Large-scale simultaneous inference under dependence

Tian, J.J.; Chen, X.; Katsevich, E.; Goeman, J.; Ramdas, A.

Received: 19 July 2021 | Revised: 2 June 2022 | Accepted: 8 June 2022

DOI: 10.1111/sjos.12614

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Large-scale simultaneous inference under dependence

Jinjin Tian[1] | **Xu Chen**[2] | **Eugene Katsevich**[3] | **Jelle Goeman**[2] | **Aaditya Ramdas**[1]

[1]Department of Statistics and Data Science, Department of Machine Learning, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

[2]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands

[3]Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Jinjin Tian, Carnegie Mellon University, Pittsburgh, PA, USA.
Email: jinjint@stat.cmu.edu

Abstract

Simultaneous inference allows for the exploration of data while deciding on criteria for proclaiming discoveries. It was recently proved that all admissible post hoc inference methods for the true discoveries must employ closed testing. In this paper, we investigate efficient closed testing with local tests of a special form: thresholding a function of sums of test scores for the individual hypotheses. Under this special design, we propose a new statistic that quantifies the cost of multiplicity adjustments, and we develop fast (mostly linear-time) algorithms for post hoc inference. Paired with recent advances in global null tests based on generalized means, our work instantiates a series of simultaneous inference methods that can handle many dependence structures and signal compositions. We provide guidance on the method choices via theoretical investigation of the conservativeness and sensitivity for different local tests, as well as simulations that find analogous behavior for local tests and full closed testing.

KEYWORDS

closed testing, multiple testing, simultaneous inference

1 | INTRODUCTION

In large-scale hypothesis testing problems, choosing the criteria for proclaiming discoveries, or even picking an error metric, can be tricky before researchers look at their data. A much more flexible approach is simultaneous (and thus post hoc) inference, which allows the researcher to examine the whole dataset and compare data-dependent guarantees on any subsets that they like before finally rejecting a set of null hypotheses along with the associated guarantee. Simultaneous inference methods are typically designed to control the false discovery proportion (FDP) for all possible choices of selections simultaneously (Blanchard et al., 2020; Goeman & Solari, 2011; Katsevich & Ramdas, 2020). It was recently proved that optimal post hoc methods must be based on closed testing (Genovese & Wasserman, 2006; Goeman et al., 2019; Marcus et al., 1976). Nevertheless, one big obstacle that prevents closed testing from being popular in practice is its exponential computation time in the worst case. Further, the complex nature of the closure process makes it hard to theoretically quantify conservativeness and power.

The key to dealing with these obstacles lies in the building block of closed testing, which is a local test for every subset of hypotheses, that tests for the presence of a signal in at least one of the hypotheses in the subset (in other words, global null testing for each subset of hypotheses). The design of such local tests, that is, the choice of the global null test to apply, is critical, as its special structure may allow fast (quadratic, linearithmic, or even linear) time shortcuts to be derived; and its robustness to dependence and power under various settings will be largely preserved after closure.

A practical choice for such a local test is a p-value combination test, that combines the evidence against the individual hypotheses in the subset into a single test statistic. Formally speaking, consider a set of hypotheses $H_1,\ldots,H_m$, each as a collection of probability measures defined on the same space $(\Omega,\mathcal{F})$, where $Q^\star$ is the true (unknown) distribution that generates the data. A hypothesis H_i is true if $Q^\star \subseteq H_i$, and the global null hypothesis is specified by

$$\bigcap_{i=1}^{m} H_i := \{H_i \text{ is true, for all } \ i \in \{1,2,\ldots,m\}\}. \tag{1}$$

Assume that we construct some test statistic, or score, T_i which captures evidence refuting H_i, and satisfying

$$\sup_{Q^\star \in H_i} \Pr_{Q^\star}\{T_i \le C_i(x)\} \le x, \quad \forall x \in [0,1], \tag{2}$$

for some corresponding critical value C_i. (The scores are high when H_i is true.) One common choice is a p-value, where $T_i = P_i$, with P_i being a valid p-value for H_i, and $C_i(x) \equiv x$. Then, global null testing can be done in the following way: combine those scores using a function f and find a calibration function C such that

$$\sup_{Q^\star \in \bigcap_{i=1}^{m} H_i} \Pr_{Q^\star}\{f(T_1,\ldots,T_m) \le C(m,x)\} \le x, \quad \forall x \in [0,1], \tag{3}$$

is true under the assumed dependence structure (if any) among the scores[1]. We call a global null test in the form of (3) as *monotonic* if f is monotonic in each of its arguments; *symmetric* if f remains unchanged on permuting its arguments. Monotonicity and symmetry are two rather

common features of a global null test. Given both monotonicity and symmetry of local tests, quadratic time shortcuts for finding simultaneous FDP confidence bounds (FDP shortcuts) have been developed by Goeman and Solari (2011) and later a quadratic time variant for simultaneous family wise error rate (FWER) control (FWER shortcuts) was presented by Dobriban (2020).

In this paper, we investigate how inference can benefit from a more specific structure of the local test, that of *separability* (see Appendix B for formal definitions of the aforementioned terms). In particular, we consider the following special case of (3):

$$\sup_{Q^\star \in \bigcap_{i=1}^m H_i} \Pr_{Q^\star}\left\{ \sum_{i=1}^m h(T_i) \le C(m,x) \right\} \le x, \quad \text{for all } x \in [0,1], \tag{4}$$

where h is a monotonic function of scores. Given the local tests of form (4), we show that both FDP and FWER shortcuts can be reduced to linear time, after an initial sorting step (Theorems 2 and 3).

Design (4) applies to a majority of existing global null tests, including famous examples like Fisher's combination test (Fisher, 1992), Stouffer's combination method (Stouffer et al., 1949), Rüschendorf's results (Rüschendorf, 1982) about the arithmetic mean of p-values; as well as recent advances like the harmonic mean (Wilson, 2019), Cauchy (Liu & Xie, 2020) and Lévy (Wilson, 2021) combinations. A particular work that is closely related to ours is a summary of all the above-mentioned global null tests: the generalized mean-based combination methods (Vovk & Wang, 2020). The fast shortcuts we developed allow bringing those canonical and new global null tests, to post hoc large-scale real-world applications. Consequently, we obtain a class of novel methods for simultaneous inference, which we found rich enough to contain powerful solutions that adapt to various dependence assumptions and signal distributions.

We further study the adaptivity in a subclass of our methods via careful quantification of the balance between conservativeness caused by the need to protect against unknown dependence and test power. Specifically, we calibrate against the intermediate setting of arbitrary Gaussian correlation (rather than the two extremes, independence and arbitrary dependence), and investigate the asymptotic power under our derived calibration. The theoretical findings regarding local tests are then empirically confirmed to be preserved after closure.

One result of independent interest is the following: if $P_1,\ldots,P_m$ are one-sided Gaussian p-values derived from the coordinates of an arbitrary m-dimensional Gaussian, then their arithmetic average P behaves like a p-value for small thresholds, satisfying $\Pr(P \le t) \le t$ for $t \le \frac{1}{2m}$.

The paper outline is as follows. In Section 2, we derive linear time algorithms for three kinds of tasks for closed testing using a local test of form (4): (1) simultaneity assessment (e.g., compute the cost of simultaneity for a single subset of hypotheses chosen pre hoc or post hoc), (2) simultaneous inference (e.g., type-I error bounds and FDP error bounds calculation for a single subset of hypotheses), and (3) automatic post hoc selection (e.g., selection of the largest set of hypotheses with a predefined error level for its post hoc FDP bound). Then we focus on the multivariate Gaussian setting to formally evaluate a class of local tests satisfying our requirements based on generalized means. Specifically, in Section 3.2, we derive the asymptotic valid calibrated threshold for positively equicorrelated Gaussians, which allows us to calculate the price paid to protect against different levels of dependence using different combinations choices. Then we calculate closed-form asymptotic power expressions under different signal settings in Section 3.3, and reason about the sweet spot for different combination methods. Finally, we confirm that our

qualitative conclusions about local tests are preserved after closure, using simulations in Section 4. A conclusion including takeaways for practitioners and future directions is provided in Section 5.

2 | SIMULTANEOUS INFERENCE VIA CLOSED TESTING

Recall that we are interested in testing hypotheses $H_1,\ldots,H_m$, each represented by a collection of probability measures on the measurable space $(\Omega,\mathcal{F})$, where $Q^\star$ is the true (unknown) measure that generates the data. We call a hypothesis H_i *null* if $Q^\star \in H_i$, and *nonnull* otherwise. We denote $H_S := \bigcap_{i\in S} H_i$ as the intersection hypotheses corresponding to index set S, which is *null* if and only if H_i is *null* for all $i \in S$. In particular, we let $H_\emptyset$ equals the set of all probability measures on $(\Omega,\mathcal{F})$, so the null hypothesis $H_\emptyset$ is always true. Let $\mathcal{H}_0 := \{i : Q^\star \in H_i\}$ denote the (unknown) set of null hypotheses that are true.

The nonnull hypotheses are usually of more interest, often serving as an important reference for variable selection and scientific discovery. Therefore we often call the nonnull hypotheses as *signals*. A common goal is to identify a large set of hypotheses that contains mostly signals. In other words, we wish to proclaim a set of "discoveries" while controlling the number or fraction of false discoveries (i.e., the null hypotheses that were incorrectly proclaimed as discoveries).

For a set $S \subseteq [m] := \{1,2,\ldots,m\}$ indexing the hypotheses, define its (unknown) number of false and true discoveries as

$$\epsilon(S) := |S \cap \mathcal{H}_0|, \quad \delta(S) := |S \setminus \mathcal{H}_0|, \tag{5}$$

respectively. We wish to find $t_\alpha(S) \in \{0,1\}$ and $e_\alpha \in \{0,1,\ldots,|S|\}$ such that:

$$\text{Type-I error control:} \quad \Pr\{\delta(S) \geq t_\alpha(S)\} \geq 1-\alpha, \tag{6}$$

$$\text{FDP control:} \quad \Pr\{\epsilon(S) \leq e_\alpha(S)\} \geq 1-\alpha, \tag{7}$$

where $t_\alpha(S)$ indicates whether we reject H_S or not, and $e_\alpha(S)$ provides the upper bound of the number of nonsignals in S. Specifically, Type-I error control guarantees that, with high probability, S is not rejected if it contains only nulls, while the FDP control guarantees that, with high probability, the number of false discoveries in set S is upper bounded. Naturally, we prefer $t_\alpha(S)$ to be one if possible and $e_\alpha(S)$ to be as small as possible. The slightly odd formalism for (6) is simply to draw parallels with the definitions that follow.

To freely examine several arbitrary sets S and then select a set, we need extra corrections to ensure post hoc validity of error guarantees. In other words, we would need to convert the above high probability guarantees for an individual set S into one for all possible sets simultaneously. Formally, we desire

$$\text{Simultaneous Type-I error control:} \quad \Pr\left\{\delta(S) \geq \bar{t}_\alpha(S) \ \text{ for all } \ S \subseteq [m]\right\} \geq 1-\alpha, \tag{8}$$

for some $\bar{t}_\alpha(S) \in \{0,1\}$ as before, and we would like to design an $\bar{e}_\alpha(S) \in \{0,1,\ldots,|S|\}$ such that

$$\text{Simultaneous FDP control:} \quad \Pr\left\{\epsilon(S) \leq \bar{e}_\alpha(S) \ \text{ for all } \ S \subseteq [m]\right\} \geq 1-\alpha. \tag{9}$$

Closed form expressions for $\bar{t}(\cdot)$ and $\bar{e}(\cdot)$ can be derived in special cases (Katsevich & Ramdas, 2020), but only bounds based on closed testing can be admissible (Goeman et al., 2021). Closed testing was initially proposed by Marcus et al. (1976), who suggested using

$$\bar{t}_\alpha(S) = \mathbf{1}\{t_\alpha(J) = 1 \text{ for all } J \supseteq S\}. \tag{10}$$

It was later noticed by Goeman and Solari (2011) that the same procedure also yields an expression for $\bar{e}_\alpha(S)$:

$$\bar{e}_\alpha(S) = \max\{|I| : I \subseteq S,\ \bar{t}_\alpha(S) = 0\}, \tag{11}$$

which is the size of the largest subset of S that is not rejected by closed testing. In this closed testing framework, t_α defined in (6) is also called as a *local test*, which is just a valid α-level test of the composite hypothesis H_S, while $\bar{t}_\alpha$ is the corresponding post hoc version. We denote the set of composite hypotheses rejected locally (before closure) as $\mathcal{U}_\alpha$, and as $\mathcal{X}_\alpha$ after closure, that is

$$\mathcal{U}_\alpha = \{S \subseteq [m] : t_\alpha(S) = 1\}, \quad \text{and} \quad \mathcal{X}_\alpha = \{S \subseteq [m] : \bar{t}_\alpha(S) = 1\}. \tag{12}$$

In this paper, we focus on the case when local test t_α is of the following form:

$$t_\alpha(S) = \mathbf{1}\left\{\sum_{i=1}^{|S|} h(T_i) \le C(|S|, \alpha)\right\}, \tag{13}$$

where $h(\cdot)$ is a monotonically increasing function.

Remark 1. In fact, the form (13) satisfies three common and reasonable designs of global null test, which are *symmetry*, monotonicity, and *separability*. Specifically the summation structure corresponds to *separability*; the monotonicity of h corresponds to *monotonicty*; and the index-invariant fact about h corresponds to *symmetry*. We refer the interested readers to Appendix B for details, and definitions of the aforementioned terms.

Before we proceed, we introduce a special class of local tests based on generalized means as discussed by Vovk and Wang (2020), since we will repeatedly use them as motivating examples. Consider the following combinations of p-values $p_1, \ldots, p_m$, indexed by $r \in [-\infty, \infty]$:

$$M_r(p_1, \ldots, p_m) := \begin{cases} \max_{i\in[m]} p_i, & \text{if } r = \infty; \\ \left(\prod_{i=1}^m p_i\right)^{1/m}, & \text{if } r = 0; \\ \left(\frac{1}{m}\sum_{i=1}^m p_i^r\right)^{1/r}, & r \in (-\infty, 0) \cup (0, \infty); \\ m \min_{i\in[m]} p_i, & \text{if } r = -\infty, \end{cases} \tag{14}$$

which corresponds to the arithmetic mean when $r = 1$; geometric mean when $r = 0$; and harmonic mean when $r = -1$. For simplicity, we use $M_{r,m}$ to stand for $M_r(p_1, \ldots, p_m)$ throughout the paper. Denote

$$t_\alpha^{(r)}(S) := \mathbf{1}\{M_r((p_i)_{i\in S}) \le c_r(|S|, \alpha)\}, \tag{15}$$

where $c_r(|S|, \alpha)$ is a critical value that depends only on $|S|$, α for different r. Then according to Vovk and Wang (2020), $t_\alpha^{(r)}(S)$ is a valid local test, and the corresponding class

$$\mathcal{T}_\alpha := \left\{ t_\alpha^{(r)} : r \in [-\infty, \infty] \right\}, \tag{16}$$

is rich enough to contain many famous local test choices like the Bonferroni ($r = -\infty$) method, the Fisher's combination ($r = 0$), and the recent harmonic mean combination method ($r = -1$); and its members also have simple enough structure such that we can summarize their nature with a univariate parameter r.

2.1 | The cost of multiplicity adjustment arising from post hoc inference

In practice, one may be concerned that simultaneity has a large statistical cost (paid in power). To address this concern, we propose a novel statistic called *coma*, which stands for the *COst of Multiplicity Adjustment* arising from requiring valid post hoc inference. The statistic is invariant to the testing level α, and only costs linear time to compute.

To construct *coma* such that it is invariant to test level α, we intentionally use the adjusted p-value, which is defined as the smallest α under which the test would be rejected. Formally, for a certain set S among a series of hypotheses $H_1, \ldots, H_m$, denote the adjusted p-value based on S using local testing rule t_α as

$$p(S) := \inf\{\alpha \in [0, 1] : t_\alpha(S) = 1\},$$

and the adjusted p-value for S after going through the closed testing procedure as

$$\bar{p}(S) := \inf\{\alpha \in [0, 1] : \bar{t}_\alpha(S) = 1\}. \tag{17}$$

Then *coma* is defined as follows.

Definition 1 (cost of multiplicity adjustment). For any $S \subseteq [m]$, define

$$\text{coma}(S) := \bar{p}(S)/p(S), \tag{18}$$

as the cost of multiplicity adjustment when testing H_S.

Note that coma(S) is a data-dependent quantity that depends on the choice of local test. As for a quick example, $\text{coma}(S) = \frac{m}{|S|}$ if t_α is Bonferroni and $p(S)$ is small enough. This example concurs with the intuition that the cost of multiplicity grows with the total dimension m; however, it decreases with the subset dimension $|S|$. The following result presents a more general expression for *coma*.

Theorem 1. *For any $S \subseteq [m]$, if the local test is of form (13), then we have a linear time expression*

$$\bar{p}(S) = \max_{0 \le i \le |S^c|} p(S \cup J_i^\star), \tag{19}$$

where $J_i^\star$ is the set of indices of hypotheses associated with the i largest p-values in S^c.

Theorem 1 is proved in Appendix A. In the following, we examine how *coma* varies with the size of S for local tests based on generalized means $\mathcal{T}_\alpha$ in (16), using calibration derived by Vovk and Wang (2020) under arbitrary dependence. Figure 1 plots *coma* versus the choice of local test for a set S of size 20 out of 200 hypotheses in total, using equicorrelated Gaussian data. We can see that *coma* with local test $t_\alpha^{(r)}$ with positive r is generally smaller than that with negative r, while the order statistics-based procedures, Simes and Bonferroni, behave similarly. This indicates one would prefer to use $t_\alpha^{(r)}$ with positive r if one does not want too different results on changing from pre hoc to post hoc. On the other hand, Figure 2 plots *coma* versus the size of the target set S (with the total number of hypotheses remaining as 200). Except for the consistent observation that positive r have lower *coma*, we can additionally see that coma(S) generally decreases with the size of S, which agrees with our intuition that lower resolution post-hoc inference should cost less.

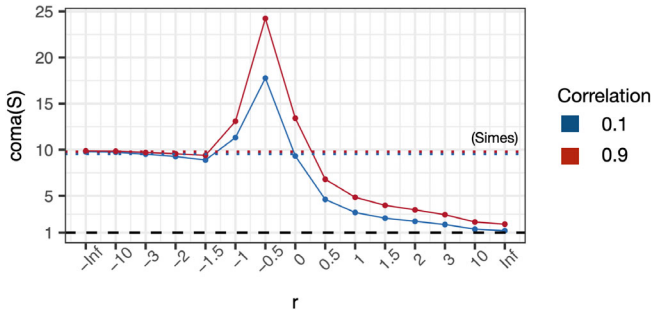

FIGURE 1 coma(S) versus different local test procedures under different extend of dependency. The dashed horizontal lines represents the value of coma(S) with the local test as Simes, whereas the solid lines plot the value of coma(S) with the local test as $t_\alpha^{(r)}$ versus different r. When $r = -\infty$ (written as $-Inf$), $t_\alpha^{(r)}$ recovers Bonferroni. We simulate the data to follow equicorrelated Gaussian, where we set total number of hypotheses $m = 200$, and size of set S as 20. We set signal proportion outside S as 0.3, signal proportion inside S as 0.7 with signal strength (i.e. the mean of Gaussian) $\mu = 2$. The results are averaged over 5×10^3 trials.

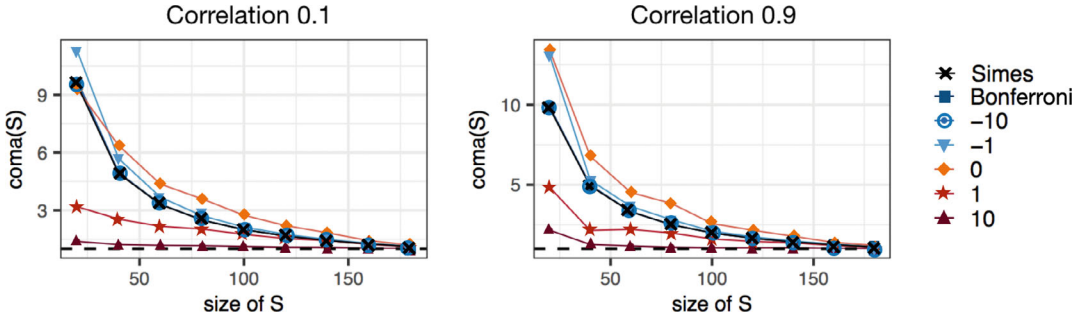

FIGURE 2 coma(S) versus the size of S using different local test procedures under We simulate the data to follow equicorrelated Gaussian, where we set total number of hypotheses $m = 200$, and size of set S as 20. We set signal proportion outside S as 0.3, signal proportion inside S as 0.7 with signal strength (i.e. the mean of Gaussian) $\mu = 2$. The results are averaged over 5×10^3 trials. We can see that, the lines for Simes, Bonferroni, and $r = -10$ almost overlap with each other, while the line for $r = -1$ is slightly higher (and the line for $r = 0$ is on the top). These observations are consistent with results in Figure 1.

2.2 | Fast shortcuts for post hoc inference and selection

Another practical concern with regard to imposing simultaneity is the heavy computation time, which is exponential in m in general. In this section, we present fast (linear time shortcuts for calculating both $\bar{t}_\alpha$ and $\bar{e}_\alpha$, for local tests of form (13).

Theorem 2. *Consider testing m hypotheses with presorted scores post-hoc via closed testing using local test t_α of form (13). For a set $S \subseteq [m]$,* Algorithm 1 *returns the simultaneous FDP bound $\bar{e}_\alpha(S)$ in (11), with at most $O(m)$ computation.*

Note that we sort the scores in ascending order in Algorithm 1 in order to have easier tracking of indices since the algorithm is a step-down procedure. The proof for Theorem 2 is in Appendix C.

Algorithm 1. Shortcut for evaluating post hoc false discoveries bound $\bar{e}_\alpha(S)$

Input: A sequence of sorted scores $T_1, \dots, T_m$ which satisfies $T_1 \geq \cdots \geq T_m$;a local test rule of form (13) with a monotonically increasing transformation function h and thresholding function C;confidence level α;candidate rejection set $S = \{i_1, i_2, \dots, i_s\}$ and its complement $S^c = \{j_1, j_2, \dots, j_{m-s}\}$ with $i_1 < i_2 < \dots i_s, \quad j_1 < j_2 < \dots j_{m-s}$.

Output: High probability $(1-\alpha)$ simultaneous bound $\bar{e}_\alpha(S)$ on the number of false discoveries in S.

Initialization:
transformed candidate set scores: $u_1, \dots, u_s$, where $u_d = h(T_{i_d})$ for $1 \leq d \leq s$;
transformed complementary set scores: $v_1, \dots, v_{m-s}$, where $v_d = h(T_{j_d})$ for $1 \leq d \leq m-s$;
ill-defined transformed scores: $v_0 = \max(u_1, v_1)$; $v_{m-s+t}, u_{s+t} \equiv \min(u_s, v_{m-s}) - 1$, $\forall 1 \leq t \leq m-s$;
iteration related indices $k \leftarrow 1; \quad b \leftarrow -1$;
accumulated scores $Q = 0$.
for $a = 1, \dots, m$ **do**
  if $u_{k+b+1} \geq v_{a-k-b}$ ***or*** $a = 1$ **then**
    $Q = Q + u_{k+b+1}$
    $b = b + 1$
  else
    $Q = Q + v_{a-k-b}$
  end
  while $k \leq \min(s, a)$ ***and*** $Q > C(a, \alpha)$ **do**
    if $b > 0$ **then**
      $b \leftarrow b - 1$
    else
      $Q \leftarrow Q + u_{k+1} - v_{a-k}$
    end
    $k \leftarrow k + 1$
  end
end
return $k - 1$

Remark 2. Note that local test $t_\alpha^{(r)}$ does not admit form (13) when $r = \pm\infty$, therefore the shortcut in Theorem 2 for evaluating corresponding $\bar{e}_\alpha^{(r)}$ is not applicable. However, they lead to

consonant[2] closed testing as proved by Lemma 2 in Appendix D, and one interesting fact pointed out by Goeman and Solari (2011) is that if for consonant closed testing, the simultaneous FDP bound for a given set reduces to finding the number of its elementary hypotheses that the closed testing cannot reject, therefore reducing to identifying the set of elementary hypotheses being rejected after closure. For $r = -\infty$, this is just Holm's method, while for $r = \infty$, this is just checking whether we can reject the largest p-value to decide either to reject all or nothing.

We have presented procedures for fast inference on a single set S picked freely by users, which in turn, enables effective post hoc selection among multiple sets of interest: linear and quadratic shortcuts for automatic selection of the largest set S with a prespecified bound $\bar{e}_\alpha$ can also be developed. For users who have no idea of which candidate set to evaluate, Theorem 3 allows them for efficient automatic selection among a sequence of incremental sets: finding the largest one among them with FDP bounded by $\gamma \in [0, 1)$. Specifically, the main ingredient is Algorithm 3 (Appendix E), which works for any kind of closed testing-based post hoc inference given incremental candidate sets. We also provide a faster alternative, Algorithm 2 (Appendix E), given $\gamma = 0$ and certain constrains of local tests.

Theorem 3. *Consider testing m hypotheses post hoc via closed testing at level α, and a series of incremental candidate sets to reject*: $S_1 \subset S_2 \cdots \subset S_n \subseteq [m]$ *with* $|S_i| = i$ *for all* $i \in [n]$. *Then we have*:

(a) *Given any desired FDP bound* $\gamma \in [0, 1)$, *Algorithm* 3 *returns the largest set* S_k *such that* $\bar{e}_\alpha(S_k) \le \gamma |S_k|$.

If we additionally require local test to be of form (13), then

(b) *Algorithm* 3 *costs at most* $O(mn)$ *computation*;
(c) *Algorithm* 3 *reduces to Algorithm* 2 *if* $\gamma \equiv 0$ *and* S_k *is the indexes of hypotheses with* k *smallest scores, which cost at most* $O(m)$ *computation with presorted scores.*

The validity of Algorithm 3 in Theorem 3 does not require any assumption on local test or presorting p-values, and needs m iterations in the worst case. In practice we expect fewer iterations will be needed as the false discoveries are ruled out in batches quickly. Particularly, for the special case stated in part (c) in Theorem 3, the task costs only at most linear time. The proof of Theorem 3 is in Appendix F.

Remark 3. Algorithm 2 in Theorem 3 is also the shortcut for finding the largest hypotheses set to reject with strong FWER control among all m hypotheses.

3 | CALIBRATION OF LOCAL TESTS FOR MULTIVARIATE GAUSSIANS

The performance of closed testing-based post hoc inference largely depends on the building blocks—local tests. Therefore, in order to provide better guidance of applying our newly derived shortcuts introduced in Section 2, we look into the properties of different global null tests, particularly the generalized mean-based ones (i.e., $t_\alpha^{(r)}$ defined in (15)) since our shortcuts apply to these.

Vovk and Wang (2020) first summarized the class of generalized mean-based combination methods, and derived closed form calibration under arbitrary dependence for different combination choice, using results based on robust risk aggregation. Now we specifically summarize the results for calibrating under arbitrary dependence (Vovk & Wang, 2020) in the following Lemma 1, as it will be our benchmark to compare with.

Remark 4. Though a refined version of Lemma 1 (which gives best possible calibration) can be found in Vovk et al. (2022, proposition 8.1), it does not admit closed-form expression as Lemma 1 does. Therefore we adopt Lemma 1 throughout the paper for simpler theoretical analysis.

Lemma 1 (Vovk and Wang (2020)). *For m hypotheses, $\alpha/\alpha_{r,m}$ is a valid critical value for the global null test $t_\alpha^{(r)}$ defined in (15), where*

$$\alpha_{r,m} := \begin{cases} (r+1)^{1/r}, & \text{if } r \in (-1,\infty]; \\ ((y_m+m)^2/(y_m+1))\mathbf{1}\{m \geq 3\} + m\mathbf{1}\{m \leq 2\}, & \text{if } r = -1; \\ \frac{r}{r+1} m^{1+1/r}, & r \in [-\infty,-1), \end{cases} \tag{20}$$

and y_m is the unique strictly positive solution of $y^2 = m\big((y+1)\log(y+1) - y\big)$. Particularly, for $r \in \{-\infty, 0, \infty\}$, we define $\alpha_{r,m}$ as $\lim_{r\to\infty}(r+1)^{1/r} = 1$, $\lim_{r\to 0}(r+1)^{1/r} = e$, and $\lim_{r\to-\infty}\frac{r}{r+1}m^{1+1/r} = m$.

Follow-up work (Chen et al., 2020; Wilson, 2020) explored the conservativeness of such calibration under some special dependence structures: Wilson (2019) derived asymptotic valid (in the sense of $m \to \infty$) calibration under independence using generalized central limit theorem, and empirically studied their performance when the independence assumption is broken; Chen et al. (2020) compared the generalized mean-based combination with order statistics-based combination, and proved that only Cauchy combination (and its analog harmonic mean) and Simes combination pay no price for calibration to achieve validity under assumptions from independence to full dependence (i.e. correlation one); Vesely et al. (2021) studied the special case that permutation tests can be used. Figure 3 summarizes all the cases (including ours) where theoretically valid calibration has been derived. Note that, before our work, almost no results have derived in cases other than the two extremes—the independence case and the arbitrary dependence case: Chen et al. (2020) provided some theoretical justification in the pairwise Gaussian scenario but only for harmonic mean. As for common intermediate dependence structures like multivariate Gaussian case, most work only explored experimentally. Therefore, as shown in Figure 3, we work toward filling in the gap by deriving calibration under one of the intermediate cases, the equicorrelated Gaussian setting, which contains both two extremes as well as different dependence levels. Later, we also investigate the performance of our calibration by analyzing the asymptotic type-I error and power under different settings. Particularly, our calibration recovers existing work in scenarios where independence provably has the highest inflated type-I error to be calibrated among others. At the same time, our theoretical performance investigation justifies the interesting behaviors noticed in early experimental studies (Chen et al., 2020; Wilson, 2019), that is, for the generalized mean-based methods, choice of positive r performs better under heavy dependence and calibrating under independence gives a high false-positive rate overall. In contrast, the choice of negative r performs poorly under heavy dependence, and calibrating under independence gives a low false-positive rate overall.

Numerical Closed-form

	Independence	Equicorrelated Gaussian	Arbitrary Dependence
$r \geq 1$	$\alpha \leq \alpha_0$ (Chen et al.) $\quad m \to \infty$ (Wilson) = (Chen et al.)	**(ours)**	(Vovk and Wang)
$0 < r < 1$		$m \to \infty$ **(ours)**	
$r = 0$	(Fisher)		
$-0.5 < r < 0$	$\alpha \to 0$ (Chen et al.) $\quad m \to \infty$ (Wilson) = (Chen et al.)		
$r = -0.5$	$\alpha \to 0$ (Chen et al.) $\quad m \to \infty$ (Wilson), (Chen et al.)		
$-1 < r \leq -0.5$	$\alpha \to 0$ (Chen et al.) $\quad m \to \infty$ (Wilson) = (Chen et al.) $m \to \infty$ & $\alpha \to 0$ (Wilson) = (Chen et al.)		
$r = -1$	$\alpha \to 0$ (Chen et al.) $\quad m \to \infty$ (Wilson) = (Chen et al.) $m \to \infty$ & $\alpha \to 0$ (Wilson) = (Chen et al.) = **(ours)**	$m \to \infty$ & $\alpha \to 0$ **(ours)**	(Vovk and Wang)
$r < -1$	$\alpha \to 0$ (Chen et al.) $\quad m \to \infty$ (Wilson) = (Chen et al.) $m \to \infty$ & $\alpha \to 0$ (Wilson) = (Chen et al.) = **(ours)**	$m \to \infty$ & $\alpha \to 0$ **(ours)**	(Vovk and Wang)

FIGURE 3 Summary of regimes in which we know how to calibrate generalized means of p-values. We omit explicit expressions as there is sometimes no analytical formula, but thresholds can be calculated numerically (blue text). We refer readers to the corresponding references mentioned in the text for explicit expressions.

3.1 | Model setup

Before presenting the main results, we first motivate our choice of the equicorrelated Gaussian model. Consider a Gaussian sequence model for the observations:

$$(X_{m1}, X_{m2}, \ldots, \ldots X_{mm}) \sim N_m(\boldsymbol{\mu}_m, \Sigma_m), \tag{21}$$

where $\boldsymbol{\mu}_m = (\mu_{m1}, \ldots, \mu_{mm})$, and each entry $\mu_{mi} \overset{iid}{\sim} \mu_m B_m$, with $\mu_m > 0$ as a scalar, and B_m as a Bernoulli random variable with parameter π_m. Additionally, we assume $\Sigma_m \in \mathcal{M}_m$, where $\mathcal{M}_m$ is the set of all $m \times m$ positive semidefinite correlation matrices. We denote the (i,j)th entry of Σ_m as ρ_{ij}. Additionally, we denote the set of all equicorrelation matrices as $\mathcal{M}_m^E$, which is the subset of $\mathcal{M}_m$ with all equal nondiagonal elements.

Suppose we are testing the global null hypothesis.

$$\bigcap_{i=1}^{m} H_{mi} := \{\mu_{mi} = 0, \ \forall\, i\},$$

at level α. We consider a one-sided p-value $p_{mi} = \Phi(-X_{mi})$ for each elementary hypothesis (where Φ is the cumulative density function (CDF) of a standard normal), and combine them using a

generalized mean $\overline{t}_\alpha^{(r)}$ (15). Denote the corresponding type-I error given the correlation matrix Σ with respect to different r as follows:

$$\tilde{\alpha}_m(\Sigma, r, c) := \Pr_{\bigcap_{i=1}^m H_{mi}}\left\{\left(\frac{1}{m}\sum_{i=1}^m p_{mi}^r\right)^{1/r} \le c\right\}, \tag{22}$$

where $\Pr_{\bigcap_{i=1}^m H_{mi}} := \sup_{Q^\star \in \bigcap_{i=1}^m H_{mi}} \Pr_{Q^\star}$, and c is a correction/calibration threshold to account for dependence, which could be an absolute constant or potentially depends on r, m, and α.

Remark 5. From the monotonicity of the generalized mean with respect to r, one can easily verify by contradiction that $\tilde{\alpha}_m(\Sigma, r, c) \le \alpha$ implies $c \le \alpha$.

Proposition 1. *Fix any $m \ge 1$ and any $r \ge 1$. If $c < \frac{1}{2m}$, then*

$$\sup_{\Sigma \in \mathcal{M}_m} \tilde{\alpha}_m(\Sigma, r, c) = \sup_{\Sigma \in \mathcal{M}_m^E} \tilde{\alpha}_m(\Sigma, r, c) = \tilde{\alpha}_m(\mathbf{1}_m \mathbf{1}_m^T, r, c), \tag{23}$$

where $\mathbf{1}_m$ is the m-dimensional vector of all ones.

Proposition 1 indicates that, for all $r \ge 1$ and appropriately small α, we only need to calibrate against the fully dependent case to have validity across the whole correlation space $\mathcal{M}_m$. The proof of Proposition 1 is in Appendix G, where we used the convexity of function $\Phi(-x)^r$ when $r \ge 1$ and $x > 0$, and the fact that (multivariate) Gaussianity is preserved under linear transformations. It is unclear whether the restriction on α can be entirely removed, but it could perhaps be slightly relaxed by constant factors. A special case that could be particularly interesting is when $r = 1$, which we record below for emphasis.

Corollary 1. *Let $\Sigma \in \mathcal{M}_m$ be an arbitrary positive semidefinite Gaussian correlation matrix (with possibly negative entries). Let $X \sim N(0, \Sigma)$ and let $P_i = \Phi(-X_i)$ for $i = 1, \ldots, m$. Then, the generalized mean of the p-values, $\overline{P}^{(r)} := (\frac{1}{m}\sum_{i=1}^m P_i^r)^{\frac{1}{r}}$ with $r \ge 1$ satisfies*

$$\sup_{\Sigma \in \mathcal{M}_m} \Pr(\overline{P}^{(r)} \le \alpha) \le \alpha, \quad \text{for any } \alpha < \frac{1}{2m}.$$

It is possible that a variant of Proposition 1 also holds for $r < 1$, but we have found it to be technically intractable to prove currently. Nevertheless, for the sake of simplicity and interpretability, we next consider an intermediate case of equicorrelated Gaussians, which we observed to be worse than the other commonly used correlation structures when $r < 1$ in extensive simulations, while also encompassing the perfectly correlated case in Proposition 1. In particular, we consider only positive correlation as the semi-positive definite requirement on the correlation matrix forces the range of negative ρ to be in $(-\frac{1}{m}, 0)$, which vanishes as $m \to \infty$.

Definition 2. Positively equicorrelated Gaussian: For each m, the observations $X_{m1}, \ldots, X_{mm}$ follow the model in (21) but with a positive equicorrelated Σ_m having its elementary entry $\rho_{ij} \equiv \rho \in [0, 1]$ for all $i \ne j \in [m]$. We denote such data distribution as $G_{\mu_m, \pi_m, \rho}$.

Formally, in the following Sections 3.2 and 3.3, we consider the model defined in Definition 2, and we write $\tilde{\alpha}_m(\Sigma, r, c)$ in (22) as $\tilde{\alpha}_m(\rho, r, c)$ for simplicity. We intend to study the asymptotic ($m \to \infty$) behavior of calibrated $\tilde{\alpha}_m(\rho, r, c)$ given fixed α. We would like to investigate how their power varies as a function of correlation ρ with respect to different r, and different signal settings.

3.2 | Calibration derivation

In this subsection, we derive the asymptotic calibration under the positively equicorrelated Gaussian model in Definition 2. First, we formally define the asymptotic calibration of our concern, and then we present our closed-form solution under the positively equicorrelated Gaussian model.

Typically, the asymptotic ($m \to \infty$) Type-I error would be defined as

$$\mathcal{A}^{\star}(r) := \limsup_{m\to\infty} \sup_{\rho\in[0,1]} \tilde{\alpha}_m(\rho, r, c). \tag{24}$$

However, we found (24) to be intractable; specifically, before taking the outer limit, we found taking the supremum with respect to ρ for fixed m to be analytically infeasible under the positively equicorrelated Gaussian model. Therefore, we settle for an alternative (weaker) definition of target type-I error as the following surrogate limit:

$$\mathcal{A}(r) := \sup_{\rho\in[0,1]} \limsup_{m\to\infty} \tilde{\alpha}_m(\rho, r, c). \tag{25}$$

Note that $\mathcal{A}^{\star}(r) \geq \mathcal{A}(r)$ deterministically[3], that is control over the surrogate asymptotic type-I error is weaker. Denote the highest calibrated threshold c that achieves $\mathcal{A}(r) \leq \alpha$ as $c_r(m, \alpha)$, that is

$$c_r(m, \alpha) := \sup\left\{ c : \sup_{\rho\in[0,1]} \limsup_{m\to\infty} \tilde{\alpha}_m(\rho, r, c) \leq \alpha \right\}, \tag{26}$$

and the corresponding limiting type-I error as

$$\tilde{\alpha}(\rho, r, \alpha) := \limsup_{m\to\infty} \tilde{\alpha}_m(\rho, r, c_r(m, \alpha)). \tag{27}$$

In the following, we derive a closed-form expression for $c_r(m, \alpha)$, and the corresponding $\tilde{\alpha}(\rho, r, \alpha)$ under the positively equicorrelated Gaussian model. Note that in this setting, the observations can be written as

$$X_i = \sqrt{\rho}\, Z_0 + \sqrt{1-\rho}\, Z_i, \quad \text{for all } i = 1, 2, \ldots, m, \tag{28}$$

where $Z_0 \sim N(0,1)$, $Z_i \overset{\text{iid}}{\sim} N(0,1)$ for all $i = 1, 2, \ldots, m$, and $Z_0 \perp\!\!\!\perp \{Z_i\}_{i=1}^m$. The corresponding one-sided p-values are

$$p_i = \Phi(-X_{mi}) = \Phi\left(-\sqrt{\rho}\, Z_0 - \sqrt{1-\rho}\, Z_i\right). \tag{29}$$

Here we drop index m as the distribution of X does not change with m and the same holds true for p-values. An important note from this decomposition is the following conditional independence,

$$p_1, p_2, \ldots, p_m \text{ are i.i.d. conditional on } Z_0, \tag{30}$$

which allows us to utilize generalized law of large numbers and obtain Theorem 4. We write the expectation of p_i^r when conditioning $Z_0 = z_0$ as a function of z_0 that is

$$g_{\rho,r}(z_0) := \mathbb{E}\left[p_i^r \,\middle|\, Z_0 = z_0\right] = \int \Phi(-\sqrt{\rho}\, z_0 - \sqrt{1-\rho}\, x)^r \phi(x) dx, \tag{31}$$

noting that $g_{0,r}(z_0)$ is a constant, we enforce $g_{0,r}^{-1}(\cdot) \equiv \infty$.

Theorem 4. *Under the positively equicorrelated Gaussian setting, we have that, given $\alpha \in (0,1)$,*

(a) if $r > 0$, then $\tilde{\alpha}(\rho, r, \alpha) = \Phi(-g_{\rho,r}^{-1}(\alpha^r))$, and $c_r(m,\alpha) = \min\left\{\alpha, \left(\frac{r}{r+1}\right)^{\frac{1}{r}}\right\}$;

(b) if $-1 < r \le 0$, then $\tilde{\alpha}(\rho, r, \alpha) = \Phi(-g_{\rho,r}^{-1}(c_r(m,\alpha)))$, and $c_r(m,\alpha) = \left(\sup_{\rho\in[0,1]} g_{\rho,r}\left(-\Phi^{-1}(\alpha)\right)\right)^{\frac{1}{r}}$ is not a function of m, where $g_{\rho,r}$ is defined in (31);

(c) if $r = -1$, then $\tilde{\alpha}(\rho, r, \alpha) = \alpha \mathbf{1}\{\rho = 0\}$, and $c_r(m,\alpha) = \frac{\alpha}{1+\alpha\log m}$ as $\alpha \to 0$;

(d) if $r < -1$, then $\tilde{\alpha}(\rho, r, \alpha) = \alpha \mathbf{1}\{\rho = 0\}$, and $c_r(m,\alpha) = \alpha m^{\frac{1}{|r|}-1}$ as $\alpha \to 0$.

In all four cases, we have that $c_r(m,\alpha) \le \alpha$.

The proof of Theorem 4 is in Appendix H, where we mainly use the decomposition described above and generalized law of large numbers. From Theorem 4, we can see that the calibrated threshold $c_r(m,\alpha)$ under positively equicorrelated Gaussian is less conservative than that under arbitrary correlation in Lemma 1; particularly, for $r > 0$, by a factor of $(r+1)^{1/r}$; for $r = 0$, by a factor of $\frac{|\log\alpha|+1}{|\log\alpha|}$; for $r = -1$, by a factor[4] of $\frac{\log m}{\alpha\log m+1}$; for $r < -1$, by a factor of $\frac{|r|}{|r|-1}$. Figure 4 displays these ratios for $r \in [-10, 20]$. We can see that, as $|r| \to \infty$, our positively equicorrelated Gaussian calibration is almost the same as the calibration derived by Vovk and Wang (2020, Lemma 1), for arbitrary dependence, indicating that we do not pay much price for calibrating against positive equicorrelation to arbitrary dependence; while as $|r| \to 1$, the positively equicorrelated Gaussian calibration is much tighter than that of arbitrary dependence in Vovk and Wang (2020).

Next, we conduct some large-scale simulations (in terms of m) to justify the surrogate control in Theorem 4. Explicitly, we compare our derived asymptotic type-I error $\tilde{\alpha}(\rho, r, \alpha)$ under the surrogate calibration (i.e. $\mathcal{A}(r) \le \alpha$) with the type-I error $\tilde{\alpha}_m^\star(\rho, r, \alpha)$ under the ideal calibration (i.e. $\mathcal{A}^\star(r) \le \alpha$), that is

$$\tilde{\alpha}_m^\star(\rho, r, \alpha) := \Pr_{\bigcap_{i=1}^m H_{mi}}\left\{\left(\frac{1}{m}\sum_{i=1}^m p_{mi}^r\right)^{\frac{1}{r}} \le c_r^\star(m,\alpha)\right\}. \tag{32}$$

$$\text{with} \quad c_r^\star(m,\alpha) := \sup\left\{c : \sup_{\rho\in[0,1]} \tilde{\alpha}_m(\rho, r, c) \le \alpha\right\}. \tag{33}$$

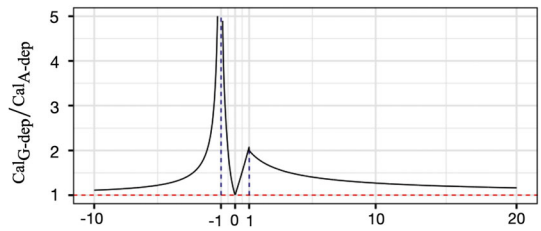

FIGURE 4 The theoretical ratio of calibrated threshold under positively equicorrelated Gaussian ($\text{Cal}_{\text{G-dep}}$) and under arbitrary dependence ($\text{Cal}_{\text{A-dep}}$) versus different r

Figure 5 empirically shows that, for $r > -1$, $c_r(m, \alpha)$ can also approximately achieve control $\sim \mathcal{A}^\star(r) \le \alpha$ in the sense that $\tilde{\alpha}(\rho, r, \alpha) \approx \tilde{\alpha}^\star_m(\rho, r, \alpha)$ for each $\rho \in [0, 1]$ if m is large enough.

For $r \le -1$ the approximation is much looser as the convergence (for point-wise ρ) in generalized law of large numbers is much slower and out of feasible simulation scope. Nevertheless, from Figure 6 one may see a trend of convergence with regard the current limited magnitude of m: the "worst case" correlation given fixed m, r, α, that is $\arg\max_{\rho\in[0,1]} \tilde{\alpha}^\star_m(\rho, r, \alpha)$ slowly approaches 0 as m grows, at different rate given different α (faster for α away from 0), and the point-wise Type-I error $\tilde{\alpha}^\star_m(\rho, r, \alpha)$ slowly approaches $\tilde{\alpha}(\rho, r, \alpha) = \alpha \mathbf{1}\{\rho = 0\}$ as m grows.

3.3 | Power analysis

In this subsection, we study the power using the calibrated threshold $c_r(m, \alpha)$ derived in Section 3.2. We look at the case $r > 0$ and $r \le -1$ separately, under different alternative settings. Given α, denote the power function for different r under distribution $G_{\mu_m, \pi_m, \rho}$ in Definition 2 as

$$\beta_{\mu_m, \pi_m, \rho}(r, \alpha) := \mathbb{P}_{G_{\mu_m, \pi_m, \rho}} \left\{ \left(\frac{1}{m} \sum_{i=1}^{m} p_{mi}^r \right)^{\frac{1}{r}} \le c_r(m, \alpha) \right\}, \tag{34}$$

with $c_r(m, \alpha)$ is specified in Theorem 4.

In the following, we are interested in the asymptotic behaviour of $\beta_{\mu_m, \pi_m, \rho}(r, \alpha)$ under different settings for μ_m and π_m.

Theorem 5. *Fix $r \ge 0$, and consider the positive equicorrelated Gaussian model in Definition 2, where*

$$\lim_{m\to\infty} \mu_m = \mu \in [0, \infty], \quad \lim_{m\to\infty} \pi_m = \pi \in [0, 1].$$

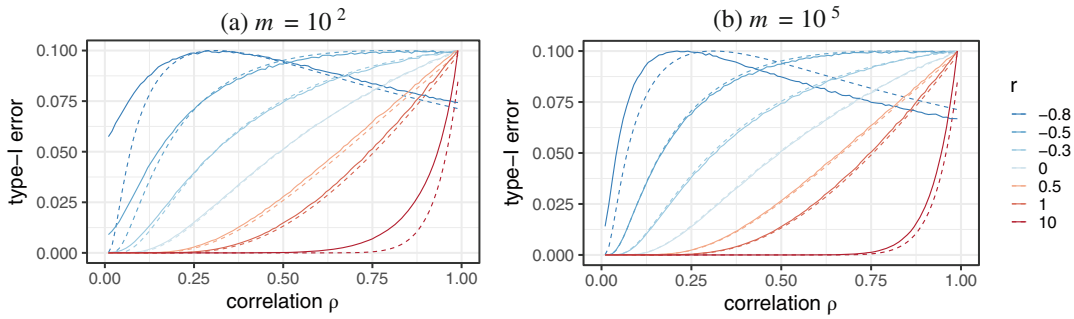

FIGURE 5 The asymptotic uniform type-I error $\tilde{\alpha}(\rho, r, \alpha)$ using the calibrated positively equicorrelated Gaussian test (dotted line, identical in both subplots), and the empirical point-wise type-I error $\tilde{\alpha}^\star_m(\rho, \alpha, r)$ (solid line) with $m = 10^2$ (left) and $m = 10^5$ (right) of different method for adjusting dependence via simulation (averaging over 10^6 trials) versus correlation given different $r > -1$ at confidence level $\alpha = 0.1$

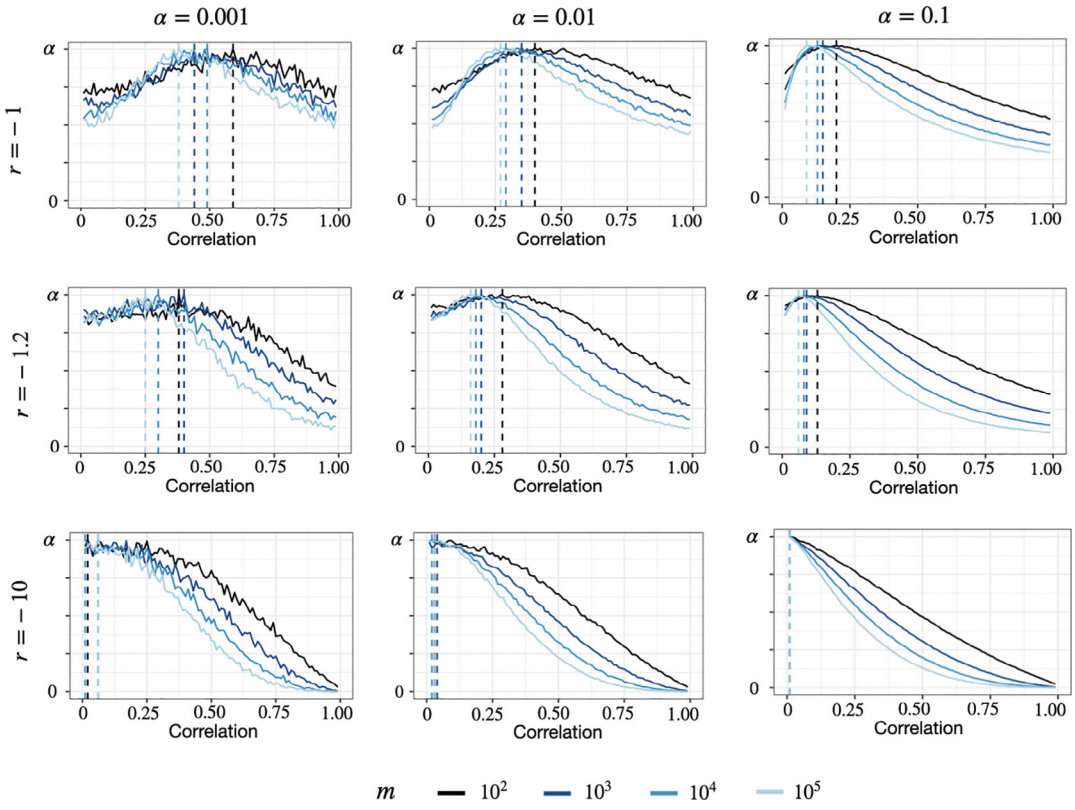

FIGURE 6 The empirical point-wise Type-I error $\tilde{\alpha}^{\star}_{m}(\rho, r, \alpha)$ (solid line) with different $m \in \{10^2, 10^3, 10^4, 10^5\}$ of different method for adjusting dependence via simulation (averaging over 10^6 trials) versus correlation given different $r \leq -1$ at confidence level $\alpha = 0.001, 0.01, 0.1$ in first, second, and third columns, respectively. The dashed vertical lines indicate the "worst case" correlation given fixed m, r, α, that is $\arg\max_{\rho \in [0,1]} \tilde{\alpha}^{\star}_{m}(\rho, r, \alpha)$.

Then for any $\alpha \in (0, (\frac{r}{r+1})^{\frac{1}{r}})$, $\rho \in [0, 1]$, *we have that the asymptotic power*

$$\lim_{m\to\infty} \beta_{\mu_m,\pi_m,\rho}(r,\alpha) = \lim_{m\to\infty} \Pr\left\{\pi_m g_{\rho,r}\left(Z_0 + \frac{\mu_m}{\sqrt{\rho}}\right) + (1-\pi_m) g_{\rho,r}(Z_0) \leq \alpha^r\right\} \in [\tilde{\alpha}(\rho, r, \alpha), 1], \tag{35}$$

with $g_{\rho,r}$ *defined in (31),* $\tilde{\alpha}(\rho, r, \alpha) \in [0, \alpha]$ *defined in (27), and* Z_0 *as a standard Gaussian random variable. In particular,*

- *if* $\pi = 1$, *then*

$$\lim_{m\to\infty} \beta_{\mu_m,\pi_m,\rho}(r,\alpha) = \begin{cases} 1, & \text{if } \mu = \infty; \\ \Phi\left(-g^{-1}_{\rho,r}(\alpha^r) + \frac{\mu}{\sqrt{\rho}}\right), & \text{if } 0 < \mu < \infty; \\ \tilde{\alpha}(\rho, r, \alpha), & \text{if } \mu = 0. \end{cases} \tag{36}$$

- *if* $0 < \pi < 1$, *then*

$$\lim_{m\to\infty} \beta_{\mu_m,\pi_m,\rho}(r,\alpha) = \begin{cases} \Phi\left(-g_{\rho,r}^{-1}\left(\frac{\alpha^r}{1-\pi}\right)\right), & \text{if } \mu = \infty; \\ \Pr\left\{\pi g_{\rho,r}\left(Z_0 + \frac{\mu}{\sqrt{\rho}}\right) + (1-\pi) g_{\rho,r}(Z_0) \le \alpha^r\right\}, & \text{if } 0 < \mu < \infty; \\ \tilde{\alpha}(\rho, r, \alpha), & \text{if } \mu = 0. \end{cases} \tag{37}$$

- *if* $\pi = 0$, *then* $\lim_{m\to\infty} \beta_{\mu_m,\pi_m,\rho}(r,\alpha) = \tilde{\alpha}(\rho, r, \alpha)$, *no matter what value that* μ *takes.*

The proof of Theorem 5 is in Appendix I, where we use a triangular-array version of the generalized law of large numbers. As a quick sanity check, expressions in (36) are always in $[\tilde{\alpha}(\rho, r, \alpha),\ 1]$, while expressions in (37) are always in $\left[\tilde{\alpha}(\rho, r, \alpha),\ \Phi\left(-g_{\rho,r}^{-1}\left(\frac{\alpha^r}{1-\pi}\right)\right)\right]$. Also we can verify some other intuitive facts: the power asymptotically goes to one when $\pi = 1$ and $\mu = \infty$, while it drops to the lower bound $\tilde{\alpha}(\rho, r, \alpha)$ in the worst case (i.e. $\pi = 0$ or $\mu = 0$). Theorem 5 also indicates some surprising findings that are somewhat exclusive to the case of $r > 0$: the asymptotic power will not go to one even when $\pi = 1$ if the signal strength is finite, that is $\mu < \infty$; a similar phenomenon occurs when $\mu = \infty$ but $\pi < 1$. These counter-intuitive behaviors happen due to the nature of the combination choices with $r > 0$, where the combination is essentially a weighted average of p-values with a monotonic increasing transformation, thus will be dominated by large p-values. Therefore, in the case of $r > 0$, as long as there is a nondiminishing part of observations that are most likely to generate large p-values (e.g., nonsignals or weak signals), then we will lose power.

To see an explicit example of Theorem 5, we consider the case that $r = 1$, $\pi \in (0, 1)$, with $\mu = \infty$ for simplicity. In this case the combination becomes $1 - \pi$ times the arithmetic mean of p-values $\left(\text{i.e. } \frac{1-\pi}{m}\sum_{i=1}^{m} p_{mi}\right)$. Therefore, from the definition the asymptotic power, we know it equals $\Phi\left(-g_{\rho,r}^{-1}\left(\frac{\alpha^r}{1-\pi}\right)\right)) = \Phi(-\infty) = 0$ from the definition of $g_{0,r}^{-1}$, which agrees with Theorem 5. This example indicates a more general phenomenon: when $r \ge 1$, as long as there are some nonsignals (i.e. $\pi < 1$), the asymptotic power under independence will always be 0 no matter how strong the signal is (i.e. how large μ is). On the other hand, if $\rho = 1$, then we have the asymptotic power equals $\frac{\alpha}{1-\pi}$, which will be close to one if and only if $\pi \to 1$. From this simple example, we justify for the behavior many (Vovk & Wang, 2020; Wilson, 2020) observed in experiments: that is the combination choice with $r \ge 1$ will be powerless unless there are many strong signals with heavy dependence.

In the following, we study the case when $r \le -1$. Firstly, we look at the setting with moderate signal strength, that is $\mu_m = o\left(\sqrt{\log m}\right)$. The following Theorem 6 shows that, as long as the signal is not strong enough, the asymptotic power for $r \le -1$ will always degenerate, no matter how dense the signal is.

Theorem 6. *Consider the positive equicorrelated Gaussian model in Definition 2 where*

$$\mu_m = o\left(\sqrt{\log m}\right), \quad \lim_{m\to\infty} \pi_m = \pi \in [0, 1],$$

For $r \le -1$, *we have that for all* $\rho \in [0, 1]$ *and* $\alpha > 0$,

$$\lim_{m\to\infty} \beta_{\mu_m,\pi_m,\rho}(r,\alpha) = \alpha \mathbf{1}\{\rho = 0\}. \tag{38}$$

The proof of Theorem 6 is in Appendix J, where we mainly use results about limitation of infinitely divisible random variables. Theorem 6 indicates that, as long as the signal is not strong enough, the combination with $r \leq -1$ will be powerless unless under independence. Despite of the observed robustness under dependency for cases of $r \leq -1$ in experiments conducted by Wilson (2019), the robustness will diminish as the number of hypotheses goes to infinity, and the method actually becomes highly sensitive to the dependence in the end. This phenomenon arises from the fact that the gap between the calibrated threshold grows at least sublinearly (i.e. $O(m^{\epsilon})$ with $\epsilon > 0$) for different ρ, therefore the conservativeness from calibration grows with the number of hypotheses, and results in high sensitivity to dependence in the end.

In the following, we study the setting when the signal is strong enough for the test to have power one. Specifically, that is when $\mu_m \geq O(\sqrt{\log m})$.

Theorem 7. *For $r \leq -1$, consider the positive equicorrelated Gaussian model in Definition 2 where*

$$\mu_m = \sqrt{2c \log m}, \quad \text{with } c > 0.$$

For all $\rho \in [0, 1]$, if further one of the following is satisfied:

(a) $\lim_{m\to\infty} \pi_m = \pi > 0$, *and* $\sqrt{c} > 1 - \sqrt{(1-\rho)}$;
(b) $\pi_m = m^{\gamma-1}$, *where* $0 < \gamma < 1$, *and* $\sqrt{c} > 1 - \sqrt{\gamma(1-\rho)}$,

then we have that, for all $\alpha > 0$,

$$\lim_{m\to\infty} \beta_{\mu_m, \pi_m, \rho}(r, \alpha) = 1. \tag{39}$$

The proof of Theorem 7 is in Appendix K, where we use a sandwiching argument, similar to Liu and Xie (2020). Theorem 7(a) indicates that, in order to achieve full power at different ρ, the signal strength needs to be stronger under heavy dependence. This conclusion agrees with the intuition since the power is fundamentally related to the tail of transformed p-value p_{mi}^r, which is thinner under heavy dependence while heavier under weak dependence. Theorem 7(b), following the sparse setting (Donoho & Jin, 2004), states that the asymptotic power will achieve one with probability one, as long as the signal strength c and signal sparsity γ achieve the detection boundary defined by ρ: $\sqrt{c} > 1 - \sqrt{\gamma(1-\rho)}$. Note that the detection boundary in (b) grows with ρ, which indicates that the signal needs to be stronger/denser to achieve the sweet spot under heavier dependence.

4 | EXPERIMENTS

In Section 3 we derive theoretical results of local test t_α^r (15) under equicorrelated Gaussian model (Definition 2) in an asymptotic setting ($m \to \infty$), which indicate behaviors that positive r works better for heavy dependence and weak dense signals, while negative r works better for weak dependence and strong sparse signals. In this section, we empirically present the evidence that the above behaviors of local test are largely preserved after going through closed testing.

As the theoretical results for local test in Section 3 are only asymptotic, while in closed testing, we need to consider all subsets of $[m]$; henceforth, our theoretical results will not be applicable for a large proportion of them. On the other hand, calibrating for subsets of size 1 to m is

computationally expensive; therefore, we use the following approximation, that is calibrating for a few sizes in $[m]$ and then interpolating to the whole $[m]$ (see Figure 7 for the case when $m = 1000$, $\alpha = 0.05$). This empirical calibration with interpolation works well in terms of maintaining correct error control and gives nontrivial power (see Figures 8–11).

Remark 6. Figure 7 provides reasonable evidence that for $r \geq 1$, the worst-case dependence is not achieved by $\rho = 1$ (perfect correlation) because if that was the case, there would be no violation of type-I error, but we observe above that the achieved error is larger than $\alpha = 0.05$.

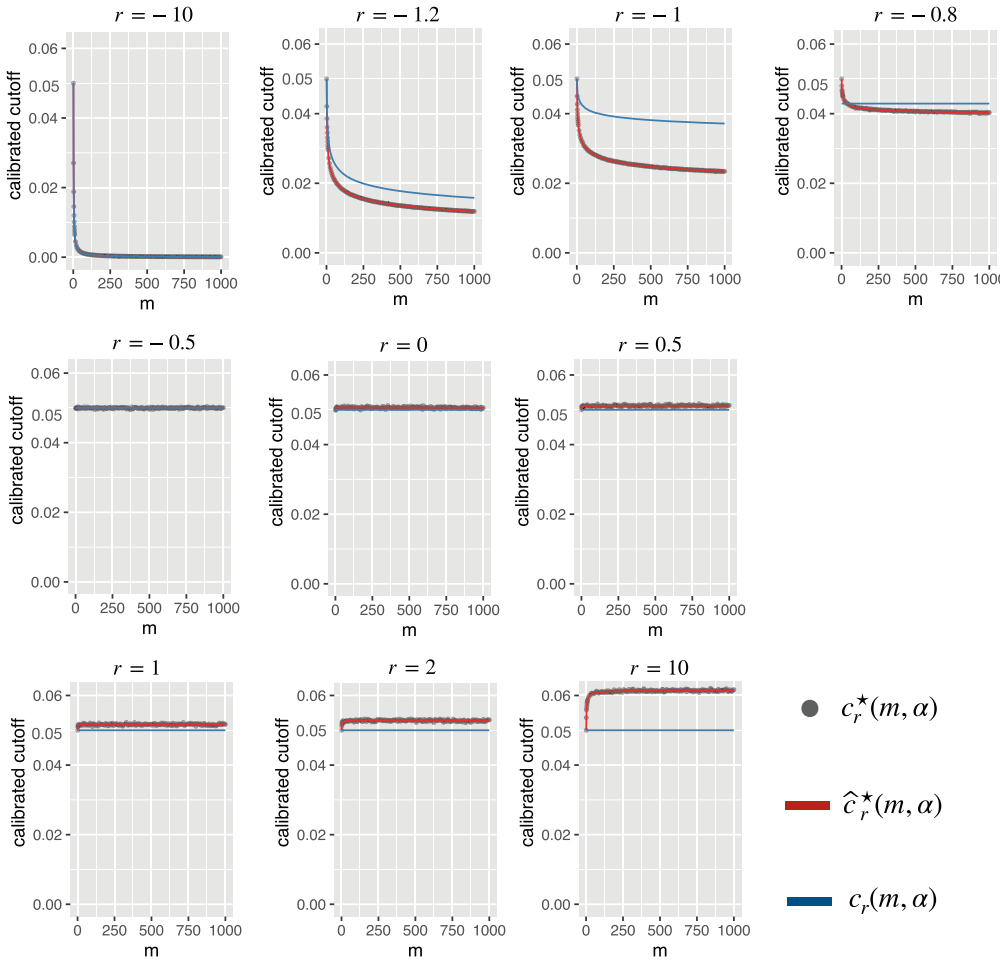

FIGURE 7 Calibration under ρ-equicorrelated Gaussian for $m \in [1000]$ for a worst case $\rho \in [0, 1]$ (calculated using a grid of width 0.01). We compute the empirical α level calibrated cutoff $c_r^\star(m, \alpha)$ with $\alpha = 0.05$ (black dots) for $\left(\frac{1}{m}\sum_{i=1}^{m} p_i^r\right)^{\frac{1}{r}}$ with grid points $m \in \{1, 2, \ldots, 10, 15, 20, \ldots, 1000\}$ and $r \in \{-10, -1.2, -1, -0.8, -0.5, 0, 0.5, 1, 2, 10\}$ via averaging over 10^6 trials. Then we approximate $c_r^\star(m, \alpha)$ for all $m \in [1000]$ via fitting a smooth line $\widehat{c}_m^\star(\alpha, r)^{\frac{1}{r}}$ (red line) for the whole $m \in [1000]$, and use the fitted value as our final calibration. As for comparison, we also plot the theoretical calibrated cutoff $c_r(m, \alpha)$ (see Definition (26)) derived for pointwise asymptotic type-I control in Section 3. In addition, for small m ($1 \sim 10$), we just use empirical calibration for accurateness.

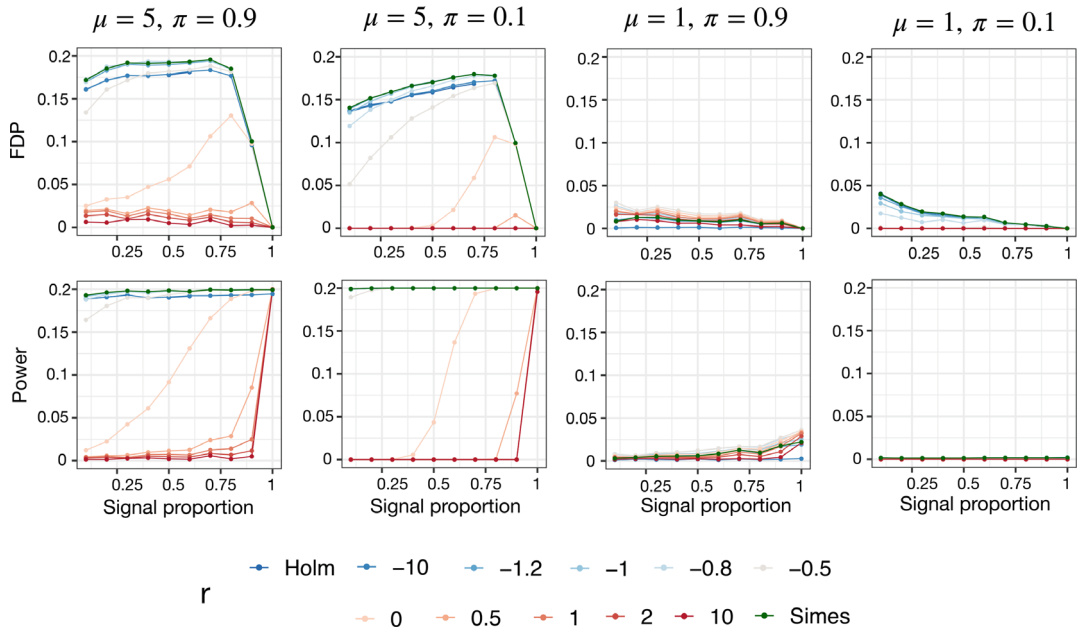

FIGURE 8 The empirical false discovery proportion and power versus the signal proportion π under different settings using fitted calibration in Figure 7 and algorithm in Theorem 3, with $\alpha = 0.05$, $\gamma = 0.2$, $m = 200$, averaging over 1000 trials

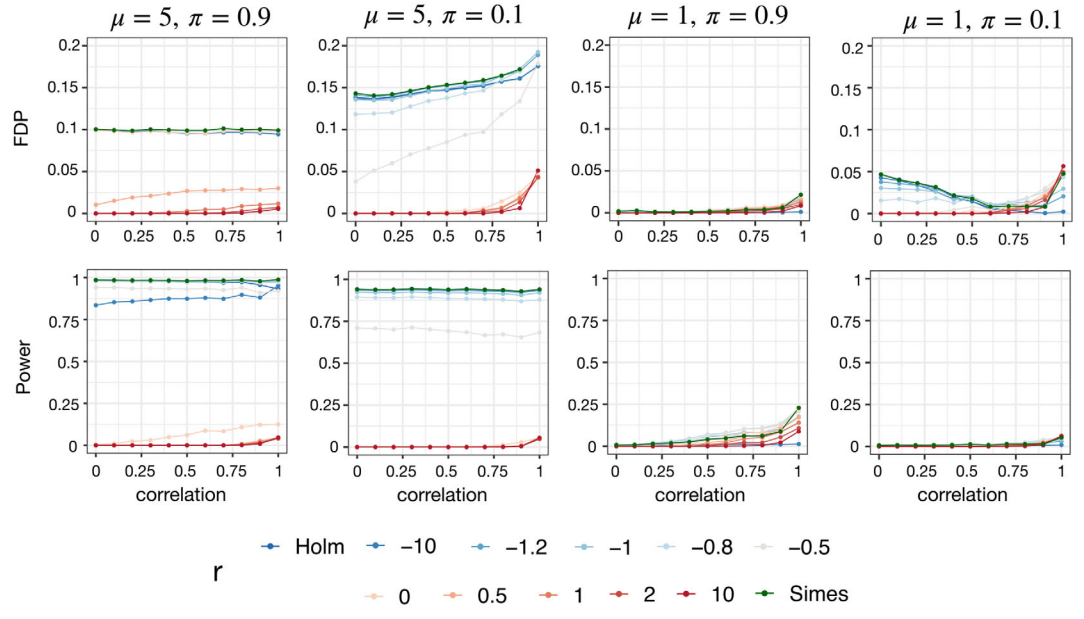

FIGURE 9 The empirical false discovery proportion and power versus correlation ρ under different settings using fitted calibration in Figure 7 and algorithm in Theorem 3, with $\alpha = 0.05$, $\gamma = 0.2$, $m = 200$, averaging over 1000 trials

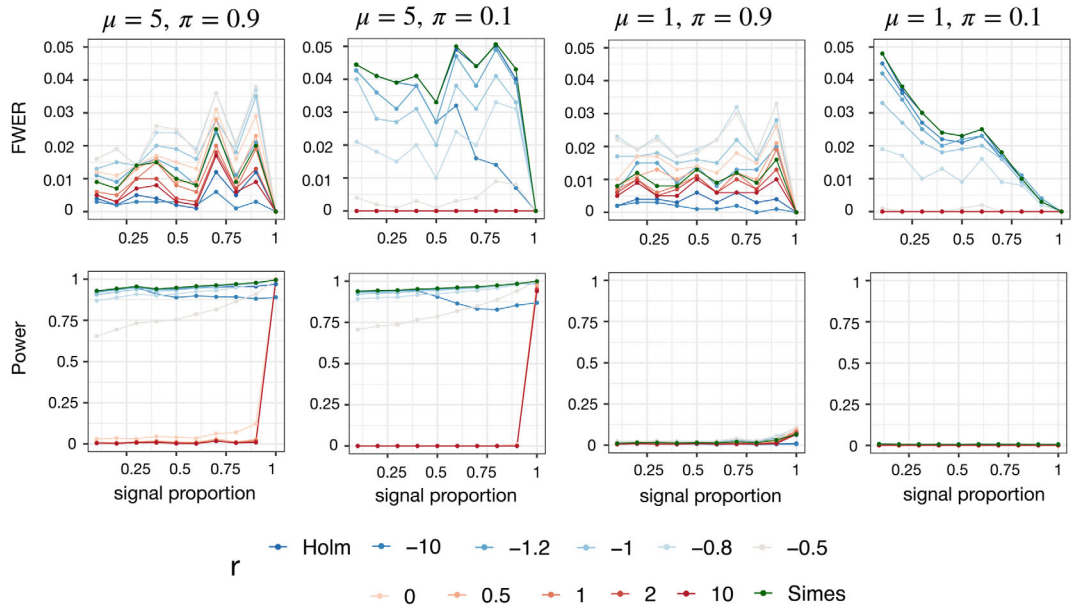

FIGURE 10 The empirical FWER and power versus signal proportion π under different settings using fitted calibration in Figure 7 and algorithm in Theorem 3, with $\alpha = 0.05$, $m = 200$, averaging over 1000 trials

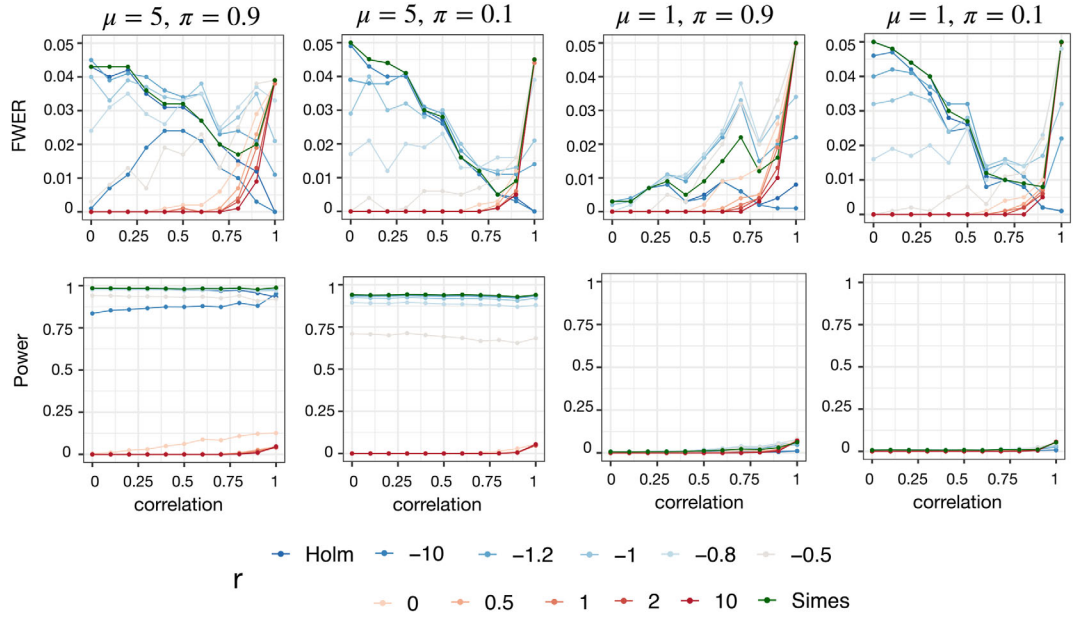

FIGURE 11 The empirical FWER and power versus correlation ρ under different settings using fitted calibration in Figure 7 and algorithm in Theorem 3, with $\alpha = 0.05$, $m = 200$, averaging over 1000 trials

In the rest of this section, we consider $m = 200$ tests, each based on different samples which are not necessarily independent, particularly, we assume the positively equicorrelated Gaussian model (Definition 2) among samples of the m tests and test whether a given set of data has zero mean. In particular, we consider $\mu_m \equiv \mu$ for all m, and $\pi_m \equiv \pi$ for all m. In Figures 8–11 we investigate the algorithms presented in Theorem 3, that is Algorithm 1 for finding the largest subset of m such that FDP is controlled under γ, and Algorithm 2 for finding the largest subset of $[m]$ such that FWER is controlled under α. Specifically, Figures 8 and 9 show the results for finding the largest subset of $[m]$ with FDP controlled under $\gamma = 0.2$, and Figures 10 and 11 show the results for finding the largest subset of $[m]$ with FWER controlled under $\alpha = 0.05$. We can see that, both FDP and FWER are controlled as we wanted, while $r < 0$ often has non-trivial power (close to one) comparing with $r \geq 0$ when signals are strong enough (i.e. μ large enough), while they are both powerless otherwise (i.e. μ not large enough). For weak signals specifically, we observe that $r > 0$ have higher power comparing to $r < 0$, especially under strong dependence (i.e. $\rho \gg 0$) and high signal density (i.e. $\pi \gg 0$). These findings generally agree with our asymptotic theory for local test in Section 3, that is $r < 0$ achieves almost perfect power under setting with sparse strong signals, but is powerless when signals are not strong enough, in which case $r > 0$ works better (especially given heavy dependence and dense signals).

5 | CONCLUSION

In this paper, we investigate the general case of closed testing with local tests that adopt a special property we called *separability*, that is, the test is a function of summation of test scores for the individual hypothesis. With *separability*, *symmetry* and *monotonicity* in local tests, we derive a class of novel, fast algorithms for various types of simultaneous inference. These algorithms have been implemented in the R package `sumSome`[5]. We pair our algorithms with recent advances in separable global null test, that is, the generalized mean-based methods summarized (Vovk & Wang, 2020), and obtain a series of simultaneous inference methods that are sufficient to handle many complex dependence structures and signal compositions. We provide guidance on choosing from these methods adaptively via theoretical investigation of the conservativeness and sensitivity for different choices of local tests in an equicorrelated Gaussian model. Specifically, we found that within the family of simultaneous inference methods using local tests introduced in (16):

- when signals are weak, all methods are powerless, while the ones with positive r perform a bit better when signals are dense and highly correlated.
- when signals are strong, methods with negative r are often able to achieve full power, and methods with positive r are often still powerless; except in the case when signals are also dense, they are comparable.

We leave the following problems for future work. First, we note that the surrogate type-I error in (27) sometimes does not agree very well with the true type-I error in simulations (see the discussion of Figure 6). We think this arises from the fact that the surrogate type-I error is based on asymptotic approximation. That is, the number of hypotheses m goes to infinity. Meanwhile, we have only tried limited dimensions empirically (i.e., $m = 10^5$). We think that the story will be more coherent when m is much larger because we suspect that the convergence occurs very slowly. Nevertheless, how to derive tight and efficient calibration explicitly for small m may be worth more attention. Secondly, in this work, we mainly focus on the equicorrelated Gaussian case, while the

derivation of a tight calibration under arbitrarily correlated Gaussians will be intriguing, though much harder. Finally, the theoretical power analysis is conducted only for the local test; formal theoretical analysis after closure would be desirable, though we expect that to be much harder.

ENDNOTES

[1]Note that generally the functions f and C can also depend on the scores themselves, however in this paper we consider specifically the case when f and C is fixed, and f as function the scores only, and C as function of the cardinality of hypotheses set only. These cases already consist of a large proportion of existed global null tests, and simplify the analysis throughout the paper.

[2]A closed testing is *consonant* if the local tests for every composite hypothesis $S \in 2^{[m]}$ are chosen in such a way that rejection of S after closure implies a rejection of at least one of its elementary hypothesis after closure.

[3]To see this, observe that $\sup_{\rho\in[0,1]} \tilde{\alpha}_m(\rho, r, c) \geq \tilde{\alpha}_m(\rho, r, c)$ for all $\rho \in [0, 1]$ and all m. Taking $limsup_m$ on both sides maintains the inequality, as does taking a further $\sup_\rho$ on both sides.

[4]Here we use the approximation $\alpha_{-1,m} \sim \log m$ as $m \to \infty$ in Vovk and Wang (2020, Proposition 6) to make the expression cleaner. But in Figure 4, we use the numerical solution as stated in Lemma 1.

[5]https://github.com/annavesely/sumSome

[6]To see this, note that $\frac{1}{m}\sum_{i=1}^{m} h_r(X_i) \leq C$ implies that $\max_{i\in[m]} h_r(X_i) \leq mC$, which happens if and only if $\min_{i\in[m]} X_i \geq -\Phi^{-1}(mC^{\frac{1}{r}}) > 0$, where the last inequality is because $C < \frac{1}{2^r m^r}$.

ORCID

Jinjin Tian https://orcid.org/0000-0003-3537-6430

REFERENCES

Blanchard, G., Neuvial, P., & Roquain, E. (2020). Post hoc confidence bounds on false positives using reference families. *Annals of Statistics*, *48*(3), 1281–1303.

Chen, Y., Liu, P., Tan, K. S., & Wang, R. (2020). Trade-off between validity and efficiency of merging *p*-values under arbitrary dependence. *arXiv preprint arXiv:2007.12366*.

Dobriban, E. (2020). Fast closed testing for exchangeable local tests. *Biometrika*, *107*(3), 761–768.

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

Fisher, R. A. (1992). *Statistical methods for research workers*. In *Breakthroughs in statistics* (pp. 66–70). Springer.

Genovese, C. R., & Wasserman, L. (2006). Exceedance control of the false discovery proportion. *Journal of the American Statistical Association*, *101*(476), 1408–1417.

Gnedenko, B. V., & Kolmogorov, A. N. (1954). *Limit distributions for sums of independent random variables Addison-Wesley Mathematics Series* (). Addison-Wesley.

Goeman, J. J., Hemerik, J., & Solari, A. (2021). Only closed testing procedures are admissible for controlling false discovery proportions. *The Annals of Statistics*, *49*(2), 1218–1238.

Goeman, J. J., Meijer, R. J., Krebs, T. J. P., & Solari, A. (2019). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *Biometrika*, *106*(4), 841–856.

Goeman, J. J., & Solari, A. (2011). Multiple testing for exploratory research. *Statistical Science*, *26*(4), 584–597.

Gordon, R. D. (1941). Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *The Annals of Mathematical Statistics*, *12*(3), 364–366.

Katsevich, E., & Ramdas, A. (2020). Simultaneous high-probability bounds on the false discovery proportion in structured, regression and online settings. *Annals of Statistics*, *48*(6), 3465–3487.

Liu, Y., & Xie, J. (2020). Cauchy combination test: A powerful test with analytic *p*-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, *115*(529), 393–402.

Marcus, R., Eric, P., & Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, *63*(3), 655–660.

Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability*, *14*(3), 623–632.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star S. A., & Williams Jr R. M. (1949). *The American soldier: Adjustment during army life*. (studies in social psychology in world war ii). Prince university press.

Uchaikin, V. V., & Zolotarev, V. M. (2011). *Chance and stability: Stable distributions and their applications*. Walter de Gruyter.

Vesely, A., Finos, L., & Goeman, J. J. (2021). Permutation-based true discovery guarantee by sum tests. *arXiv preprint arXiv:2102.11759*.

Vovk, V., Wang, B., & Wang, R. (2022). Admissible ways of merging p-values under arbitrary dependence. *The Annals of Statistics*, *50*(1), 351–375.

Vovk, V., & Wang, R. (2020). Combining p-values via averaging. *Biometrika*, *107*(4), 791–808.

Wilson, D. J. (2021). The Lévy combination test. *arXiv preprint arXiv:2105.01501*.

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

Wilson, D. J. (2020). Generalized mean *p*-values for combining dependent tests: Comparison of generalized central limit theorem and robust risk analysis. *Wellcome Open Research*, *5*(55), 55.

How to cite this article: Tian, J., Chen, X., Katsevich, E., Goeman, J., & Ramdas, A. (2023). Large-scale simultaneous inference under dependence. *Scandinavian Journal of Statistics*, *50*(2), 750–796. https://doi.org/10.1111/sjos.12614

APPENDIX A. PROOF FOR THEOREM 1

We have $t_\alpha(S) = 1$ if and only if $p(J) \le \alpha$ for all $S \subseteq J \subseteq [m]$, which happens if and only if $\alpha \ge \max_{S\subseteq J\subseteq[m]} p(J)$. Therefore,

$$\overline{p}(S) = \max_{S\subseteq J\subseteq[m]} p(J).$$

By monotonicity, we have that if $J \supseteq S$ then $p(J) \ge p(S \cup J^\star_{|J|-|S|})$, where $J^\star_i$, is the set of the indices of the i largest p-values in S^c. Therefore,

$$\overline{p}(S) = \max_{|S|\le i\le m} p(S \cup J^\star_{i-|S|}) = \max_{0\le i\le|S^c|} p(S \cup J^\star_i).$$

Since $J^\star_0 \subseteq \ldots \subseteq J^\star_{|S^c|}$, it is clear that this expression can be calculated in $O(|S^c|) = O(m)$ time.

APPENDIX B. MORE GENERAL FRAMEWORK OF LOCAL TESTS DESIGN

We start with defining some terminology. Recall that a local test is an indicator function of whether to reject S or not:

$$t_\alpha(S) : \quad \mathbb{R}^{|S|} \to \{0, 1\}. \tag{B1}$$

We consider t_α of the following form:

$$t_\alpha(S) = \mathbf{1}\{f(|S|; (T_i)_{i\in S}) \le \alpha\}, \tag{B2}$$

where the function f depends on the size of S and the vector of scores. Given local tests of this form, there are two commonly satisfied conditions—*symmetry* and *monotonicity*—which makes the computation manageable: quadratic time shortcuts for simultaneous FDP inference and FWER control have been developed by Goeman and Solari (2011) and Dobriban (2020), respectively, under these two conditions.

Condition 1 (Monotonicity). A local test of form (B2) is called monotonic if for any $s \geq 1$, any two sets of scores $(T_1,\ldots,T_s)$ and $(T'_1,\ldots,T'_s)$ with $T_i \leq T'_i$ for all $i = 1,\ldots,s$, we have

$$f(s;T_1,\ldots,T_s) \geq f(s;T'_1,\ldots,T'_s). \tag{B3}$$

Monotonicity is a reasonable requirement for a local test. Several well-known global null tests are monotonic, for example, Fisher's test, Simes' test, Bonferroni test, etc; as well as the tests based on generalized means Vovk and Wang (2020).

Condition 2 (Symmetry) 2. A local test t_α of form (B2) is called symmetric if for any $s \geq 1$, any set of scores $(T_1,\ldots,T_s)$, and any permutation $(i_1,\ldots,i_s)$ of $(1,\ldots,s)$, we have

$$f(s;T_1,\ldots,T_s) = f(s;T_{i_1},\ldots,T_{i_s}). \tag{B4}$$

Another condition that we find could further reduce computation time is *separability*. This condition is relatively under-emphasized in the past closed testing literature; however, it has a long history in the global null testing literature with an increased recent interest (Chen et al., 2020; Vovk & Wang, 2020; Wilson, 2019).

Condition 3 (Separability) 3. A local test of form (B2) is called separable if for $s \geq 1$, and a set of scores $(T_1,\ldots,T_s)$, there exists a series of transformation functions on $\mathbb{R}$ $\{h_i\}_{j=1}^s$ and a function g on $\mathbb{R}^2$ such that

$$t_\alpha(s;T_1,\ldots,T_s) = \mathbf{1}\left\{\sum_{j=1}^{s} h_j(T_j) \leq g(|S|,\alpha)\right\}. \tag{B5}$$

Recall the class of local tests T_α defined in (16) of the main paper. It is easy to check that each of its element $t_\alpha^{(r)}$ is *monotonic*, *symmetric* for all r, and *separable* iff $r \neq \pm\infty$.

Remark 7. A local test t_α of form (B2) with both *symmetry* and *separability* must admit the following form:

$$t_\alpha(S) = \mathbf{1}\left\{\sum_{i=1}^{|S|} h(T_{s_i}) \leq g(|S|,\alpha)\right\}, \tag{B6}$$

that is the transformation functions in the summation are the same for each hypothesis.

APPENDIX C. PROOF FOR THEOREM 2

Without loss of generality, we assume that all the scores (e.g. p-values) are already sorted in a descending order, that is $T_1 \geq T_2 \geq \ldots \geq T_m$. Denote $S = \{i_1, i_2,\ldots,i_s\}$, with $1 \leq s \leq m$ and $i_1 < i_2, < \cdots < i_s$; and $S^c = \{j_1, j_2,\ldots,j_{m-s}\}$ as the complement set of S, with $j_1 < j_2, < \cdots < j_{m-s}$.

To prove the validity of Algorithm 1, we first focus on the crucial line 9, and claim that when event

$$\mathcal{E} := \{\text{line 9 is evaluated with } \ k \leq \min\{s,a\}\}, \tag{C1}$$

happens, the following four statement are true.

(i) $Q = \sum_{j=1}^{k+b} u_j + \sum_{l=1}^{a-k-b} v_l$;
(ii) $b \geq 0$;
(iii) $v_{a-k-b} \geq u_{k+b+1}$;
(iv) $u_{k+b} \geq v_{a-k-b+1}$, if $b > 0$.

Note that $\sum_{l=1}^{0} v_l$ is defined to be 0, and not $v_1 + v_0$; and $\sum_{j=1}^{0} u_j$ is defined to be 0, and not $u_1 + u_0$.

We show this by induction. The first time event $\mathcal{E}$ defined in (C1) happens, we have $k = a = 1, b = 0$ and $Q = u_1$, so (i) and (ii) hold, (iii) holds since $v_0 \geq u_2$, and for (iv), since $b = 0$ there is nothing to prove.

Additionally, we need to prove that when $b > 0$ for the first time, (iv) is satisfied. Note that the only way that $b > 0$ for the first time, is through the satisfaction of the condition in line 3, specifically $u_{k+b_0+1} \geq v_{a-k-b_0}$ with $b_0 = 0$. The reason is that a always increases by one in each while-loop of line 2, and after the first while-loop, b can be at most 0, and we have $a = 2$. Therefore under the condition that $u_{k+b_0+1} \geq v_{a-k-b_0}$ with $b_0 = 0$, the algorithm will go through line 5, and has $b = b_0 + 1 > 0$, and $u_{k+b} \geq v_{a-k-b+1}$. That is (iv) is satisfied, when b satisfies $b > 0$ for the first time.

Now assume that (i)–(iv) hold the previous time the event $\mathcal{E}$ happens. Let a_0, k_0, b_0, and Q_0 be the value of a, k, b, and Q during that previous step. There are five routes for event $\mathcal{E}$ to happen again, which we can characterize by the way a, k, b are updated. We will discuss these routes one by one.

1. **Line *9* → *10* → *11* → *15* → $\mathcal{E}$**. In this case we update $a = a_0; b = b_0 - 1; k = k_0 + 1$. We have $b \geq 0$ since $b_0 > 0$. We have $Q = Q_0$, and (i) holds since $k + b = k_0 + b_0$. By the induction hypothesis,

$$v_{a-k-b} = v_{a_0-k_0-b_0} \geq u_{k_0+b_0+1} = u_{k+r+1}.$$

   If $b > 0$, then also $b_0 > 0$, so, by the induction hypothesis,

$$u_{k+b} = u_{k_0+b_0} \geq v_{a_0-k_0-b_0+1} = v_{a-k-b+1}.$$

2. **Line *9* → *10* → *13* → *15* → $\mathcal{E}$**. In this case we update $a = a_0$; $b = b_0 = 0$; $k = k_0 + 1$. Clearly, (ii) holds. We have

$$Q = \sum_{j=1}^{k_0+1} u_j + \sum_{l=1}^{a_0-k_0-1} v_l,$$

   which reduces to (i). By the induction hypothesis,

$$v_{a-k-b} = v_{a_0-k_0-b_0-1} \geq v_{a_0-k_0-b_0} \geq u_{k_0+b_0+1} = u_{k+b} \geq u_{k+b+1},$$

   so (iii) follows. Since $b = 0$ there is nothing to prove (iv).
3. **Line *9* → *2* → *3* → *4*, *5* → $\mathcal{E}$**. In this case we update $a = a_0 + 1; b = b_0 + 1; k = k_0$. We get (ii) from the induction assumption since $b = b_0 + 1 \geq 1$. We obtain (i) since

$$Q = \sum_{j=1}^{k_0+b_0+1} u_j + \sum_{l=1}^{a_0-k_0-b_0} v_l.$$

   By the induction hypotheses,

$$v_{a-k-b} = v_{a_0-k_0-b_0} \geq u_{k_0+b_0+1} = u_{k+b} \geq u_{k+b+1}.$$

   Also, $b > 0$ and

$$u_{k+b} = u_{k_0+b_0+1} \geq v_{a-k_0-b_0} = v_{a-k-b+1}.$$

4. **Line *9* → *2* → *3* → *7* → $\mathcal{E}$**. We update $a = a_0 + 1$, $k = k_0$ and $b = b_0$. We get (ii) trivially, and (i) since

$$Q = \sum_{j=1}^{k_0+b_0} u_j + \sum_{l=1}^{a_0-k_0-b_0+1} v_l.$$

We have (iii) since

$$v_{a-k-r} = v_{a-k_0-b_0} \geq u_{k_0+b_0+1} = u_{k+b+1}.$$

By the induction hypothesis, we get (iv), since,

$$u_{k+b} = u_{k_0+b_0} \geq v_{a_0-k_0-b_0+1} = v_{a-k-b} \geq v_{a-k-b+1}.$$

5. **Line *9* → *10* → *13* → *15* → *9* → *2* → *7* → $\mathcal{E}$**. First we use proof by contradiction that this case happens only if $k_0 = \min\{a_0, s\}$: Otherwise, for step "***9* → *2***" to happen in the route, we need $Q_0 + u_{k_0+1} - v_{a_0-k_0} \leq C(a_0, \alpha)$ to break the while loop in line 9. This cannot happen since we need $b_0 = 0$ to reach line 13 in this route, which indicates b_0 was not updated in line 5 (otherwise $b_0 > 0$ from our induction hypothesis (ii)), that is, $u_{k_0+b_0+1} > v_{a_o-k_0-b_0}$ in line 3. Therefore we in fact have $Q_0 + u_{k_0+1} - v_{a_0-k_0} > Q_0 > C(a_0, \alpha)$, which is a contradiction to what we require.
Consequently, we update $a = a_0 + 1$, $k = k_0 + 1$, and, since $u_{k+1} \leq v_0$ by definition of v_0, $b = b_0 = 0$. So first (ii) holds, and also we get (i) since

$$Q = \sum_{j=1}^{k_0} u_j + \sum_{l=1}^{a_0-k_0} v_l + u_{k_0+1} - v_{a_0-k_0} + v_{a-k} = \sum_{j=1}^{k} u_j + \sum_{l=1}^{a-k} v_l = \sum_{j=1}^{k+b} u_j + \sum_{l=1}^{a-k-b} v_l.$$

Moreover, since $a = k + b$ and $b = 0$, we have (iii) because $v_{a-k-b} = v_0 \geq u_1 \geq u_{k+b+1}$. There is nothing to prove (iv) since $b = 0$.

Since we have exhausted the possibilities to get from one happening of event $\mathcal{E}$ to the next, the above analysis proves (i)–(iv). It follows from (i)-(iv) that, in line 9, $Q = W_{a,k}$, where

$$W_{a,k} = \max\{Q_I : |I \cap S| \geq k, |I| = a\}.$$

To see why this is true, note that (i) Q is a sum of a terms, of which at least k terms are from S. The sum is the largest possible such sum since the k largest scores in S are used, and by (iii) and (iv), of the $a - k$ remaining scores, the smallest score that is included in the sum is larger than or equal to the largest score that is not included. Note that, if $k \leq k'$, $W_{a,k} \geq W_{a,k'}$.

Now, suppose $\bar{e}_\alpha(S) = e > 0$. Then there exists some $I \subseteq S$ with $|I| = e$ and some $J \supseteq I$ such that $Q_J > g(|J|, \alpha)$. In the algorithm, if $a = |J|$ and $k \leq e$, we have $Q = W_{a,k} \geq W_{a,e} \geq Q_J > g(|J|, \alpha)$, so the algorithm enters the *while* loop in line 9, incrementing k while keeping a fixed. This step is repeated at least until $k \geq e + 1$. Since k is non-decreasing in the steps of the algorithm, it returns $k - 1 \geq e$. The same holds trivially if $e = 0$.

If $\bar{e}_\alpha(S) = e < s$, then for every $I \subset S$ with $|I| > e$ we have $\bar{t}_\alpha(I) = 1$, so for all $J \supseteq I$, we have $Q_J \leq g(|J|, \alpha)$. In particular, this holds for the worst case set, so for every $e + 1 \leq a \leq m$, we have $W_{a,e+1} \leq g(a, \alpha)$. If $k = e + 1$, therefore, the algorithm never enters the *while* loop in line 9, and

consequently never increments k further. The algorithm therefore ends with $k \le e+1$ and returns $k-1 \le e$. The same holds trivially if $e = s$.

To sum up, since $k-1 \ge e$ and $k-1 \le e$, we have $k-1 = e$.

Finally, we prove that the algorithm takes $O(m)$ time to run. First, note that there are m for-loop iterations. In each for-loop iteration, it is obvious that apart from the while-loop, the algorithm takes constant time. For the while-loop part, we can additionally show that the total calculations that it takes over all the for-loop iteration is at most s. A key observation for the proof is that k can only be updated through the while-loop part, and k always increases by one each time going through one while-loop iteration. Since the global upper bound for k is s, the number of total iterations for the while-loop over all for-loop iterations is at most s. Since each while-loop iteration also takes constant time, the algorithm takes at most $m+s$ steps of $O(1)$ calculation. In conclusion, the algorithm takes $O(m)$ time to run.

APPENDIX D. DISCUSSION OF CONSONANCE

Firstly we formally state the definition of consonance.

Definition 3 (Consonance). A closed testing procedure is *consonant* if the local tests for every composite hypothesis $S \in 2^{[m]}$ are chosen in such a way that rejection of S after closure implies rejection of at least one of its elementary hypotheses after closure.

Lemma 2. *The closed testing procedure using local test $t_\alpha^{(r)}$ defined in (15) is consonant if and only if $r = \pm\infty$.*

Proof. We will prove this proposition by analyzing different r case by case. Firstly, we show that closed testing using local test $t_\alpha^{(r)}$ (15) is consonant when $r = \pm\infty$.

1. When $r = \infty$, note that

$$t_\alpha^{(\infty)}(S) = \mathbf{I}\left\{\max_{i\in S} p_i \le \alpha\right\}. \tag{D1}$$

   Therefore, rejecting S after closure implies rejecting all the sets containing it locally, including the set $[m]$, which in turn indicates rejection of all the sets locally, and after closure as well, therefore trivially, we have all the subsets of S being rejected after closure. In conclusion, the corresponding closed testing when $r = \infty$ is consonant.
2. When $r = -\infty$, note that

$$t_\alpha^{(-\infty)}(S) = \mathbf{I}\left\{|S|\min_{i\in S} p_i \le \alpha\right\}. \tag{D2}$$

   Therefore $\overline{t_\alpha}^{-\infty}(S) = 1$ implies

$$t_\alpha^{-\infty}(B) = 1 \quad \text{for all } B \in \mathcal{B} := \{I \subseteq [m] : S \subset I\}, \tag{D3}$$

   and particularly $t_\alpha^{-\infty}(S) = 1$, which in turn gives us

$$t_\alpha^{-\infty}(A) = 1 \quad \text{for all } A \in \mathcal{A} := \left\{I \subseteq [m] : I \subset S, \min_{i\in S} p_i \in I\right\}, \tag{D4}$$

   from the expression in (D2).

   On the other hand, note that for some $A \in \mathcal{A}$, and $J \supseteq A$, we have either $J \supseteq S$, or $|J| \not\supseteq S$. In the former case, J is rejected locally due to fact (D3). In the later case, if $|J| \le |S|$, we have

$|J| \min_{i\in J} p_i \le |S| \min_{i\in A} p_i = |S| \min_{i\in S} p_i \le \alpha$; if $|J| > |S|$, then there must exist $B \in \mathcal{B}$ such that $|B| = |J|$, which implies $|J| \min_{i\in J} p_i = |B| \min_{i\in B} p_i \le \alpha$ due to fact (D4). Therefore J is still rejected locally. Hence there exists at least one subset A of S that is rejected after closure, that is the corresponding closed testing is consonant.

Then we use counter examples to show that closed testing using local test $t_\alpha^{(r)}$ (15) is not consonant when $r \neq \pm\infty$.

1. When $0 < r < \infty$, note that
$$t_\alpha^{(r)} = \mathbf{I}\left\{ \sum_{i=1}^m p_i^r \le \frac{m\alpha^r}{2(r+1)} \right\}.$$
Let $\beta_{r,\alpha} = \frac{\alpha^r}{2(r+1)}$, the local testing rule becomes $\sum_{i=1}^m p_i^r \le m\beta_{r,\alpha}$. Note that $0 \le \beta_{r,\alpha} \le \frac{1}{2(r+1)} \le 1/2$ for any $r > 0$. For the case $m = 3$, and $p_1^r = \beta_{r,\alpha}/3$, $p_2^r = \beta_{r,\alpha}/2$, $p_3^r = 2\beta_{r,\alpha}$, we have $p_1^r + p_2^r + p_3^r < 3\beta_{r,\alpha}$, $p_1^r + p_2^r < 2\beta_{r,\alpha}$, while $p_1^r + p_3^r > 2\beta_{r,\alpha}$ and $p_1^r < \beta_{r,\alpha}$. Therefore, we reject $H_1 \cap H_2 \cap H_3$, $H_1 \cap H_2$ but neither H_1 nor H_2 after closure, therefore the rejection $H_1 \cap H_2$ is not consonant.
2. When $r = 0$, note that
$$t_\alpha^{(0)} = \mathbf{I}\left\{ \sum_{i=1}^m \log\frac{1}{p_i} > m \log\frac{e}{\alpha} \right\}.$$
Let $\beta_\alpha = \log\frac{e}{\alpha}$, and $q_i = \log(\frac{1}{p_i})$, then the local testing rule becomes $\sum_{i=1}^m q_i \ge m\beta_\alpha$. Note that $1 \le \beta_\alpha < \infty$, and $1 \le q_i \le \infty$. For $m = 3$, let $\alpha = e^{0.9}$, and $q_1 = 1.6\beta_\alpha$, $q_2 = 1.4\beta_\alpha$, $q_3 = 0.1\beta_\alpha$, therefore we will reject $H_1 \cap H_2 \cap H_3$ and $H_1 \cap H_2$ after closure, but neither H_1 nor H_2, which indicates that the rejection $H_1 \cap H_2$ is not consonant.
3. When $-1 < r < 0$, note that
$$t_\alpha^{(r)} = \mathbf{I}\left\{ \sum_{i=1}^m p_i^r \ge \frac{m\alpha^r}{2(r+1)} \right\}.$$
Let $\beta_{r,\alpha} = \frac{\alpha^r}{2(r+1)}$, then the local testing rule becomes $\sum_{i=1}^m p_i^r \ge m\beta_{r,\alpha}$. Note that $1/2 \le \frac{1}{2(r+1)} \le \beta_{r,\alpha} < \infty$ for any $-1 < r < 0$. For the case $m = 3$, let $\alpha = \sqrt[r]{20(r+1)}$, and $p_1^r = 1.6\beta_{r,\alpha}$, $p_2^r = 1.4\beta_{r,\alpha}$, $p_3^r = 0.1\beta_{r,\alpha}$. Note the fact that $\beta_{r,\alpha} \ge 10$, we have that $p_1^r + p_2^r + p_3^r > 3\beta_{r,\alpha}$, $p_1^r + p_2^r > 2\beta_{r,\alpha}$, while $p_1^r + p_3^r < 2\beta_{r,\alpha}$ $p_2^r + p_3^r < 2\beta_{r,\alpha}$. Therefore, we reject $H_1 \cap H_2 \cap H_3$, $H_1 \cap H_2$ but neither H_1 nor H_2 after closure, therefore the rejection $H_1 \cap H_2$ is not consonant.
4. When $r = -1$, note that
$$t_\alpha^{(-1)} = \mathbf{I}\left\{ \sum_{i=1}^m \frac{1}{p_i} \ge \frac{em\log m}{\alpha} \right\}.$$
Let $\beta_\alpha = \frac{e}{\alpha}$, $q_i = 1/p_i$, then the testing rule becomes $\sum_{i=1}^m q_i \ge m\log m\beta_\alpha$. For the case $m = 5$, let $q_1 = q_2 = 2\beta_\alpha \log 5$, $q_3 = q_4 = q_5 = \frac{1}{3}\beta_\alpha \log 5$, then we have
$$\sum_{i=1}^5 q_i = 5\log 5\beta_\alpha, \quad \sum_{i=1}^4 q_i = 4\frac{2}{3}\log 5\beta_\alpha \ge 4\log 4\beta_\alpha, \quad \text{and } \sum_{i=1, i\neq 2}^5 q_i = 3\log 5\beta_\alpha \le 4\log 4\beta_\alpha.$$
Therefore, we must locally reject $\cap_{i=1}^5 H_i$, $\cap_{i=1}^4 H_i$, but not $\cap_{i=1, i\neq 2}^5 H_i$. Therefore, after closure, we will reject $\cap_{i=1}^5 H_i$ and $\cap_{i=1}^4 H_i$, but we will reject neither H_1, nor H_2, H_3 or H_4, therefore the rejection $\cap_{i=1}^4 H_i$ is not consonant.

4. When $-\infty < r < -1$, note that

$$t_\alpha^{(r)} = \mathbf{I}\left\{ \sum_{i=1}^{m} p_i^r \geq m^{-r}\alpha^{r+1} \right\}.$$

Let $t = -r, \beta_{t,\alpha} = \alpha^{-t+1}, q_i = 1/p_i$, then the local test becomes $\sum_{i=1}^m q_i^t \geq m^t \beta_{t,\alpha}$. For the case $m \geq \max\{3, \frac{\sqrt[t]{3}}{\sqrt[t]{3}-1}\}$, let $q_1^t = q_2^t \frac{1}{2} m^t \beta_\alpha$, and $q_3^t = \cdots = q_m^t = \frac{1}{6(m-2)} m^t \beta_\alpha$ (choose α such that $\beta_{t,\alpha} > 6$), then we have that, $\sum_{i=1}^m q_i^t \geq m^t \beta_\alpha$, $\sum_{i=1}^{m-1} q_i^t \geq m^t \beta_\alpha$, $\sum_{i=1, i\neq 2}^m q_i^t < m^t \beta_\alpha$. Therefore, we will reject $\cap_{i\in[m-1]} H_i$ after closure, but we cannot reject H_1 after closure, therefore the rejection $\cap_{i\in[m], i\neq 2} H_i$ is not consonant. ∎

APPENDIX E. ALGORITHMS FOR POST HOC AUTO-SELECTION SHORTCUTS

Algorithm 2. Shortcut for auto-selection of the largest rejection set with zero FDP

Input: A sequence of sorted scores $T_1, \ldots, T_m$ which satisfies $T_1 \geq \cdots \geq T_m$;a local test rule of form (13) with a monotonically increasing transformation function h and thresholding function g; confidence level α.

Output: Largest set S with zero false discoveries among all possible subsets of $[m]$, equivalently, the set of hypotheses with strong FWER control of level α.

Initialization:
transformed scores $u_1, \ldots, u_m$ where $u_i = h(T_i)$ for $1 \leq i \leq m$;
iteration related indices $k \leftarrow 1$; $s \leftarrow 1$;
accumulated scores $Q \leftarrow u_k$;
layer-wise thresholding$c \leftarrow g(s, \alpha)$.
while $k < m$ ***and*** $s < m$ **do**
  if $Q > c$ **then**
    if $s \geq k$ **then**
      $c \leftarrow c - u_k$
      $Q \leftarrow Q - u_k$
    else
      $Q \leftarrow Q - u_k + u_{k+1}$;
    end
    $k \leftarrow k + 1$
  else
    $c \leftarrow c + g(s+1, \alpha) - g(s, \alpha)$
    if $s \geq k$ **then**
      $Q \leftarrow Q + u_{k+1}$
    else
      $c \leftarrow c - u_s$
    end
    $s \leftarrow s + 1$
  end
end
return $S = \{k, \ldots, m\}$

Algorithm 3. Shortcut for auto-selection of the largest rejection set with bounded FDP

Input: confidence level $\alpha \in (0,1)$; desired FDP bound $\gamma \in (0,1)$; incremental candidate sets $S_1 \subset S_2 \cdots \subset S_n$ with $|S_k| = k$.

Output: the largest S_k such that $\overline{e}_\alpha(S_k) \le \gamma |S_k|$.

Initialization: $k \leftarrow 1$
while $k \ge 1$ **do**
  $\overline{e} \leftarrow \overline{e}_\alpha(S_k)$
  if $\overline{e}/k \le \gamma$ **then**
    return S_k
  else
    $k \leftarrow \lfloor \frac{k-\overline{e}}{1-\gamma} \rfloor$
  end
end
return $\emptyset$

APPENDIX F. PROOF FOR THEOREM 3

We first prove part (a) of Theorem 3. Denote $e_k = \overline{e}(S_k)$, and $d_k = k - e_k$. Consider we are at the iteration when $k = i$ with $i \in [n]$. If $\frac{e_i}{i} \le \gamma$, then we return i, otherwise, we need to look for $j < i$ such that $\frac{e_j}{j} \le \gamma$. Note that, for $j < i$, if $\frac{d_i}{1-\gamma} < j$, we have

$$\frac{d_j}{j} \overset{(*)}{\le} \frac{d_i}{j} < 1 - \gamma, \quad \text{and in turn} \quad \frac{e_j}{j} > \gamma, \tag{F1}$$

where (*) is true from lemma 3 in Goeman et al. (2021). Therefore we cannot have $\frac{e_j}{j} \le \gamma$ for $j < i$ if $j > \frac{d_i}{1-\gamma}$, so we directly skip those iterations in batches, and that gives us Algorithm 3.

Next, we prove part (b) and (c) which requires additional assumptions about the local tests. Part (b) is follows immediately from Theorem 2. Hence, we only prove part (c) of the theorem in the following. Without loss of generality, assume that all the m scores are sorted as $T_1 \ge T_2 \ge \ldots \ge T_m$.

Firstly, we claim that the largest set $S \subseteq [m]$ with $\overline{e}_\alpha(S) = 0$ admits strong FWER control at level α. Note, from the definition of $\overline{e}_\alpha(S)$ in (11), for any S such that $\overline{e}_\alpha(S) = 0$, each of its elementary subset is rejected by closed testing at level α; and conversely, for any hypotheses set S that is a collection of elementary hypotheses rejected by closed testing at level α, each of its subsets is also rejected by closed testing, and hence $\overline{e}_\alpha(S) = 0$. Therefore, the largest set $S \subseteq [m]$ with $\overline{e}_\alpha(S) = 0$ is the collection of all the elementary hypotheses that are rejected by closed testing at level α. Then recalling the well-known fact that the collection of all elementary hypotheses rejected by closed testing at level α is a hypothesis set with strong FWER control at level α, we have proved our claim.

Then we show that finding the collection of all the elementary hypotheses rejected by closed testing at level α is equivalent to finding a cutoff in the ordered scores. From the monotonicity of the local test, it is easy to see that, for any $k \in [m]$, if closed testing rejects T_k, it must also reject T_i for all $i > k$. Therefore, the final rejection sets must be of the form $\{T_{k^\star}, \ldots, T_m\}$, where $k^\star$ is a cutoff we are interested in finding in the ordered scores.

Finally, we show that Algorithm 2 is constructed in a way to find the correct cutoff, which is realized via searching from the largest score and stopped at the first one (which is our cutoff $k^\star$)

rejected by closed testing. Note that we reject H_k via closed testing if and only if each composite hypothesis containing it can be rejected locally. Using the monotonicity of the local test, this is saying that, for each $s = 1,\ldots, m$, we have:

$$\begin{cases} \sum_{i=1}^{s} h(T_i) \le g(s,\alpha), & \text{if } s \ge k; \\ h(T_k) + \sum_{i=1}^{s-1} h(T_i) \le g(s,\alpha), & \text{otherwise.} \end{cases} \tag{F2}$$

With simple rearrangement, one may see that Algorithm 2 starts with $k = 1$, increase k by 1 in each of its updates with k, and stops at the first time that (F2) is satisfied, when k is the cutoff $k^\star$ of our interests. Therefore, we have finished the proof for part (c).

APPENDIX G. PROOF FOR PROPOSITION 1

We call X_{mi} as just X_i in this proof for brevity. Note that, for $r \ge 1$,

$$\tilde{\alpha}_m(\Sigma, r, c) := \Pr_{\cap_{i=1}^m H_{mi}} \left\{ \left(\frac{1}{m} \sum_{i=1}^m p_{mi}^r \right)^{\frac{1}{r}} \le c \right\} = \Pr \left\{ \frac{1}{m} \sum_{i=1}^m h_r(X_i) \le C \right\}, \tag{G1}$$

where $h_r(x) := \Phi(-x)^r$, and $C := c^r$. Note that h_r is a convex function for $x \ge 0$ when $r \ge 1$. Indeed, taking second derivative of h_r with respect to x, we have

$$\frac{d^2 h_r(x)}{dx^2} = r\Phi(-x)^{r-2}\phi^2(x)\left[x\Phi(-x) + r - 1\right].$$

When $C < \frac{1}{2^r m^r}$, the event $E_1 := \left\{ \frac{1}{m}\sum_{i=1}^m h_r(X_i) \le C \right\}$ implies[6] the event $E_2 := \{X_i > 0, \forall i \in [m]\}$. Then, E_1 and E_2 together imply the event $E_3 := \left\{ h_r\left(\frac{1}{m}\sum_{i=1}^m X_i\right) \le C \right\}$ due to convexity of $h_r(x)$ for $x \ge 0$ and $r \ge 1$. Therefore,

$$\begin{aligned} \Pr \left\{ \frac{1}{m} \sum_{i=1}^m p_i^r \le C \right\} &= \Pr \left\{ \frac{1}{m} \sum_{i=1}^m h_r(X_i) \le C \right\} \le \Pr \left\{ h_r\left(\frac{1}{m} \sum_{i=1}^m X_i \right) \le C \right\} \\ &= \Pr \left\{ \Phi\left(-\frac{1}{m} \sum_{i=1}^m X_i \right)^r \le C \right\} = \Pr \left\{ \frac{1}{m} \sum_{i=1}^m X_i \ge -\Phi^{-1}(C^{\frac{1}{r}}) \right\} \\ &= \Pr \left\{ \frac{1}{m} \sum_{i=1}^m X_i \ge C_2 \right\} = \Pr_{\mathbf{X} \sim N(0,\Sigma)} \left\{ \frac{1}{m} \mathbb{I}_m^T X \ge C_2 \right\} \\ &\overset{(*)}{=} \Pr_{Z \sim N(0,\sigma_\Sigma^2)} \{ Z \ge C_2 \} = 1 - \Phi(C_2/\sigma_\Sigma), \end{aligned} \tag{G2}$$

where $C_2 = -\Phi^{-1}(C^{\frac{1}{r}}) > 0$, and $\sigma_\Sigma^2 = \frac{1}{m^2}\mathbb{I}_m^T \Sigma \mathbb{I}_m \in \mathbb{R}$, where $\mathbb{I}_m$ is vector of all ones in $\mathbb{R}^m$. Particularly, (*) is true due to the fact that Gaussianity is preserved under affine transformations.

On the other hand, under full dependence (i.e. $\rho_{ij} \equiv 1$ for all i, j), we have

$$\begin{aligned} \Pr \left\{ \frac{1}{m} \sum_{i=1}^m p_i^r \le C \right\} &= \Pr \left\{ p_1^r \le C \right\} = \Pr \left\{ \Phi(-X_1)^r \le C \right\} \\ &= \Pr \{ X_1 \ge C_2 \} = 1 - \Phi(C_2). \end{aligned} \tag{G3}$$

Therefore combining (G2) and (G3), and the fact that $C_2 > 0$, we have

$$1 - \Phi(C_2) \leq \sup_{\Sigma \in \mathcal{M}_m} \Pr\left\{\frac{1}{m}\sum_{i=1}^{m} p_i^r \leq C\right\} \leq \sup_{\Sigma \in \mathcal{M}_m} 1 - \Phi(C_2/\sigma_\Sigma) \tag{G4}$$

$$\overset{(a)}{=} \sup_{\Sigma \in \mathcal{M}_m^E} 1 - \Phi(C_2/\sigma_\Sigma)$$

$$\overset{(b)}{=} \sup_{\rho \in [-\frac{1}{m}, 1]} 1 - \Phi\left(\frac{C_2}{\frac{1}{m} + \frac{m-1}{m}\rho}\right) = 1 - \Phi(C_2), \tag{G5}$$

where $\mathcal{M}_m$ is the class of all correlation matrix, and $\mathcal{M}_m^E$ is the class of all equicorrelation matrices with correlation $\rho \in [-\frac{1}{m}, 1]$. Specifically, (a) is true since σ_Σ only depends on the average of all entries in Σ, and (b) is true since $\sigma_\Sigma = \frac{1}{m} + \frac{m-1}{m}\rho$ for any Σ in $\mathcal{M}_m^E$. In conclusion, we have

$$\sup_{\Sigma \in \mathcal{M}_m} \Pr\left\{\frac{1}{m}\sum_{i=1}^{m} p_i^r \leq C\right\} = 1 - \Phi(C_2), \tag{G6}$$

for all $r \geq 1$ and the supremum is achieved at full dependence. Transforming back to the original representation in (22), we have completed our proof.

APPENDIX H. PROOF FOR THEOREM 4

Recalling decomposition (28) in the main paper, we can rewrite $\tilde{\alpha}_m(\rho, r, c)$ as the following, which makes the link to the Generalized Law of Large Numbers clearer:

$$\tilde{\alpha}_m(\rho, r, c) = \mathbb{E}_{Z_0}\left[\Pr\left\{\left(\frac{1}{m}\sum_{i=1}^{m} p_i^r\right)^{\frac{1}{r}} \leq c) \,\middle|\, Z_0 = z_0\right\}\right] = \mathbb{E}_{Z_0}\left[\Pr\left\{\operatorname{sign}(r)\frac{1}{m}\sum_{i=1}^{m} p_i^r \leq \operatorname{sign}(r) \cdot C) \,\middle|\, Z_0 = z_0\right\}\right], \tag{H1}$$

where we use the conditional independence amongst $\{p_i\}_{i=1}^m$, and replace $C := c^r$. Then we have

$$\limsup_{m\to\infty} \tilde{\alpha}_m(\rho, r, c) = \limsup_{m\to\infty} \mathbb{E}_{Z_0}\left[\Pr\left\{\operatorname{sign}(r)\frac{1}{m}\sum_{i=1}^{m} p_i^r \leq \operatorname{sign}(r) \cdot C \,\middle|\, Z_0 = z_0\right\}\right] = \mathbb{E}_{Z_0}\left[\limsup_{m\to\infty} \Pr\left\{\operatorname{sign}(r)\frac{1}{m}\sum_{i=1}^{m} p_i^r \leq \operatorname{sign}(r) \cdot C \,\middle|\, Z_0 = z_0\right\}\right],$$

where the last equality is true by applying dominance convergence theorem since the inner probability is integrable. In the following, we focus on quantifying the limitation of the conditional probability

$$\Pr\left\{\operatorname{sign}(r)\frac{1}{m}\sum_{i=1}^{m}p_i^r \le \operatorname{sign}(r)\cdot C \;\middle|\; Z_0 = z_0\right\}, \tag{H2}$$

for which we need Lemma 3 to characterizes the distribution of p_i^r for $r \neq \pm\infty$.

Lemma 3. *Denote the CDF of p_i^r (with p_i defined in (29)) conditioning on $Z_0 = z_0$ as F_{r,ρ,z_0}, and the corresponding density as f_{r,ρ,z_0}, we have that*:

$$F_{r,\rho,z_0}(y) = \Phi\left(\operatorname{sign}(r)\frac{\Phi^{-1}(y^{\frac{1}{r}}) + \sqrt{\rho}z_0}{\sqrt{1-\rho}}\right), \quad \text{and} \quad f_{r,\rho,z_0}(y) = O\left(y^{-(\frac{1}{(\rho-1)r}+1)}\right) \quad \text{as } y^r \to 0, \tag{H3}$$

where we take $y^{-(\frac{1}{(\rho-1)r}+1)} = \exp\left(-\frac{y}{\rho-1}\right)$ *when* $r = 0$.

Proof. Without loss of generality, we only prove for the case $r \geq 0$. Firstly, when $r > 0$, we have:

$$\begin{aligned} F_{r,\rho,z_0}(y) &= \Pr\left\{p_i^r \le y \;\middle|\; Z_0 = z_0\right\} = \Pr\left\{\Phi(-\sqrt{\rho}z_0 - \sqrt{1-\rho}Z_i) \le y^{\frac{1}{r}}\right\} \\ &= \Pr\left\{Z_i \le \frac{\Phi^{-1}(y^{\frac{1}{r}}) + \sqrt{\rho}z_0}{\sqrt{1-\rho}}\right\} = \Phi\left(\frac{\Phi^{-1}(y^{\frac{1}{r}}) + \sqrt{\rho}z_0}{\sqrt{1-\rho}}\right), \end{aligned} \tag{H4}$$

and also the density

$$f_{r,\rho,z_0}(y) = \frac{dF_{r,\rho,z_0}(y)}{dy} \propto \frac{y^{\frac{1}{r}-1}}{r\sqrt{1-\rho}}\phi\left(\frac{\Phi^{-1}(y^{\frac{1}{r}}) + \sqrt{\rho}Z_0}{\sqrt{1-\rho}}\right)\Big/\phi(\Phi^{-1}(y^{\frac{1}{r}})) \tag{H5}$$

$$\propto \frac{y^{\frac{1}{r}-1}}{r\sqrt{1-\rho}}\exp\left(-\frac{\rho\Phi^{-1}(y^{\frac{1}{r}})^2 + \sqrt{\rho}Z_0\Phi^{-1}(y^{\frac{1}{r}})}{2(1-\rho)}\right). \tag{H6}$$

Using the approximation

$$\Phi^{-1}(x) = O\left(-\sqrt{\log\frac{1}{x^2}}\right) \quad \text{when } x \to 0,$$

we have that,

$$f_{r,\rho,z_0}(y) = O\left(y^{-(\frac{1}{(\rho-1)r}+1)}\right) \quad \text{as } y \to 0. \tag{H7}$$

For $r = 0$, we have:

$$\begin{aligned} F_{r,\rho,z_0}(y) &= \Pr\left\{\log p_i \le y \;\middle|\; Z_0 = z_0\right\} = \Pr\left\{\Phi(-\sqrt{\rho}z_0 - \sqrt{1-\rho}Z_i) \le \exp(y)\right\} \\ &= \Pr\left\{Z_i \le \frac{\Phi^{-1}(\exp(y)) + \sqrt{\rho}z_0}{\sqrt{1-\rho}}\right\} = \Phi\left(\frac{\Phi^{-1}(\exp(y)) + \sqrt{\rho}z_0}{\sqrt{1-\rho}}\right), \end{aligned} \tag{H8}$$

and also the density

$$f_{r,\rho,z_0}(y) = \frac{dF_{r,\rho,z_0}(y)}{dy} \propto \frac{\exp(y)}{r\sqrt{1-\rho}}\phi\left(\frac{\Phi^{-1}(\exp(y)) + \sqrt{\rho}Z_0}{\sqrt{1-\rho}}\right) \Big/ \phi(\Phi^{-1}(\exp(y))) \tag{H9}$$

$$\propto \frac{\exp(y)}{r\sqrt{1-\rho}} \exp\left(-\frac{\rho\Phi^{-1}(\exp(y))^2 + \sqrt{\rho}Z_0\Phi^{-1}(\exp(y))}{2(1-\rho)}\right). \tag{H10}$$

Again using the approximation

$$\Phi^{-1}(x) = O\left(-\sqrt{\log\frac{1}{x^2}}\right) \quad \text{when } x \to 0,$$

we have that,

$$f_{r,\rho,z_0}(y) = O\left(\exp\left(\frac{y}{1-\rho}\right)\right) \quad \text{as } y \to -\infty, \text{i.e. } \log y \to 0. \tag{H11}$$

■

(a) and (b) $r > -1$:

When $r > -1$, using Lemma 3 we have that $\mathbb{E}\left[p_1^r \,\middle|\, Z_0 = z_0\right] < \infty$ for any $\rho \in [0,1]$, therefore by the Law of Large Numbers, we have

$$\frac{1}{m}\sum_{i=1}^{m} p_i^r \,\middle|\, Z_0 = z_0 \xrightarrow{d} \mathbb{E}\left[p_1^r \,\middle|\, Z_0 = z_0\right], \tag{H12}$$

where $\xrightarrow{d}$ means converge in distribution. Therefore,

$$\limsup_{m\to\infty} \Pr\left\{\text{sign}(r)\frac{1}{m}\sum_{i=1}^{m} p_i^r \le \text{sign}(r)\cdot C \,\middle|\, Z_0 = z_0\right\}$$

$$= \Pr\left\{\text{sign}(r)\mathbb{E}\left[p_1^r \,\middle|\, Z_0 = z_0\right] \le \text{sign}(r)\cdot C\right\}, \tag{H13}$$

and hence

$$\limsup_{m\to\infty} \tilde{\alpha}_m(\rho, r, c) = \mathbb{E}_{Z_0}\left[\Pr\left\{\text{sign}(r)\mathbb{E}\left[p_1^r \,\middle|\, Z_0 = z_0\right] \le \text{sign}(r)\cdot C\right\}\right] := h(\rho, r, C). \tag{H14}$$

Recall that the conditional mean $g_{\rho,r}(z_0) := \mathbb{E}\left[p_i^r \,\middle|\, Z_0 = z_0\right]$ in (31), we have

$$g_{\rho,r}(z_0) = \int \Phi\left(-\sqrt{\rho}\, z_0 - \sqrt{1-\rho}\, x\right)^r \phi(x)dx$$

$$= \frac{1}{\sqrt{1-\rho}} \int \phi\left(\frac{y - \sqrt{\rho}z_0}{\sqrt{1-\rho}}\right) \Phi(-y)^r dy, \tag{H15}$$

where ϕ as the standard normal p.d.f. From expression in (H15), it is easy to see that $g_{\rho,r}(z_0)$ is monotonically nonincreasing in z_0 when $r \ge 0$, while monotonically nondecreasing in z_0 when

$r < 0$. Therefore, using this monotonicity, we have explicit expression

$$h(\rho, r, C) = \Phi\left(-g_{\rho,r}^{-1}(C)\right). \tag{H16}$$

Recall the relationship $C \equiv c^r$, and the definition of $c_r(m, \alpha)$ that

$$c_r(m, \alpha) := \sup\left\{c : \sup_{\rho\in[0,1]} \limsup_{m\to\infty} \tilde{\alpha}_m(\rho, \alpha, c) \le \alpha\right\},$$

where the supremum over c is taking over $\mathbb{R}$, we omit it for simplicity.

For $r > 0$, plugging in expression (H16) in (H14), we have the following closed expression

$$c_r(m, \alpha) = \left(\sup\left\{C : \sup_{\rho\in[0,1]} \Phi\left(-g_{\rho,r}^{-1}(C)\right) \le \alpha\right\}\right)^{\frac{1}{r}}.$$

Denote $C_\rho := \sup\left\{C : \Phi(-g_{\rho,r}^{-1}(C)) \le \alpha\right\}$, we claim that $c_r(m, \alpha)$ is equivalent to $\left(\inf_{\rho\in[0,1]} C_\rho\right)^{\frac{1}{r}}$. To prove this claim, first note that

$$\sup\left\{C : \sup_{\rho\in[0,1]} \Phi\left(-g_{\rho,r}^{-1}(C)\right) \le \alpha\right\} = \sup \bigcap_{\rho\in[0,1]} \left\{C : \Phi\left(-g_{\rho,r}^{-1}(C)\right) \le \alpha\right\}. \tag{H17}$$

This is true due to the following simple reasoning. For each c such that $\sup_{\rho\in[0,1]} \Phi(-g_{\rho,r}^{-1}(c)) \le \alpha$, we have

$$c \in \bigcap_{\rho\in[0,1]} \left\{C : \Phi\left(-g_{\rho,r}^{-1}(C)\right) \le \alpha\right\}.$$

On the other hand, for each $c \in \bigcap_{\rho\in[0,1]}\{C : \Phi(-g_{\rho,r}^{-1}(C)) \le \alpha\}$, we have

$$\sup_{\rho\in[0,1]} \Phi\left(-g_{\rho,r}^{-1}(c)\right) \le \alpha.$$

Therefore

$$\left\{C : \sup_{\rho\in[0,1]} \Phi\left(-g_{\rho,r}^{-1}(C)\right) \le \alpha\right\} = \bigcap_{\rho\in[0,1]} \left\{C : \Phi\left(-g_{\rho,r}^{-1}(C)\right) \le \alpha\right\},$$

and taking supremum on both sides we have (H17).

Then, using the monotonicity of $g_{\rho,r}(z)$ with regard z, and the fact that $g_{\rho,r}(z)$ decreases with ρ when $\rho > 0, z > 0$ and $r \ge 0$ (easy to verify that the derivative with regard ρ is always negative), we further have

$$\begin{aligned}\sup \bigcap_{\rho\in[0,1]} \left\{C : \Phi\left(-g_{\rho,r}^{-1}(C)\right) \le \alpha\right\} &= \sup \bigcap_{\rho\in[0,1]} \left\{C : C \le g_{\rho,r}\left(-\Phi^{-1}(\alpha)\right)\right\} \\ &= \sup\left\{C : C \le \inf_{\rho\in[0,1]} g_{\rho,r}\left(-\Phi^{-1}(\alpha)\right)\right\} = \inf_{\rho\in[0,1]} g_{\rho,r}\left(-\Phi^{-1}(\alpha)\right)\end{aligned}$$

Combining with (H17), we have

$$c_r(m,\alpha) = \left(\inf_{\rho\in[0,1]} g_{\rho,r}\left(-\Phi^{-1}(\alpha)\right)\right)^{\frac{1}{r}}. \tag{H18}$$

Using properties of $g_{\rho,r}$ we can further simplify the above expression. We find that $g_{\rho,r}(z)$ decreases with ρ when $z>0$ but increases with ρ when $z\le 0$. Since we assume only $\alpha\in(0,1)$, we are not sure about the sign of $\Phi^{-1}(\alpha)$. Therefore the minima can be achieved at both boundaries, that is

$$\inf_{\rho\in[0,1]} g_{\rho,r}(-\Phi^{-1}(\alpha)) = \min\left\{g_{0,r}(-\Phi^{-1}(\alpha)), g_{1,r}(-\Phi^{-1}(\alpha))\right\}.$$

Note that

$$g_{0,r}(-\Phi^{-1}(\alpha)) = \frac{r}{r+1}; \quad g_{1,r}(-\Phi^{-1}(\alpha)) = \alpha^r,$$

finally we have $c_r(m,\alpha) = \min\{\alpha, \left(\frac{r}{r+1}\right)^{\frac{1}{r}}\}$ for $r>0$.

Similarly, for $-1<r\le 0$, as $c := C^{\frac{1}{r}}$ decreases with C, we have the closed expression

$$c_r(m,\alpha) = \left(\inf\left\{C : \sup_{\rho\in[0,1]} \Phi(-g_{\rho,r}^{-1}(C)) \le \alpha\right\}\right)^{\frac{1}{r}} = \left(\sup_{\rho\in[0,1]} C_\rho\right)^{\frac{1}{r}},$$

where $C_\rho := \inf\left\{C : \Phi(-g_{\rho,r}^{-1}(C)) \le \alpha\right\}$. Using the monotonicity of $g_{\rho,r}(z)$ with regard z, and the fact that $g_{\rho,r}(z)$ decreases with ρ when $\rho>0, z<0$ and $r<0$ (easy to verify that the derivative with regard ρ is always negative), we further have

$$C_\rho = \inf\left\{C : C \ge g_{\rho,r}\left(-\Phi^{-1}(\alpha)\right)\right\} = g_{\rho,r}\left(-\Phi^{-1}(\alpha)\right). \tag{H19}$$

Therefore,

$$c_r(m,\alpha) = \left(\sup_{\rho\in[0,1]} g_{\rho,r}\left(-\Phi^{-1}(\alpha)\right)\right)^{\frac{1}{r}}. \tag{H20}$$

Finally, we have that,

$$\begin{aligned}\tilde{\alpha}(\rho,r,\alpha) &= \limsup_{m\to\infty} \tilde{\alpha}_m(\rho,r,c_r(m,\alpha)) \\ &= \begin{cases} \Phi\left(-g_{\rho,r}^{-1}(\alpha^r)\right), & \text{if } r>0; \\ \Phi(-g_{\rho,r}^{-1}(c_r(m,\alpha)^r), & \text{if } -1\le r\le 0. \end{cases}\end{aligned} \tag{H21}$$

And

$$c_r(m,\alpha) = \begin{cases} \min\left\{\alpha, \left(\frac{r}{r+1}\right)^{\frac{1}{r}}\right\}, & \text{if } r>0; \\ \left(\sup_{\rho\in[0,1]} g_{\rho,r}\left(-\Phi^{-1}(\alpha)\right)\right)^{\frac{1}{r}} & \text{if } -1\le r\le 0. \end{cases} \tag{H22}$$

(c) and (d): $r \leq -1$

When $r \leq -1$, things get a bit tricky, since according to Lemma 3, $\mathbb{E}\left[p_i^r \,\middle|\, Z_0 = z_0\right]$ may not exist. In the following, we will use the stable law stated in Lemma 4 to derive the asymptotic behaviour of $\tilde{\alpha}_m(\rho, r, c)$ for $r < 0$.

Lemma 4. *(Generalized LLN* (Uchaikin & Zolotarev, 2011)*) Consider a sequence of i.i.d random variables $X_1, X_2, \ldots, X_m$ which shares the same distribution with X, where X has support on $[1, \infty]$ and density f satisfying the following*:

$$f(x) = O(x^{-(\beta+1)}), \quad \text{as } x \to \infty \text{ with } \beta > 0.$$

Denote $\overline{X}_m := \frac{1}{m}\sum_{i=1}^m X_i$, we have that

(a) if $0 < \beta < 1$, then $m^{1-\frac{1}{\beta}}\overline{X}_m \xrightarrow{d} Y$;
(b) if $\beta = 1$, then $\overline{X}_m - \log m \xrightarrow{d} Y$;
(c) if $1 < \beta < 2$, then $m^{1-\frac{1}{\beta}}(\overline{X}_m - \mathbb{E}[X]) \xrightarrow{d} Y$;
(d) if $\beta \geq 2$, then $\overline{X}_m \xrightarrow{d} \mathbb{E}[X]$,

where Y is some random variable that shares the same tail behaviour with X.

Then, for $r \leq -1$, from Lemma 3, we have $\beta = \frac{1}{(\rho-1)r}$. Let

$$C(\alpha, r, m, \rho) := \begin{cases} C_{\alpha,r} m^{-1+(\rho-1)r}, & \text{if } 0 \leq \rho < 1 + \frac{1}{r}; \\ C_{\alpha,r} + \log m, & \text{if } \rho = 1 + \frac{1}{r}; \\ C_{\alpha,r} m^{-1+(\rho-1)r} + \mathbb{E}\left[p_1^r \,\middle|\, Z_0 = z_0\right], & \text{if } 1 + \frac{1}{r} < \rho \leq 1 + \frac{1}{2r}; \\ C_{\alpha,r} + \mathbb{E}\left[p_1^r \,\middle|\, Z_0 = z_0\right], & \text{if } 1 + \frac{1}{2r} < \rho \leq 1, \end{cases} \tag{H23}$$

where $C_{\alpha,r}$ is some constant that depends only on α and r that we will specify later. Using Lemma 4, we have

$$\begin{aligned} &\lim_{m\to\infty} \Pr\left\{\frac{1}{m}\sum_{i=1}^m p_i^r \geq C(\alpha, r, m, \rho) \,\middle|\, Z_0 = z_0\right\} \\ &= \Pr\left\{Y \geq C_{\alpha,r} \,\middle|\, Z_0 = z_0\right\} \overset{(*)}{=} \Pr\left\{p_1^r \geq C_{\alpha,r} \,\middle|\, Z_0 = z_0\right\} + o(1) \\ &= F_{r,\rho,z_0}(C_{\alpha,r}) + o(1) \quad \text{as } \alpha \to 0, \end{aligned} \tag{H24}$$

where Y is the random variable comes from the limitation in Lemma 4, which shares the same tail behaviour of p_1^r, therefore we have the approximation $(*)$.

Recalling definitions in (22), (27) of the main paper, our goal is to find c such that

$$\sup_{\rho\in[0,1]} \limsup_{m\to\infty} \tilde{\alpha}_m(\rho, r, c) \leq \alpha,$$

or equivalently find C such that

$$\sup_{\rho\in[0,1]} \limsup_{m\to\infty} \tilde{\alpha}_m(\rho, r, C^{\frac{1}{r}}) \le \alpha.$$

Note that $C(\alpha, r, m, \rho)$ is monotonically nonincreasing in ρ, and $C(\alpha, r, m, 0)$ dominates $C(\alpha, r, m, \rho)$ for any $0 < \rho \le 1$. Therefore, to calibrate for arbitrary $\rho \in [0, 1]$, that is to find a critical value that does not depend on ρ, we have no choice but let $C = C(\alpha, r, m, 0)$, and hence

$$\begin{aligned}&\sup_{\rho\in[0,1]} \limsup_{m\to\infty} \tilde{\alpha}_m(\rho, r, C(\alpha, r, m, 0)^{\frac{1}{r}}) = \tilde{\alpha}_m(\rho, r, C(\alpha, r, m, 0)^{\frac{1}{r}})\\ &= \mathbb{E}_{Z_0}\left[1 - F_{r,0,Z_0}(C_{\alpha,r})\right] = \mathbb{E}_{Z_0}\left[\Phi\left(\Phi^{-1}(C_{\alpha,r}^{\frac{1}{r}})\right)\right] = C_{\alpha,r}^{\frac{1}{r}} \le \alpha,\end{aligned}$$

which indicates we should set $C_{\alpha,r} = \alpha^r$ to achieve the upper bound.

Therefore we have

$$c_r(m, \alpha) = (C(\alpha, r, m, 0))^{\frac{1}{r}} = \begin{cases} \alpha m^{\frac{1}{|r|}-1}, & \text{if } r < -1; \\ \frac{\alpha}{1+\alpha m}, & \text{if } r = -1, \end{cases} \tag{H25}$$

and correspondingly

$$\tilde{\alpha}(\rho, r, \alpha) = \limsup_{m\to\infty} \tilde{\alpha}_m(\rho, r, c_r(m, \alpha)) = \alpha \mathbb{I}\{\rho = 0\},$$

where the last equality is true due to the nature of stable law, where the tail behavior determines the rate of growth, and the mismatch of the growth rate leads to degenerate asymptotic probability.

Here we finish the proof for Theorem 4.

APPENDIX I. PROOF FOR THEOREM 5

In the following, we are interested in calculating the asymptotic power using the calibrated threshold $c_r(m, \alpha)$ derived in Theorem 4. In particular, the power can rewritten as

$$\beta_{\mu_m,\pi_m,\rho}(r, \alpha) := \Pr\left\{ \text{sign}(r)\frac{1}{m}\sum_{i=1}^{m} p_{mi}^r \le \text{sign}(r) C_r(m, \alpha) \right\}, \tag{I1}$$

where $C_r(m, \alpha) = c_r(m, \alpha)^r$, and $p_{mi} = \Phi^{-1}(-X_{mi})$ for all i. Using similar decomposition as in the proof of Theorem 4, we have that, for all $i = 1, 2, \ldots, m$,

$$\begin{aligned} X_{mi} &= \mu_{mi} + \sqrt{\rho}\, Z_0 + \sqrt{1-\rho}\, Z_i, \\ p_{mi} &= \Phi(-X_{mi}) = \Phi\left(-\mu_{mi} - \sqrt{\rho}\, Z_0 - \sqrt{1-\rho}\, Z_i\right), \end{aligned} \tag{I2}$$

where variable $Z_0 \sim N(0, 1)$, $Z_i \overset{\text{iid}}{\sim} N(0, 1)$, $\{Z_0\} \perp\!\!\!\perp \{\mu_{mi}, Z_i\}_{i=1}^m$, and $\mu_{mi} \overset{iid}{\sim} \mu_m B_{mi}$ with $B_{mi} \overset{\text{iid}}{\sim} \text{Bernoulli}(\pi_m)$ for all $i = 1, 2, \ldots, m$. Also, we have the conditional independence:

$$p_{m1}, p_{m2}, \ldots, p_{mm} \text{ are independent conditioning on } Z_0. \tag{I3}$$

Then the asymptotic power is given by

$$\lim_{m\to\infty}\beta_{\mu_m,\pi_m,\rho}(r,\alpha)=\mathbb{E}_{Z_0}\left[\lim_{m\to\infty}\Pr\left\{\operatorname{sign}(r)\frac{1}{m}\sum_{i=1}^{m}p_{mi}^{r}\le\operatorname{sign}(r)C_r(m,\alpha)\;\middle|\;Z_0\right\}\right]. \tag{I4}$$

When $r>0$, we can use the law of large numbers of triangular array, that is,

$$\sup_m\mathbb{E}\left[p_{mi}^{2r}\;\middle|\;Z_0=z_0\right]<\infty\quad\Rightarrow\quad\frac{1}{m}\sum_{i=1}^{m}p_{mi}^{r}-\mathbb{E}\left[p_{mi}^{r}\;\middle|\;Z_0=z_0\right]\xrightarrow{p}0, \tag{I5}$$

almost surely for all possible value of z_0. Then we get

$$\begin{aligned}\lim_{m\to\infty}\Pr\left\{\frac{1}{m}\sum_{i=1}^{m}p_{mi}^{r}\le C_r(m,\alpha)\;\middle|\;Z_0\right\}&=\lim_{m\to\infty}\Pr\left\{\mathbb{E}\left[p_{m1}^{r}\;\middle|\;Z_0\right]\le\alpha^r\right\}\\&=\lim_{m\to\infty}\Pr\left\{\pi_m\mathbb{E}\left[p_{m1}^{r}\;\middle|\;Z_0,\mu_{m1}=\mu_m\right]+(1-\pi_m)\mathbb{E}\left[p_{m1}^{r}\;\middle|\;Z_0,\mu_{m1}=0\right]\le\alpha^r\right\}.\end{aligned} \tag{I6}$$

Combining (I4) and (I6), we have that, when $r>0$,

$$\lim_{m\to\infty}\beta_{\mu_m,\pi_m,\rho}(r,\alpha)=\Pr\left\{\pi g_{\rho,r}\left(Z_0+\frac{\mu}{\sqrt{\rho}}\right)+(1-\pi)g_{\rho,r}(Z_0)\le\alpha^r\right\}, \tag{I7}$$

where $g_{\rho,r}$ is defined in (31). From this expression, the following cases can be specified,

- if $\pi=1$, then

$$\lim_{m\to\infty}\beta_{\mu_m,\pi_m,\rho}(r,\alpha)=\begin{cases}1, & \text{if } \mu=\infty;\\ \Phi\left(-g_{\rho,r}^{-1}(\alpha^r)+\frac{\mu}{\sqrt{\rho}}\right), & \text{if } 0<\mu<\infty;\\ \tilde{\alpha}(\rho,r,\alpha), & \text{if } \mu=0.\end{cases} \tag{I8}$$

- if $0<\pi<1$, then

$$\lim_{m\to\infty}\beta_{\mu_m,\pi_m,\rho}(r,\alpha)=\begin{cases}\Phi\left(-g_{\rho,r}^{-1}\left(\frac{\alpha^r}{1-\pi}\right)\right), & \text{if } \mu=\infty;\\ \Pr\left\{\pi g_{\rho,r}\left(Z_0+\frac{\mu}{\sqrt{\rho}}\right)+(1-\pi)g_{\rho,r}(Z_0)\le\alpha^r\right\}, & \text{if } 0<\mu<\infty;\\ \tilde{\alpha}(\rho,r,\alpha), & \text{if } \mu=0.\end{cases} \tag{I9}$$

- if $\pi=0$, then $\lim_{m\to\infty}\beta_{\mu_m,\pi_m,\rho}(r,\alpha)\equiv\tilde{\alpha}(\rho,r,\alpha)$, no matter what value that μ takes.

Therefore, we complete the proof.

APPENDIX J. PROOF FOR THEOREM 6

When $r\le-1$, We utilize the following results: as long as the triangular array $\{Y_{mi},i=1,\ldots,i_m\}$ satisfy the uniformly asymptotically negligible (UAN) condition, that is for any $\epsilon>0$,

$$\lim_{m\to\infty}\max_i \Pr\ \{|Y_{mi}|>\epsilon\}=0, \tag{J1}$$

then we have that, $\lim_{m\to\infty}\sum_i Y_{mi}$ converge to an infinitely divisible distribution under certain conditions. The specific argument is formally stated in the following Lemma 5.

Lemma 5. *(Theorem 3.2.2 in* Gnedenko & Kolmogorov, 1954*) Consider an triangular array* $\{Y_{mk}, k=1,\dots,k_m\}$*, such that the UAN condition is fulfilled, that is for any* $\epsilon>0$

$$\lim_{m\to\infty}\max_k \mu_{mk}\{|y|>\epsilon\}=0, \tag{J2}$$

where μ_{mk} *is the distribution function for* Y_{mk}*, and denote* $S_m := Y_{m1}+\cdots+Y_{m,k_m}$.

Then there exists a deterministic sequence a_m *such that sequence* S_m-a_m *converges weakly to an infinitely divisible random variable* Y *if and only if the following conditions are fulfilled*:

1. *for any* $A=(-\infty,x)$ *with* $x<0$*, and* $A=(x,\infty)$ *with* $x>0$ *such that* $\nu(\partial A)=0$,

$$\nu(A) := \lim_{m\to\infty}\sum_{k=1}^{k_m}\mu_{mk}(A), \tag{J3}$$

is a Lévy measure, that is, a σ*-finite Borel measure on* $\mathbb{R}\setminus 0$ *such that* $\int_{\mathbb{R}\setminus 0}\min\{1,x^2\}\nu(dx)<\infty$.

2. *moreover*,

$$\begin{aligned}&\lim_{\tau\to 0}\limsup_{m\to\infty}\sum_{k=1}^{k_m}\mathrm{Var}\,(Z_{mk}\mathbf{1}\{|Z_{mk}|<\tau\})\\ &\quad=\lim_{\tau\to 0}\liminf_{m\to\infty}\sum_{k=1}^{k_m}\mathrm{Var}\,(Z_{mk}\mathbf{1}\{|Z_{mk}|<\tau\})=\sigma^2<\infty.\end{aligned} \tag{J4}$$

Particularly, Y has the characteristic exponent

$$\phi(t)=-\frac{1}{2}\sigma^2t^2+\int_{\mathbb{R}\setminus\{0\}}(e^{itx}-1-itx\mathbf{1}\{|x|\le 1\})\nu(dx), \tag{J5}$$

and a_m *can be chosen by*

$$a_m=\sum_{k=1}^{k_m}\int_{|x|<1}x\mu_{nk}(dx)+o(1), \tag{J6}$$

given that $\nu(\{x:|x|=1\})=0$.

In our case, let

$$Y_{mi}=\frac{1}{m^{-r}}(p_{mi}^r-a_{r,m})|Z_0,$$

where $a_{r,m}=0$ if $r<-1$, and $a_{r,m}=\log m$ if $r=-1$. We firstly check the UAN condition (J1). Note that

$$
\begin{aligned}
&\lim_{m\to\infty}\max_i \Pr\left\{\left|\frac{1}{m^{-r}}\left(p_{mi}^r-a_{r,m}\right)\right|>\epsilon \;\middle|\; Z_0=z_0\right\}\\
&=\lim_{m\to\infty}\Pr\left\{\left|\frac{1}{m^{-r}}\left(p_{mi}^r-a_{r,m}\right)\right|>\epsilon \;\middle|\; Z_0\right\}\\
&=\lim_{m\to\infty}\Pr\left\{p_{mi}^r>m^{-r}\epsilon+a_{r,m}\;\middle|\; Z_0\right\}+\Pr\left\{p_{mi}^r<-m^{-r}\epsilon+a_{r,m}\;\middle|\; Z_0\right\}\\
&=\lim_{m\to\infty}\pi_m\Pr\left\{p_{mi}^r>m^{-r}\epsilon+a_{r,m}\;\middle|\; Z_0,\mu_{mi}=\mu_m\right\}\\
&\quad+(1-\pi_m)\Pr\left\{p_{mi}^r>m^{-r}\epsilon+a_{r,m}\;\middle|\; Z_0,\mu_{mi}=0\right\}\\
&\quad+\pi_m\Pr\left\{p_{mi}^r<-m^{-r}\epsilon+a_{r,m}\;\middle|\; Z_0,\mu_{mi}=\mu_m\right\}\\
&\quad+(1-\pi_m)\Pr\left\{p_{mi}^r<-m^{-r}\epsilon+a_{r,m}\;\middle|\; Z_0,\mu_{mi}=0\right\}\\
&=\lim_{m\to\infty}\pi_m\Phi\left(\frac{\Phi^{-1}((m^{-r}\epsilon+a_{r,m})^{\frac{1}{r}})+\mu_m+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)\\
&\quad+(1-\pi_m)\Phi\left(\frac{\Phi^{-1}((m^{-r}\epsilon+a_{r,m})^{\frac{1}{r}})+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)\\
&\quad+\pi_m\Phi\left(-\frac{\Phi^{-1}((-m^{-r}\epsilon+a_{r,m})^{\frac{1}{r}})+\mu_m+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)\\
&\quad+(1-\pi_m)\Phi\left(-\frac{\Phi^{-1}((-m^{-r}\epsilon+a_{r,m})^{\frac{1}{r}})+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right).
\end{aligned}
\tag{J7}
$$

For $r<-1$, we have $a_{r,m}=0$, and thus, (J7) can be simplified as

$$
\lim_{m\to\infty}\pi_m\Phi\left(\frac{\Phi^{-1}((m^{-r}\epsilon)^{\frac{1}{r}})+\mu_m+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)=\lim_{m\to\infty}\pi_m\Phi\left(\frac{\Phi^{-1}(\epsilon^{\frac{1}{r}}\frac{1}{m})+\mu_m+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right),
\tag{J8}
$$

while on the other hand, for $r=-1$, we have $a_{r,m}=\log m$, and (J7) can also be simplified as

$$
\lim_{m\to\infty}\pi_m\Phi\left(\frac{\Phi^{-1}((m\epsilon+\log m)^{\frac{1}{r}})+\mu_m+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)=\lim_{m\to\infty}\pi_m\Phi\left(\frac{\Phi^{-1}(\epsilon^{\frac{1}{r}}\frac{1}{m})+\mu_m+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right).
\tag{J9}
$$

Therefore, in order to make (J8) and (J9) goes to zero, we only need to make μ_m grows slower than $|\Phi^{-1}(\frac{1}{m})|=O(\sqrt{\log m})$, that is $\mu_m=o(\sqrt{\log m})$. Returning to the proof of the theorem, we first consider the case $\rho>0$, under which we will prove that for each i, $Y_{mi}=o_p(1)$ when $r<-1$, and $Y_{mi}=o(\log m)$ when $r=-1$, as $m\to\infty$. We prove this by applying Lemma 5, during which we check the conditions 1 and 2 in it.

As for condition 1 in Lemma 5 for $r\le -1$, defining $\nu(x):=1-\lim_{m\to\infty} m\Pr\{Y_{mi}>x\}$ for all $x>0$, it can be simplified to checking that

$$1-\nu(1)+\int_{0<x<1}x^2\nu(dx)<\infty. \tag{J10}$$

Note that

$$\begin{aligned}
&\Pr\ \{Y_{mi}>x\}\\
&=\Pr\ \left\{m^rP_{mi}^r>x\ \middle|\ Z_0=z_0\right\}=\Pr\ \left\{P_{mi}<\frac{x^{\frac{1}{r}}}{m}\ \middle|\ Z_0=z_0\right\}\\
&=\pi_m\Pr\ \left\{P_{mi}<\frac{x^{\frac{1}{r}}}{m}\ \middle|\ Z_0=z_0,\mu_{mi}=\mu_m\right\}+(1-\pi_m)\Pr\ \left\{P_{mi}<\frac{x^{\frac{1}{r}}}{m}\ \middle|\ Z_0=z_0,\mu_{mi}=0\right\}\\
&=\pi_m\Phi\left(\frac{\Phi^{-1}((x^{\frac{1}{r}}/m)+\mu_m+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)+(1-\pi_m)\Phi\left(\frac{\Phi^{-1}((x^{\frac{1}{r}}/m)+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)\\
&=\pi_m\Phi\left(\frac{-\sqrt{2\log(mx^{-\frac{1}{r}})}+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)+(1-\pi_m)\Phi\left(\frac{-\sqrt{2\log(mx^{-\frac{1}{r}})}+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)+o(1)\\
&=\Phi\left(\frac{-\sqrt{2\log(mx^{-\frac{1}{r}})}+\sqrt{\rho}z_0}{\sqrt{1-\rho}}\right)+o(1)=O(m^{-\frac{1}{1-\rho}}x^{\frac{1}{(1-\rho)r}}),
\end{aligned} \tag{J11}$$

therefore, we have

$$\nu(x)=1-\lim_{m\to\infty}m\Pr\ \{Y_{mi}>x\}=1-x^{\frac{1}{(1-\rho)r}}\lim_{m\to\infty}m^{-\frac{\rho}{1-\rho}}=0, \tag{J12}$$

since $\rho>0$. Therefore (J10) is true, and in particular $\nu(x)=0$ when $\rho>0$.

Afterwards, we check condition 2 in Lemma 5, which simplifies to verifying

$$\lim_{\tau\to0}\lim_{m\to\infty}m\mathrm{Var}(Y_{mi}\mathbf{1}\{Y_{mi}<\tau\})<\infty, \tag{J13}$$

in our setting. Using the similar technique that we will use to calculate a_m, that is, the truncated first moment, we have the following about the truncated second moment for any fixed truncation position $\tau>0$,

$$\begin{aligned}
&\mathrm{Var}(Y_{mi}\mathbf{1}\{Y_{mi}<\tau\})\le m\mathbb{E}\left[Y_{mi}^2\mathbf{1}\{Y_{mi}<\tau\}\right]\\
&\quad=o\left(m^{1-\frac{1}{1-\rho}}\log^{\frac{1-2r}{2}}(m)\right)\to0,\quad \text{as } m\to\infty,\ \text{since }\ \rho>0.
\end{aligned} \tag{J14}$$

Therefore, the limit distribution does not have a normal term when $\rho>0$. Lastly, we compute a_m via (J6), that is

$$\begin{aligned}
a_m&=m\mathbb{E}\left[Y_{mi}\mathbf{1}\{Y_{mi}<1\}\right]=m^{r+1}\mathbb{E}\left[P_{mi}^r\mathbf{1}\left\{P_{mi}^r<\frac{1}{m^r}\right\}\right]\\
&=-\frac{m^{r+1}}{r\sqrt{1-\rho}}\int_1^{m^{-r}}y^{\frac{1}{r}}\exp\left(-\frac{\rho\Phi^{-1}\left(y^{\frac{1}{r}}\right)^2+2A_m\Phi^{-1}\left(y^{\frac{1}{r}}\right)+A_m^2}{2(1-\rho)}\right)dy,
\end{aligned} \tag{J15}$$

where $A_m = \sqrt{\rho} z_0 + \mu_m = o(\sqrt{\log m})$. Let $x = \Phi^{-1}(y^{\frac{1}{r}})$, we have that (J15) equals

$$\frac{m^{r+1}}{\sqrt{1-\rho}} \int_{\Phi^{-1}(\frac{1}{m})}^{\infty} \Phi(x)^r \exp\left(-\frac{x^2 + 2A_m x + A_m^2}{2(1-\rho)}\right) dx$$
$$= \frac{m^{r+1}}{\sqrt{1-\rho}} \left(\int_{\Phi^{-1}(\frac{1}{m})}^{1} + \int_{1}^{\infty}\right) \Phi(x)^r \exp\left(-\frac{x^2 + 2A_m x + A_m^2}{2(1-\rho)}\right) dx = \frac{m^{r+1}}{\sqrt{1-\rho}} (\mathbb{I}_1 + \mathbb{I}_2). \quad \text{(J16)}$$

Using the following well-known Mill's inequality (Gordon, 1941), that is for any $u > 0$,

$$\frac{u}{1+u^2}\phi(u) \le \Phi(-u) \le \frac{1}{u}\phi(u), \quad \text{(J17)}$$

we have that

$$\mathbb{I}_1 \le \int_{\Phi^{-1}(\frac{1}{m})}^{1} \left(\frac{-x}{1+x^2}\right)^r \phi(-x)^r \exp\left(-\frac{x^2 + 2A_m x + A_m^2}{2(1-\rho)}\right) dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{1}^{-\Phi^{-1}(\frac{1}{m})} \left(\frac{x}{1+x^2}\right)^r \exp\left(-\frac{[r(1-\rho)+1]x^2 - 2A_m x + A_m^2}{2(1-\rho)}\right) dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{1}^{-\Phi^{-1}(\frac{1}{m})} \left(\frac{1}{x} + x\right)^{-r} \exp\left(-\frac{[r(1-\rho)+1]x^2 - 2A_m x + A_m^2}{2(1-\rho)}\right) dx$$
$$\le \frac{2^{-r}}{\sqrt{2\pi}} \int_{1}^{-\Phi^{-1}(\frac{1}{m})} x^{-r} \exp\left(-\frac{[r(1-\rho)+1]x^2 - 2A_m x + A_m^2}{2(1-\rho)}\right) dx$$
$$= \frac{2^s}{\sqrt{2\pi}\exp(c_m)} \int_{1}^{-\Phi^{-1}(\frac{1}{m})} x^s \exp\left(\frac{a}{2}x^2 + b_m x\right) dx, \quad \text{(J18)}$$

and

$$\mathbb{I}_1 \ge \int_{\Phi^{-1}(\frac{1}{m})}^{1} \left(\frac{1}{-x}\right)^r \phi(-x)^r \exp\left(-\frac{x^2 + 2A_m x + A_m^2}{2(1-\rho)}\right) dx$$
$$= \frac{1}{\sqrt{2\pi}\exp(c_m)} \int_{1}^{-\Phi^{-1}(\frac{1}{m})} x^s \exp\left(\frac{a}{2}x^2 + b_m x\right) dx, \quad \text{(J19)}$$

where $s = -r \ge 1$; $a = \frac{r(\rho-1)-1}{1-\rho} = s - \frac{1}{1-\rho}$; $b_m = \frac{A_m}{1-\rho} > 0$; $c_m = \frac{A_m^2}{2(1-\rho)}$. Combining (J18) and (J19), we have that

$$\mathbb{I}_1 = O\left(\exp(c_m) \int_{1}^{-\Phi^{-1}(\frac{1}{m})} x^s \exp\left(\frac{a}{2}x^2 + b_m x\right) dx\right). \quad \text{(J20)}$$

In the following, we first consider $a > 0$, under which case we demonstrate the rate of $\mathbb{I}_1$. Then we argue that the case with $a \le 0$ will only lead to a slower rate.

Let $h_m(x) = x^s \exp\left(\frac{a}{2}x^2 + b_m x\right)$. When $x > 1$, we have

$$\frac{\partial^2 h_m(x)}{\partial x^2} = \left[(ax + b_m)^2 x^s + a(2s+1)x^s + 2sb_m x^{s-1} + s(s-1)x^{s-2}\right] \exp\left(\frac{a}{2}x^2 + b_m x\right) \geq 0, \quad \text{(J21)}$$

that is, h_m is convex in x for $x > 1$. Plugging into (J20), we have

$$\begin{aligned}\mathbb{I}_1 &\lesssim \frac{2^{s-1}}{\sqrt{2\pi}\exp(c_m)} \left|\Phi^{-1}\left(\frac{1}{m}\right)\right| \left[\exp\left(\frac{a}{2} + b_m\right) + \left|\Phi^{-1}\left(\frac{1}{m}\right)\right|^s \exp\left(\frac{a}{2}\Phi^{-1}\left(\frac{1}{m}\right)^2 + b_m\Phi^{-1}\left(\frac{1}{m}\right)\right)\right] \\ &= o\left(m^{-r-\frac{1}{1-\rho}} \log^{\frac{1-r}{2}}(m)\right) \quad \text{as} \quad m \to \infty. \end{aligned} \quad \text{(J22)}$$

On the other hand, we have

$$\mathbb{I}_2 \leq 2^{-r} \int_1^\infty \exp\left(-\frac{x^2 + 2A_m x + A_m^2}{2(1-\rho)}\right) < 2^{-r} \int_{-\infty}^\infty \exp\left(-\frac{(x + A_m)^2}{2(1-\rho)}\right) = 2^{-r}\sqrt{2\pi(1-\rho)}, \quad \text{(J23)}$$

using the fact

$$\int_{-\infty}^\infty \exp(-ax^2)dx = \sqrt{\frac{\pi}{a}}, \quad (a > 0).$$

Finally, plugging (J22) and (J23) into (J16), we have that

$$a_m = o\left(m^{1-\frac{1}{1-\rho}} \log^{\frac{1-r}{2}}(m)\right) \to 0 \quad \text{as } m \to \infty, \text{ since } \rho > 0. \quad \text{(J24)}$$

Based on the above calculations, we can finally apply Lemma 5 and have

$$\sum_{i=1}^m Y_{mi} - a_m \xrightarrow{p} 0 \quad \text{for all } r \leq -1. \quad \text{(J25)}$$

Therefore, when $r < -1$,

$$\begin{aligned}\lim_{m\to\infty} \beta_{\mu_m,\pi_m,\rho}(r,\alpha) &= \lim_{m\to\infty} \beta_m(\rho, r, \alpha) = \lim_{m\to\infty} \mathbb{E}\left[\Pr\left\{\frac{1}{m}\sum p_{mi}^r \geq C_r(m,\alpha) \,\middle|\, Z_0\right\}\right] \\ &= \lim_{m\to\infty} \mathbb{E}\left[\Pr\left\{\frac{1}{m^{-r}}\sum p_{mi}^r \geq m^{1+r} C_r(m,\alpha) \,\middle|\, Z_0\right\}\right] \\ &= \mathbb{E}\left[\lim_{m\to\infty} \Pr\left\{\sum Y_{mi} \geq m^{r+1}\alpha^r m^{-1-r}\right\}\right] \\ &= \lim_{m\to\infty} \Pr\left\{\sum Y_{mi} - a_m \geq \alpha^r - a_m\right\} = 0; \end{aligned} \quad \text{(J26)}$$

and similarly when $r = -1$,

$$\begin{aligned}\lim_{m\to\infty} \beta_{\mu_m,\pi_m,\rho}(r,\alpha) &= \lim_{m\to\infty} \beta_m(\rho, \alpha, r) = \lim_{m\to\infty} \mathbb{E}\left[\Pr\left\{\frac{1}{m}\sum p_{mi}^r \geq \frac{1}{\alpha} + \log m \,\middle|\, Z_0\right\}\right] \\ &= \lim_{m\to\infty} \Pr\left\{\sum Y_{mi} - a_m \geq \frac{1}{\alpha} + \log m - a_m\right\} = 0. \end{aligned} \quad \text{(J27)}$$

In conclusion, for $r \leq -1$, we have that, $\beta(\rho, r, \alpha) = 0$ as long as $\mu_m = o(\sqrt{\log m})$ and $\rho > 0$.

On the other hand, recall that in Theorem 4 we derive that the calibrated threshold under equicorrelation when $r \le -1$ in fact equals to that under independence. Therefore when $\rho = 0$, for all $r \le -1$ we have

$$\begin{aligned}\lim_{m\to\infty} \beta_{\mu_m,\pi_m,\rho}(r,\alpha) &= \lim_{m\to\infty} \beta_m(\rho,\alpha,r) = \lim_{m\to\infty} \Pr\left\{\frac{1}{m}\sum p_{mi}^r \ge C_r(m,\alpha)\right\} \\ &= \lim_{m\to\infty} \Pr\left\{\frac{1}{m}\sum p_{mi}^r \ge C_r(m,\alpha)\right\} = \alpha. \end{aligned} \tag{J28}$$

Here we finish the proof for Theorem 6.

APPENDIX K. PROOF FOR THEOREM 7

Using the calibrated threshold $c_r(m,\alpha)$ derived in Theorem 4, we have that

$$\beta_{\mu_m,\pi_m,\rho}(r,\alpha) = \Pr\left\{\frac{1}{m}\sum_{i=1}^m p_{mi}^r \ge C_r(m,\alpha)\right\} = \Pr\left\{\sum_{i=1}^m m^r(p_{mi}^r - a_{rm}) \ge \alpha^r\right\}, \tag{K1}$$

where $C_r(m,\alpha) = c_r(m,\alpha)^r$, $a_{rm} = 0$ for $r < -1$, and $a_{rm} = \log m$ for $r = -1$.

Therefore, we only need to prove that $\sum_{i=1}^m m^r(P_{mi}^r - a_{rm}) \to \infty$ with probability one, where $a_{rm} = 0$ for $r < -1$, and $a_{rm} = \log m$ for $r = -1$. Since

$$\sum_{i=1}^m m^r(P_{mi}^r - a_{rm}) \ge \max_i\left\{m^r(P_{mi}^r - a_{rm})\right\} = (m\min\{P_{mi}\})^r - m^r a_{rm}, \tag{K2}$$

and with part (a) we have

$$\begin{aligned}\min_i\{P_{mi}\} &= \Phi(-\sqrt{1-\rho}\max_i\left\{Z_i + \mu_{mi}/\sqrt{1-\rho}\right\} - \sqrt{\rho}Z_0) \\ &= \Phi(-\sqrt{1-\rho}\sqrt{2\log m} - \mu_m - \sqrt{\rho}Z_0) + o_p(1) \\ &= O_p(m^{-(\sqrt{(1-\rho)}+\sqrt{c})^2}). \end{aligned} \tag{K3}$$

Therefore, we have that

$$(m\min\{P_{mi}\})^r - m^r a_{rm} = O_p\left(m^{-r\left(\left(\sqrt{(1-\rho)}+\sqrt{c}\right)^2-1\right)}\right) \to \infty, \tag{K4}$$

with probability one, since $\sqrt{c} > 1 - \sqrt{(1-\rho)}$. Hence we have proved the argument for part (a). Similarly, as for part (b) we have

$$\min_i\{P_{mi}\} = \Phi\left(-\sqrt{1-\rho}\max_i\left\{Z_i + \mu_{mi}/\sqrt{1-\rho}\right\} - \sqrt{\rho}Z_0\right). \tag{K5}$$

$$\le \Phi\left(-\sqrt{1-\rho}\sqrt{2\gamma\log m} - \mu_m - \sqrt{\rho}Z_0\right) + o_p(1). \tag{K6}$$

$$= O_p\left(m^{-(\sqrt{\gamma(1-\rho)}+\sqrt{c})^2}\right). \tag{K7}$$

Therefore, we have that

$$(m \min\{P_{mi}\})^r - m^r a_{rm} = O_p\left(m^{-r\left((\sqrt{\gamma(1-\rho)}+\sqrt{c})^2-1\right)}\right) \to \infty, \tag{K8}$$

with probability one, since $\sqrt{c} > 1 - \sqrt{\gamma(1-\rho)}$. Hence we have concluded the proof.