66.

He, Y., E. Hooker, E.-J. Yu, H. Wu, G. R. Cunha, and Z. Sun (2018). An indispensable role of androgen receptor in wnt responsive cells during prostate development, maturation, and regeneration. *Stem Cells 36*(6), 891–902.

Ma, T., M. A. Trinh, A. J. Wexler, C. Bourbon, E. Gatti, P. Pierre, D. R. Cavener, and E. Klann (2013). Suppression of eIF2α kinases alleviates Alzheimer's disease–related plasticity and memory deficits. *Nature Neuroscience 16*(9), 1299–1305.

Mandal, P. K., S. Saharan, M. Tripathi, and G. Murari (2015). Brain glutathione levels–a novel biomarker for mild cognitive impairment and Alzheimer's disease. *Biological Psychiatry 78*(10), 702–710.

Maynard, S., S. H. Schurman, C. Harboe, N. C. de Souza-Pinto, and V. A. Bohr (2009). Base excision repair of oxidative dna damage and association with cancer and aging. *Carcinogenesis 30*(1), 2–10.

Mazurkiewicz-Kwilecki, I. and S. Nsonwah (1989). Changes in the regional brain histamine and histidine levels in postmortem brains of Alzheimer patients. *Canadian Journal of Physiology and Pharmacology 67*(1), 75–78.

Meynial-Denis, D. (2016). Glutamine metabolism in advanced age. *Nutrition Reviews 74*(4), 225–236.

Mori, H., R. D. Cardiff, and A. D. Borowsky (2018). Aging mouse models reveal complex tumor-microenvironment interactions in cancer progression. *Frontiers in Cell and Developmental Biology 6*, 35.

Parrish, A. R. (2017). The impact of aging on epithelial barriers. *Tissue Barriers 5*(4), e1343172.

Porcellini, E., E. Calabrese, F. Guerini, M. Govoni, M. Chiappelli, E. Tumini, K. Morgan, S. Chappell, N. Kalsheker, M. Franceschi, et al. (2007). The hydroxy-methyl-glutaryl CoA reductase promoter polymorphism is associated with Alzheimer's risk and cognitive deterioration. *Neuroscience Letters 416*(1), 66–70.

Ren, W.-Y., K.-F. Wu, X. Li, M. Luo, H. C. Liu, S. C. Zhang, and Y. Hu (2014). Age-related changes in small intestinal mucosa epithelium architecture and epithelial tight junction in rat models. *Aging Clinical and Experimental*

Research 26(2), 183–191.

Rey, R. A. (2021). The role of androgen signaling in male sexual development at puberty. *Endocrinology 162*(2).

Salcedo, C., J. V. Andersen, K. T. Vinten, L. H. Pinborg, H. S. Waagepetersen, K. K. Freude, and B. I. Aldana (2021). Functional metabolic mapping reveals highly active branched-chain amino acid metabolism in human astrocytes, which is impaired in iPSC-derived astrocytes in Alzheimer's disease. *Frontiers in Aging Neuroscience 13*.

Sandiford, O. A., C. A. Moore, J. Du, M. Boulad, M. Gergues, H. Eltouky, and P. Rameshwar (2018). Human aging and cancer: role of miRNA in tumor microenvironment. *Exosomes, Stem Cells and MicroRNA*, 137–152.

Terao, A., T. L. Steininger, S. R. Morairty, and T. S. Kilduff (2004). Age-related changes in histamine receptor mRNA levels in the mouse brain. *Neuroscience Letters 355*(1-2), 81–84.

Van der Goot, A. T. and E. A. Nollen (2013). Tryptophan metabolism: entering the field of aging and age-related pathologies. *Trends in Molecular Medicine 19*(6), 336–344.

Vieira, O. V., R. J. Botelho, and S. Grinstein (2002). Phagosome maturation: aging gracefully. *Biochemical Journal 366*(3), 689–704.

Yan, J., J. K. Thomson, W. Zhao, X. Wu, X. Gao, D. DeMarco, W. Kong, M. Tong, J. Sun, M. Bakhos, et al. (2018). The stress kinase JNK regulates gap junction Cx43 gene expression and promotes atrial fibrillation in the aged heart. *Journal of Molecular and Cellular Cardiology 114*, 105–115.

Yeh, H.-I., H.-M. Chang, W.-W. Lu, Y.-N. Lee, Y.-S. Ko, N. Severs, and C.-H. Tsai (2000). Age-related alteration of gap junction distribution and connexin expression in rat aortic endothelium. *Journal of Histochemistry and Cytochemistry 48*(10), 1377–1389.

Zhou, Y., E. C. Bolton, and J. O. Jones (2015). Androgens and androgen receptor signaling in prostate tumorigenesis. *Journal of Molecular Endocrinology 54*(1), R15–R29.