이학박사 학위논문

Cauchy Combination Test with Thresholding Under Arbitrary Dependency Structures

임의의 의존성 구조하에서 분계점을 이용한 코쉬 결합 검정 연구

2021년 8월

서울대학교 대학원

통계학과

김 준 식

Cauchy Combination Test with Thresholding Under Arbitrary Dependency Structures

지도교수 장 원 철

이 논문을 이학박사 학위논문으로 제출함

2021년 6월

서울대학교 대학원

통계학과

김 준 식

김준식의 이학박사 학위논문을 인준함

2021년 6월

위 원 장	임	요	한
부위원장	장	원	철
위 원	PARK	JUNYONG	
위 원	정	석	오
위 원	최	동	석

Cauchy Combination Test with Thresholding Under Arbitrary Dependency Structures

by

Junsik Kim

A Thesis

Submitted in fulfilment of the requirements

for the degree of

Doctor of Philosophy

in Statistics

Department of Statistics

College of Natural Sciences

Seoul National University

August, 2021

Abstract

Cauchy Combination Test with Thresholding Under Arbitrary Dependency Structures

Junsik Kim
The Department of Statistics
The Graduate School
Seoul National University

Combining individual p-values to aggregate sparse and weak effects is a substantial interest in large-scale data analysis. The individual p-values or test statistics are often correlated, although many p-values combining methods are developed under i.i.d. assumption. The Cauchy combination test is a method to combine p-values for arbitrary dependence structures, but in practice, the type I error increases as the correlation increases. In this thesis, we propose a global test that extends the Cauchy combination test by thresholding arbitrarily dependent p-values. Under arbitrary dependence structures, we show that the tail probability of the proposed method is asymptotically equivalent to that of the Cauchy distribution. In addition, we show that the power of the proposed test achieves the optimal detection boundary asymptotically in a strong sparsity condition. Extensive simulation results show that the power of the proposed test is robust to correlation structures and more powerful under a sparse situation. As a case study, we apply the proposed test to GWAS of Inflammatory bowel disease (IBD).

key words: Combining p-values, Cauchy distribution, Global hypothesis testing, GWAS

Student Number: 2016-30092

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the field of large scale multiple testing, detecting sparse or/and weak signals is a major interest problem. Jin (2008) presents following three interconnected topics for testing a large number of signals of particular interest in this field.

(a) Global testing : Is there any signal at all?

(b) Estimating the proportion of signals: How many signals are there?

(c) Simultaneous testing : Which are signals and which are noises?

For the large scale multiple testing setting, testing of signals, or significant factors, among many hypotheses, is a well-known topic and the plenty of research is under way. If it is confirmed that there is a signal through global testing, we need more information on how many signals there are, which is the second topic. Then, in the third topic, studies can be conducted on which are the signals and which are the noises. The first topic has been studied in Tippett (1931), Berk and Jones (1979), Donoho and Jin (2004, 2008) and Hall and Jin (2010). The second topic has been studied in Meinshausen and Rice (2006), Langaas et al.

(2005), Jin (2008) and Storey (2002). And the third topic related to FDR methods has been studied in Benjamini and Hochberg (1995), Efron (2010), Storey (2002) and Genovese and Wasserman (2004).

In this thesis, we mainly consider the first topic and investigate the second topic incidently. Specifically, we combine p-values as a component for the global testing and use the combined statistic to verify the existence of weak or/and sparse signals.

For motivating example, in genome-wide association study (GWAS), there are hundreds of thousands of single-nucleotide polymorphisms (SNPs). The SNP is a variation of a single nucleotide at a specific position in the genome. Although most SNPs do not affect on diseases or phenotypes, some of these genetic differences have proven to be very significant in the study of diseases. They can act as biological markers that are associated with a certain type of disease. The goal in GWAS is to test if any of these SNPs are associate with some disease or a phenotype of interest. In GWAS, a commonly used method is the set-based analysis which partitions the SNPs into genes based on the biological information and then tests associations between the disease and each gene. Each gene consists of several SNPs, and the global testing of the SNPs of the gene can be used to determine whether each gene affects a certain disease.

In this thesis, we use the data of a Crohn's disease GWAS as in Duerr et al. (2006) for a real-data application. The goal of the case study is to find genes that have associations with inflammatory bowel disease (IBD). We test each gene whether each gene itself affects the IBD.

1.1 Combining p-values

In this thesis, we address methods combining p-values as a way to solve the problem associated with the first topic, the global testing. Combining p-values involves an averaging of transformed p-values by means of specific transformations of p-values, such as the log-transformation (Fisher, 1934) and the probit transformation (Stouffer et al., 1949), as well as the minimum p-value (Tippett, 1931), in that they utilize informations of p-values. Combining p-values aggregates non-detecting signals, thus increasing the testing power. Therefore, methods for combining p-values are ongoing research tasks, and several methods have been proposed.

First, consider the case where p-values are independent. By denoting p-values obtained from the null distribution as the null p-values, the null p-values follow the i.i.d. uniform distribution, $\mathcal{U}[0, 1]$. As a result, it is a relatively easy case and many methods have been proposed. Methods such as the Higher criticism (Donoho and Jin, 2004, 2008), the Berk-Jones test (Berk and Jones, 1979) in addition to the minimum p-value (Tippett, 1931) and averaging type methods (Vovk and Wang, 2020) can combine p-values. The Higher criticism and the Berk-jones test actually measure the distance from the uniform distribution by using the order statistics of random variables obtained from the uniform distribution.

In contrast, practically, the assumption of independence between p-values is too strong, since in most real data, p-values are often correlated. The dependence between p-values keeps p-values from following an independent distribution or the uniform distribution. Indeed, if p-values are not independent, especially there is a certain type of the dependence in data from which p-values are generated, the parametrically calculated p-values can not be uniformly distributed or unknown sometimes, making it particularly difficult to combine

p-values. If the dimension of the given sample data is not large, the sample can be used to estimate the dependence structure of the data and we can decorrelate the data to apply the global testing methods that are developed for an independent case. But in many cases as in GWAS, the dimension of the data is very large, and in some cases, even worse, only p-values are provided rather than raw data. The Higher criticism and the Berk-Jones statistics defined under the independence assumption in which p-values follow the uniform distribution are also disabled. Hence combining the dependent p-values is not easy, and some methods of combining the p-values have been proposed for cases where the correlation structure is given as a certain form.

Although the Cauchy combination test that combines arbitrarily dependent p-values is proposed recently, the Cauchy combination test has the disadvantage for controlling the type I error when correlations between p-values increase as we can see this in simulation studies. Therefore, to reduce the effects of correlations, we propose a more robust and powerful testing method by thresholding null effect signals or equivalently p-values and using only small p-values.

1.2 Main Contributions

In this thesis, we propose a global test statistic for detecting the signals by combining p-values when the dependence between p-values exists. Specifically, arbitrarily dependent p-values are thresholded and then transformed to follow the uniform distribution marginally. Thereafter the p-values are Cauchy transformed by using the tangent function then combined with weights. We show that the tail probability of the proposed statistic is equivalent to that of the Cauchy distribution. The equivalence holds non-asymptotically for finite dimension case and also asymptotically for infinite dimension case. To choose the thresholding value, we use $2\hat{\pi}_1$, where $\hat{\pi}_1$ is an estimate of the proportion of non-null hypotheses. To estimate $\hat{\pi}_1$, we propose to use the Storey method (Storey, 2002) which uses the fact that large *p*-values are obtained from the null distribution. We extend the Storey's method to the case of arbitrary dependent *p*-values. The proposed method is adaptive to the observed *p*-values in contrast to the methods that are developed for the theoretical null distribution, $\mathcal{U}[0, 1]$. Extensive numerical simulations show that the proposed method estimates the proportion of non-null hypotheses more stably than other competing methods.

We propose a more powerful and robust *p*-value-based testing procedure by thresholding null *p*-values under arbitrary dependence structure. We prove that the power of the proposed method achieves the optimality in the asymptotic sense under a strong sparsity setting. In general, the thresholding method improves the power when high-dimensional alternative hypotheses are tested since it uses only significant data. Moreover, for the highly correlated case where the effective sample size is small, the thresholding method using a small amount of data improves the power. Despite these enhancements, test statistics based on the thresholding are restrictively used in practice, in that the null distribution of thresholded data is hard to find or the convergence rate of the null distribution of thresholded data is very slow. For example, Fan (1996) uses the adaptive Neyman test by thresholding data, however, its convergence rate of the test statistic under the null is very slow. Fan (1998) presents a table that shows the finite sample distribution for the adaptive Neyman test statistic under null hypotheses. On the other hand, the thresholding method proposed in this thesis has a self-contained null tail probability. The tail probability of null distributions for the thresholded test statistic is equivalent to that of the un-thresholded test statistic. The fast convergence rate of null distribution allows the proposed method to be well applied even in the high-dimensional setting.

Extensive simulation results in Chapter 5 present that the type I error of the proposed combining method is well controlled with respect to the correlation between p-values than the Cauchy combination test of Liu and Xie (2020). Analyzing the testing power of the proposed method implies that as the dimension of data increases, the power to separate the null and alternative hypotheses increases. By using a small proportion of significant p-values, numerical studies show that the proposed method tends to be more powerful than other p-values combining methods.

1.3 Outline of the Thesis

The remainder of this thesis is organized as follows. In chapter 2, we review the methods that combine p-values to aggregate information for the global test. Methods developed under an independent case are briefly reviewed as well as those under the dependent case. The Cauchy combination test under arbitrary dependency structure is reviewed in detail. The Cauchy combination test with thresholding is presented in chapter 3. In analyzing the Cauchy combination test with the thresholding method, establishing the threshold value contains the problem of estimating the proportion of non-null hypotheses, and it is presented in chapter 4. We present extensive simulations results of type I error and power in chapter 5. In chapter 6, We use our method to demonstrate the effectiveness by applying real data of SNPs of Inflammatory bowel disease. We end up this thesis with concluding remarks in Chapter 7. The proofs of the main theorems are in Appendix.

Chapter 2

Literature Review

Combining of p-values is a common method in a high-dimensional setting that combines individual p-values to integrate weak signals that cannot be detected to increase power for a global test. When the dimension of data is much larger than the number of sample subjects, such as in the case of GWAS, by combining the information contained in individual p-values, a powerful testing procedure can be constructed.

Suppose that d-dimensional data follow the normal distribution with a covariance matrix Σ. Define $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T$ as an effect of signal.

$$X \sim N_d(\boldsymbol{\mu}, \Sigma). \tag{2.1}$$

If there are no signals in the data, the mean of each covariate is zero. The main concern of this setting is to test that if there are any signals in the data. It can be expressed as a hypothesis testing as follows.

$$H_0 : X \sim N_d(0, \Sigma) \quad \text{vs.} \quad H_1 : X \sim N_d(\boldsymbol{\mu}, \Sigma) \tag{2.2}$$

Testing the hypothesis (2.2) is equivalent to test the effect of signals is zero or non-zero under the setting (2.1) which is a two-sided hypothesis testing,

$$H_0 : \boldsymbol{\mu} = 0 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu} \neq 0. \tag{2.3}$$

For such a two-sided test procedure, *p*-values, $p_1, \ldots, p_d$, are defined by

$$p_i = 2[1 - \Phi(|X_i|)], \quad i = 1, \ldots, d,$$

where Φ is the cumulative distribution function of the normal distribution.

To alleviate the condition (2.1) which is the normality of data, we assume that the *p*-values are obtained from *z*-scores. For example, if we use a two-sample *t*-statistic to compare the means of two samples, the test statistic follows the *t* distribution with some degree of freedom, say df. Let t_i be the *t* test statistic for i-th covariate. Then the *z*-score of X_i is defined by

$$X_i = \Phi^{-1}(F_{df}(t_i)),$$

where F_{df} is the cumulative distribution function of the *t* distribution with df degree of freedom. In general, for observed data Y of some certain form, let $T_i(Y)$, $i = 1, \ldots, d$ be the i-th test statistic of Y for testing hypothesis for the mean comparison and F be the cumulative distribution function of the test statistic. Define

$$X_i = \Phi^{-1}(F(T_i(Y))), \quad i = 1, \ldots, d$$

as the *z*-scores. Then by *z*-scoring the test statistic, under the null, X_i follows the standard normal distribution. Then we can use the *z*-score X_i to construct a valid *p*-value, p_i. By the constructing, the null distribution of the test statistic must have a known distribution. As a consequence, we can always standardize each test statistic so that $\mathsf{Var}(X_i) = 1$. From now, without loss of generality, we assume that diagonal elements of Σ in (2.1) is 1, which implies that Σ is a correlation matrix.

Combining $p_1, \ldots, p_d$ to test (2.2) is the main topic of this thesis. This chapter provides the detailed review of methods that combine p-values in the weak and/or sparse signal setting. In Chapter 2.1, we review methods that combine p-values where the independence between p-values is assumed and in Chapter 2.2, methods of combninig p-values under an assumption of the dependence between p-values are reviewed.

2.1 Combining p-values Under Independence Structure

Suppose that p-values are independent that is, $\Sigma = I_d$ in (2.1) and (2.2), where I_d is a $d \times d$ identity matrix. In this case, the independent p-values, obtained from null, follow the independent identically uniform distribution, $\mathcal{U}[0, 1]$, that is, under the complete null which is $\boldsymbol{\mu} = \mathbf{0}$,

$$p_1, \ldots, p_d \overset{i.i.d.}{\sim} \mathcal{U}[0, 1].$$

This implies that the null p-values follow the i.i.d. uniform distribution. Using the property of the uniform distribution, we can combine p-values to proceed with a global testing procedure by measuring the distance between the null and alternative. In other words, the effect of alternative can be determined by measuring how far the combining statistic of the complete null p-values are compared to those of the actual observed p-values.

For that reasons, the combining statistic of the p-values constructed under the null must be obtained in a form that is easy to use or implement. A common method to combine p-values when the null p-values follow the i.i.d. uniform is the empirical distribution function,

$$F_d(u) = \frac{1}{d} \sum_{i=1}^{d} I(p_i \leq u),$$

where $I(A)$ is an indicator function that is one if the event A is true and zero otherwise. If the null p-values follow the i.i.d. uniform distribution, $d \cdot F_d(u)$ follows a binomial distribution, $\mathcal{B}(d, u)$, which can be used to measure the distance between the null and alternative. On th other hand, from the fact that the null p-values follow a uniform distribution, we can consider a method using the ordered statistics of p-values which follow a beta distributions marginally.

In Chapter 2.1.1, we review methods that use the empricial distribution function of p-values and in Chapter 2.1.2 we review methods that use the combination of p-values.

2.1.1 Based on Empirical Distribution of p-values

Donoho and Jin (2004) proposed the Higher criticism method that is closely related to functionals of the standard uniform empirical process. Define the uniform empricial process,

$$U_d(u) = \sqrt{d}(F_d(u) - u), \;\; 0 < u < 1,$$

and the normalized uniform empirical process

$$W_d(u) = \frac{U_d(u)}{\sqrt{u(1-u)}}.$$

Then the Higher criticism statistic, HC_d^* is defined by

$$HC_d^* = \max_{0<u<\alpha_0} W_d(u),$$

for some $\alpha_0 \in (0, 1)$. If all p-values are the null p-values, for fixed u, $W_n(u)$ follows the standard normal distribution asymptotically. Accordingly, a limiting behavior under the alternative is expected to be different to that of the null. Let $p_{(i)}$ be the i-th order statistic of p-values in the increasing order, HC_d^* can be expressed as follows.

$$HC_d^* = \max_{0<i<\alpha_0 \cdot d} \frac{\sqrt{d}(i/d - p_{(i)})}{\sqrt{p_{(i)}(1 - p_{(i)})}}. \tag{2.4}$$

Donoho and Jin (2004) proved that for fixed $\alpha > 0$, under the null,

$$P(HC_d^* > h(d, \alpha)) \leq \alpha,$$

where $h(d, \alpha) \approx \sqrt{2 \log \log(d)}$. Donoho and Jin (2004) also proved that the Higher criticism has a full power to detecting signals that is, under the alternative, as $d \to \infty$,

$$P(HC_d^* > h(d, \alpha_d)) \to 1,$$

where $h(d, \alpha_d) = \sqrt{2 \log \log(d)}(1 + o(1))$.

From the properties of the sorted null p-values including the asymptotic normality,

$$p_{(i)} \sim_{\text{approx}} N(i/d, i/d(1 - i/d)),$$

Donoho and Jin (2008) defined a different type of the Higher criticism such that

$$HC_d^* = \max_{0 < i < \alpha_0 \cdot d} \frac{\sqrt{d}(i/d - p_{(i)})}{\sqrt{i/d(1 - i/d)}}.$$

Similar to (2.4), the large discrepancy between the expected behavior under null and the observed behavior reflects the distance between the null and alternative.

However, Barnett et al. (2017) pointed out that the Higher criticism which is viewed as a supremum of the normalized empirical process converges asymptotically to a Gumbel distribution with a very slow rate. Ditzhaus and Janssen (2017) also showed that the Higher criticism test has a trivial power that is, no power on the boundary for various settings.

Another global testing procedure is the Berk-Jones statistic (Berk and Jones, 1979) which is based on the likelihood ratio statistic. For the uniformly distributed null p-values, the Berk-Jones statistic, BJ_d, is defined by

$$BJ_d = \sup_{-\infty < u < \infty} K(F_d(u), u), \tag{2.5}$$

where

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

Wellner and Koltchinskii (2003) represented the Berk-Jones statistic (2.5) by using the liklihood ratio statistic. Since $d \cdot F_d(u) \sim \mathcal{B}(d, u)$, the likelihood ratio statistic is expressed as follows.

$$\lambda_d(u) \quad = \quad \frac{L(F_d(u))}{L(\mathcal{U}(u))} = \left(\frac{F_d(u)}{\mathcal{U}(u)}\right)^{d \cdot F_d(u)} \left(\frac{1 - F_d(u)}{1 - \mathcal{U}(u)}\right)^{d(1 - F_d(u))}, \tag{2.6}$$

where L is the likelihood function and $\mathcal{U}(u) = P(U \leq u)$ for an uniform distribution random variable U. By taking the log function to λ_d, the Berk-Jones statistic (2.5) can be expressed as follows.

$$BJ_d = \sup_{0 \leq u \leq 1} \frac{1}{d} \log \lambda_d(u).$$

Using the sorted p-values as in the Higher criticism (2.4), the Berk-Jones statistic also can be expressed by

$$BJ_d = \sum_{1 \leq i \leq d} d \cdot \left\{ (i/d) \log \left(\frac{i/d}{p_{(i)}}\right) + \left(1 - \frac{i}{d}\right) \log \left(\frac{1 - i/d}{1 - p_{(i)}}\right) \right\}.$$

Wellner and Koltchinskii (2003) proved that, under the null, the Berk-Jones statistic converges to the null distribution Y_4, such that

$$nR_d - p_d \xrightarrow{d} Y_4 \sim E_v^4,$$

where $E_v^4 = P(Y_4 \leq x) = \exp(-4 \exp(-x))$ and $p_d = \log_2 d + (1/2) \log_3 d - (1/2) \log(4\pi)$. Li and Siegmund (2015) showed that over a wide range of sample sizes, the Berk-Jones statistics are much powerful than the Higher criticism to detect a sparse mixture setting.

2.1.2 Based on Combination Statistic of p-values

For the independent and identically uniformly distributed p-values, we can consider the order statistic of the p-values. The order statistic uses all informations of p-values in that it requires orders of each p-value. For $p_1, \ldots, p_d \overset{i.i.d}{\sim} \mathcal{U}[0,1]$, it is known that for each $i = 1, \ldots, d$,

$$p_{(i)} \sim \text{Beta}(i, d+1-i). \tag{2.7}$$

Tippett (1931) proposed to use the minimum p-value, that is, $p_{(1)}$. Using the minimum p-value is equivalent to using only the maximum test statistic. Arias-Castro et al. (2011) refered to the procedure using the minimum p-value as the Max test and proved the optimality for using the Max test.

Combinating informations by weighted averaging p-values, in general, can be expressed as follows.

$$S = \sum_{i=1}^{d} w_i f(p_i), \tag{2.8}$$

where w_i's are weights and f is a certain type of transformation of p-values. This form of combination statistics include Fisher's method (Fisher, 1934), Pearson's method (Pearson, 1933) and Stouffer's method (Stouffer et al., 1949). The Fisher's method is defined by $w_i = 1$ and $f(p_i) = \log(p_i)$ for all $i = 1, \ldots, d$ in (2.8). If the p-values follow i.i.d. uniform distribution, the Fisher's statistic,

$$S_F = -2 \sum_{i=1}^{d} \log p_i$$

follows the Chi-square distribution with $2d$ degree of freedom. The Pearson's statistic and

Stouffer's statistic are defined by

$$S_P = -\sum_{i=1}^{d} \log(1 - p_i), \quad S_T = \sum_{i=1}^{d} \Phi^{-1}(p_i),$$

following the Chi-square distribution and the normal distribution, respectively.

If the p-values are uniformly distributed, the null distribution of such summation type statistics is obtained by a known distribution, so the test statistics can be used to proceed with the global testing procedure. According to Heard and Rubin-Delanchy (2018), the Fisher's method and the minimum p-value are sensitive to the smallest p-value, and the Pearson's method is sensitive to the largest p-value.

2.2 Combining p-values Under Dependence Structure

Methods reviewed in Chapter 2.1.1 and 2.1.2 are based on the fact that the null p-values follow the uniform distribution independent and identically. Under a dependence structure, the assumption of uniform distribution is not valid anymore. Unlike the methods reviewed in Chapter 2.1, if the p-values are dependent on each other, the null distribution of combined test statistics is hard to obtain so that these methods are difficult to use practically. Instead, many methods such as decorrelating or estimating the dependence structure methods are used. In addition, a testing method using the Cauchy combination has recently been proposed for arbitrary dependency structures, which will be introduced in the following chapters.

2.2.1 Based on Decorrelating or Estimating Correlation Structure

Suppose that Σ in (2.1) is known. Hall and Jin (2010) proposed the innovated Higher criticism to test (2.3). Specifically, the goal of the innovated Higher criticism is testing following two hypotheses.

$$H_0 : \boldsymbol{\mu} = 0 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu} \text{ is a sparse vector .}$$

To represent a sprasity of the mean vector $\boldsymbol{\mu}$, the number of nonzero entries of $\boldsymbol{\mu}$ can be expressed as d^{γ}, where $\gamma \in (0, 1/2)$. Denote the inverse matrix of Cholesky factorization of Σ as $U = (u_{jk})_{1 \leq j,k \leq d}$. Then the innovated Higher criticism is a standard Higher criticism for the decorrelated data by using U with a fixed bandwidth b_d depending on the dimension d, which is used to construct a decorrelating matrix. The resulting innovated Higher criticism statistic is defined by

$$iHC = \frac{1}{\sqrt{2b_d - 1}} \sup_{j:1/d \leq p_{(j)} \leq 1/2} \left\{ \sqrt{d} \cdot \frac{j/d - p_{(j)}}{\sqrt{p_{(j)}(1 - p_{(j)})}} \right\}.$$

The innovated Higher criticism can be viewed as a supremum of the normalized empirical process as the Higher criticism and Barnett et al. (2017) pointed out that the innovated Higher criticism converges asymptotically, under the null, to an extreme value distribution with a very slow rate of $O((\log d)^{-1/2})$.

If Σ is unknown, an estimate of Σ can be used to define the innovated Higher criticism. However, for the setting in this thesis where only z-scores of test statistics are given, estimating Σ is not possible since we only have 1-sample data.

2.2.2 Based on Merging p-values

Wilson (2019) proposed the harmonic mean of p-values to combine dependent p-values. If

p-values are independent and identically distributed, the harmonic mean of p-values,

$$R_H = \frac{1}{d} \sum_{j=1}^{d} \frac{1}{p_j},$$

converges to a Landau distribution which is the stable distribution by the generalized central limit theorem. Wilson (2019) showed that the approximation of the null distribution using the generalized central limit theorem is robust to dependencies between the p-values. An average of reciprocals of p-values weights a large value to the small p-value and small value to the large p-value so that the harmonic mean statistic, R_H is a valid global testing statistic. However, Goeman et al. (2019) pointed out that the harmonic mean statistic loses controlling type I error under an arbitrary dependence structure.

Vovk and Wang (2020) generalized the combination of p-values using the combining function given by

$$M_{r,K(p_1,\ldots,p_d)} = \left(\frac{p_1^r + \cdots + p_d^r}{d} \right)^{1/r}, \tag{2.9}$$

for $r \in \mathbb{R} \setminus \{0\}$ and where K is a combining function. The harmonic mean is a specific form of (2.9) when $r = -1$. Limit cases of (2.9) include that

$$\begin{aligned} M_{0,K(p_1,\ldots,p_d)} &= \left(\prod_{j=1}^{d} p_j \right)^{1/d}, && \text{as } r \to 0 \\ M_{\infty,K(p_1,\ldots,p_d)} &= \max(p_1, \ldots, p_d), && \text{as } r \to \infty \\ M_{-\infty,K(p_1,\ldots,p_d)} &= \min(p_1, \ldots, p_d), && \text{as } r \to -\infty. \end{aligned}$$

Vovk and Wang (2020) proved that the merging functions $M_{\infty,K(p_1,\ldots,p_d)}$ and $d \cdot M_{-\infty,K(p_1,\ldots,p_d)}$ are precise and $e \cdot M_{0,K(p_1,\ldots,p_d)}$ is asymptotically precise.

Now, consider a compound symmetry correlation structure,

$$\mathsf{Cov}(X_i, X_j) = \rho, \quad \text{for } i \neq j, \ \ i, j = 1, \ldots, d,$$

where $-1/(d-1) \leq \rho \leq 1$. Hartung (1999) proposed the weighted inverse normal statistic under the null, which is given as follows.

$$t(\rho) = \frac{\sum_{i=1}^{d} \lambda_i X_i}{\sqrt{(1-\rho)\sum_{i=1}^{d} \lambda_i^2 + \rho \left(\sum_{i=1}^{d} \lambda_i\right)^2}} \sim N(0,1) \quad \text{under } H_0.$$

Hartung (1999) also defined an estimator of $t(\rho)$ by

$$t(\hat{\rho}^*, \kappa) = \frac{\sum_{i=1}^{d} \lambda_i X_i}{\sqrt{\sum_{i=1}^{d} \lambda_i^2 + \left[\left(\sum_{i=1}^{d} \lambda_i\right)^2 - \sum_{i=1}^{d} \lambda_i^2\right] \cdot \left\{\hat{\rho}^* + \kappa \cdot \sqrt{\frac{2}{d+1}}(1-\hat{\rho}^*)\right\}}}, \quad (2.10)$$

where $\hat{\rho}^* = \max\{-1/(d-1), \hat{\rho}\}$, $\kappa > 0$ and

$$\hat{\rho} = 1 - \frac{1}{d-1} \sum_{i=1}^{d} \left(X_i - \frac{1}{d}\sum_{i=1}^{d} X_i\right)^2,$$

which is an unbiased estimator of ρ. Then, under the null, $t(\hat{\rho}^*, \kappa)$ is approximately standard normally distributed so that it can be used as a global testing procedure.

Demetrescu et al. (2006) extended (2.10) to allow for a certain type of correlation matrix. Let $\Sigma = (\mathsf{Cov}(X_i, X_j))_{i,j=1,\ldots,d} = (\rho_{ij})_{i,j=1,\ldots,d}$ and assume that

$$\lim_{d\to\infty} \frac{1}{d(d-1)} \sum\sum_{i\neq j} \rho_{ij} = \tilde{\rho},$$

where $\tilde{\rho} \in (0,1)$ and that

$$\lim_{d\to\infty} \frac{1}{d(d-1)} \sum\sum_{i\neq j} (\rho_{ij} - \tilde{\rho})^2 = 0.$$

Demetrescu et al. (2006) proved that if $\lambda_i = \lambda$ for all $i = 1, \ldots, d$ in (2.10), then $t(\hat{\rho}^*, \kappa)$ is still approximately standard normally distributed. When there are a finite number of outliers and the average of elements of the correlation matrix Σ converges to a constant as the dimension grows, the Hartung's weighted inverse normal method can be used as a global testing procedure.

2.3 Cauchy Combination Test

Recall that the weighted averaging of transformation of *p*-values, (2.8),

$$S = \sum_{i=1}^{d} w_i f(p_i).$$

Weighted averaging methods such as the Fisher's method, Pearson's method and Stouffer's method reviewed in Chapter 2.1.2 are all the global testing procedures under an independence assumption. These methods may not only be difficult to control type I errors but also hard to obtain the null distribution of test statistics unless the independence assumption is satisfied. Recently, Liu and Xie (2020) proposed a global testing procedure that uses a Cauchy transform as f in (2.8) under arbitrary dependency structures. Pillai and Meng (2016) showed that when $Y = (Y_1, \ldots, Y_d)^T$ and $Z = (Z_1, \ldots, Z_d)^T$ are i.i.d. $N_d(0, \Sigma)$ where Σ is an arbitrary correlation matrix, the convex combination of Y_i/Z_i, that is,

$$\sum_{i=1}^{d} w_i Y_i / Z_i,$$

where $0 \leq w_i \leq 1$ for $i = 1, \ldots, d$ and $\sum_{i=1}^{d} w_i = 1$, follows a standard Cauchy distribution. The idea behind the Cauchy combination is that since the Cauchy distribution has heavy tails, it is robust to arbitrarily dependent random variables. In other words, in that the tail behavior of Cauchy combination applied to arbitrary dependent *p*-values is analogous to that of standard Cauchy distribution, Liu and Xie (2020) proposed the Cauchy combination test procedure. Despite the robustness of the Cauchy combination, as seen in Chapter 5, it can not control the type I error well for strongly correlated *p*-values. In the following chapters, we review, in detail, the Cauchy combination test procedure which is the one of main ingredients in this thesis.

2.3.1 Cauchy Combination Test Under Independence Structure

Let $\Sigma = I_d$ to consider the case of independence structure, where I_d is an identity matrix of size $d \times d$. Note that the independence between *p*-values implies that, under the null,

$$p_1, \dots, p_d \overset{i.i.d.}{\sim} \mathcal{U}[0, 1].$$

Define a Cauchy transformation, $h : [0, 1] \to \mathbb{R}$, of each *p*-value, by

$$h(p_i) = \tan\{(1/2 - p_i)\pi\}, \quad i = 1, \dots, d.$$

By the definition of h, note that each $h(p_i)$ follows a standard Cauchy distribution. Williams (1969) showed that the convex combination of independent $h(p_i)$ for $i = 1, \dots, d$ follows a standard Cauchy distribution exactly.

Lemma 1 (Williams, 1969)**.** *Let* $\min_{1 \le i \le d} w_i \ge c_0/d$ *and* $\sum_{i=1}^{d} w_i = 1$*. Then the convex combination of* $h(p_i) := \tan\{[1/2 - p_i]\pi\}$*, for* $i = 1, \dots, d$*, follows a standard Cauchy distribution,*

$$T(P) = \sum_{i=1}^{d} w_i h(p_i) \sim \text{Cauchy}(0, 1). \tag{2.11}$$

Proof. If the a_i possess no finite limit-point then for $U \sim \text{Cauchy}(0, 1)$,

$$\sum_{i=1}^{d} w_i \cdot \frac{1 + a_i U}{a_i - U} \sim \text{Cauchy}(0, 1).$$

If $a_i = \infty$, then

$$\sum_{i=1}^{d} w_i \cdot \frac{1 + a_i U}{a_i - U} = \sum_{i=1}^{d} w_i \cdot U \sim \text{Cauchy}(0, 1).$$

$\square$

Using the above lemma 1, the global testing prodecure using $T(P)$ under the independence structure can be applied by rejecting the null hypothesis if

$$T(P) > t_C(\alpha),$$

where $t_C(\alpha)$ is the α-quantile of a standard Cauchy distribution.

2.3.2 Cauchy Combination Test Under Arbitrary Dependency Structure

For arbitrarily dependent p-values, Liu and Xie (2020) studied the behavior of the convex combination of Cauchy transformations. By the definition of p-values, that is,

$$p_i = 2[1 - \Phi(|X_i|)], \quad i = 1, \ldots, d,$$

the Cauchy transformation function $h(p_i)$ is equivalently expressed in terms of X_i such that

$$h(p_i) = \tan\{(1/2 - p_i)\pi\} = \tan\{(2\Phi(|X_i|) - 3/2)\pi\}.$$

Now we denote h_p as the Cauchy transformation function in terms of p-values and h_X as the Cauchy transformation function in terms of statistics $X_1, \ldots, X_d$, that is,

$$\begin{aligned} h_p(p_i) &= \tan\{(1/2 - p_i)\pi\} \\ h_X(X_i) &= \tan\{(2\Phi(|X_i|) - 3/2)\pi\}. \end{aligned}$$

Using the Cauchy transformation function h_X, let T_X be the convex combination of $h_X(X_i)$ for $i = 1, \ldots, X$ as follows.

$$T_X(X) = \sum_{i=1}^{d} w_i h_X(X_i) = \sum_{i=1}^{d} w_i \tan\{(2\Phi(|X_i|) - 3/2)\pi\},$$

where $\min_{1 \leq i \leq d} w_i \geq c_0/d$ and $\sum_{i=1}^{d} w_i = 1$.

Liu and Xie (2020) showed the tail probability of a standard Cauchy distribution is proportional to the reciprocal of a critical value asymptotically.

Lemma 2 (Liu and Xie, 2020). *Let W_0 have a standard Cauchy distribution. Then*

$$P\left(W_0 > t\right) = \frac{1}{t\pi} + O(t^{-3}).$$

Under the arbitrary depdendency structures, Liu and Xie (2020) proved that the tail probability of T_X, which is $P(T_X(X) > t)$ for large t, is equivalent to that of a standard Cauchy distribution if the bivariate normality assumption holds.

Definition 1 (Bivariate normality).
For any $1 \leq i < j \leq d$, $(X_i, X_j)^T$ follows a bivariate normal distribution.

Theorem 1 (Liu and Xie, 2020). *Under the bivariate normality assumption, if the correlation matrix of X is positive semi-definite, then*

$$\lim_{t \to \infty} \frac{P(T_X(X) > t)}{P(W_0 > t)} = 1,$$

where W_0 denotes a generic standard Cauchy random variable.

Liu and Xie (2020) also proved that as the dimension d increases, the tail probability of T_X is equivalent to that of a standard Cauchy distribution asymptotically.

Theorem 2 (Liu and Xie, 2020). *Under the bivariate normality assumption and additional regular conditions of the arbitrary correlation structure, for $d = o(t^c)$ for any constant $0 < c < 1/2$,*

$$\lim_{t \to \infty} \frac{P(T_X(X) > t)}{P(W_0 > t)} = 1,$$

where W_0 denotes a generic standard Cauchy random variable.

While the Cauchy combination test is somewhat robust to arbitrary dependence structures, simulation results in Chpater 5 show that the Cauchy combination test fails to control the type I error when p-values are strongly correlated. In addition, although, under the sparse alternative, the asymptotic power of the Cuacy combination test converges to 1 for any significance level, for highly correlated case with a finite sample, the power is shown to decrease. Therefore, in the next chapter, we propose a more robust and powerful method than the Cauchy combination test by adaptively thersholding p-values.

Chapter 3

Cauchy Combination Test with Threshold Under Arbitrary Dependency Structures

Again, consider a hypothesis testing problem such that

$$H_0 : X \sim N_d(0, \Sigma) \quad \text{vs.} \quad H_1 : X \sim N_d(\boldsymbol{\mu}, \Sigma).$$

As mentioned in Chapter **??**, we consider X as a normally distributed data or z-scored test statistic

$$X_i = \Phi^{-1}(F(T_i(Y))), \quad i = 1, \ldots, d$$

where $T_i(Y)$ is a test stastistic standardized to have variance 1 and F is the cumulative distribution function of the test statistic. Based on X_i, define p-values as follows.

$$p_i = 2[1 - \Phi(|X_i|)], \;\; i = 1, \ldots, d.$$

Suppose that sparse signals are concentrated in few locations, say first m covariates are signals and let m be much smaller than the dimension d. The Cauchy combnation test, which is defiend by

$$T_X(X) = \sum_{j=1}^{d} w_j h(X_j),$$

can be decomposed into a signal part, S, and noise part, N, as follows.

$$T_X(X) = \sum_{j=1}^{m} w_j h_X(X_j) + \sum_{j=m+1}^{d} w_j h_X(X_j) =: S + N.$$

In the noise part N, X_j's follow a normal distribution so that *p*-values are uniformly distributed marginally. Without a priori information about weights w_i's, let the weights be equal, that is, $w_i = 1/d$ for $i = 1, \ldots, d$. For small m, since the test statistic T_X is an average of the noise part and the signal part, the noise part N that has a negative effect on T_X tends to dominate the signal part S, which implies that the power of T_X for detecting the signals would decrease as m decreases. Moreover, if X_i's are highly correlated, the effective sample size decreases and using all X_i's is not conducive to the statistical power. Therefore, by thresholding X_i or equivalently p_i, we can enhance the power of the Cauchy combination test. This is a main motivation of the proposed method.

In general, there are different types of the thresholding method. The first type is thresholding data or statistics X_i's which are larger than some critical value. The second type is thresholding first few sorted data. The former thresholding is called the hard thresholding and the latter type is called the Neyman's truncation (Neyman, 1937). Fan (1996) showed that hard thresholding X_i's ourperforms the Neyman's truncation method and proved the stastistic constructed by hard thresholding converges to a standard normal distribution. However, the convergence is so slow and it is hard to practically used.

We use, in this thesis, the hard thresholding type of *p*-values and propose a test statistic combining the thresholded *p*-values with an easily obtainable null distribution of the test statistic. Let

$$\mathcal{M} = \{i \in \{1, \ldots, d\} : p_j \in [0, \delta]\},$$

where $0 < \delta \leq 1$ is given. Define

$$q_j = \frac{p_j}{\delta}, \ \ j \in \mathcal{M}, \tag{3.1}$$

so that $q_j \in [0, 1]$ for any $j \in \mathcal{M}$. Since all p_j's follow the uniform distribution marginally under the null, each q_j is also uniformly distributed under the null. By using such q_j's for $j \in \mathcal{M}$, define the proposed test statistic as follows.

$$T_q^*(q) = \sum_{j \in \mathcal{M}} w_j h_q(q_j) = \sum_{j \in \mathcal{M}} w_j \tan\{(1/2 - q_j)\pi\}, \tag{3.2}$$

where the weigths $w_j \geq 0$, $j = 1, \ldots, d$. By the duality relation between data X and *p*-values p_j's, we can define the statistic with respect to the data X which is equivalent to T_q^*.

$$T_X^*(X) = \sum_{j \in \mathcal{M}} w_j h_X(X_j) = \sum_{j \in \mathcal{M}} w_j \tan\{(2\Phi(|X_j|) - 3/2)\pi\}, \tag{3.3}$$

where the weights $w_j \geq 0$, $j = 1, \ldots, d$.

3.1 Null Distribution

Obtaining the null distribution of test statistics plays an important role in controlling the type I error. That the null distribution of a test statistic converges to a known distribution at a fast rate is one of the necessary conditions of the test statistic. In general, testing procedures that are constructed when there is an arbitrary dependency structure assume that the dependency structure is known in advance, or can be estimated from the data to decorrelate the test statistic (Hall and Jin, 2010). However, estimating the dependency structure is not easily implemented, especially in high dimensional settings, and sometimes estimating the arbitrary dependency structure requires many constraints. In addition, there are cases where the null distribution constructed under a dependency structure is different from that of the independent situation, or the null distribution does not converge to a known distribution. For example, the null distribution of the minimum p-value (Tippett, 1931) under an arbitrary dependency structure is hard to obtain without knowledge of the dependency structure, although the null distribution under independence is the beta distribution. Generally, the null distribution depending on correlations between variates requires information about the dependency structure, and it is not easy to use in that it is unknown in advance.

On the other hand, the convergence rate of the null distribution constructed by using a threshold method is known to be very slow. The null distribution of a thresholding method often converges to a normal distribution rather than other known distribution (Fan, 1996, Kim and Akritas, 2010). This convergence rate is very slow, making it difficult to the control type I error in practice. In addition, it is often different from the original un-thresholded null distribution, which also requires additional information about the dependency structure.

3.1.1 Approximation of Tail Probability for The Null Distribution in Finite Dimension

When we use the threshold method in constructing the Cauchy combination test statistic, the convergence rate of the null distribution is still fast, especially non-asymptotic. Moreover, the null distribution of the thresholded combined test statistic still has the similar behavior of the Cauchy distribution. In other words, the null distribution of the proposed test statistic can be obtained with robustness to the dependency structure.

The following theorem presents the tail behavior of the proposed test statistic under the null, in the finite dimensional setting. Specifically, theorem 3 shows that under the bivariate normality condition defined in the definition 1 in Chapter 2.1.2, the tail probability of the null distribution of T_q with arbitrary dependency structures is equivalent to the tail probability of a standard Cauchy distribution non-asymptotically.

Theorem 3. *Let $X = (X_1, \ldots, X_d)^T \sim N(0, \Sigma)$. Suppose that X is bivariate normally distributed, i.e., for any $1 \le i \ne j \le d$, $(X_i, X_j)^T \sim N(0, \Sigma_{ij})$ where diagonal elements of Σ_{ij} are 1 and off-diagonals are σ_{ij}. Let q_j be the transformed p-value defined in (3.1) and $T^*_q(q)$ be the test statistc defined in (3.2). Denote $C_{(a,b)}$ as a random variable following the Cauchy distribution with the location paramater a and the scale parameter b. Then we have*

$$\lim_{t \to \infty} \frac{P(T^*_q(q) > t)}{P(C_{(0, w^*\delta)} > t)} = 1,$$

where $w^* = \sum_{j=1}^{d} w_j$.

Proof. See Appendix. □

Theorem 3 implies that the tail probability of the null distribution of the proposed method is robust to the impact of the dependency structures in *p*-values or marginal test statistics. As each Cauchy transformation of *p*-values are weighted averaged, the effect of dependencies between *p*-values on the convergence of the null distribution is reduced. In fact, the tail probability of a Cauchy distribution obtained asymptotically to be proportional to the reciprocal of t, as in lemma 2, appears to be a polynomial order inducing a very heavy form of the tail probability. This heavy tail probability causes the impact of dependency structures to converge to zero asymptotically in the null distribution.

3.1.2 Approximation of Tail Probability for The Null Distribution in Infinite Dimension

When the dimension d grows infinitely, the tail probability of the null distribution of the proposed test statistic is asymptotically equivalent to that of a Cauchy distribution. Before we state the theorem, first consider the following assumptions on Σ.

(A.1) $\lambda_{\max}(\Sigma) \leq C_0$ for a constant C_0 where $\lambda_{\max}(\Sigma)$ is the largest eigenvalue of Σ.

(A.2) Let $\Sigma = (\sigma_{ij})_{i,j=1}^{d}$ and $\max_{1 \leq i < j \leq d} \sigma_{ij}^2 \leq \sigma_{\max}^2 < 1$ for some constant $\sigma_{\max}^2 \in (0, 1)$.

The first assumption (A.1) implies that the largest eigenvalue of Σ is bounded by a constant. The second assumption (A.2) guarantees that Σ is a well-defined correlation matrix. If $\sigma_{s,t} = 1$ for $1 \leq s \neq t \leq d$, the correlation matrix becomes ill-conditioned. Condition (A.1) and (A.2) are commomly used assumptions in the high-dimensional setting. Now, we state the theorem that show the tail probabilities of the proposed method and a Cauchy distribution are asymptoticall equivalent, as $d \to \infty$.

Theorem 4. *Let $X = (X_1, \ldots, X_d)^T \sim N(0, \Sigma)$. Suppose that X is bivariate normally distributed, i.e., for any $1 \leq i \neq j \leq d$, $(X_i, X_j)^T \sim N(0, \Sigma_{ij})$ where diagonal elements of Σ_{ij} are 1 and off-diagonals are σ_{ij}. Let q_j be the transformed p-value defined in (3.1) and $T_q^*(q)$ be the test statistc defined in (3.2). Denote $C_{(a,b)}$ as a random variable following the Cauchy distribution with the location paramater a and the scale parameter b. For $d = o(t^\gamma)$ where $0 < \gamma < 1/2$, under the conditions (A.1) and (A.2), we have,*

$$\lim_{d \to \infty} \frac{P(T_q^*(q) > t)}{P(C_{(0,w^*\delta)} > t)} = 1,$$

where $w^* = \sum_{j=1}^{d} w_j$.

Proof. See Appendix. $\square$

Remark 1. *We defined δ_j^* such that $\delta < \delta_j^* < 1$ in the proof. By using the symmetry property of the sine function,* $\sin(x) = \sin(\pi - x)$*, we can define*

$$\delta_j^{**} = \frac{1}{q_j} - \delta_j^*,$$

so that $\sin(\delta_j^* q_j \pi) = \sin(\pi - \delta_j^{**}) = \sin(\delta_j^{**} q_j \pi)$ *and* $0 < \delta_j^{**} < \delta$*. In this re-defining, we obtain*

$$\sin(\delta_j^{**} q_j \pi) = \delta_j^{**} q_j \pi + O\left((\delta_j^{**} q_j)^2\right).$$

As the dimenstion d increases, $\delta \to 0$. Then

$$\frac{\sin(\delta_j^{**} q_j \pi)}{\delta_j^{**} q_j \pi} \to 1.$$

Therefore we have

$$\lim_{t\to\infty} \frac{P(T^*(X) > t)}{P(C_{(0,w^*)} > t)} = 1.$$

Sketch of proof for theorem 3 and 4.

Step 1. The random set $\mathcal{M}$ can be expressed as random weights $\tilde{w}_j$'s :

$$T_q^* = \sum_{j\in\mathcal{M}} w_j h_q(q_j) = \sum_{j\in\{j:p_j\le\delta\}} w_j h_q(q_j) = \sum_{j=1}^{d} w_j I(q_j \le 1) h_q(q_j) =: \sum_{j=1}^{d} \tilde{w}_j h_q(q_j)$$

Step 2. Decompose the joint distribution with arbitrary dependence structure to marginal case by using Bonferroni inequality :

$$\begin{aligned} P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t\right) &= P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \bigcup_{j=1}^{d} \{\tilde{w}_j h_q(q_j) > t\}\right) \\ &\quad + P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \bigcap_{j=1}^{d} \{\tilde{w}_j h_q(q_j) \le t\}\right) \end{aligned}$$

$$
\begin{aligned}
&= P\left(\bigcup_{j=1}^{d}\left\{\tilde{w}_j h_q(q_j) > t, \sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t\right\}\right) \\
&\quad + P\left(\bigcap_{j=1}^{d}\left\{\tilde{w}_j h_q(q_j) \le t, \sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t\right\}\right) \\
&=: P\left(\bigcup_{j=1}^{d} A_j\right) + P\left(\bigcap_{j=1}^{d} B_j\right)
\end{aligned}
$$

$$
\sum_{j=1}^{d} P\left(A_j\right) - \sum\sum_{j\neq i} P\left(A_j \cap A_i\right) \le P\left(A_j\right) \le \sum_{j=1}^{d} P\left(\bigcup_{j=1}^{d} A_j\right)
$$

Step 3. Calculate order each decomposition parts

□

3.1.3 Approximation of Tail Probability for The Null Distribution with Random Weights

Note that if we set $w_j^{-1} = d\delta$ for all $j = 1, \ldots, d$, the tail probability of the proposed test statistic is still equivalent to that of the standard Cauchy distribution asymptotically. However, since $\mathcal{M}$ is a random set, the number of summands in the test statistic is a random number as well, although the weights are fixed number $1/d\delta$. So, practically a correction is needed to obtain the appropriate tail probability of the null distribution and improve the power.

To handle the correction, we can set the weight w_i as a random variable. Indeed, instead of $w_j^{-1} = d\delta$, we define $w_j^{-1} = |\mathcal{M}|$ to reflect the number of *p*-values contained in $\mathcal{M}$. When there are many observed *p*-values in $\mathcal{M}$, we assign more weights to the thresholded *p*-values. To this end, consider specific types of dependencies between *p*-values which are contained in $\mathcal{M}$, that is, $p_j \le \delta$.

Definition 2 (Type of Dependencies)**.**

1. *Heyde (2004)*

$$P(p_k \leq \delta \mid \mathcal{F}_k) = (1-\theta)\delta + \theta S_k/k,$$

where θ is a depdendence parameter with $0 \leq \theta < 1$ and $\mathcal{F}_k = \sigma(p_1, \ldots, p_k)$ is the σ-field generated by $\{X_1, \ldots, X_k\}$.

2. *Gava and Rezende (2021)*

$$P(p_k \leq \delta \mid \mathcal{F}_k) = \delta + d_k S_k/k,$$

where $S_k = I(p_1 \leq \delta) + \cdots + I(p_{k-1} \leq \delta)$ and $\mathcal{F}_k = \sigma(p_1, \ldots, p_k)$ is the filtration, $0 \leq \delta + d_k < 1$ for $k \geq 1$.

3. *Exponentially decaying dependency*

$$\mathsf{Cov}(I(p_i \leq \delta), I(p_j \leq \delta)) \leq c^{|i-j|},$$

for $i, j \leq d$ and $|c| < 1$.

Consider the case where there is a special dependency structure defined in the definition 2 between the *p*-values included in $\mathcal{M}$. Then, random weights can be used as shown in corollary 1 below.

Corollary 1. *If p-values are weak dependent to each other or there is a specific correlation structure in p-values as in definition 2, we can set $w_j^{-1} = \#\{j : p_j \leq \delta\}$ to obatin the result in Theorem 4.*

Proof. Under correlation structures in definition 2, it follows that

$$\lim_{d \to \infty} \frac{\#\{j : p_j \leq \delta\}}{d\delta} = 1.$$

Hence we have

$$T_q^*(q) = \sum_{j\in\mathcal{M}} \frac{1}{d\delta}\,(1+o(1))\,h_q(q_j).$$

It follows from Theorem 4 that

$$\lim_{d\to\infty} \frac{P(T_q^*(q)>t)}{P(C_{(0,1)}>t)} = 1.$$

$\square$

As corollary 1, for large d with a specific dependence structure, $\#\{i : p_i \le \delta\} = w_j^{-1}$ for any $j = 1,\ldots,d$. On the other hand, for the finite dimension d, the law of large number in corollary 1 contains an approximation error. Practically, we are given only observed *p*-values and $\mathcal{M}$ is also given. To reduce the approximation error, we use instead a finiteness correction such that,

$$T_q^{**}(q) = \sum_{j\in\mathcal{M}} \frac{1}{\#\{j : p_j \le \delta\}} h_q(q_j).$$

Then the null tail probability of T_q^{**} can be adjusted. If $\#\{j : p_j \le \delta\}$ is given, then from the proof of Theorem 4, for large t,

$$\begin{aligned}
P(T_q^{**}(q)>t) &= E\left[P(T_q^{**}(q)>t \mid \#\{j : p_j \le \delta\})\right] \\
&= E\left[P\left(\sum_{j\in\mathcal{M}} \frac{1}{\#\{j : p_j \le \delta\}} h_q(q_j) > t \;\Big|\; \#\{j : p_j \le \delta\}\right)\right] \\
&= E\left[P\left(\sum_{j\in\mathcal{M}} \frac{1}{d\delta} h_q(q_j) > \frac{\#\{j : p_j \le \delta\}\cdot t}{d\delta} \;\Big|\; \#\{j : p_j \le \delta\}\right)\right] \\
&= E\left[\frac{d\delta}{\#\{j : p_j \le \delta\}}\cdot\frac{1}{t\pi} + o(1/t)\right].
\end{aligned}$$

If we let

$$w^{**} = \frac{d\delta}{\#\{j : p_j \le \delta\}},$$

we have that

$$\lim_{t\to\infty} \frac{P(T_q^{**}(q)>t)}{P(C_{(0,w^{**})}>t)} = 1.$$

3.2 Power Analysis of Cauchy Combination Test with Thresholding

In this section, we analyze the power of the proposed test statistic. Recall the null and alternative hypotheses.

$$H_0 : X \sim N_d(0, \Sigma) \quad \text{vs.} \quad H_1 : X \sim N_d(\boldsymbol{\mu}, \Sigma). \tag{3.4}$$

We consider a sparse mean vector as an alternative. Let $\mathcal{S} = \{j \in \{1, \ldots, d\} : \mu_j \neq 0\}$ be an index set of nonzero mean elements and $|\mathcal{S}|$ be the cardinality of $\mathcal{S}$, that is, the number of signals. Assume that $|\mathcal{S}| = d^\gamma$ where $0 < \gamma < 1$. The sparsity of the mean vector $\boldsymbol{\mu}$ is defined by the parameter γ. If $0 < \gamma < 1/2$, we call the mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T$ a sparse mean vector. To test the sparse alternative, suppose that $X \sim N_d(\boldsymbol{\mu}, \Sigma)$. Then we can consider the following hypotheses as in Hall and Jin (2010), instead of (3.4).

$$H_0 : \boldsymbol{\mu} = 0 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu} \text{ is a sparse vector .} \tag{3.5}$$

Assume that signals have the same magnitude and let $\mu_i = \sqrt{2r \log d}$ for $i \in \mathcal{S}$ where r is a constant. The form of the signal magnitude is decided to consider the optimal detection boundary in Donoho and Jin (2004). The following theorem shows that the power of the proposed adaptive thresholding Cauchy combination test statistic to test hypothesis (3.5) converges to 1.

Theorem 5. *Denote $\mathcal{S} = \{j \in \{1, \ldots, d\} : \mu_j \neq 0\}$ as the set of signals. Suppose that the number of signals $|\mathcal{S}| = d^\gamma$, where $0 < \gamma < 1/2$. Assume that $\mu_j = \sqrt{2r \log d}$ for all $j \in \mathcal{S}$, where $r > (1 - \sqrt{\gamma})^2$. Consider the Cauchy combination test statistic by thresholding at value $\delta \in (0, 1)$. Then, for α-quantile t_α of $\mathrm{Cauchy}(0, w^*\delta)$ where w^* is the sum of all* *weights, as* $d \to \infty$*,*

$$P_{H_1}\left(T_q^*(q) \geq t_\alpha\right) \to 1,$$

where P_{H_1} *is a probability measure under the alternative.*

Theorem 5 shows that the power of the proposed method converges to 1 as the dimension d increases. This means that the summation of the type I error and the type II error converges to 0 equivalently. The condition $0 < \gamma < 1/2$ in theorem 5 implies the sparsity of the mean vector. Hence, for the sparse signal, by using the proposed method, H_0 and H_1 of (3.5) can be separated asymptotically. We can also show that, with the result in theorem 5, the proposed method attains the optimal detection boundary defined in Donoho and Jin (2004). Note that the detection boundary is defined as follows.

$$\rho^*(\gamma) = \begin{cases} (1-\sqrt{\gamma})^2, & 0 < \gamma < \frac{1}{4}, \\ \frac{1}{2} - \gamma, & \frac{1}{4} < \gamma \leq \frac{1}{2} \end{cases},$$

where r and γ are given in theorem 5. Since we have the condition $r > (1-\sqrt{\gamma})^2$, the proposed method achieves the optimal detection boundary when $0 < \gamma < 1/4$, which is the strong sparsity situation. Note that

$$(1-\sqrt{\gamma})^2 > \frac{1}{2} - \gamma \quad \Leftrightarrow \quad 2\left(\sqrt{\gamma} - \frac{1}{2}\right)^2 > 0.$$

For $1/4 < \gamma < 1/2$, we have $r > (1-\sqrt{\gamma})^2 > 1/2 - \gamma$ so that the proposed method also attains the optimal detection boundary. Therefore the proposed method achieves the optimal detection boundary for the sparse signal, which is represented by $0 < \gamma < 1/2$.

Sketch of proof for theorem 5.

Step 1. Decompose T_q^* into two parts, a null part and a signal part :

$$T_q^* \;=\; \sum_{j \in \mathcal{M}} w_j h_q(q_j) \;=\; \sum_{j \in \mathcal{M} \cap N} w_j h_q(q_j) + \sum_{j \in \mathcal{M} \cap S} w_j h_q(q_j)$$

Step 2. Claim that the null part is lower bounded by a finite constant with probability tending to 1:

$$\sum_{j \in \mathcal{M} \cap N} w_j h_q(q_j) > -\infty$$

Step 3. Claim that the maximum of signal part is transformed by using the mean value theorem (lemma 5) and is lower bounded by an increasing order:

$$\sum_{j \in \mathcal{M} \cap S} w_j h_q(q_j) = \sum_{j \in \mathcal{M} \cap S} w_j^* h_X(X_j) \geq w_{i^*}^* h_X(X_{i^*}) + (|\mathcal{M} \cap S| - 1) \cdot w_{k^*}^* h_X(X_{k^*}) \to \infty,$$

where $i^* = \arg\max_{j \in \mathcal{M} \cap S} h_X(X_j)$ and $k^* = \arg\min_{j \in \mathcal{M} \cap S} h_X(X_j)$.

□

3.3 Choosing the thresholding value δ

In this section, we propose a plug-in estimator $\hat{\delta}$ for the thresholding value δ. First, decompose the test statistic, $T_q^*(q)$, into the null part $\mathcal{N}$ and the signal part $\mathcal{S}$ as follows.

$$T_q^*(q) = \sum_{j \in \mathcal{M}} w_j h_q(q_j) = \sum_{j \in \mathcal{M} \cap N} w_j h_q(q_j) + \sum_{j \in \mathcal{M} \cap S} w_j h_q(q_j)$$

Suppose that the weights are equal that is, for all $i = 1, \ldots, d$, $w_i^{-1} = d\delta$ for convenience. From the results and proofs of theorem **??** and **??**, the tail behavior of the null part statistic,

$$\sum_{j \in \mathcal{M} \cap N} w_j h_q(q_j)$$

is asymptotically equivalent to that of a Cauchy distribution, regardless of δ. Indeed, we have

$$\lim_{t \to \infty} \frac{P(\sum_{j \in \mathcal{M} \cap N} (1/d\delta) h_q(q_j) > t)}{P(C_{(0, |\mathcal{N}|/d)} > t)} = 1.$$

On the other hand, we can show that the signal part statistic is an increasing function with respect to δ.

$$\sum_{j\in\mathcal{M}\cap S} w_j h_q(q_j) = \sum_{j\in\mathcal{M}\cap S} \frac{1}{d\delta} \tan\left\{\left(\frac{1}{2} - \frac{p_j}{\delta}\right)\pi\right\} = \sum_{j\in\mathcal{M}\cap S} \frac{1}{d\delta} \frac{\cos\left(\frac{p_j\pi}{\delta}\right)}{\sin\left(\frac{p_j\pi}{\delta}\right)}. \tag{3.6}$$

By lemma 3, note that the right hand side of (3.6) increases as δ increases.

Lemma 3. *The signal part of t_q^* is an increasing function of δ.*

Proof. Define

$$f(\delta) := \frac{1}{\delta} h\left(\frac{p}{\delta}\right),$$

where

$$h\left(\frac{p}{\delta}\right) = \tan\left(\left(\frac{1}{2} - \frac{p}{\delta}\right)\pi\right) = \frac{\sin\left(\frac{\pi}{2} - \frac{p_j\pi}{\delta}\right)}{\cos\left(\frac{\pi}{2} - \frac{p_j\pi}{\delta}\right)} = \frac{\cos\left(\frac{p_j\pi}{\delta}\right)}{\sin\left(\frac{p_j\pi}{\delta}\right)}.$$

Then, we have

$$\begin{aligned} f'(\delta) &= \frac{\frac{p\pi}{\delta^2}\sin^2\left(\frac{p\pi}{\delta}\right)\delta - \left[\sin\left(\frac{p\pi}{\delta}\right) - \frac{p\pi}{\delta}\cos\left(\frac{p\pi}{\delta}\right)\right]\cos\left(\frac{p\pi}{\delta}\right)}{(\delta\sin\left(\frac{p\pi}{\delta}\right))^2} \\ &= \frac{\frac{p\pi}{\delta} - \sin\left(\frac{p\pi}{\delta}\right)\cos\left(\frac{p\pi}{\delta}\right)}{(\delta\sin\left(\frac{p\pi}{\delta}\right))^2}. \end{aligned}$$

Since $x > \sin(x)\cos(x)$ for $0 < x < 1$, we can show that $f'(\delta) > 0$ for $0 < \delta < 1$. □

Therefore, as δ increases, $\mathcal{M}$ contains more *p*-values corresponding to the signals and the signal statistic increases. At the same time, if

$$\frac{1}{2} > \frac{p_j}{\delta}$$

in the middle equation of (3.6) is greater than 0, we have

$$\tan\left\{\left(\frac{1}{2}-\frac{p_j}{\delta}\right)\pi\right\} > 0,$$

which can make the signal part of the statistic larger. In other words, it can be concluded that setting the $\delta = 2|\mathcal{S}|/d$ is a trivial way to increase the power. We choose δ such that $\mathcal{M} = \{j : p_j \leq \delta\}$ contains as many as signal *p*-values with $p_j/\delta \leq 1/2$ or equivalently $p_j \leq \delta/2$. Then we use a plug-in estimator $\hat{\delta} = 2\hat{\pi}_1$, where π_1 is the proportion of signals and $\hat{\pi}_1$ is an estimator. We propose a method to estimate the proportion of signals, π_1, in the next chapter.

Chapter 4

Estimating The Proportion of Non-null Hypotheses

In Chapter **??**, we proposed the estimator of the proportion of signals as an critical value in thresholding p-values. In this chapter, we propose an estimator of the proportion of signals when arbitrarily dependent p-values are given. In terms of the hypotheses testing procedure, the proportion of signals is consistent with the proportion of non-null hypotheses so that we can approach the problem in the form of a mixture from the perspective of estimating the proportion of the non-null or alternative hypotheses.

When p-values obtained from the null H_0 and alternative H_1 hypotheses are distinguished into the null p-values and alternative p-values, respectively, the distribution of p-values in the following mixture of p-values can be considered.

$$f(p) = \pi_0 f_0(p) + \pi_1 f_1(p), \tag{4.1}$$

where f is a marginal density function of observed p-values and f_0 and f_1 are density functions of null and alternative p-values, respectively. π_0 and π_1 are the mixing coefficients that is, the proportions of the null and alternative with $\pi_0 + \pi_1 = 1$. Estimating the proportion of signals is then equivalent to estimating the mixing coefficient, π_1, in (4.1).

To estimate the proportion of signals, there are conditions usually assumed. First, it is assumed that the *p*-values are independent. If the *p*-values are independent, then the *p*-values obtained from the null follow an uniform distribution identically. In the notations in (4.1), we can assume that under the null,

$$f_0 \sim \mathcal{U}(0, 1).$$

As in Efron (2010), we also assume that the zero-assumption that any *p*-value greater than some λ is assumed to be the null *p*-value, that is, for large $\lambda \in (0, 1)$,

$$p|\{p > \lambda\} \sim f_0.$$

The alternative density f_1 is assumed to be a decreasing function, so that if *p*-values are close to 1, then the value of f_1 is close to 0 which means that $f_1(p) \approx 0$ for $p \approx 1$.

In the following section, we briefly review existing methods estimating π_1 or π_0. Thereafter, we propose a method to estimating π_1 for the case that an arbitrary dependency structure between *p*-values exists.

4.1 Literature Reviews

In this section, we review methods estimating the proportion of signals, π_1, or that of null, $\pi_0 = 1 - \pi_1$. First we review methods that estimate π_1 under the independence assumption. And then we review methods estimating π_1 under the arbitrary dependency structures in Chapter 4.1.2.

4.1.1 Methods That Estimating The Proportion of Signals for Independent p-values

Suppose that p-values are independent. Estimating π_1 is equivalent to estimating the mixing coefficient of a mixture model of p-values as mentioned in the previous section.

$$f(p) = \pi_0 f_0(p) + \pi_1 f_1(p), \tag{4.2}$$

Since $\pi_1 = 1 - \pi_0$, estimating π_0 is equivalent to estimating π_1. The observed p-values are independent and identically distributed and we have

$$p_1, \ldots, p_d \overset{i.i.d.}{\sim} f.$$

The independence between p-values implies that the null density of p-values, f_0 in (4.2), is the density function of the uniform distribution so that

$$f_0(p) = 1, \quad \text{for any } p \in (0, 1)$$

Then (4.2) can be expressed as follows.

$$f(p) = \pi_0 + (1 - \pi_0) f_1(p). \tag{4.3}$$

Storey (2002) assumed that the density function f_1 of the alternative p-values is a decreasing function with $f_1(1) = 0$. He further assumed the zero-assumption, that is, if p-values are close to 1, then the value of f_1 is close to 0 which means that $f_1(p) \approx 0$ for $p \approx 1$. Integrating the both sides of (4.3) with respect to p from λ to 1 implies that

$$\int_{\lambda}^{1} f(p) dp \approx \pi_0 (1 - \lambda),$$

where $\lambda \in (0, 1)$ is a given constant which is close to 1. Storey (2002) proposed an estimator of π_0 as follows.

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_j > \lambda\}}{d(1 - \lambda)},$$

where λ can be chosen to minimize the mean square error, MSE, by using a bootstrap method.

Let $p_{(1)} \leq \cdots \leq p_{(d)}$ be the ordered observed *p*-values from the d hypothesis tests. Langaas et al. (2005) proposed an estimator of π_0 based on the ordered *p*-values.

$$\hat{\pi}_0^p = \min_{l \leq d-1} \left\{ \frac{1 - l/d}{1 - p_{(l)}} \right\}.$$

With an additional stronger assumption that f_1 is a decreasing and convex function, Langaas et al. (2005) used an iterative method based on the steepest descent algorithm. In (4.3), the assumption that f_1 is a decreasing and convex function implies that f is also a decreasing and convex function. From the fact that any twice continuously differentiable convex decreasing density f on $[0, 1]$ can be represented as a mixture

$$f(p) = \int_0^1 f_\theta(p)\mu(d\theta),$$

where f_θ is a triangular density. Let $\hat{f}$ be the nonparametric maximum likelihood estimate of f. Langaas et al. (2005) proposed a method estimating π_0 by

$$\hat{\pi}_0 = \hat{f}(1).$$

Genovese and Wasserman (2004) converted the mixture setting into a stochastic process setting:

$$F(t) = \pi_0 F_0(t) + \pi_1 F_1(t), \tag{4.4}$$

where F_0 is an uniform distribution, that is, $F_0(t) = t$. To estimate π_1, Genovese and Wasserman (2004) assumed that F is concave and hence $f = F'$ is a decreasing function. For a finite sample confidence envelope $[\gamma^-(\cdot), \gamma^+(\cdot)]$ for a density f, Genovese and Wasserman (2004) proposed an estimator of π_1 by

$$\hat{\pi}_1 = 1 - \min\{h(1) : \gamma^- \leq h \leq \gamma^+\},$$

and proved that $\hat{\pi}_1$ converges to π_1.

Meinshausen and Rice (2006) considered a lower confidence bound of π_1. The goal is, then, to construct a lower bound of

$$\hat{\pi}_1 = \frac{1}{d}\sum_{i=1}^{d} \mathbf{1}\{\mu_i \neq 0\}$$

with the property of

$$P(\hat{\pi}_1 \leq \pi_1) \geq 1-\alpha, \tag{4.5}$$

for a specified confidence level $1-\alpha$. From the stochastic process (4.4), Meinshausen and Rice (2006) defined an estimate of π_1 as follows.

$$\hat{\pi}_1 = \sup_{t\in(0,1)} \frac{F_n(t) - t - \beta_{n,\alpha}\delta(t)}{1-t},$$

where $\beta_{n,\alpha}$ is a bounding sequence for $\delta(t)$ at level α, $\delta(t)$ is a bounding function and F_n is the empirical distribution of *p*-values. Then Meinshausen and Rice (2006) proved the (4.5).

In addition to using the mixture model of *p*-values, methods using observed *p*-values also had been proposed. Under the assumption that *p*-values are independent so that null *p*-values are uniformly distributed, Schweder and Spjøtvoll (1982) proposed a method to estimate π_0 by using the approximation such that

$$\mathsf{E}\left[\sum_{i=1}^{d} I(p_i > t)\right] \approx \pi_0(1-t).$$

By plotting the number of *p*-values greater than t against $1-t$, Schweder and Spjøtvoll (1982) estimated π_0 based on *p*-values deviating from a straight line.

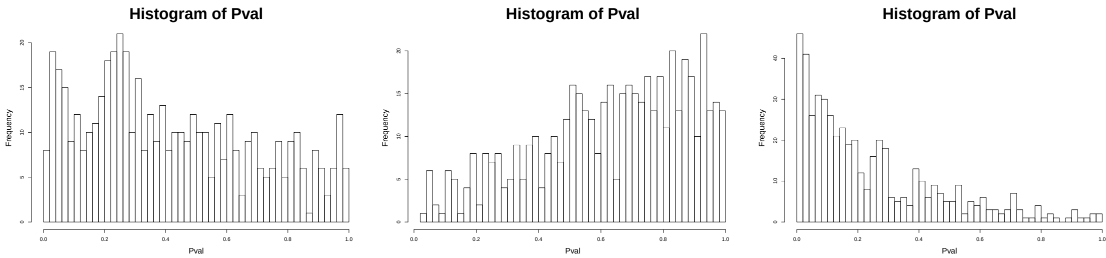

Figure 1: Historgrams for p-values obtained for $\Sigma = (0.5)_{i=1}^{500}$ with diagonal elements 1.

4.1.2 Methods That Estimating The Proportion of Signals for Arbitrarily Dependent p-values

If p-values are arbitrarily dependent, then the assumptions that observed p-values follow the mixture model (4.1) and null p-values follow independently uniform distribution can not be used to estimate the proportion of signals. Indeed, observed p-values are not uniformly distributed but increasingly or decreasingly distributed. Even worse there are no visible patterns for p-values. Figure 1 shows that histograms of null p-values obtained from

$$X \sim N_d(0, \Sigma),$$

where $d = 500$ and

$$\Sigma = (\sigma_{ij})_{i,j=1}^{d} = \begin{cases} 1 & i = j \\ 0.5 & i \neq j. \end{cases}$$

Recall the hypotheses testing procedure.

$$X_{H_0} \sim N(0, \Sigma), \quad X_{H_1} \sim N(\theta, \Sigma)$$

Meinshausen and Bühlmann (2005) estimated π_1 by

$$\hat{\pi}_1 = \sup_{\gamma \in [0,1]} \{R(\gamma) - G_\alpha(\gamma)\},$$

where $R(\gamma)$ is the number of p-values which are less than $\gamma \in (0, 1)$ and G_α is a bounding function and the quantile function of the number of false rejections under permutation.

Tong, Feng, Hilton and Zhao (2013) used a linear fitting under the assumption that the pattern of true null p-values is linear or unimodal. For example, only p-values in the middle and right plots in Figure 1 are observed. Let $W_0(\lambda)$ be the total number of the true null p-values in $(\lambda, 1]$ and

$$f(\lambda) = \frac{W_0(\lambda)}{d\pi_0}$$

be the proportion of the true null p-values in $(\lambda, 1]$. From the idea of Storey (2002), that is, for $W(\lambda) = \sum_{j=1}^{d} I(p_j > \lambda)$,

$$\hat{\pi}_0(\lambda) = \frac{W(\lambda)}{d(1-\lambda)},$$

we have $W_0(\lambda) \approx W(\lambda)$ and

$$\pi_0 = \frac{W_0(\lambda)}{df(\lambda)} \approx \frac{W(\lambda)}{df(\lambda)}.$$

Tong, Feng, Hilton and Zhao (2013) chose λ by using a bootstrap algorithm to minimize the mean square error.

However the linear fitting of f denoting the proportion of true null p-values in $(\lambda, 1]$ too simplify the patterns of observed p-values. As shown in numerical studies in chapter 5, the estimate $\hat{pi}_1$ has greater variations.

4.2 Proposed Method

As Storey (2002), from the independence assumption point of view, the mixture of p-values considers

$$f(p) = \pi_0 + (1 - \pi_0) f_1(p),$$

where $f_1(p) \approx 0$ if p is close to 1. In contrast to the independent case, we adapt the null density of p-values, f_0 to accommodate patterns of null p-values when dependency between p-values exits. Hence we consider instead that

$$f(p) = \pi_0 f_0(p) + (1 - \pi_0) f_1(p).$$

Tong et al. (2013) pointed out that althoguh regardless of the dependence of p-values, the marginal distribution of the null p-values remains a standard uniform distribution, variations of the expected frequency of observed p-values increase as correlations of p-values are increase. Hence they concluded that if we handle the pattern of the observed p-values, we can improve the estimation of the proportion of true nulls.

From this motivation, we propose an estimation of proportion of null hypotheses, based on Storey (2002) and Tong et al. (2013), such that

$$\hat{\pi}_0(\lambda) = \frac{W(\lambda)}{d(1 - \hat{F}_0(\lambda))}, \tag{4.6}$$

where $F_0(\lambda) = \int_0^\lambda f_0(x)dx$ and $f_0(x)$ is a density of null p-values. Then $\hat{F}_0$ is a estimated F_0 by using the kernel density estimation. Indeed, we have

$$\hat{\pi}_0(\lambda) = \frac{\int_\lambda^1 \hat{f}(p)dp}{\int_\lambda^1 \hat{f}_0(p)dp} = \frac{1 - \hat{F}(\lambda)}{1 - \hat{F}_0(\lambda)} = \frac{\sum_{j=1}^d I(p_j \leq \lambda)}{1 - \hat{F}_0(\lambda)}, \tag{4.7}$$

where $\hat{F}_0$ is estimated by the using kernel density estimation. We call it "dependence-corrected Storey method". Storey (2002) proposed the bootstrap method that optimal λ is chosen as follows.

1. Set $\mathcal{R} = \{0, 0.05, \ldots, 0.95\}$ to be the grid of λ.

2. For each $\lambda \in \mathcal{R}$, generate the bootstrap copies $\hat{\pi}_0^b$, $b = 1, \ldots, B$ for predetermined B from (4.6).

3. Estimate MSE(λ) by

$$\widehat{\mathrm{MSE}}(\lambda) = \sum_{b=1}^{B} \frac{(\hat{\pi}_0^b(\lambda) - \hat{\pi}_0^p)^2}{B},$$

   where $\hat{\pi}_0^p = \min_{\lambda \in \mathcal{R}} \hat{\pi}_0$.

4. Define $\hat{\lambda} = \arg\min_{\lambda \in \mathcal{R}} \widehat{\mathrm{MSE}}(\lambda)$ and estimate $\hat{\pi}_0 = \hat{\pi}_0(\hat{\lambda})$.

In the next section, extensive simulation studies support the accuracy of the proposed method.

Chapter 5

Numerical Studies

In this chapter, we evaluate the performances of the proposed method comparing other competitive methods. To investigate the robustness to dependency structures, we construct *p*-values from various dependency structures including the compound symmetry, exponentially decaying and polynomially decaying dependency structures. Specifically, data or the normalized test statistic X are constructed from

$$H_0 : X \sim N_d(0, \Sigma) \quad \text{vs.} \quad H_1 : X \sim N_d(\boldsymbol{\mu}, \Sigma).$$

The dimension d has a different range from 20 to 500 through simulations in this chapter. All the signals for an alternative have the same strength $\mu_i := \mu_0 = \sqrt{3\log(d)}/s^{1/5}$ for $i = 1, \ldots, d$, where s is the number of signals. The magnitude of signals is defined to consider the settings of the sparse signal. The dependency structure or the correlation matrix $\Sigma = (\sigma_{ij})_{i,j=1,\ldots,d}$ is defined as follows.

1. **Compound Symmetry**: For $0 < \rho < 1$,

$$\sigma_{ij} = \begin{cases} \rho & i \neq j \\ 1 & i = j \end{cases}$$

2. **Polynomial decay**: For $\rho > 0$,

$$\sigma_{ij} = \begin{cases} \frac{1}{0.7+|i-j|^{\rho}} & i \neq j \\ 1 & i = j \end{cases}$$

3. **Exponential decay, AR(1)**: For $0 < \rho < 1$,

$$\sigma_{ij} = \rho^{|i-j|}.$$

For each dependency structure, values of the correlation coefficient ρ decide the strength of the dependency. Thus we vary the correlation coefficient ρ to investigate effects of the dependencies through following simulations.

In Chapter 5.1, we compare the proposed method to other methods of estimating the proportion of signals. Methods of Storey (2002), Langaas et al. (2005) and Tong et al. (2013) are used for the comparison. In Chapter 5.2, we evaluate the type I error of the proposed method. For convenience, we use the equal weights, that is, $w_i^{-1} = d\delta$ for $i = 1, \ldots, d$. Since the tail probability of the proposed method converges to that of a standard Cauchy distribution, we use a level α-quantile of the standard Cauchy distribution. In Chapter 5.3, under different dependency structures, we investigate the power of the proposed method compared to other global testing procedures, the Higher criticism (Donoho and Jin, 2004), the Berk-Jones (Berk and Jones, 1979), the minimum p-value (Tippett, 1931) as well as the Cauchy combination test (Liu and Xie, 2020).

5.1 Simulation Studies of Estimating the proportion of non-null hypotheses

In this section, we investigate the accuarcy of the proposed method for estimating the proportion of signals which is analyzed in Chapter 4. To constuct the test statistic, we define p-values by

$$p_i = 2(1 - \Phi(|X_i|)), \quad i = 1, \ldots, d$$

where X_i is i-th element of $X \sim N_d(\boldsymbol{\mu}, \Sigma)$. The dimension considered in this simulation is $d = 100$. Suppose that all the signals have the same strength $\mu_0 = \sqrt{3\log(d)}/s^{1/5}$, where s is the number of signals and $1 \leq s \leq d$. Thus we denote $\pi_1 = s/d$. To examine the dependency structures, we impose the compound symmetry structure on $\Sigma = (\sigma_{ij})_{i,j=1,\ldots,d}$ with different values of $\sigma_{ij} = 0, 0.1, \ldots, 0.5$ for all $i \neq j = 1, \ldots, d$ and $\sigma_{ij} = 1$ for $i = j$. The proportion of signals varies $\pi_1 = 0.05, 0.1, 0.15$ to apply both sparse and dense signal settings. Numerical studies for different settings are similar to the following results and are presented in the appendix.

For the case of different proportions of signals and compound symmetry correlation matrices with different correlation coefficients, Figure 2, Figure 3 and Figure 4 show the boxplots of the proportions 0.05, 0.1 and 0.15, respectively. As shown in Figure 2, 3 and 4 the dependence-corrected Storey method has a much smaller standard error with accurate estimation, while the method of Langaas et al. (2005) is accurate only for the independent case but standard errors increase as correlation coefficients increase, as mentioned in their Figure 5.

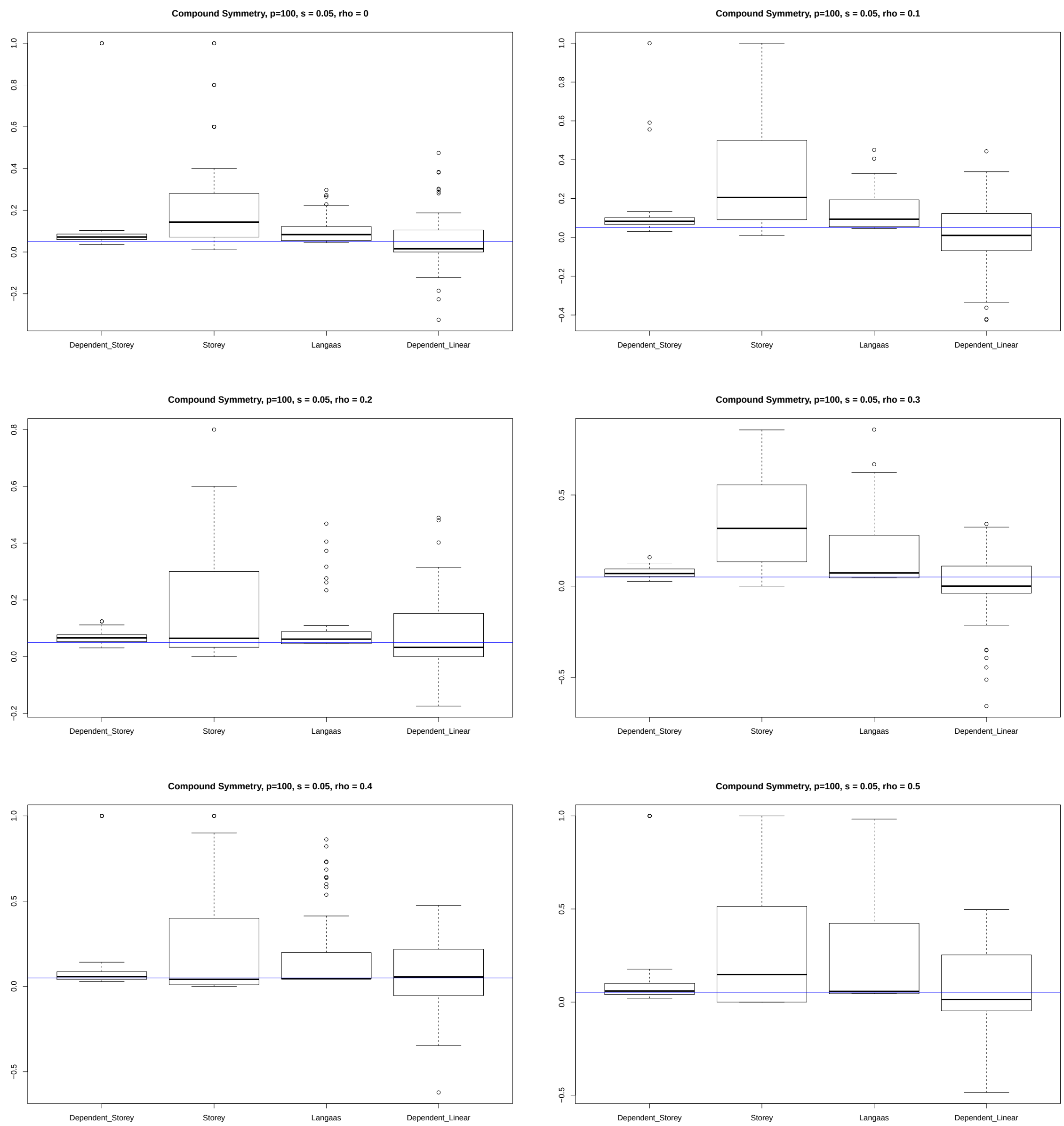

Figure 2: Simulation results of the case of compound symmetry dependence structure with different correlaation coeeficients. The proportion of signals is 0.05 and dimension is 100. Blue horizontal lines indicate 0.05.

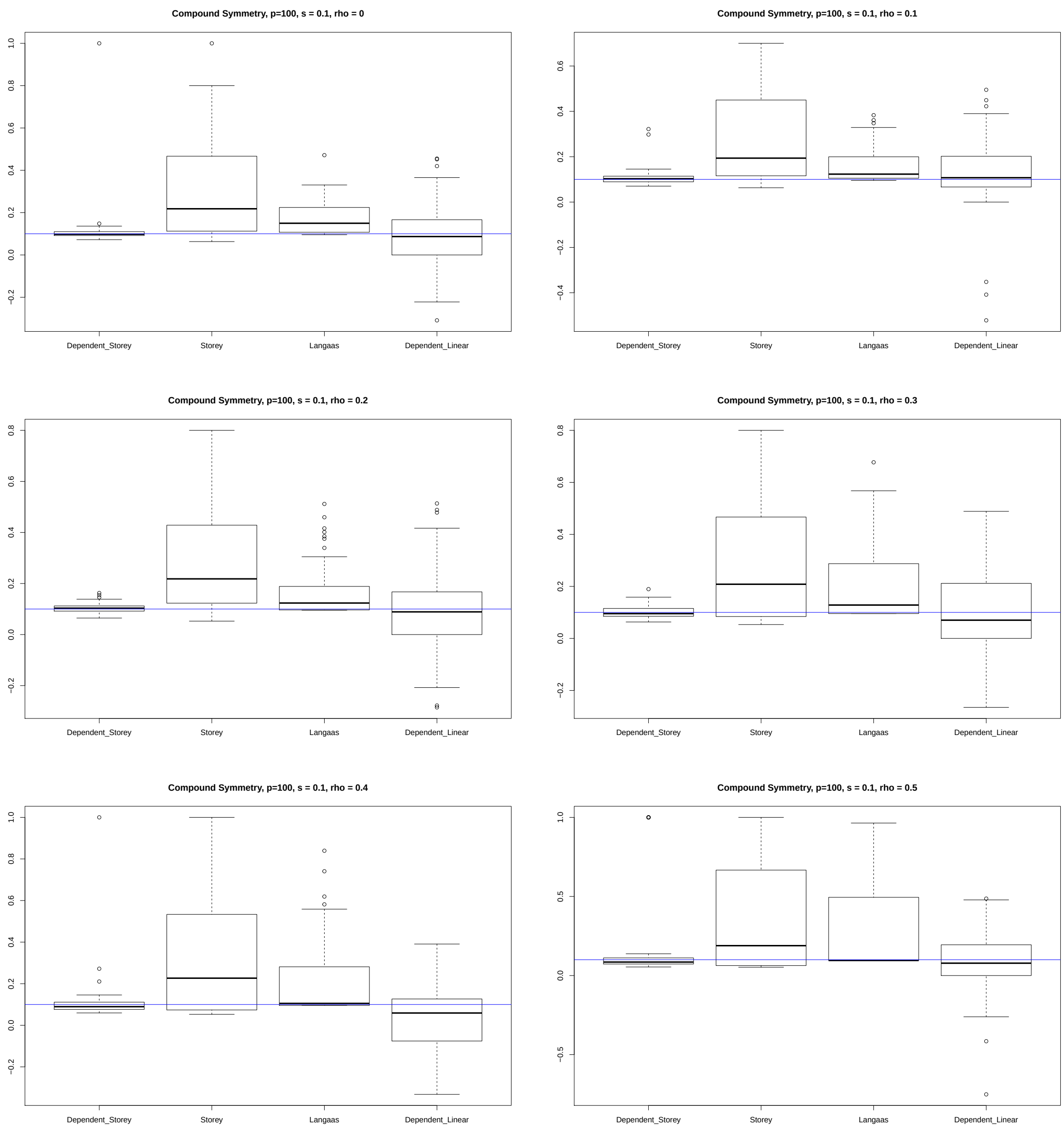

Figure 3: Simulation results of the case of compound symmetry dependence structure with different correlation coeeficients. The proportion of signals is 0.1 and dimension is 100. Blue horizontal lines indicate 0.1.

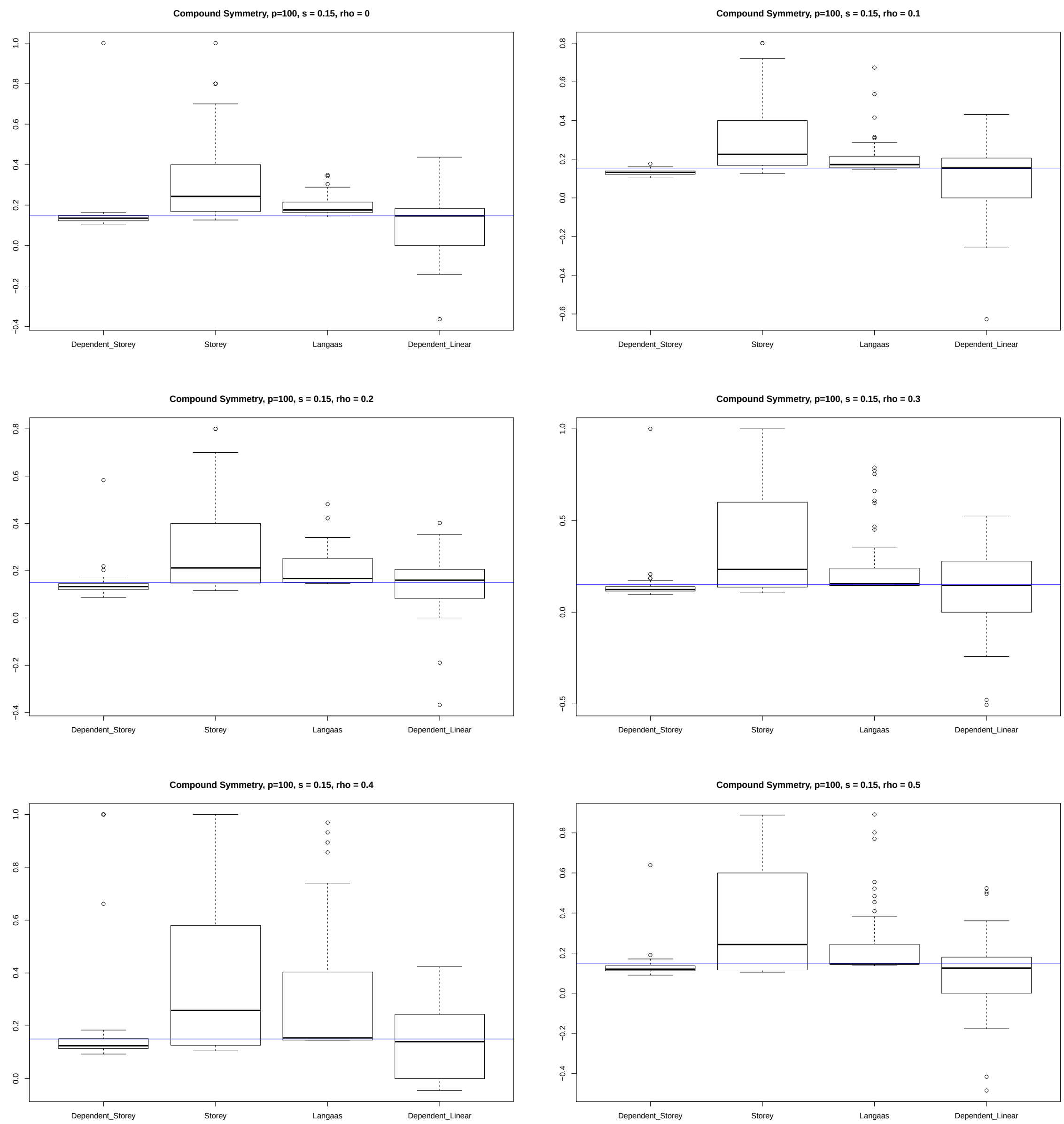

Figure 4: Simulation results of the case of compound symmetry dependence structure with different correlation coeeficients. The proportion of signals is 0.15 and dimension is 100. Blue horizontal lines indicate 0.15.

5.2 Simulation Studies of Type I Error

In this section, we examine the type I error of the proposed method and compare the empirical type I error to that of the Cauchy combination test (Liu and Xie, 2020) as a benchmark. We compare the type I error at a given significance level for different dependency structures, the compound symmetry structure, the polynomially decaying structure and the exponentially decaying structure. The dimension is fixed to $d = 300$. Define $\Sigma = (\sigma_{ij})_{i,j=1}^{d}$. For the compound symmetry dependency structure, which is the case that *p*-values are highly correlated, the correlation coefficient $\sigma := \sigma_{ij}$ varies from 0 (independent *p*-values) to 0.5. For the polynomially decaying dependency structure, consider

$$\sigma_{ij} = I(i = j) + \{1/(0.7 + |i - j|^{\rho})\} \cdot I(i \neq j), \tag{5.1}$$

and conduct simulations with $\rho = 0, 0.1, \ldots, 0.9$. Lastly, for the exponentially decaying dependency structure, we set $\sigma_{ij} = \rho^{|i-j|}$ with $\rho = 0.4, 0.5, \ldots, 0.9$. Since the type I error is defined by the number of false rejections under the null, thresholding value δ can not be defined by the proportion of signals. So we use $\delta = 0.1$ in this simulation.

Figure 7 shows boxplots of the type I error for the proposed method and the Cauchy distribution test at level $\alpha = 0.05$, under the exponentially decaying dependency structure. The boxplot shows that the Cauchy combination test fails to control the type I error as the correlation coefficient increases. On the other hand, we can see that the type I error of the proposed method is controlled quite well especially in the highly correlated case.

For a more strong dependency structure, the polynomially decaying correlation, Figure 6 shows that the Cauchy distribution test can not control the type I error and it is much larger than 0.05 for highly correlated *p*-values. When $\rho = 0.1$ in (5.1) which reflects the highly correlated case, the type I error of the Cauchy combination test is larger than 0.07, that is, above level $\alpha = 0.05$ for as much as 40%. This result implies that if *p*-values are highly correlated, the tail probability of the Cauchy combination test is severely affected on the correlation unlike the proposed method that thresholds large *p*-values.

The failure of the Cauchy combination test to control type I error under highly correlated cases also can be found in Figure 7 where *p*-values are obtained from the compound symmetry structure. On the contrary, the proposed method can control the type I error well even under the highly correlated case. Since controlling type I error guarantees the reliability of the test statistics, simulation results in Figure 7, 6 and 5 indicate that the power of our proposed method is more reliable than that of the Cauchy combination test. We present results of the power comparison in the next section.

5.3 Simulation Studies of Power Analysis

In this section, we investigate the powers of the proposed method compared to other global testing procedures, the Higher criticism (Donoho and Jin, 2004), the Berk-Jones (Berk and Jones, 1979), the minimum *p*-value (Tippett, 1931) as well as the Cauchy combination test (Liu and Xie, 2020) under different dependency structures. We compare the power with varying dimension $d = 20, 40, 60, 100, 200, 300, 500$. All the signals have the same amplitude $\mu_i = \sqrt{3\log(d)}/s^{1/5}$ for all $i \in \mathcal{S} = \{j \in \{1, \ldots, d\} : \mu_i \neq 0\}$ and $|\mathcal{S}| = s$. We use a different range of the correlation coefficient ρ. For the compound symmetry dependency structure, we use $\rho = 0, 0.05, 0.1, \ldots, 0.55, 0.6$. For the polynomially decaying case, $\rho = 0, 0.05, 0.1, \ldots, 0.55, 0.6$ are used to compare the powers. And we use $\rho = 0.4, 0.45, 0.5, \ldots, 0.95, 0.99$ for the AR(1) structure. Figure 8, 9 and 10 show the powers of the proposed method and other global testing procedures. Rows of the power plots indicate

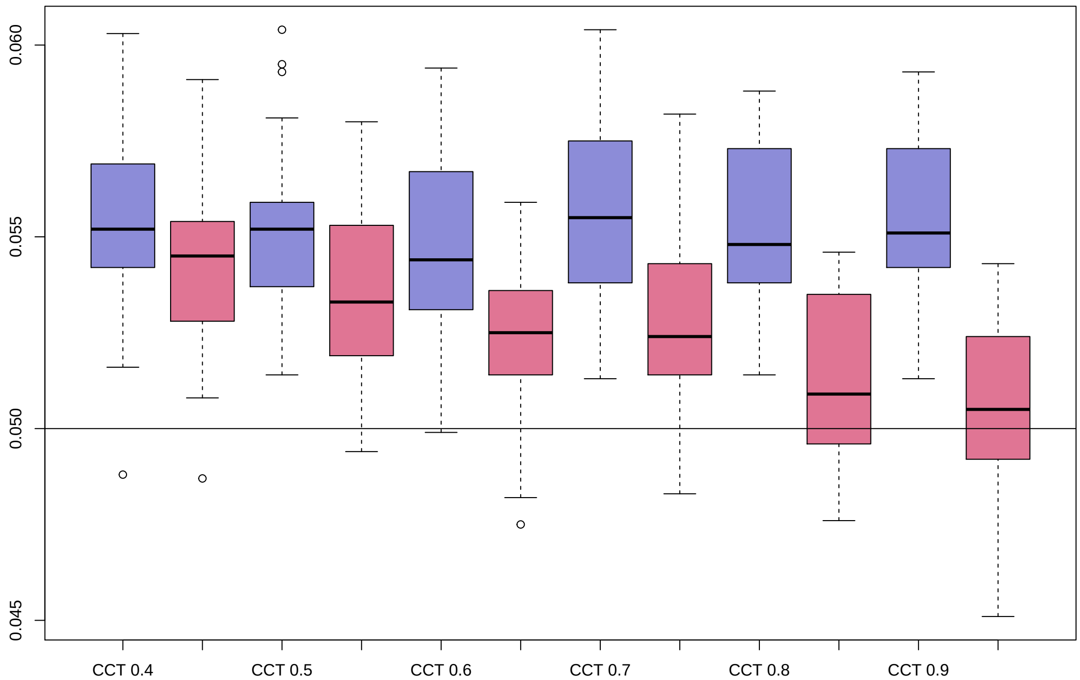

Figure 5: Box plots of type I error at level 0.05 for dimension $d = 300$ and Exponentially decaying with correlation range 0.4 to 0.9. Red boxes are Adaptive thresholding CCT with thresholding 0.1 and blue boxes are CCT.

the proportion of signals, $0.01, 0.05, 0.1$, and 0.15. Columns of the plots indicate varying dimensions $d = 20, 40, 60, 100, 200, 300, 500$. For sparse signal settings, $\pi_1 = 0.01$, there are no signals if $d = 20, 40$ and 60. Hence we assume that there is only one signal for $d = 20, 40$ and 60. For each dependency structure, we simulate by using the Monte Carlo sampling methods to calculate the critical values of each test procedures at the significance level of 0.05.

Since constructing the test statistic of the Higher criticism and the Berk-Jones method is based on independent p-values, the powers of those methods are expected to decrease

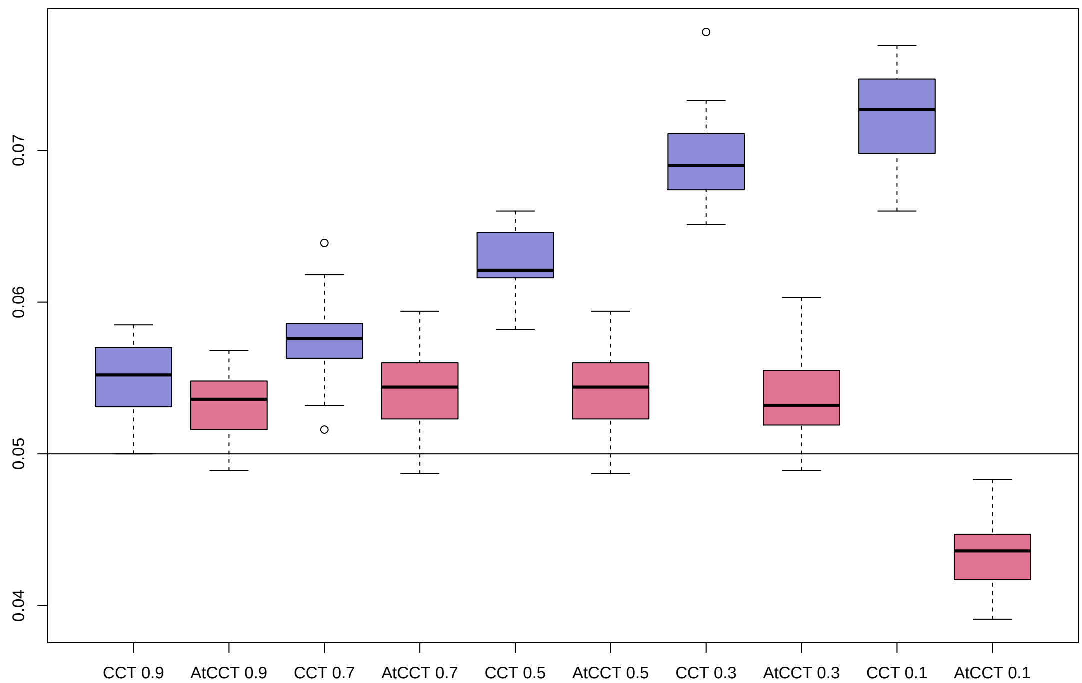

Figure 6: Box plots of type I error at level 0.05 for dimension $d = 300$ and Polynomially decaying with correlation range 0.1 to 0.9. Red boxes are Adaptive thresholding CCT with thresholding 0.1 and blue boxes are CCT.

as correlations between *p*-values increase. Figure 8, 9 and 10 support this expectation. The minimum *p*-value method is robust to the magnitude of correlation coefficients although it loses the power when dependency is weak or *p*-values are independent as shown in Figure 9 and 10.

The Cauchy combination test has a comparable power compared to other tests. However, we have seen in Chapter 5.2 that the Cauchy combination test fails to control the type I error. By thresholding *p*-values, the proposed method has a comparable power compared to all other methods and the proposed method is robust to the magnitude of correlation coeffi-

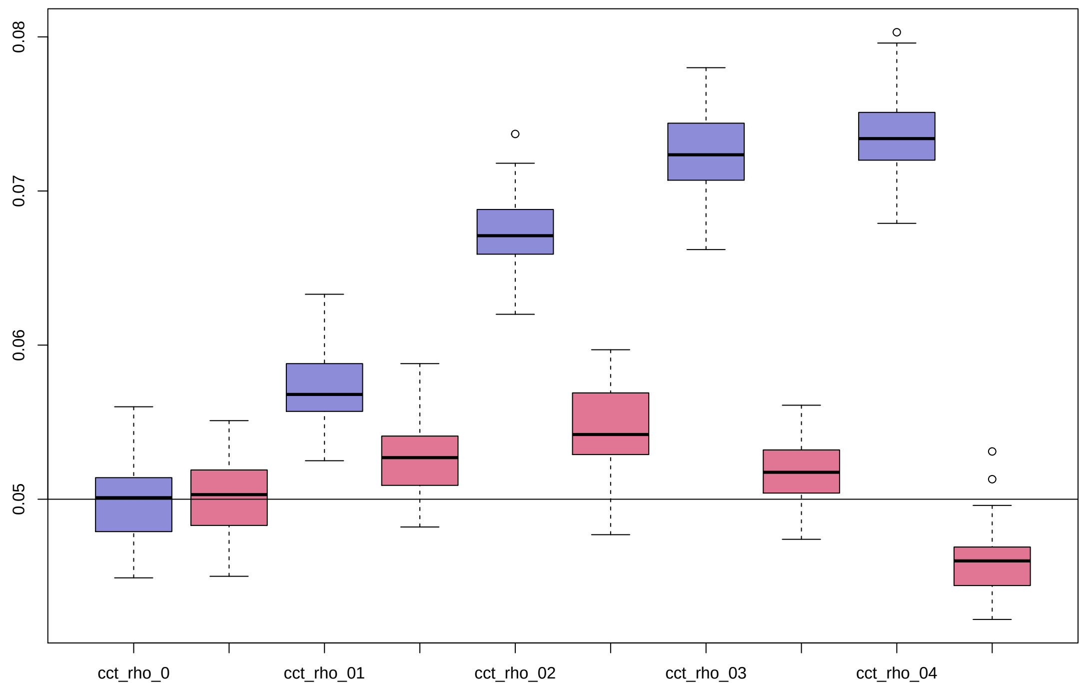

Figure 7: Box plots of type I error at level 0.05 for dimension $d = 300$ and CS with correlation range 0 to 0.5. Red boxes are Adaptive thresholding CCT with thresholding 0.1 and blue boxes are CCT.

cients and the proportion of signals as well as dimensions. Therefore we can conclude that the proposed method outperforms all other competing methods, in that it has a comparable power even if there is no prior information about dependency structures or the proportion of signals.

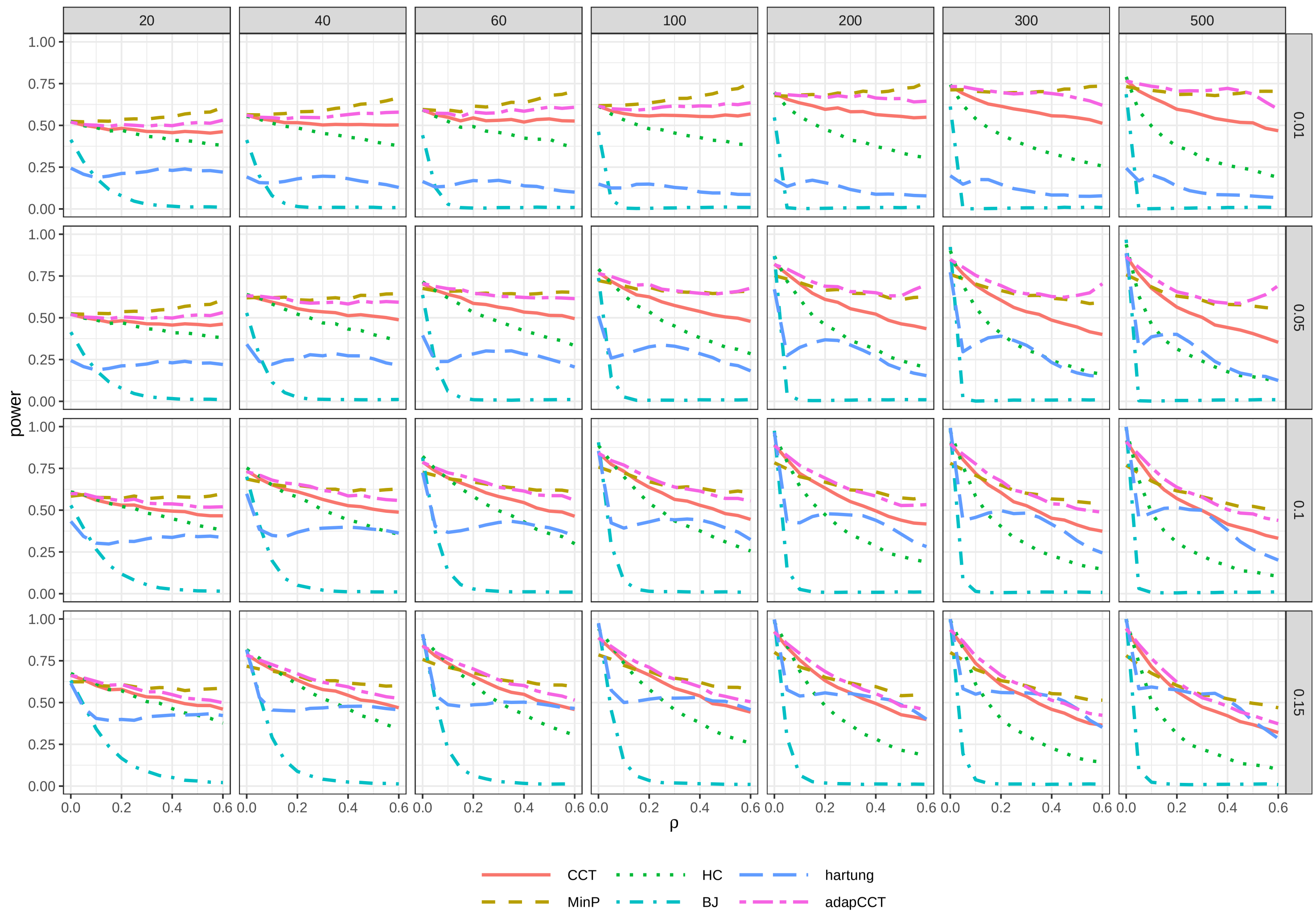

Figure 8: Power comparison of Adaptive thresholding CCT, MinP, HC, BJ, Hartung and CCT. The x-axis is the correlation strength of the compound symmetry structure. The columns from left to right correspond to the dimension $d = 20, 40, 60, 100, 200, 300, 500$. The rows from top to bottom correspond to the signal proportion $5\%, 10\%, 15\%$.

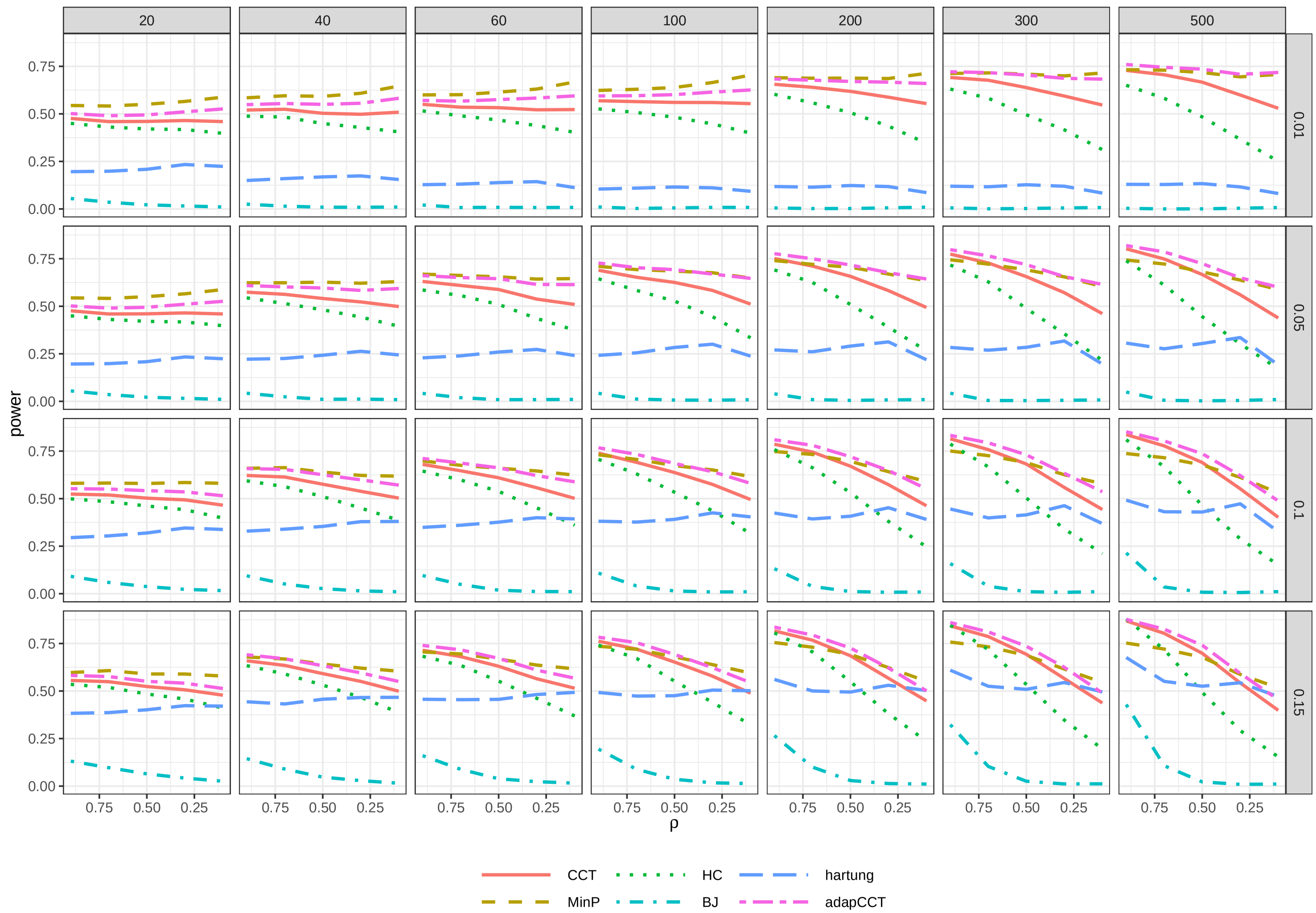

Figure 9: Power comparison of Adaptive thresholding CCT, MinP, HC, BJ, Hartung and CCT. The x-axis is the correlation strength of the polynomially decaying structure. The columns from left to right correspond to the dimension $d = 20, 40, 60, 100, 200, 300, 500$. The rows from top to bottom correspond to the signal proportion $5\%, 10\%, 15\%$.

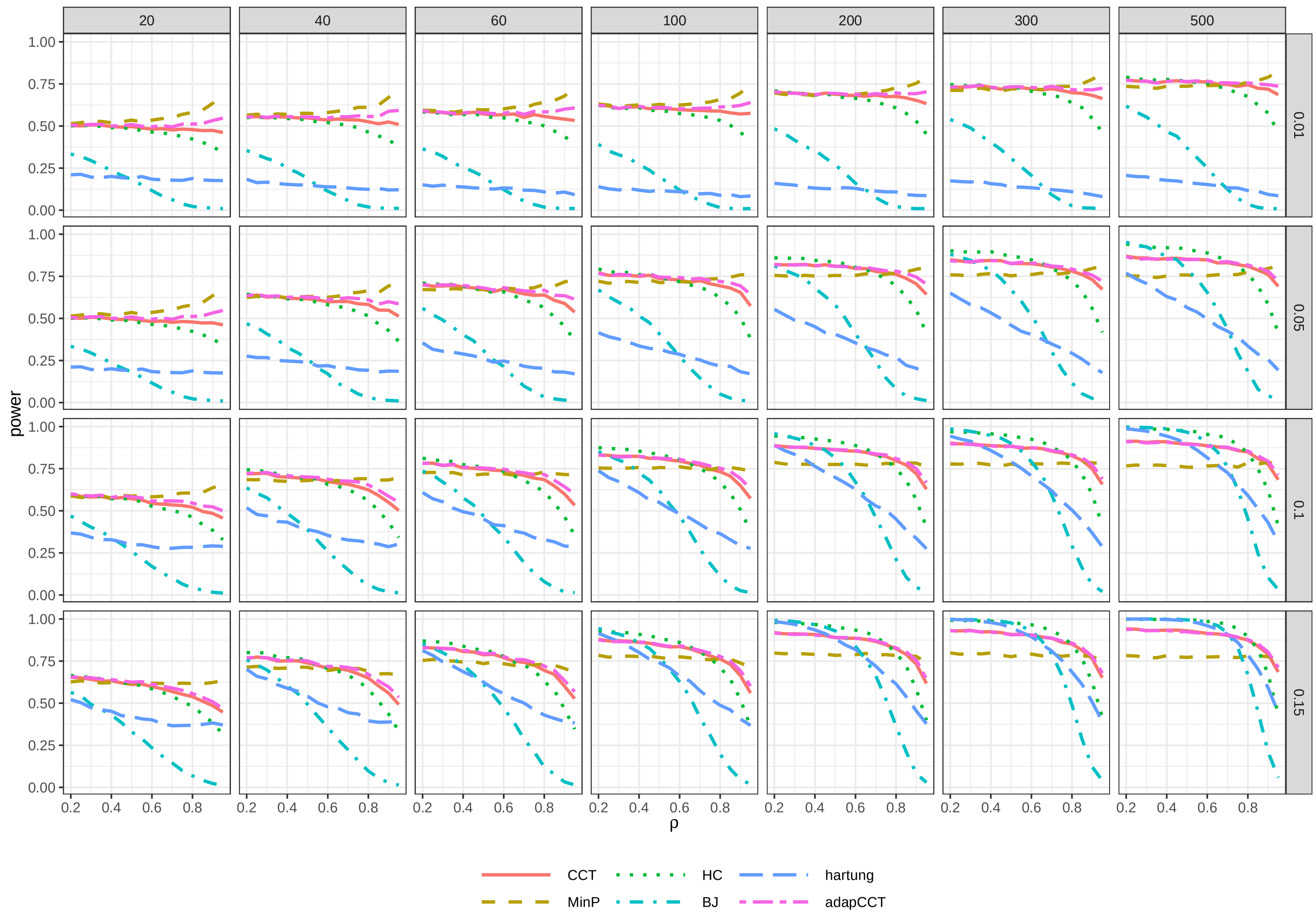

Figure 10: Power comparison of Adaptive thresholding CCT, MinP, HC, BJ, Hartung and CCT. The x-axis is the correlation strength of the exponentially decaying structure. The columns from left to right correspond to the dimension $d = 20, 40, 60, 100, 200, 300, 500$. The rows from top to bottom correspond to the signal proportion $5\%, 10\%, 15\%$.

Chapter 6

Case Studies

We apply the Cauchy combination with thresholding test procedure to the Inflammatory bowel disease (IBD) genome-wide association study (GWAS) for the real data analysis. Crohn's disease is a type of IBD and the goal of Crohn's disease GWAS is to find genes associated with the IBD. The data consists of p-values of each SNP analyzed to test the case and control association among both the Jewish and non-Jewish cohorts. p-values are constructed by using a Cochran-Mantel Haenszel chi-square test computed separately for each SNP. The dataset is downloaded from the Genotypes and Phenotypes (dbGaP).

The dataset contains 968 IBD-affected cases and 995 unrelated controls by using the Illumina HumanHap300 Genotyping BeadChip. The cases were selected to have Crohn's disease with ileal involvements, and controls were matched to the cases based on sex and year of birth. Subjects were drawn from two cohorts:

1. persons with non-Jewish, European ancestry (561 cases and 563 controls)

2. persons with Jewish ancestry (407 cases and 432 controls)

To analyze the association between genes and IBD, we group SNPs to genes via GBiB of

UCSC. Each SNP can be contained in multiple genes. As a result, all 242,535 SNPs are grouped into 19,769 genes. The number of SNPs in each gene ranges from 1 to 676. Among all genes, only 4,969 genes having more than 10 SNPs are used to improve the accuracy of estimating the proportion of signals in each gene.

Figure 6 shows histograms of p-values of SNPs in two genes, “NLGN1” gene and “CDH4” gene. As shown in Figure 6, p-values in each gene deviate from the uniform distribution which means that the p-values are highly correlated. Since the dataset consists of only p-values, estimating the dependency structure can not be applied so that global testing procedures constructed under an independence assumption or based on an estimate of dependency structure can not be applied.

Table 1 presents 10 smallest global p-values obtained by the proposed Cauchy combination test with thresholding. We first estimate the proportion of signals in each gene and then use the twice estimate of the proportion of signals as the thresholding value. The gene with the smallest global p-value, “C1orf141”, is identified to have an association with IBD in Dinu et al. (2012) and the gene with the second smallest p-value, “IL23R” is identified in Duerr et al. (2006).

Table 2 shows genes that are significant at the level of 0.05 by using the proposed method but not significant by using the Cauchy combination test. There are no genes identified by using the Cauchy combination test while unidentified by the proposed method. This result indicates that our proposed method can find more significant genes and is more powerful, although further biological research is required.

Figure 12 and 13 show histograms of the global p-values of genes identified by our method. From the figures, we can confirm that if there is are strong and spase signals, as a global testing procedure, our method can identify such genes.

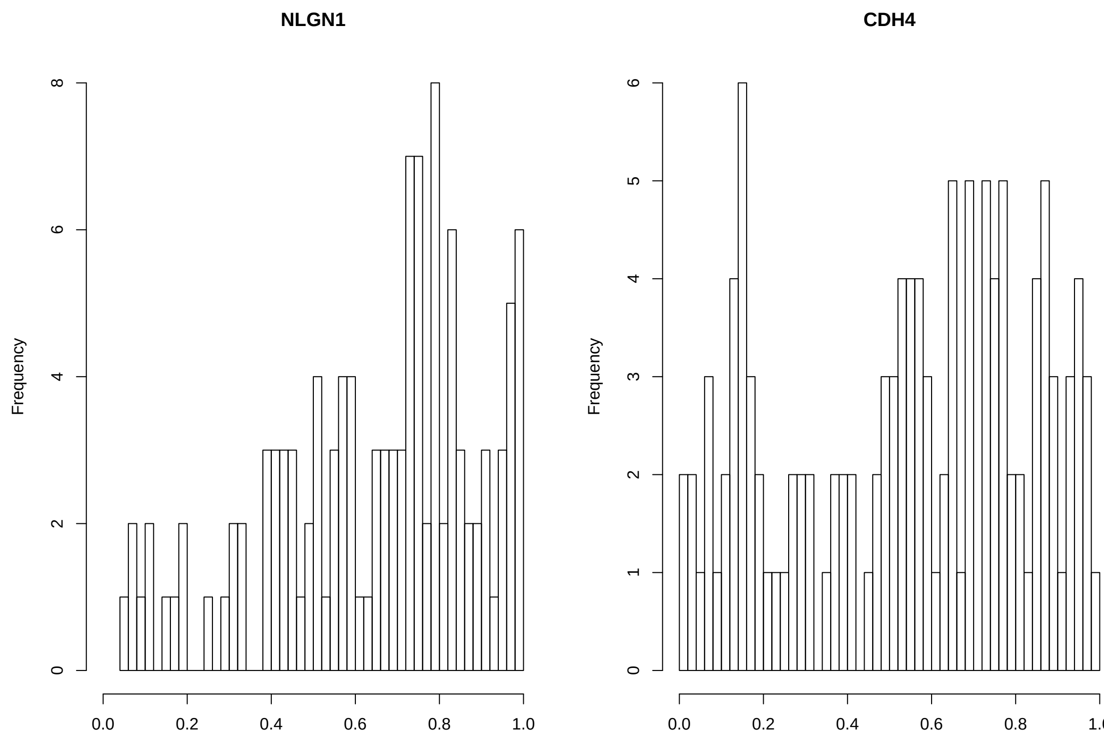

Figure 11: Histogram of p-values of SNPs for the "NLGN1" gene and "CDH4" gene

GENE	d	Test Statistics	p-values
C1orf141	20	5.35×10^{10}	5.95×10^{-12}
IL23R	23	4.79×10^{10}	6.65×10^{-12}
ZNF532	18	6.24×10^{3}	3.39×10^{-5}
CACNA1C	100	1.84×10^{3}	1.04×10^{-4}
AC067751.1	16	1.63×10^{3}	1.95×10^{-4}
NCF4	14	1.34×10^{3}	2.39×10^{-4}
NCF4-AS1	17	1.10×10^{3}	2.89×10^{-4}
CLSTN2	108	1.03×10^{3}	3.09×10^{-4}
STXBP4	12	7.91×10^{2}	4.02×10^{-4}
SULF2	22	7.67×10^{2}	4.15×10^{-4}

Table 1: 10 smallest p-values for the Cauchy combination Test with thresholding.

GENE	adaptive Cauchy	Cauchy	d
SORBS2	0.04308	0.99142	72
FRMD6-AS2	0.03371	0.98582	34
TANC2	0.02523	0.97815	17
AC008591.1	0.02951	0.96717	28
AC026369.1	0.01263	0.94029	13
IQSEC3	0.02623	0.93082	27
AC008691.1	0.04671	0.78029	59
USH1C	0.04447	0.73130	15
AC114311.1	0.02907	0.48239	12
AC026765.3	0.03911	0.41990	17
TEX26-AS1	0.04919	0.17864	14
SSBP3	0.04629	0.05536	31

Table 2: Genes that are significant at a level of 0.05 by using the Cauchy combination test with thresholding but not the Cauchy combination test.

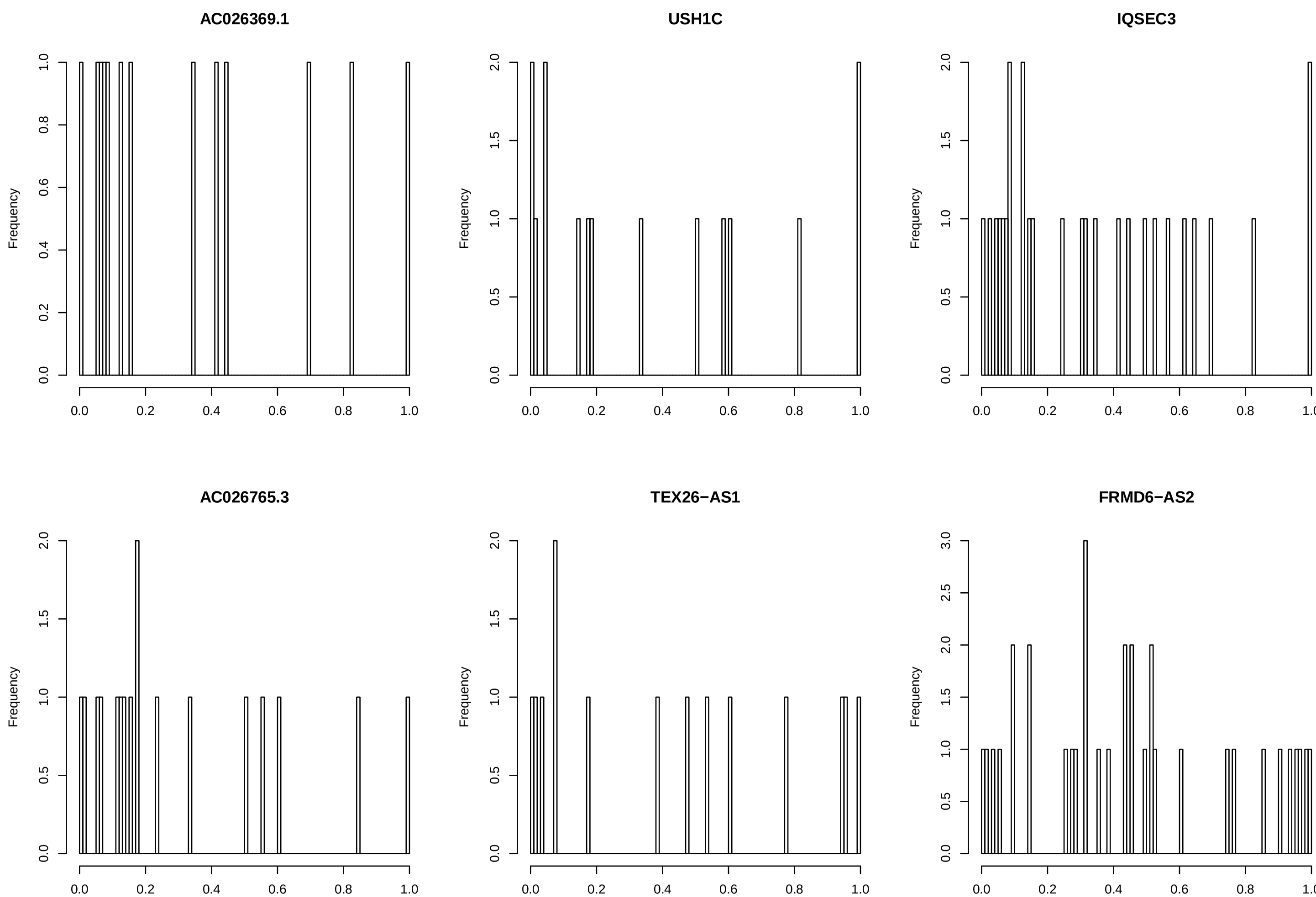

Figure 12: Histograms for p-values of “AC026369.1”, “USH1C”, “IQSEC3”, “AC026765.3”, “TEX26-AS1”, “FRMD6-AS2” and “AC091078.1”

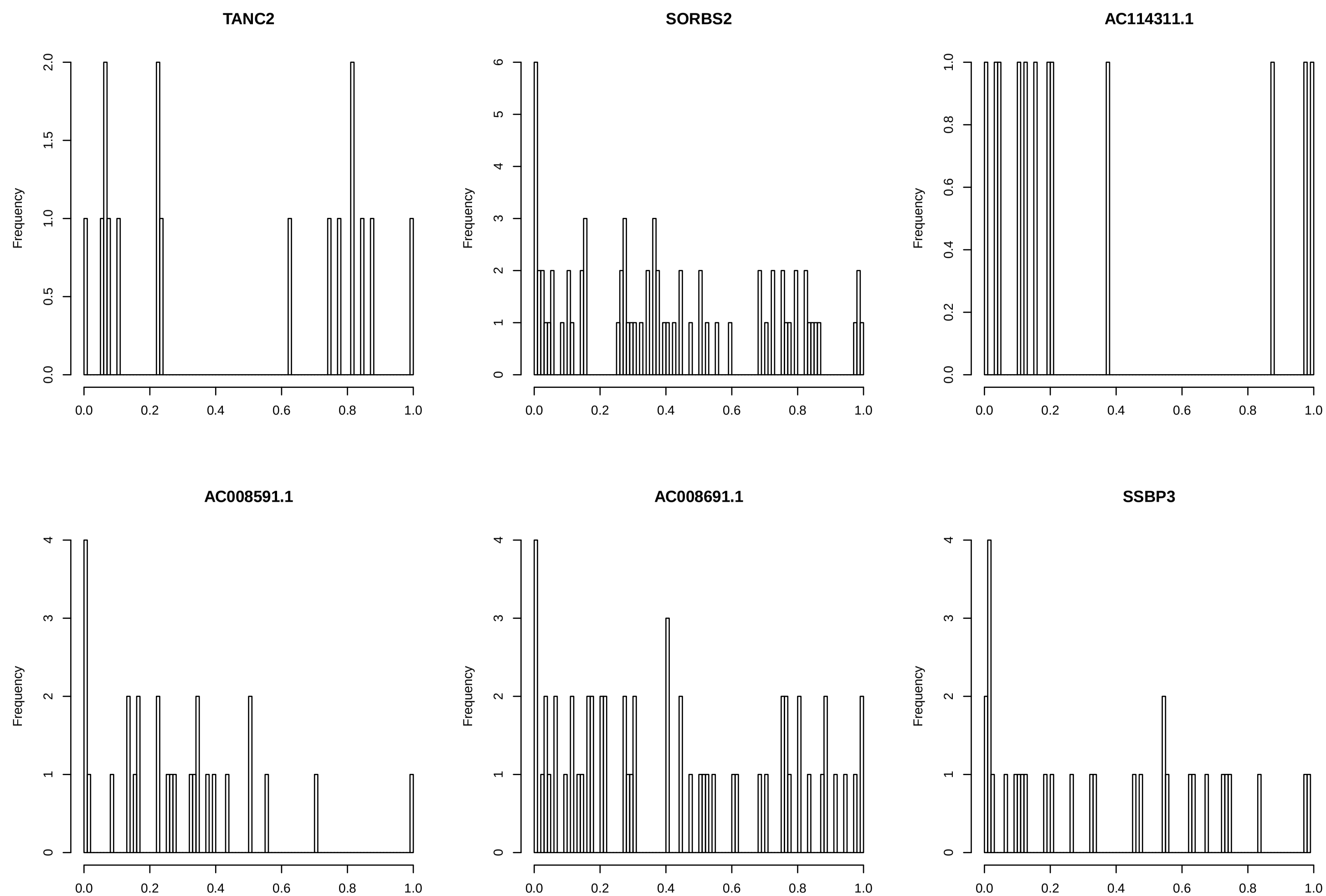

Figure 13: Histograms for p-values of “TANC2”, “SORBS2”, “AC114311.1”, “AC008591.1”, “AC008691.2” and “SSBP3”

Chapter 7

Conclusions

In this thesis, we proposed a global testing procedure for detecting a sparse mean vector under arbitrary dependency structures based on observed p-values. By thresholding the p-values, we extend the Cauchy combination test to a more powerful and robust method for the dependency structures without any prior information of data structures. We proved the tail probability of the proposed test statistic is equivalent to that of a Cauchy distribution under the null. The equivalence is proved both asymptotically or non asymptotically. We also proved that the power of the proposed method converges to 1. In addition, we found that the proposed method achieves the optimal detection boundary for sparse settings. Extensive numerical studies support the proposed method. Simulation results showed that the type I error of the proposed method is well controlled at a given significance level for a wide range of dependencies, although the type I error of the un-thresholded Cauchy combination methods increases as the correlation coefficient increases. And the power of the proposed method outperforms other competing global testing methods, especially under the sparse and highly correlated case.

For practical usage of the proposed method, we showed that the thresholding value can be chosen as twice the proportion of the signals. To estimate the proportion of the signals, we extended Storey's method to accommodate the arbitrary dependency structures and denote the method as the "dependence-corrected Storey method". Arbitrary patterns of the observed p-values are estimated by the kernel estimation. Numerical studies showed that the proposed estimate of the proportion of signals outperforms other methods. The standard error of the proposed method is much smaller and the accuracy of the estimation is comparable.

We applied the proposed method to IBD genome-wide association study. The dataset consists of only arbitrarily dependent p-values of SNPs. We used the proposed method to find genes, which are grouped by SNPs, associated with Crohn's disease. We found significant genes among 5,000 genes.

Future work will generalize the Cauchy combination with thresholding to the heavy-tailed distribution. Error rates of convergence rates of the null distribution and the power depending on dependency structures will establish a more robust testing procedure.

Appendix A

Proof of Theorems and Lemmas

A.1 Proof of Theorem 4

For given $\delta \in [0,1]$, let $\mathcal{M} = \{i \in [1,d] : p_j \in [0,\delta]\}$. From the equivalence of (3.2) and (3.3), T_q^* can be considered.

$$P\left(\sum_{j\in\mathcal{M}} w_j h_q(q_j) > t\right) = P\left(\sum_{j=1}^{d} w_j I(p_j \leq \delta) h_q(q_j) > t\right). \tag{A.1}$$

Since p_j's follow marginally uniform distribution for all $j = 1, \dots, d$, for $j \in \mathcal{M}$, trivially note that $q_j = p_j/\delta \sim U[0,1]$. Let $\tilde{w}_j = w_j I(p_j \leq \delta)$. Then we can decompose (A.1) as follows.

$$P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t\right) \tag{A.2}$$

$$= P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \quad \bigcup_{j=1}^{d} \{\tilde{w}_j h_q(q_j) > (1+\delta_t)t\}\right) \tag{A.3}$$

$$+ P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \quad \bigcap_{j=1}^{d} \{\tilde{w}_j h_q(q_j) \leq (1+\delta_t)t\}\right), \tag{A.4}$$

where δ_t is a constant only depends on t satisfying $\delta_t > 0$, $\delta_t \to 0$ and $\delta_t t \to \infty$ as $t \to \infty$. Letting

$$A = \bigcup_{j=1}^{d} A_j = \bigcup_{j=1}^{d} \left\{ \tilde{w}_j h_q(q_j) > (1+\delta_t)t, \quad \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t \right\},$$

the probability in the right hand side of (A.3) can be expressed by $P(\cup_{j=1}^{d} A_j)$. By the Bonferroni inequality,

$$\sum_{j=1}^{d} P(A_j) - \sum_{1 \le i < j \le d} P(A_i \cap A_j) \le P\left(\bigcup_{j=1}^{d} A_j\right) \le \sum_{j=1}^{d} P(A_j). \tag{A.5}$$

First consider $P(A_j)$ in (A.5). It can be decomposed as follows.

$$P(A_j) = P(\tilde{w}_j h_q(q_j) > (1+\delta_t)t) - P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t, \quad \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) \le t\right).$$

Then

$$\begin{aligned} P(\tilde{w}_j h_q(q_j) > (1+\delta_t)t) &= P(w_j I(p_j \le \delta) h_q(q_j) > (1+\delta_t)t) \\ &= P(w_j h_q(q_j) > (1+\delta_t)t \mid p_j \le \delta) \cdot P(p_j \le \delta) \\ &= \frac{w_j \delta}{(1+\delta_t)t\pi} + o(1/t). \end{aligned} \tag{A.6}$$

The last equality holds from lemma 2 and the fact that p_j follows the uniform distribution marginally. To conclude that $P(A_j) = (w_j\delta)/((1+\delta_t)t\pi) + o(1/t)$, it suffices to show that $P(\tilde{w}_j h_q(q_j) > (1+\delta_t)t, \ \ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) \le t) = o(1/t)$. Note that

$$P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t, \quad \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) \le t\right) \le P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t, \quad \sum_{i \ne j} \tilde{w}_i h_q(q_i) \le -\delta_t t\right).$$

Recall the condition that the largest eigenvalue of Σ is bounded by C_0. Let $\mathcal{J}_j = \{i \ne j : \sigma_{ij}^2 \ge \sigma_0^2\}$ for some constant $0 < \sigma_0 < 1$. Then the cardinality of $\mathcal{J}_j$ should be less than or equal to C_0/σ_0^2. Define $\delta_t = (1 + C_0/\sigma_0^2)t^{-\epsilon}$ for $0 < \epsilon < 1$. Then

$$
\begin{aligned}
&\left\{\sum_{i\neq j}\tilde{w}_i h_q(q_i) \leq -\delta_t t\right\} \\
&= \left\{\sum_{i\in\mathcal{J}_j}\tilde{w}_i h_q(q_i) + \sum_{i\in\mathcal{J}_j^c}\tilde{w}_i h_q(q_i) \leq -\delta_t t\right\} \\
&\subseteq \left\{\left[\sum_{i\in\mathcal{J}_j}\tilde{w}_i h_q(q_i) \leq -\left(\frac{C_0}{\sigma_0^2}\right)t^{1-\epsilon}\right] \cup \left[\sum_{i\in\mathcal{J}_j^c}\tilde{w}_i h_q(q_i) \leq -\delta_t t + \left(\frac{C_0}{\sigma_0^2}\right)t^{1-\epsilon}\right]\right\} \\
&\subseteq \left\{\left[\bigcup_{i\in\mathcal{J}_j}\left\{\tilde{w}_i h_q(q_i) \leq -\left(\frac{C_0}{\sigma_0^2}\right)\frac{t^{1-\epsilon}}{|\mathcal{J}_j|}\right\}\right] \cup \left[\bigcup_{i\in\mathcal{J}_j^c}\left\{\tilde{w}_i h_q(q_i) \leq \frac{-t^{1-\epsilon}}{|\mathcal{J}_j^c|}\right\}\right]\right\}
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \sum_{j=1}^{d}\tilde{w}_j h_q(q_j) \leq t\right) \\
&\leq \sum_{i\in\mathcal{J}_j} P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \tilde{w}_i h_q(q_i) \leq -\left(\frac{C_0}{\sigma_0^2}\right)\frac{t^{1-\epsilon}}{|\mathcal{J}_j|}\right) \\
&\quad + \sum_{i\in\mathcal{J}_j^c} P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \tilde{w}_i h_q(q_i) \leq \frac{-t^{1-\epsilon}}{|\mathcal{J}_j^c|}\right) \\
&\leq \sum_{i\in\mathcal{J}_j} P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \tilde{w}_i h_q(q_i) \leq -t^{1-\epsilon}\right) \\
&\quad + \sum_{i\in\mathcal{J}_j^c} P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \tilde{w}_i h_q(q_i) \leq \frac{-t^{-\epsilon}}{d}\right) \\
&= T_1 + T_2.
\end{aligned}
$$

First consider T_1. By lemma 2, we obtain that for any $i \in \mathcal{J}_j$,

$$
\begin{aligned}
&P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \tilde{w}_i h_q(q_i) \leq -t^{1-\epsilon}\right) \\
&= P\left(w_j h_q(q_j) > (1+\delta_t)t,\ \ w_i h_q(q_i) \leq -t^{1-\epsilon} \mid p_j \leq \delta,\ p_i \leq \delta\right) \cdot p(p_j \leq \delta,\ p_i \leq \delta) \\
&\leq P\left(w_j h_q(q_j) > (1+\delta_t)t,\ \ w_i h_q(q_i) \leq -t^{1-\epsilon} \mid p_j \leq \delta,\ p_i \leq \delta\right)
\end{aligned}
$$

The equivalence between h_q and h_X yields that when $p_j \le \delta$ and $p_i \le \delta$,

$$
\begin{aligned}
& P\left(w_j h_q(q_j) > (1+\delta_t)t,\ \ w_i h_q(q_i) \le -t^{1-\epsilon}\right) \\
& = P\left(p_j < \delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right),\ \ p_i \ge \delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right) \\
& = P\left(h_q(p_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\ \ h_q(p_i) \le h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right)\right) \\
& = P\left(h_X(X_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\ \ h_X(X_i) \le h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right)\right). \quad \text{(A.7)}
\end{aligned}
$$

The bivariate normality assumption of X implies that $X_i = \sigma_{ij}X_j + \gamma_{ij}Z_{ij}$, where $\gamma_{ij}^2 = 1 - \sigma_{ij}^2$ and Z_{ij} is a random variable following a standard normal distribution and independent of X_j.

$$
\begin{aligned}
\left\{h_X(X_i) \le h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right)\right\} &= \left\{h_X(\sigma_{ij}X_j + \gamma_{ij}Z_{ij}) \le h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right)\right\} \\
&= \left\{Z_{ij} \le \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left[h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right)\right] - \sigma_{ij}X_j\right]\right\}.
\end{aligned}
$$

Since $h_X(X)$ is monotonic increasing in X and $h_q(q)$ is monotonic decreasing in q, the X_j in the event (A.7) can be lower bounded such that

$$
X_j > h_X^{-1}\left[h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right)\right] \ge h_X^{-1}\left[h_q\left(h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right)\right] = h_X^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right).
$$

Hence

$$
\begin{aligned}
& \left\{h_X(X_i) \le h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right)\right\} \\
& \subseteq \left\{Z_{ij} \le \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left[h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right)\right] - \sigma_{ij}h_X^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right]\right\}. \quad \text{(A.8)}
\end{aligned}
$$

It can be shown that $h_q^{-1}(-t^{1-\epsilon}/(w_i)) \to 1$ so that for some $\tilde{\delta}$,

$$
h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right) = h_q(\delta(1+o(1))) \le h_q(\tilde{\delta}).
$$

From properties of trigonometric functions and Mill's inequality, for $|x| > \Phi^{-1}(3/4+\tau/\pi)$ where $0 < \tau < \pi/4$,

$$\begin{aligned}
h_X(x) &= \tan([2\Phi(|x|) - 3/2]\pi) = \tan((p(x) - 1/2)\pi) = \frac{1}{\tan(p(x)\pi)} = \frac{\cos(p(x)\pi)}{\sin(p(x)\pi)} \\
&\geq \frac{\cos(p(x)\pi)}{p(x)\pi} = \frac{\cos(p(x)\pi)}{2\pi(1 - \Phi(|x|))} \\
&\geq \frac{\cos(\pi/2 - \tau)}{2\pi}\frac{|x|}{\phi(x)} \geq \frac{\cos(\pi/2 - \tau)}{2\pi}\Phi^{-1}(3/4 + \tau/\pi)e^{x^2/2} \\
&= C_\tau e^{x^2/2},
\end{aligned}$$

where we denote $C_\tau = \cos(\pi/2 - \tau)\Phi^{-1}(3/4 + \tau/\pi)/2\pi$. The inverse function of h_X can be bounded by

$$\sqrt{2\log y} \leq h_X^{-1}(y) \leq \sqrt{2\log(y/C_\tau)} \leq (1 + o(1))\sqrt{2\log y}. \tag{A.9}$$

Therefore (A.7) can be bounded as follows.

$$\begin{aligned}
&P\left(h_X(X_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\ \ h_X(X_i) \leq h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{w_i}\right)\right)\right) \\
&\leq P\left(h_q(p_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\right. \\
&\qquad\qquad \left.|Z_{ij}| \leq \frac{1}{\gamma_{ij}}\left[h_X^{-1}(h_q(\tilde{\delta})) - (1+o(1))\sigma_{ij}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right]\right) \\
&= P\left(h_q(q_j) > \frac{(1+\delta_t)t}{w_j}\right)\cdot P\left(|Z_{ij}| \leq -\frac{(1+o(1))\sigma_{ij}}{\sqrt{1-\sigma_{ij}^2}}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right) \\
&= P\left(h_q(q_j) > \frac{(1+\delta_t)t}{w_j}\right)\cdot P\left(|Z_{ij}| \geq \frac{(1+o(1))\sigma_{ij}}{\sqrt{1-\sigma_{ij}^2}}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right) \\
&\leq P\left(h_q(q_j) > \frac{(1+\delta_t)t}{w_j}\right)\cdot P\left(|Z_{ij}| \geq \frac{(1+o(1))\sigma_0}{\sqrt{1-\sigma_0^2}}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right).
\end{aligned} \tag{A.10}$$

The last inequality holds since for $i \in \mathcal{J}_j$, $\sigma_{ij}^2 \geq \sigma_0^2$. By Mill's inequality,

$$P\left(|Z_{ij}| \geq \frac{(1+o(1))\sigma_0}{\sqrt{1-\sigma_0^2}}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right)$$

$$
\begin{aligned}
&= 2\left(1-\Phi\left(\frac{(1+o(1))\sigma_0}{\sqrt{1-\sigma_0^2}}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right)\right)\\
&\leq 2\frac{\phi\left(\frac{(1+o(1))\sigma_0}{\sqrt{1-\sigma_0^2}}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right)}{\left(\frac{(1+o(1))\sigma_0}{\sqrt{1-\sigma_0^2}}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right)}.\\
&= \frac{2}{\sqrt{2\pi}}\frac{1}{f(\sigma_0)\left(\frac{(1+\delta_t)t}{w_j}\right)^{f(\sigma_0)^2}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}}.
\end{aligned}
$$

where we denote $f(\sigma_0) = (1+o(1))\sigma_0/\sqrt{1-\sigma_0^2}$. For sufficiently large t, it can be shown that

$$
P\left(|Z_{ij}| \geq \frac{(1+o(1))\sigma_0}{\sqrt{1-\sigma_0^2}}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right) \leq o(1).
$$

Then (A.10) can be bounded by

$$
P\left(h_q(q_j) > \frac{(1+\delta_t)t}{w_j}\right)\cdot P\left(|Z_{ij}| \geq \frac{(1+o(1))\sigma_0}{\sqrt{1-\sigma_0^2}}\sqrt{2\log\left(\frac{(1+\delta_t)t}{w_j}\right)}\right) \leq \left(\frac{w_j}{t\pi}+o(1/t)\right)\cdot o(1).
$$

Therefore

$$
T_1 = \left(\frac{\sum_{i\in\mathcal{J}_j} w_j}{t\pi} + o(1/t)\right)\cdot o(1). \tag{A.11}
$$

Next, consider T_2. By using a similar steps in T_1, when $p_j \leq \delta$ and $p_i \leq \delta$,

$$
\begin{aligned}
&P\left(w_j h_q(q_j) > (1+\delta_t)t,\ \ w_i h_q(q_i) \leq \frac{-t^{1-\epsilon}}{d}\right)\\
&= P\left(p_j > \delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right),\ \ p_i \leq \delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{dw_i}\right)\right)\\
&= P\left(h_q(p_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\ \ h_q(p_i) \leq h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{dw_i}\right)\right)\right)\\
&= P\left(h_X(X_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\ \ h_X(X_i) \leq h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{dw_i}\right)\right)\right).
\end{aligned}
$$

The bivariate normality of X also implies that

$$\begin{aligned}
\left\{h_X(X_i) \le h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{dw_i}\right)\right)\right\} &= \left\{h_X(\sigma_{ij}X_j + \gamma_{ij}Z_{ij}) \le h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{dw_i}\right)\right)\right\} \\
&= \left\{Z_{ij} \le \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left[h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{dw_i}\right)\right)\right] - \sigma_{ij}X_j\right]\right\}.
\end{aligned}$$

For some $\bar{\delta}$,

$$h_q\left(\delta h_q^{-1}\left(\frac{-t^{1-\epsilon}}{dw_i}\right)\right) = h_q(\delta(1+o(1))) \le h_q(\bar{\delta}).$$

Then the same arguments in T_1 implies

$$T_2 = \left(\frac{\sum_{i\in\mathcal{J}_j^c} w_j}{t\pi} + o(1/t)\right)\cdot o(1). \tag{A.12}$$

(A.11) and (A.12) imply that

$$\sum_{j=1}^{d} P(A_j) = \left(\frac{w^*}{t\pi} + o(1/t)\right)\cdot o(1) = o(1/t),$$

where $w^* = \sum_{j=1}^d w_j$. Using a similar steps of proving the case of $P(A_j)$, it is easy to show that for $1 \le i \ne j \le d$,

$$\sum_{1\le i<j\le d} P(A_i \cap A_j) = o(1/t).$$

Therefore (A.3) can be expressed by

$$P\left(\sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t,\ \ \bigcup_{j=1}^{d}\{\tilde{w}_j h_q(q_j) > (1+\delta_t)t\}\right) = \frac{w^*\delta}{t\pi} + o(1/t). \tag{A.13}$$

It remains to show that $(A.4) = o(1/t)$. The proof below is based on Liu and Xie (2020).

Let

$$B = \left\{\sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t,\ \ \bigcap_{j=1}^{d}\{\tilde{w}_j h_q(q_j) \le (1+\delta_t)t\}\right\}.$$

The event $\{\sum_{j=1}^d \tilde{w}_j h_q(q_j) > t\}$ implies that there exists at least one j such that $\tilde{w}_j h_q(q_j) > t/d$. Then it can be shown that

$$B \ = \ \bigcap_{j=1}^{d}\left\{\sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t,\ \ \tilde{w}_j h_q(q_j) \le (1+\delta_t)t\right\}$$

$$\begin{aligned}
&\subseteq \bigcap_{j=1}^{d}\left\{\sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t,\ \ \tilde{w}_j h_q(q_j) \le (1+\delta_t)t,\ \ \bigcup_{j=1}^{d}\left\{\tilde{w}_j h_q(q_j) > \frac{t}{d}\right\}\right\} \\
&= \bigcup_{j=1}^{d}\left\{\sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t,\ \ \tilde{w}_j h_q(q_j) \le (1+\delta_t)t,\ \ \tilde{w}_j h_q(q_j) > \frac{t}{d}\right\} \\
&= \bigcup_{j=1}^{d}\left\{\sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t,\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1+\delta_t)t\right\}.
\end{aligned}$$

Then we have

$$\begin{aligned}
P(B) &\le \sum_{j=1}^{d} P\left(\sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t,\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1+\delta_t)t\right) \\
&\le \sum_{j=1}^{d} P\left(\sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t,\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1-\delta_t)t\right) \\
&\quad + \sum_{j=1}^{d} P\left((1-\delta_t)t < \tilde{w}_j h_q(q_j) \le (1+\delta_t)t\right) \\
&= \sum_{j=1}^{d} B_{1j} + \sum_{j=1}^{d} B_{2j}.
\end{aligned}$$

For $j = 1, \ldots, d$,

$$B_{2j} = P\left((1-\delta_t)t < w_j h_q(q_j) \le (1+\delta_t)t \ \mid\ p_j \le \delta\right) \cdot P(p_j \le \delta).$$

If $p_j \le \delta$, by lemma 2,

$$P\left(\frac{(1-\delta_t)t}{w_j} < h_q(q_j) \le \frac{(1+\delta_t)t}{w_j}\right) = w_j\left[\frac{1}{(1-\delta_t)t} - \frac{1}{(1+\delta_t)t}\right] + O(1/t^3).$$

Hence by letting $\delta_t \to 0$,

$$\sum_{j=1}^{d} B_{2j} = w^*\delta\left[\frac{1}{(1-\delta_t)t} - \frac{1}{(1+\delta_t)t}\right] + o(1/t) = o(1/t). \tag{A.14}$$

Now consider B_{1j}.

$$P\left(\sum_{j=1}^{d}\tilde{w}_j h_q(q_j) > t,\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1-\delta_t)t\right)$$

$$\leq P\left(\sum_{i\neq j}^{d}\tilde{w}_i h_q(q_i) > \delta_t t,\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \leq (1-\delta_t)t\right).$$

As the case of A_t, define $\mathcal{J}_j = \{i \neq j : \sigma_{ij}^2 \geq \sigma_0^2\}$ where $0 < \sigma_0^2 < 1$. For $1 \leq j \leq d$, the cardinality of $\mathcal{J}_j$, $|\mathcal{J}_j| \leq C_0/\sigma_0^2$. Letting $\delta_t = (1 + C_0/\sigma_0^2)t^{-\epsilon}$ for constant $0 < \epsilon < 1$,

$$\begin{aligned}
\left\{\sum_{i\neq j}\tilde{w}_i h_q(q_i) > \delta_t t\right\} &= \left\{\sum_{i\in\mathcal{J}_j}\tilde{w}_i h_q(q_i) + \sum_{i\in\mathcal{J}_j^c}\tilde{w}_i h_q(q_i) > \delta_t t\right\} \\
&\subseteq \left\{\left[\sum_{i\in\mathcal{J}_j}\tilde{w}_i h_q(q_i) > \left(\frac{C_0}{\sigma_0^2}\right)t^{1-\epsilon}\right] \cup \left[\sum_{i\in\mathcal{J}_j^c}\tilde{w}_i h_q(q_i) > \delta_t t - \left(\frac{C_0}{\sigma_0^2}\right)t^{1-\epsilon}\right]\right\} \\
&\subseteq \left\{\left[\bigcup_{i\in\mathcal{J}_j}\left\{\tilde{w}_i h_q(q_i) > \left(\frac{C_0}{\sigma_0^2}\right)\frac{t^{1-\epsilon}}{|\mathcal{J}_j|}\right\}\right] \cup \left[\bigcup_{i\in\mathcal{J}_j^c}\left\{\tilde{w}_i h_q(q_i) > \frac{t^{1-\epsilon}}{|\mathcal{J}_j^c|}\right\}\right]\right\}
\end{aligned}$$

Then we have, since $|\mathcal{J}_j| \leq C_0/\sigma_0^2$ and $|\mathcal{J}_j^c| \leq d$,

$$\begin{aligned}
&P\left(\sum_{i\neq j}^{d}\tilde{w}_i h_q(q_i) > \delta_t t,\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \leq (1-\delta_t)t\right) \\
&\leq \sum_{i\in\mathcal{J}_j} P\left(\frac{t}{d} < \tilde{w}_j h_q(q_j) \leq (1-\delta_t)t,\ \ \tilde{w}_i h_q(q_i) > \left(\frac{C_0}{\sigma_0^2}\right)\frac{t^{1-\epsilon}}{|\mathcal{J}_j|}\right) \\
&\quad + \sum_{i\in\mathcal{J}_j^c} P\left(\frac{t}{d} < \tilde{w}_j h_q(q_j) \leq (1-\delta_t)t,\ \ \tilde{w}_i h_q(q_i) > \frac{t^{1-\epsilon}}{|\mathcal{J}_j^c|}\right) \\
&\leq \sum_{i\in\mathcal{J}_j} P\left(\frac{t}{d} < \tilde{w}_j h_q(q_j) \leq (1-\delta_t)t,\ \ \tilde{w}_i h_q(q_i) > t^{1-\epsilon}\right) \\
&\quad + \sum_{i\in\mathcal{J}_j^c} P\left(\frac{t}{d} < \tilde{w}_j h_q(q_j) \leq (1-\delta_t)t,\ \ \tilde{w}_i h_q(q_i) > \frac{t^{-\epsilon}}{d}\right) \\
&= R_1 + R_2.
\end{aligned}$$

For any $i \in \mathcal{J}_j$, note that

$$\begin{aligned}
&P\left(\frac{t}{d} < \tilde{w}_j h_q(q_j) \leq (1-\delta_t)t,\ \ \tilde{w}_i h_q(q_i) > t^{1-\epsilon}\right) \\
&= P\left(\frac{t}{d} < \tilde{w}_j h_q(q_j) \leq (1-\delta_t)t,\ \ \tilde{w}_i h_q(q_i) > t^{1-\epsilon} | p_j \leq \delta,\ p_i \leq \delta\right) \cdot P(p_j \leq \delta,\ p_i \leq \delta)
\end{aligned}$$

$$\leq P\left(\frac{t}{d} < w_j h_q(q_j) \leq (1-\delta_t)t,\ \ w_i h_q(q_i) > t^{1-\epsilon} | p_j \leq \delta,\ p_i \leq \delta\right). \tag{A.15}$$

Equivalence relation between h_q and h_X implies that, when $p_i \leq \delta$ and $p_j \leq \delta$,

$$(A.15)$$
$$\leq P\left(h_X(X_i) > h_q\left[\delta h_q^{-1}\left(\frac{t^{1-\epsilon}}{w_i}\right)\right],\ h_q\left[\delta h_q^{-1}\left(\frac{t}{w_j d}\right)\right] < h_X(X_j) \leq h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right). \tag{A.16}$$

The bivariate normality assumption can be used to bound (A.16). Specifically, it can be expressed that $X_i = \sigma_{ij} X_j + \gamma_{ij} Z_{ij}$ where $\gamma_{ij}^2 = 1 - \sigma_{ij}^2$ and Z_{ij} follows a standard normal distribution and is independent of X_i. Since we have, in the event (A.16),

$$\begin{aligned}
&\left\{h_X(X_i) > h_q\left[\delta h_q^{-1}\left(\frac{t^{1-\epsilon}}{w_i}\right)\right]\right\}\\
&= \left\{|\sigma_{ij} X_j + \gamma_{ij} Z_{ij}| > h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{t^{1-\epsilon}}{w_i}\right)\right]\right)\right\}\\
&\subseteq \left\{|Z_{ij}| > \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{t^{1-\epsilon}}{w_i}\right)\right]\right) - \sigma_{ij}|X_j|\right]\right\}\\
&\subseteq \left\{|Z_{ij}| > \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{t^{1-\epsilon}}{w_i}\right)\right]\right) - \sigma_{ij} h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right]\right\}.
\end{aligned}$$

By using bounds of the inverse function h_X^{-1}, (A.9),

$$\begin{aligned}
&h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{t^{1-\epsilon}}{w_i}\right)\right]\right) - \sigma_{ij} h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\\
&\leq \sqrt{2\log(h_q(\delta h_q^{-1}(t^{1-\epsilon}/w_i)))} - (1+o(1))\sigma_{ij}\sqrt{2\log(h_q(\delta h_q^{-1}((1-\delta_t)t/w_j)))}\\
&\leq \{\kappa_t - (1+o(1))\sigma_{ij}\} \cdot \sqrt{2\log(h_q(\delta h_q^{-1}((1-\delta_t)t/w_j)))},
\end{aligned}$$

where κ_t is a constant depending t and δ. Specifically, as t goes to infinity,

$$\frac{\sqrt{2\log(h_q(\delta h_q^{-1}(t^{1-\epsilon}/w_i)))}}{\sqrt{2\log(h_q(\delta h_q^{-1}((1-\delta_t)t/w_j)))}} \to 0,$$

which implies that there is a constant κ_t such that for sufficiently large t,

$$\frac{\sqrt{2\log(h_q(\delta h_q^{-1}(t^{1-\epsilon}/w_i)))}}{\sqrt{2\log(h_q(\delta h_q^{-1}((1-\delta_t)t/w_j)))}} \le \kappa_t \quad \text{and} \quad \kappa_t > (1+o(1))\sigma_{ij}.$$

Hence we have that

$$P\left(|Z_{ij}| > \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{t^{1-\epsilon}}{w_i}\right)\right]\right) - \sigma_{ij} h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right]\right) = o(1).$$

Then (A.16) can be bounded as follows.

$$\begin{aligned}
&(A.16)\\
&\le P\left(h_q\left[\delta h_q^{-1}\left(\frac{t}{w_j d}\right)\right] < h_X(X_j)\right)\\
&\quad \times P\left(|Z_{ij}| > \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{t^{1-\epsilon}}{w_i}\right)\right]\right) - \sigma_{ij} h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right]\right)\\
&\le \left(\frac{w_j d}{t\pi} + O(1/t^3)\right) \cdot o(1).
\end{aligned}$$

Therefore we have

$$R_1 \le \sum_{i \in \mathcal{J}_j} \left(\frac{w_j d}{t\pi} + O(1/t^3)\right) \cdot o(1) \le \max_{i \in \mathcal{J}_j} \left\{\left(\frac{C_0}{\sigma_0^2}\right)\left(\frac{w_j d}{t\pi} + O(1/t^3)\right)\right\} \cdot o(1) = o(1/t).$$

As the same way, we can obtain $R_2 \le o(1/t)$. With (A.14), we can conclude that $P(B) = o(1/t)$ and from (A.13), $P(A) = \frac{w^*\delta}{t\pi} + o(1/t)$ so that

$$\lim_{d\to\infty} \frac{P(T_X^*(X) > t)}{P(C_{(0,w^*\delta)} > t)} = 1.$$

A.2 Proof of Theorem 3

In the proof of theorem 4, it has been shown that for given $\delta \in [0,1]$,

$$P\left(\sum_{j\in\mathcal{M}} w_j h_q(q_j) > t\right) = P\left(\sum_{j=1}^{d} w_j I(p_j \le \delta) h_q(q_j) > t\right) =: P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t\right),$$

where $\tilde{w}_j = w_j I(p_j \leq \delta)$. We have shown that it is decomposed as follows.

$$P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t\right) = P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t,\ \ \bigcup_{j=1}^{d} \{\tilde{w}_j h_q(q_j) > (1+\delta_t)t\}\right) \quad \text{(A.17)}$$

$$+ P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t,\ \ \bigcap_{j=1}^{d} \{\tilde{w}_j h_q(q_j) \leq (1+\delta_t)t\}\right), \quad \text{(A.18)}$$

$$=: P(A) + P(B),$$

where δ_t is a constant only depends on t satisfying $\delta_t > 0$, $\delta_t \to 0$ and $\delta_t t \to \infty$ as $t \to \infty$.

For the $P(A)$, by using the bonferroni inequality, if we let

$$A = \bigcup_{j=1}^{d} A_j = \bigcup_{j=1}^{d} \left\{\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t\right\},$$

we have

$$\sum_{j=1}^{d} P(A_j) - \sum_{1 \leq i < j \leq d} P(A_i \cap A_j) \leq P\left(\bigcup_{j=1}^{d} A_j\right) \leq \sum_{j=1}^{d} P(A_j).$$

First consider $P(A_j)$. It can be decomposed as follows.

$$P(A_j) = P(\tilde{w}_j h_q(q_j) > (1+\delta_t)t) - P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) \leq t\right).$$

And we proved that

$$P(\tilde{w}_j h_q(q_j) > (1+\delta_t)t) = \frac{w_j \delta}{(1+\delta_t)t\pi} + o(1/t).$$

To conclude that $P(A_j) = (w_j\delta)/((1+\delta_t)t\pi) + o(1/t)$, it suffices to show that $P(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) \leq t) = o(1/t)$. Note that

$$P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) \leq t\right) \leq P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \sum_{i \neq j} \tilde{w}_i h_q(q_i) \leq -\delta_t t\right).$$

Note that since

$$\left\{\sum_{i \neq j} \tilde{w}_i h_q(q_i) \leq -\delta_t t\right\} \subseteq \bigcup_{i \neq j} \left\{\tilde{w}_i h_q(q_i) \leq -\frac{\delta_t t}{d-1}\right\},$$

we have,

$$P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\quad \sum_{i\neq j}\tilde{w}_i h_q(q_i) \leq -\delta_t t\right)$$
$$\leq \sum_{i\neq j} P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t, \tilde{w}_i h_q(q_i) \leq -\frac{\delta_t t}{d-1}\right).$$

It can be shown that

$$\begin{aligned}
&P\left(\tilde{w}_j h_q(q_j) > (1+\delta_t)t,\ \ \tilde{w}_i h_q(q_i) \leq -\frac{\delta_t t}{d-1}\right)\\
&= P\left(w_j h_q(q_j) > (1+\delta_t)t,\ \ w_i h_q(q_i) \leq -\frac{\delta_t t}{d-1} \mid p_j \leq \delta,\ p_i \leq \delta\right)\cdot p(p_j \leq \delta,\ p_i \leq \delta)\\
&\leq P\left(w_j h_q(q_j) > (1+\delta_t)t,\ \ w_i h_q(q_i) \leq -\frac{\delta_t t}{d-1} \mid p_j \leq \delta,\ p_i \leq \delta\right).
\end{aligned}$$

The equivalence between h_q and h_X yields that when $p_j \leq \delta$ and $p_i \leq \delta$,

$$\begin{aligned}
&P\left(w_j h_q(q_j) > (1+\delta_t)t,\ \ w_i h_q(q_i) \leq -\frac{\delta_t t}{d-1}\right)\\
&= P\left(p_j < \delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right),\ \ p_i \geq \delta h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right)\right)\\
&= P\left(h_q(p_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\ \ h_q(p_i) \leq h_q\left(\delta h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right)\right)\right)\\
&= P\left(h_X(X_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\ \ h_X(X_i) \leq h_q\left(\delta h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right)\right)\right).
\end{aligned} \tag{A.19}$$

The bivariate normality assumption of X implies that $X_i = \sigma_{ij}X_j + \gamma_{ij}Z_{ij}$, where $\gamma_{ij}^2 = 1 - \sigma_{ij}^2$ and Z_{ij} is a random variable following a standard normal distribution and independent of X_j.

$$\begin{aligned}
\left\{h_X(X_i) \leq h_q\left(\delta h_q^{-1}\left(\tfrac{-\delta_t t}{(d-1)w_i}\right)\right)\right\} &= \left\{h_X(\sigma_{ij}X_j + \gamma_{ij}Z_{ij}) \leq h_q\left(\delta h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right)\right)\right\}\\
&= \left\{Z_{ij} \leq \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left[h_q\left(\delta h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right)\right)\right] - \sigma_{ij}X_j\right]\right\}.
\end{aligned}$$

By the monitonicities of h_X and h_q, the X_j in the event (A.19) can be lower bounded as follows.

$$X_j > h_X^{-1}\left[h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right)\right] \geq h_X^{-1}\left[h_q\left(h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right)\right] = h_X^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right).$$

Hence we have,

$$\begin{aligned}
&\left\{h_X(X_i) \leq h_q\left(\delta h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right)\right)\right\} \\
&\subseteq \left\{Z_{ij} \leq \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left[h_q\left(\delta h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right)\right)\right] - \sigma_{ij} h_X^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right]\right\}. \qquad \text{(A.20)}
\end{aligned}$$

As $\delta_t t \to \infty$, it can be shown that

$$h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right) \to 1,$$

so that for some $\check{\delta}$,

$$h_q\left(\delta h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right)\right) \leq h_q(\check{\delta}).$$

On the other hand, as $\delta_t t \to \infty$, $(1+\delta_t)t/w_j \to \infty$ implies that

$$h_X^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right) \to \infty.$$

Therefore the probability of the event in (A.20) can be bounded by $o(1)$. By the bivariate normality assumption, (A.19) can be bounded as follows.

$$\begin{aligned}
&P\left(h_X(X_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\ h_X(X_i) \leq h_q\left(\delta h_q^{-1}\left(\frac{-\delta_t t}{(d-1)w_i}\right)\right)\right) \\
&\leq P\left(h_q(p_j) > h_q\left(\delta h_q^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right),\ Z_{ij} \leq \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left[h_q(\check{\delta})\right] - \sigma_{ij} h_X^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right]\right) \\
&= P\left(h_q(q_j) > \frac{(1+\delta_t)t}{w_j}\right) \cdot P\left(Z_{ij} \leq \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left[h_q(\check{\delta})\right] - \sigma_{ij} h_X^{-1}\left(\frac{(1+\delta_t)t}{w_j}\right)\right]\right) \\
&= o(1/t) \cdot o(1) \\
&= o(1/t).
\end{aligned}$$

Now, we can conclude that

$$\sum_{j=1}^{d} P(A_j) = \frac{w^*\delta}{t\pi} + o(1/t),$$

where $w^* = \sum_{j=1}^{d} w_j$. By using a similar stpes, it can be shown that

$$\sum_{1\le i<j\le d} P(A_i \cap A_j) = o(1/t).$$

Therefore

$$P(A) = \frac{w^*\delta}{t\pi} + o(1/t).$$

It remains to show that $P(B) = o(1/t)$. Recall that, in (A.18),

$$B = \left\{ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \quad \bigcap_{j=1}^{d} \{\tilde{w}_j h_q(q_j) \le (1+\delta_t)t\} \right\}.$$

As in the proof of theorem 4, the event B can be expressed as follows.

$$\begin{aligned}
B &= \bigcap_{j=1}^{d} \left\{ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \ \ \tilde{w}_j h_q(q_j) \le (1+\delta_t)t \right\} \\
&\subseteq \bigcap_{j=1}^{d} \left\{ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \ \ \tilde{w}_j h_q(q_j) \le (1+\delta_t)t, \ \ \bigcup_{j=1}^{d} \left\{ \tilde{w}_j h_q(q_j) > \frac{t}{d} \right\} \right\} \\
&= \bigcup_{j=1}^{d} \left\{ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \ \ \tilde{w}_j h_q(q_j) \le (1+\delta_t)t, \ \ \tilde{w}_j h_q(q_j) > \frac{t}{d} \right\} \\
&= \bigcup_{j=1}^{d} \left\{ \sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1+\delta_t)t \right\}.
\end{aligned}$$

Then we have

$$\begin{aligned}
P(B) &\le \sum_{j=1}^{d} P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1+\delta_t)t \right) \\
&\le \sum_{j=1}^{d} P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t, \ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1-\delta_t)t \right) \\
&\quad + \sum_{j=1}^{d} P\left((1-\delta_t)t < \tilde{w}_j h_q(q_j) \le (1+\delta_t)t\right)
\end{aligned}$$

$$= \sum_{j=1}^{d} B_{1j} + \sum_{j=1}^{d} B_{2j}.$$

We have shown that

$$\sum_{j=1}^{d} B_{2j} = o(1/t)$$

in the proof of theorem 4, and it does not depend on d. It means that for fixed d, it also can be shown

$$\sum_{j=1}^{d} B_{2j} = o(1/t).$$

Now consider B_{1j}.

$$\begin{aligned}
& P\left(\sum_{j=1}^{d} \tilde{w}_j h_q(q_j) > t,\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1-\delta_t)t\right) \\
& \le P\left(\sum_{i\neq j}^{d} \tilde{w}_i h_q(q_i) > \delta_t t,\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1-\delta_t)t\right) \\
& \le P\left(\bigcup_{i\neq j}^{d} \left\{\tilde{w}_i h_q(q_i) > \frac{\delta_t t}{d-1},\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1-\delta_t)t\right\}\right) \\
& \le \sum_{i\neq j}^{d} P\left(\tilde{w}_i h_q(q_i) > \frac{\delta_t t}{d-1},\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1-\delta_t)t\right) \\
& = \sum_{i\neq j}^{d} P_{1,ij}.
\end{aligned}$$

Note that

$$\begin{aligned}
P_{1,ij} &= P\left(\tilde{w}_i h_q(q_i) > \frac{\delta_t t}{d-1},\ \ \frac{t}{d} < \tilde{w}_j h_q(q_j) \le (1-\delta_t)t \,\middle|\, p_j \le \delta,\ p_i \le \delta\right) \cdot P(p_j \le \delta,\ p_i \le \delta) \\
&\le P\left(w_i h_q(q_i) > \frac{\delta_t t}{d-1},\ \ \frac{t}{d} < w_j h_q(q_j) \le (1-\delta_t)t \,\middle|\, p_j \le \delta,\ p_i \le \delta\right).
\end{aligned}$$

Now we consider only the case that $p_i \le \delta$ and $p_j \le \delta$ for $1 \le i \neq j \le d$. The equivalence relation between h_q and h_X implies that,

$P_{1,ij}$

$$\leq P\left(h_X(X_i) > h_q\left[\delta h_q^{-1}\left(\tfrac{\delta_t t}{w_i(d-1)}\right)\right],\ \ h_q\left[\delta h_q^{-1}\left(\tfrac{t}{w_j d}\right)\right] < h_X(X_j) \leq h_q\left[\delta h_q^{-1}\left(\tfrac{(1-\delta_t)t}{w_j}\right)\right]\right). \tag{A.21}$$

By the assumption of bivariate normality, $X_i = \sigma_{ij}X_j + \gamma_{ij}Z_{ij}$ where $\sigma_{ij}^2 + \gamma_{ij}^2 = 1$ and Z_{ij} follows a standard normal distribution and is independent of X_i. If $\sigma_{ij} = 0$, then X_i and X_j are independent so that

$$\begin{aligned}
\text{(A.21)} \quad &\leq \quad P\left(h_X(X_i) > h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right)\cdot P\left(h_q\left[\delta h_q^{-1}\left(\frac{t}{w_j d}\right)\right] < h_X(X_j)\right) \\
&= \quad P\left(h_q(q_i) > \frac{\delta_t t}{w_i(d-1)}\right)\cdot P\left(h_q(q_j) > \frac{t}{w_j d}\right) \\
&= \quad o(1/t).
\end{aligned}$$

Hence, suppose that $0 < |\sigma_{ij}| < 1$. In the event in (A.21),

$$\begin{aligned}
&\left\{h_X(X_i) > h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right\} \\
&= \left\{|\sigma_{ij}X_j + \gamma_{ij}Z_{ij}| > h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right)\right\} \\
&\subseteq \left\{|Z_{ij}| > \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right) - \sigma_{ij}|X_j|\right]\right\} \\
&\subseteq \left\{|Z_{ij}| > \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right) - \sigma_{ij}h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right]\right\}.
\end{aligned} \tag{A.22}$$

As $t \to \infty$, by choosing δ_t such that $\delta_t t \to \infty$, since

$$h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right) \to 0,$$

we have

$$h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right) \to \infty.$$

By lemma 4 below, as $t \to \infty$,

$$
\begin{aligned}
h_X\left(\sigma_{ij} h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right) &\le \frac{h_X\left(h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right)}{c_a\left(h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right)^2} \\
&= \frac{h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]}{c_a\left(h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right)^2}.
\end{aligned}
$$

We can consider an additional condition of δ_t in addition to $\delta_t t \to \infty$ such that

$$
\frac{h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]}{c_a\left(h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right)^2 h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]} \to 0,
$$

which implies that

$$
h_X\left(\sigma_{ij} h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{(1-\delta_t)t}{w_j}\right)\right]\right)\right) \le o\left(h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right).
$$

Then we have that

$$
\text{(A.22)} \subseteq \left\{|Z_{ij}| > \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right) - h_X^{-1}\left(o\left(h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right)\right)\right]\right\}.
$$

Therefore we can conclude that

$$
\begin{aligned}
&P_{1,ij} \\
&\le P\left(h_X(X_i) > h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right) \\
&\quad \times P\left(|Z_{ij}| > \frac{1}{\gamma_{ij}}\left[h_X^{-1}\left(h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right) - h_X^{-1}\left(o\left(h_q\left[\delta h_q^{-1}\left(\frac{\delta_t t}{w_i(d-1)}\right)\right]\right)\right)\right]\right) \\
&= o(1/t).
\end{aligned}
$$

Hence we can obtain that

$$
\sum_{j=1}^{d} B_{1j} = \sum_{j=1}^{d}\sum_{i\neq j}^{d} P_{1,ij} = o(1/t),
$$

which implies that

$$P(B) = o(1/t),$$

and the theorem is proved.

Lemma 4 (Liu and Xie, 2020)**.** *For any constant* $0 < |a| < 1$*,*

$$\lim_{x\to\infty} \frac{h_X(x)}{x^2 h_X(ax)} > c_a > 0,$$

where c_a *is a constant only depend on* a*.*

A.3 Proof of Theorem 5

Let $\mathcal{N} = \mathcal{S}^c$ be a set of null hypotheses.

$$\begin{aligned} T_q^*(q) &= \sum_{j\in\mathcal{M}} w_j h_q(q_j) \\ &= \sum_{j\in\mathcal{M}\cap\mathcal{S}} w_j h_q(q_j) + \sum_{j\in\mathcal{M}\cap\mathcal{N}} w_j h_q(q_j) \\ &= \sum_{j\in\mathcal{M}\cap\mathcal{S}} w_j h_q(q_j) + \sum_{j\in\mathcal{M}\cap\mathcal{N}} w_j h_q(q_j). \end{aligned}$$

It suffices to show that under H_1,

$$\sum_{j\in\mathcal{M}\cap\mathcal{S}} w_j h_q(q_j) \to \infty \tag{A.23}$$

and

$$\sum_{j\in\mathcal{M}\cap\mathcal{N}} w_j h_q(q_j) > -\infty. \tag{A.24}$$

By lemma 5, it can be expressed that

$$\sum_{j\in\mathcal{M}\cap\mathcal{S}} w_j h_q(q_j) = \sum_{j\in\mathcal{M}\cap\mathcal{S}} w_j^* h_q(p_j) = \sum_{j\in\mathcal{M}\cap\mathcal{S}} w_j^* h_X(X_j)$$

and

$$\sum_{j\in\mathcal{M}\cap\mathcal{N}} w_j h_q(q_j) = \sum_{j\in\mathcal{M}\cap\mathcal{N}} w_j^* h_q(p_j) = \sum_{j\in\mathcal{M}\cap\mathcal{N}} w_j^* h_X(X_j).$$

It follows from lemma 6 that (A.24) holds.

For $j \in \mathcal{M}$, $p_j < \delta$ implies that

$$|X_j| > 1 - \frac{\delta}{2}.$$

Using the fact, we have

$$\sum_{j\in\mathcal{M}\cap\mathcal{S}} w_j^* h_X(X_j) \geq w_{j'}^* h_X\left(\max|X_j|\right) + (|\mathcal{M}\cap\mathcal{S}| - 1) w_{i'}^* h_X(\min|X_j|),$$

where j' is an index with $|X_{j'}| = \max_{j\in\mathcal{M}\cap\mathcal{S}}|X_j|$ and i' is an index with $|X_{i'}| = \min_{j\in\mathcal{M}\cap\mathcal{S}}|X_j|$.

Lemma 5 (Mean Value Theorem)**.**

$$T_q^*(q) = \sum_{j=1}^{m} w_j^* h(p_j),$$

where for constants $\delta < \delta_j^* < 1$, $j = 1, \ldots, m$ *and* $1 \leq c \leq 8$,

$$w_j^* = \frac{1}{m}\left[1 - \frac{c(1-\delta)\delta}{(\delta_j^*)^2}\right].$$

Proof.

$$T_X^*(X) = T_p^*(q) = \frac{1}{m}\sum_{j=1}^{m} h_p(q_j)$$

Note that by the mean value theorem, there are constants δ_j^* for $j = 1, \ldots, m$, with $\delta < \delta_j^* < 1$ such that

$$h'(q_j^*) = \frac{h(q_j) - h(p_j)}{q_j - p_j} = \frac{h(q_j) - h(p_j)}{(1-\delta)q_j},$$

so that $h(q_j) = h(p_j) + (1-\delta)q_j h'(q_j^*)$. We can find δ_j^* satisfying $q_j^* = \delta_j^* q_j$. Since

$$h(\delta_j^* q_j) = \frac{\sin((1/2 - \delta_j^* q_j)\pi)}{\cos((1/2 - \delta_j^* q_j)\pi)}$$

and its derivative implies that

$$h'(\delta_j^* q_j) = -\frac{\pi}{\cos^2((1/2 - \delta_j^* q_j)\pi)},$$

we obatin

$$T_p^*(q) = \frac{1}{m}\sum_{j=1}^{m} h_p(p_j) - \sum_{j=1}^{m} \frac{(1-\delta)q_j\pi}{m\sin^2(\delta_j^* q_j \pi)}.$$

Also specific values of δ_j^*, $j = 1, \dots, m$ can be calculated as follows.

$$\delta_j^* = \frac{1}{q_j\pi}\arcsin\left(\sqrt{\frac{\pi(q_j - p_j)}{h(p_j) - h(q_j)}}\right) = \frac{1}{q_j\pi}\arcsin\left(\sqrt{\frac{\pi(p_j/\delta - p_j)}{h(p_j) - h(p_j/\delta)}}\right).$$

Here we use notations T^* and h instead of T_p^* and h_p for notational convenience. By using properties of trigonometric functions,

$$\begin{aligned}
T^*(q) &= \frac{1}{m}\sum_{j\in M} h(q_j) \\
&= \frac{1}{m}\sum_{j\in M} h(p_j) - \frac{1}{m}\sum_{j\in M}\frac{(1-\delta)q_j\pi}{\sin^2(\delta_j^* q_j\pi)} \\
&\le \frac{1}{m}\sum_{j\in M} h(p_j) - \frac{1}{m}\sum_{j\in M}\frac{(1-\delta)q_j\pi}{(\delta_j^* q_j\pi)^2} \\
&= \frac{1}{m}\sum_{j\in M} h(p_j) - \frac{1}{m}\sum_{j\in M}\frac{1-\delta}{(\delta_j^*)^2 q_j\pi} \\
&= \frac{1}{m}\sum_{j\in M} h(p_j) - \frac{1}{m}\sum_{j\in M}\frac{(1-\delta)\delta}{(\delta_j^*)^2 p_j\pi} \\
&\le \frac{1}{m}\sum_{j\in M} h(p_j) - \frac{1}{m}\sum_{j\in M}\frac{(1-\delta)\delta}{(\delta_j^*)^2 \tan(p_j\pi)} \\
&= \frac{1}{m}\sum_{j\in M} h(p_j) - \frac{1}{m}\sum_{j\in M}\frac{(1-\delta)\delta}{(\delta_j^*)^2} h(p_j)
\end{aligned}$$

$$= \frac{1}{m}\sum_{j\in M}\left[1-\frac{(1-\delta)\delta}{(\delta_j^*)^2}\right]h(p_j)$$

The lower bound of $T^*(q)$ can be obtained similary as follows.

$$\begin{aligned}
T^*(q) &= \frac{1}{m}\sum_{j\in M}h(p_j)-\frac{1}{m}\sum_{j\in M}\frac{(1-\delta)q_j\pi}{\sin^2(\delta_j^* q_j\pi)}\\
&\geq \frac{1}{m}\sum_{j\in M}h(p_j)-\frac{1}{m}\sum_{j\in M}\frac{(1-\delta)q_j\pi}{\tan(\delta_j^* q_j\pi/2)\sin(\delta_j^* q_j\pi)}\\
&\geq \frac{1}{m}\sum_{j\in M}h(p_j)-\frac{1}{m}\sum_{j\in M}\frac{2(1-\delta)q_j\pi}{(\delta_j^* q_j\pi)\sin(\delta_j^* q_j\pi)}\\
&= \frac{1}{m}\sum_{j\in M}h(p_j)-\frac{1}{m}\sum_{j\in M}\frac{2(1-\delta)}{\delta_j^*\sin(\delta_j^* q_j\pi)}\\
&\geq \frac{1}{m}\sum_{j\in M}h(p_j)-\frac{1}{m}\sum_{j\in M}\frac{4(1-\delta)}{\delta_j^*(\delta_j^* q_j\pi)}\\
&= \frac{1}{m}\sum_{j\in M}h(p_j)-\frac{1}{m}\sum_{j\in M}\frac{4(1-\delta)\delta}{(\delta_j^*)^2 p_j\pi}\\
&= \frac{1}{m}\sum_{j\in M}h(p_j)-\frac{1}{m}\sum_{j\in M}\frac{8(1-\delta)\delta}{(\delta_j^*)^2\sin(2p_j\pi)}\\
&\geq \frac{1}{m}\sum_{j\in M}h(p_j)-\frac{1}{m}\sum_{j\in M}\frac{8(1-\delta)\delta}{(\delta_j^*)^2\tan(p_j\pi)}\\
&= \frac{1}{m}\sum_{j\in M}h(p_j)-\frac{1}{m}\sum_{j\in M}\frac{8(1-\delta)\delta}{(\delta_j^*)^2}h(p_j)\\
&= \frac{1}{m}\sum_{j\in M}\left[1-\frac{8(1-\delta)\delta}{(\delta_j^*)^2}\right]h(p_j)
\end{aligned}$$

Hence we obtain

$$\sum_{j=1}^{m}\left(\frac{1}{m}-\frac{8(1-\delta)\delta}{m(\delta_j^*)^2}\right)h(p_j)\leq T^*(q)\leq\sum_{j=1}^{m}\left(\frac{1}{m}-\frac{(1-\delta)\delta}{m(\delta_j^*)^2}\right)h(p_j).$$

$\square$

Lemma 6. *Let $\mathcal{S}=\{i\in[d]:\mu_i\neq 0\}$ with $|\mathcal{S}|=d^\gamma$ for $\gamma\in(0,1/2)$. Let $\mathcal{N}=\mathcal{S}^c$. Then with probability tending to 1,*

$$\sum_{j\in\mathcal{M}\cap\mathcal{N}} w_j^* h_X(X_j) > -\infty.$$

Proof. Note that $|\mathcal{N}| = d - d^\gamma$.

$$\sum_{j\in\mathcal{M}\cap\mathcal{N}} w_j^* h(X_j) \geq |\mathcal{M}\cap\mathcal{N}| \cdot w_{i'}^* h_X \left(\min_{j\in\mathcal{M}\cap\mathcal{N}} |X_j|\right),$$

where i'' is an index with $|X_{i''}| = \min_{j\in\mathcal{M}\cap\mathcal{N}}|X_j|$. To obtain the lower bound, we have

$$\begin{aligned}
P\left(\min_{i\in S^c}|X_i| < \epsilon_d\right) &\leq \sum_{i\in S^c} P(|X_i| < \epsilon_d) \\
&= (d - d^\gamma) P(|X_i| < \epsilon_d) \\
&= (d - d^\gamma)\left\{\Phi(\epsilon_d) - \Phi(-\epsilon_d)\right\} \\
&\leq 2\phi(-\epsilon_d)(d - d^\gamma)\epsilon_d \\
&\leq 2\phi(0) d^\gamma \epsilon_d \\
&\leq d^\gamma \epsilon_d.
\end{aligned}$$

By letting $\epsilon_d = d^{\gamma_0}$ with $\gamma_0 < -1/2$,

$$P\left(\min_{i\in S^c}|X_i| < \epsilon_d\right) = o(1).$$

Hence with probability tending to 1,

$$h\left(\min_{i\in S^c}|X_i|\right) > -\infty,$$

which implies

$$|\mathcal{M}\cap\mathcal{N}| \cdot w_{i'}^* h_X \left(\min_{j\in\mathcal{M}\cap\mathcal{N}} |X_j|\right) > -\infty.$$

□

Appendix B

Supplementary Analysis

B.1 t-Distribution

To investigate the null distribution and the power of the proposed method when the normality assumption is violated, suppose that $X \sim t_d(\mu, \Sigma)$. The vector X following the multivariate t-distribution is defined as follows.

$$X - \mu := \frac{Y}{\sqrt{\chi_d^2/d}},$$

where $Y \sim N_d(0, \Sigma)$ and χ_d^2 is a random variable following χ^2 distribution with d degree of freedom. For the multivariate distribution, corresponding p-values are defined as follows.

$$p_i = 2[1 - T_d(|X_i|)], \quad i = 1, \ldots, d,$$

where T_d is the cumulative distribution function of t-distribution.

Figure 14 presents the type I error of the proposed method and the Cauchy combination test where Σ is a compound symmetry with $d = 300$. Similar to Figure 7, the type I error of the proposed method is well controlled, although the Cauchy combination test fails to control the type I error. The pattern of boxplots that the type I error of the Cauchy combination test increases as the correlation increases is similar to that of figure 7.

Figure 15 shows the power of the proposed method compared to other global testing procedures. Under the t-distribution assumption, the power of the proposed method is also similar to that under the normal distribution assumption. The proposed method has the comparable power compared to all other methods.

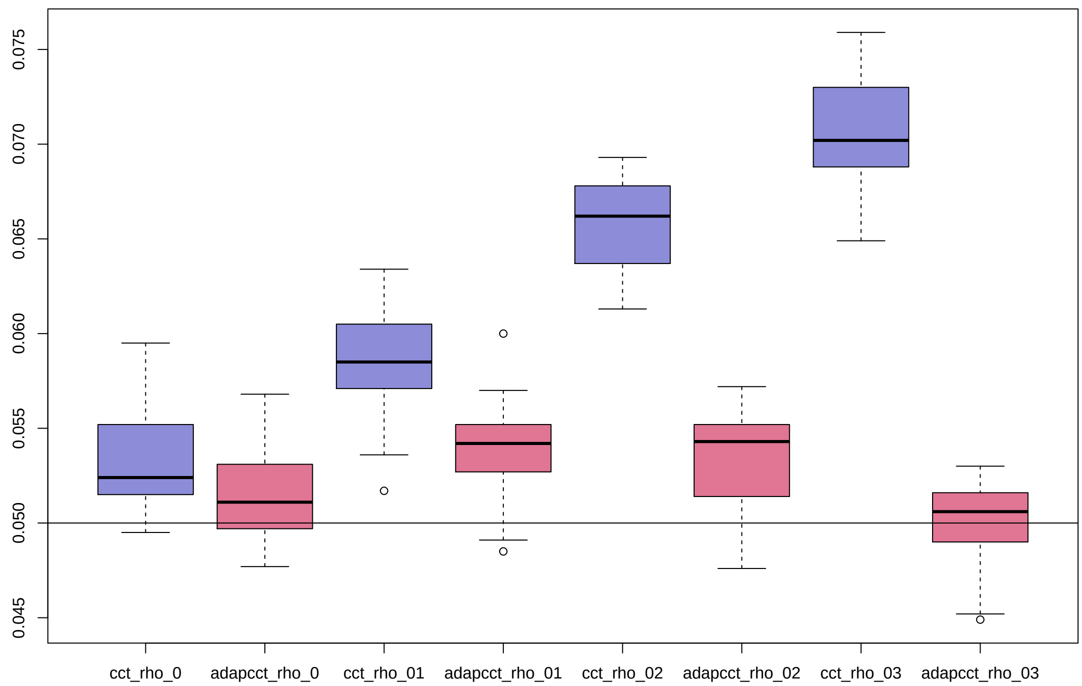

Figure 14: Box plots of type I error at level 0.05 for dimension $d = 300$ and CS with correlation range 0 to 0.5. Red boxes are Adaptive thresholding CCT with thresholding 0.1 and blue boxes are CCT.

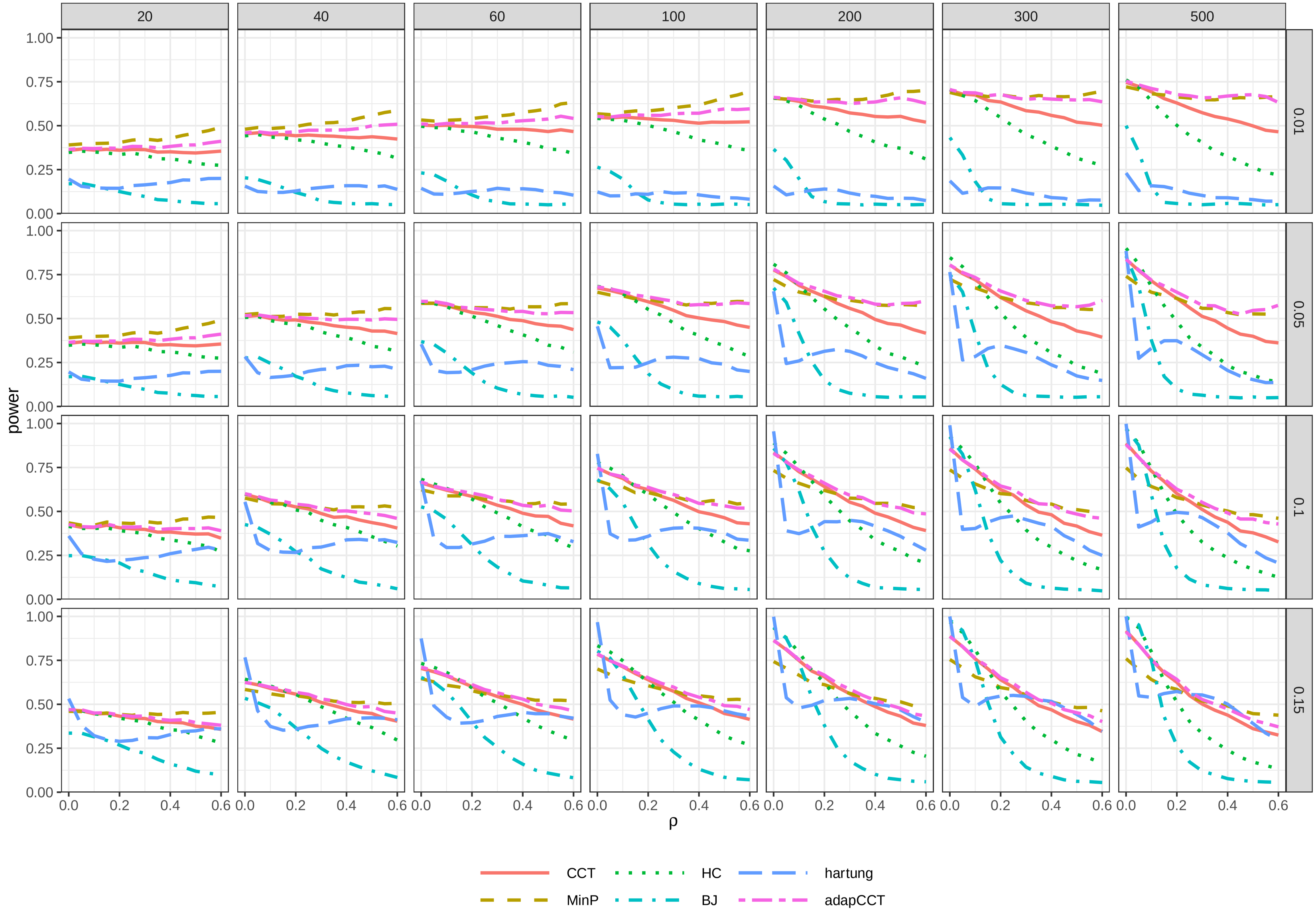

Figure 15: Power comparison of Adaptive thresholding CCT, MinP, HC, BJ, Hartung and CCT. The x-axis is the correlation strength of the compound symmetry structure. The columns from left to right correspond to the dimension $d = 20, 40, 60, 100, 200, 300, 500$. The rows from top to bottom correspond to the signal proportion $5\%, 10\%, 15\%$.

B.2 Comparison with Hartung's method

Recall the weighted inverse normal statistic proposed by Hartung (1999), which is defined in (2.10),

$$t(\hat{\rho}^*, \kappa) = \frac{\sum_{i=1}^d \lambda_i X_i}{\sqrt{\sum_{i=1}^d \lambda_i^2 + \left[\left(\sum_{i=1}^d \lambda_i\right)^2 - \sum_{i=1}^d \lambda_i^2\right] \cdot \left\{\hat{\rho}^* + \kappa \cdot \sqrt{\frac{2}{d+1}}(1-\hat{\rho}^*)\right\}}},$$

where $\hat{\rho}^* = \max\{-1/(d-1), \hat{\rho}\}$, $\kappa > 0$ and

$$\hat{\rho} = 1 - \frac{1}{d-1}\sum_{i=1}^d \left(X_i - \frac{1}{d}\sum_{i=1}^d X_i\right)^2,$$

which is an unbiased estimator of ρ. For the compound symmetry dependence structure, we mentioned that $t(\hat{\rho}^*, \kappa)$ follows the standard normal distribution approximately, under the null.

Demetrescu et al. (2006) extended Hartung's method to allow for a relaxed type of correlation matrix. They showed that if the average of elements of the correlation matrix converges to a constant and deviations from the convergence value are small, Hartung's method also can be applied. Indeed, for $\Sigma = (\mathsf{Cov}(X_i, X_j))_{i,j=1,\ldots,d} = (\rho_{ij})_{i,j=1,\ldots,d}$, assume that

$$\lim_{d\to\infty} \frac{1}{d(d-1)} \sum\sum_{i\neq j} \rho_{ij} = \tilde{\rho}, \tag{B.1}$$

where $\tilde{\rho} \in (0,1)$ and that

$$\lim_{d\to\infty} \frac{1}{d(d-1)} \sum\sum_{i\neq j} (\rho_{ij} - \tilde{\rho})^2 = 0. \tag{B.2}$$

Then if $\lambda_i = \lambda$ for all $i = 1, \ldots, d$ in $t(\hat{\rho}^*, \kappa)$, then $t(\hat{\rho}^*, \kappa)$ is still approximately standard normally distributed.

Figure 16 shows the type I errors of the Cauchy combination test, Hartung's method and the proposed method at confidence level 0.05 for $d = 300$. The correlation coefficient, ρ, of the compound symmetry structure varies from 0 to 0.4. Unlike the Cauchy combination test, Hartung's method is expected to work well for an highly correlated case than the weak dependent case in that it uses the estimated dependence structure. Figure 16 supports this expectation. As ρ increases, Hartung's method control the type I error well, although it fails to control the type I error for small ρ.

Figures 17, 18 and 19 show comparisons of the power of Hartung's method and other methods, under the compound symmetry, polynomially decaying and exponentially decaying dependence structures, respectively. For the compound symmetry structure, the power of Hartung's method is robust to the correlation coefficient. On the other hand, as it uses all data, Hartung's method is powerful on the dense signals setting, although as the sparsity of signals increases, it is less powerful than other methods. By definition, the polynomially and exponentially decaying dependence structures cannot satisfy two conditions, (B.1) and (B.2). So it is expected that Hartung's method cannot detect signals in two dependence structures and Figures 18 and 19 support this speculation.

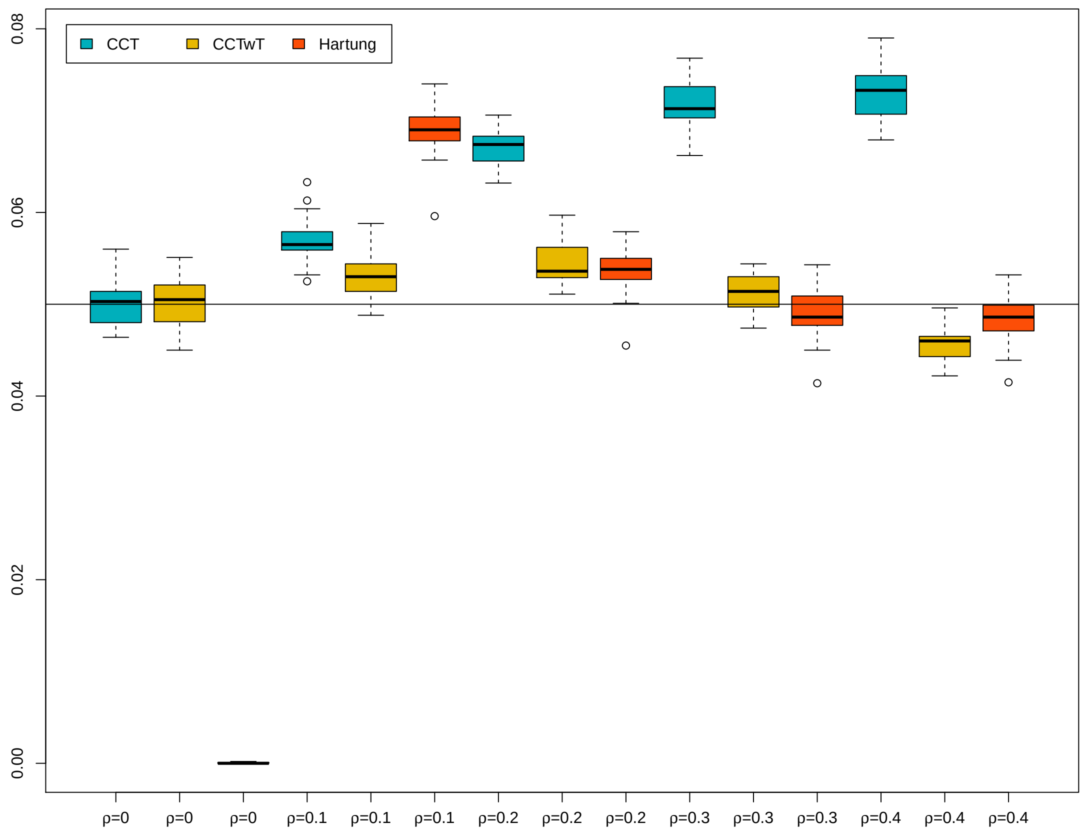

Figure 16: Box plots of type I error at level 0.05 for dimension $d = 300$ and CS with correlation range 0 to 0.5. Yellow boxes are Adaptive thresholding CCT with thresholding 0.1 and blue boxes are CCT. Red boxes indicate Hartung's method

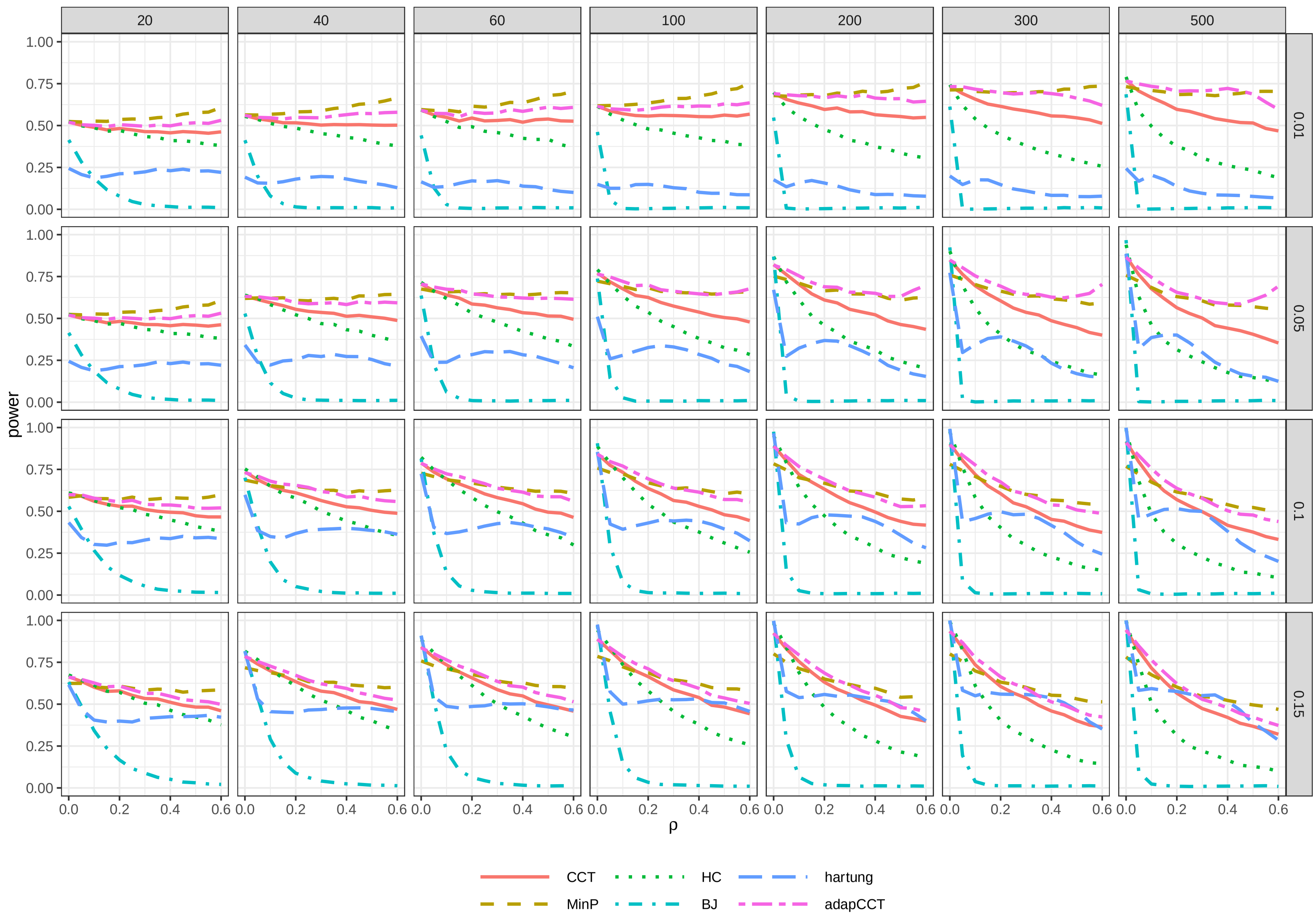

Figure 17: Power comparison of Adaptive thresholding CCT, MinP, HC, BJ, Hartung and CCT. The x-axis is the correlation strength of the compound symmetry structure. The columns from left to right correspond to the dimension $d = 20, 40, 60, 100, 200, 300, 500$. The rows from top to bottom correspond to the signal proportion $5\%, 10\%, 15\%$.

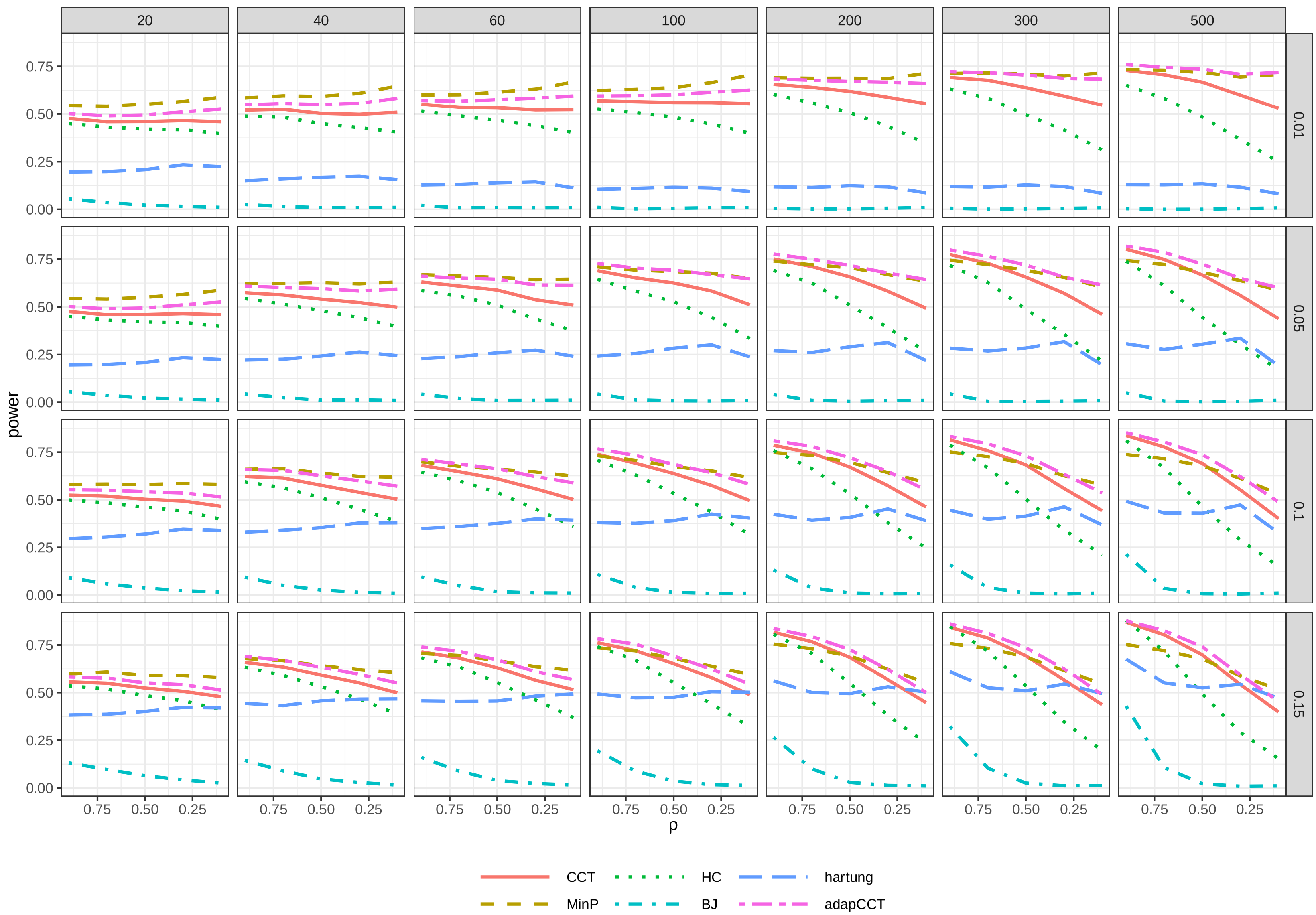

Figure 18: Power comparison of Adaptive thresholding CCT, MinP, HC, BJ, Hartung and CCT. The x-axis is the correlation strength of the compound symmetry structure. The columns from left to right correspond to the dimension $d = 20, 40, 60, 100, 200, 300, 500$. The rows from top to bottom correspond to the signal proportion $5\%, 10\%, 15\%$.

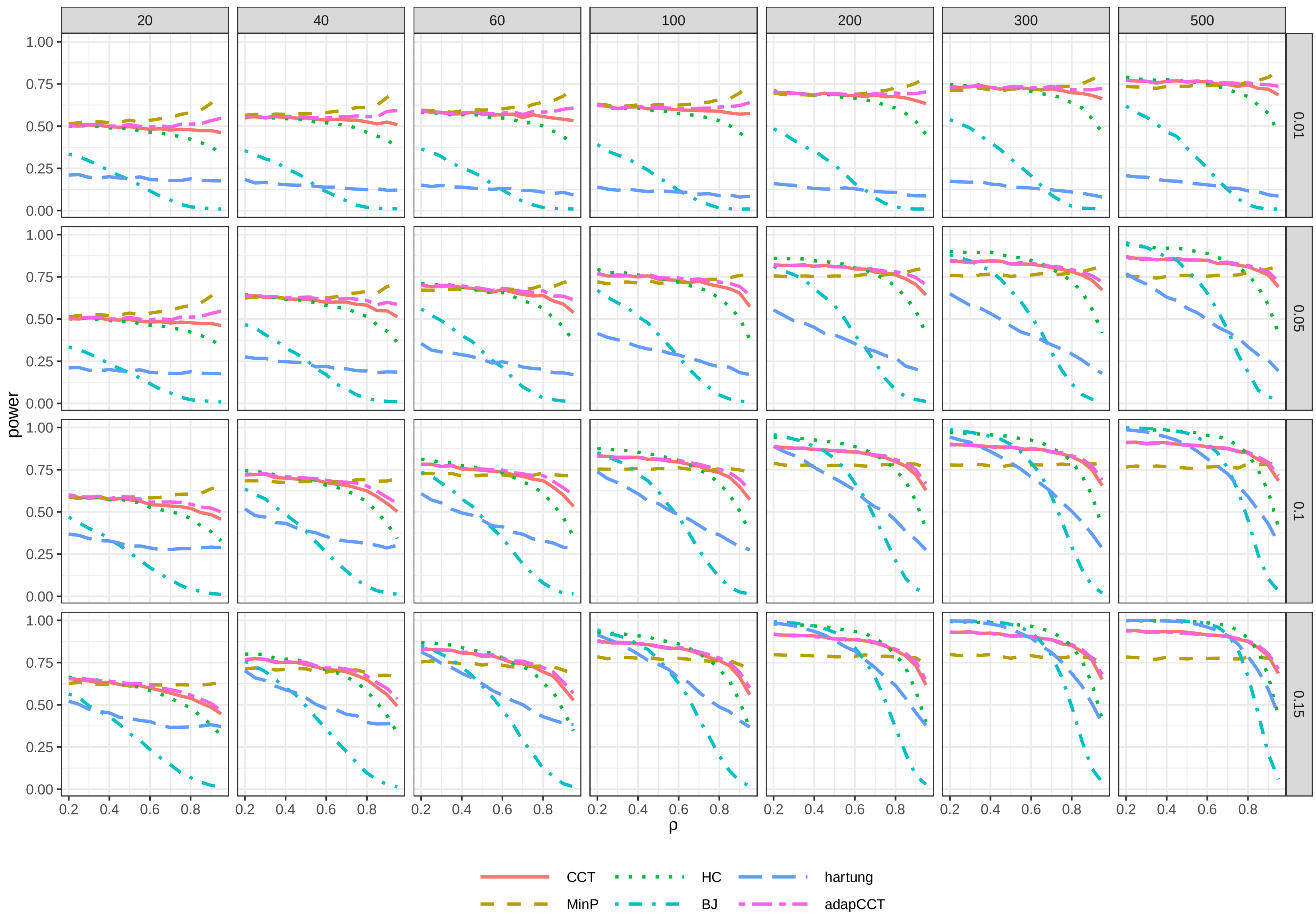

Figure 19: Power comparison of Adaptive thresholding CCT, MinP, HC, BJ, Hartung and CCT. The x-axis is the correlation strength of the compound symmetry structure. The columns from left to right correspond to the dimension $d = 20, 40, 60, 100, 200, 300, 500$. The rows from top to bottom correspond to the signal proportion $5\%, 10\%, 15\%$.

Bibliography

Arias-Castro, E., Candés, E. and Plan, Y. (2011), Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism, *The Annals of Statistics* **39(5)**, 2533–2556.

Barnett, I., Mukherjee, R. and Lin, X. (2017), The generalized higher criticism for testing snp-set effects in genetic association studies, *Journal of the American Statistical Association* **112(517)**, 64–76.

Benjamini, Y. and Hochberg, Y. (1995), Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **57(1)**, 289–300.

Berk, R. H. and Jones, D. H. (1979), Goodness-of-fit test statistics that dominate the kolmogorov statistics, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **47**, 47–59.

Demetrescu, M., Hassler, U. and Adina-Ioana, T. (2006), Combining significance of correlated statistics with application to panel data, *Oxford Bulletin of Economics and Statstics* **68(5)**, 647–663.

Dinu, I., Mahasirimongkol, S. and Liu, Q. (2012), Snp-snp interactions discovered by logic regression explain crohn's disease genetics, *PLOS one* **7(10)**, e43035.

Ditzhaus, M. and Janssen, A. (2017), The power of big data sparse signal detection tests on nonparametric detection boundaries, *ArXiv:1709.07264v1* .

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

Donoho, D. and Jin, J. (2008), Higher criticism thresholding: Optimal feature selection when useful features are rare and weak, *Proceedings of the National Academy of Sciences* **105(39)**, 14790–14795.

Duerr, R. H., Taylor, K. D. and Brant, S. R. (2006), A genome-wide association study identifies il23r as an inflammatory bowel disease gene, *Science* **314(5804)**, 1461–1463.

Efron, B. (2010), *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing and Prediction*, Cambridge university press, UK.

Fan, J. (1996), Test of significance based on wavelet thresholdling and neyman's truncation, *Journal of the American Statistical Association* **91(434)**, 674–688.

Fan, J. (1998), Test of significance when data are curves, *Journal of the American Statistical Association* **93(443)**, 1007–1021.

Fisher, R. A. (1934), *Statistical methods for research workers*, Oliver & Boyd, Edinburgh.

Gava, R. J. and Rezende, B. L. F. (2021), Some limit theorems for dependent bernoulli random variables, *Statistics and Probability Letters* **170**, 109010.

Genovese, C. and Wasserman, L. (2004), A stochastic process approach to false discovery control, *The Annals of Mathematical Statistics* **32(3)**, 1035–1061.

Goeman, J. J., Rosenblatt, J. D. and Nichols, T. E. (2019), The harmonic mean p-value: Strong versus weak control, and the assumption of independence, *Proceedings of the National Academy of Sciences* **116(47)**, 23382–23383.

Hall, P. and Jin, J. (2010), Innovated higher criticism for detecting sparse signals in correlated noise, *Annals of Statistics* **38(3)**, 1686–1732.

Hartung, J. (1999), A note on combining dependent tests of significance, *Biometrical Journal* **45(7)**, 849–855.

Heard, N. A. and Rubin-Delanchy, P. (2018), Choosing between methods of combining *p*-values, *Biometrika* **105(1)**, 239–246.

Heyde, C. C. (2004), Asymptotics and criticality for a correlated bernoulli process, *The Australian and New Zealand Journal of Statistics* **46(1)**, 53–57.

Jin, J. (2008), Proportion of non-zero normal means:universal oracle equivalences and uniformly consistent estimators, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **70(3)**, 461–493.

Kim, M. H. and Akritas, M. G. (2010), Order thresholding, *The Annals of Statistics* **38(4)**, 2314–2350.

Langaas, M., Lindqvist, B. H. and Ferkingstad, E. (2005), Estimating the proportion of true null hypotheses, with application to dna microarray data, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **67(4)**, 555–572.

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

Liu, Y. and 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.

Meinshausen, N. and Bühlmann, P. (2005), Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures, *Biometrika* **92(4)**, 893–907.

Meinshausen, N. and Rice, J. (2006), Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses, *The Annals of Statistics* **34(1)**, 373–393.

Neyman, J. (1937), Smooth test for goodness of fit, *Skandinavisk Aktuarietidskrift* **20**, 149–199.

Pearson, K. (1933), On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random, *Biometrika* **25(3/4)**, 379–410.

Pillai, N. S. and Meng, X. L. (2016), An unexpected encounter with cauchy and lévy, *The Annals of Statistics* **44(5)**, 2089–2097.

Schweder, T. and Spjøtvoll, E. (1982), Plots of *p*-values to evaluate many tests simultaneously, *Biometrika* **69(3)**, 493–502.

Storey, J. D. (2002), A direct approach to false discovery rates, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **64(3)**, 479–498.

Stouffer, S. A., Suchman, E. A., EdVinney, L. C., Star, S. A. and Williams, R. M. (1949), *The American Soldier: Adjustment During Army Life*, Princeton University Press, New Jersey.

Tippett, L. H. C. (1931), *The Methods of Statistics*, Williams and Norgate, London.

Tong, T., Feng, Z., Hilton, J. S. and Zhao, H. (2013), Estimating the proportion of true null hypotheses using the pattern of observed *p*-values, *Journal of Applied Statistics* **40(9)**, 1949–1964.

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

Wellner, J. A. and Koltchinskii, V. (2003), *A note on the asymptotic distribution of Berk-Jones type statistics under the null hypothesis, In: Hoffmann-Jørgensen J., Wellner J.A., Marcus M.B. (eds) High Dimensional Probability III. Progress in Probability, vol 55*, Birkhäuser, Basel.

Williams, E. J. (1969), Cauchy-distributed functions and a characterization of the cauchy distribution, *The Annals of Mathematical Statistics* **40(3)**, 1083–1085.

Wilson, D. J. (2019), The harmonic mean *p*-value for combining dependent tests, *Proceedings of the National Academy of Sciences* **1116(4)**, 1195–1200.

국문초록

크기가 약하고 희박한 신호들을 집합하기 위해 개별적으로 구해진 유의확률들을 결합하는 방법은 고차원 대규모 자료 분석에 있어 매우 중요한 주제 중 하나이다. 개별적으로 구해진 유의확률 또는 검정 통계량은 때때로 밀접하게 연관되어 있는 경우가 많은데, 많은 경우의 유의확률 결합 방법들은 이러한 연관성을 고려하지 않고 동일하며 독립적이라는 가정하에서 개발된 경우가 많다. 코쉬 결합 검정은 이러한 방법들과는 다르게 임의의 연관성 구조에 영향을 받지 않고 개별 유의확률들을 결합할 수 있게끔 개발된 방법이지만 실제로는 변수들 사이의 연관성이 증가함에 따라 여전히 제1종 오류가 증가한다는 단점이 있다. 본 학위논문은 임의의 연관성 구조하에서 얻어진 유의확률들의 분계점을 이용하여 코쉬 결합 검정을 확장한 새로운 전역 가설 검정 방법을 제안한다. 임의의 연관성 구조하에서, 본 학위논문에서 제안된 방법의 꼬리 확률이 점근적으로 코쉬 분포의 꼬리 확률과 일치함을 보인다. 또한 강한 희박성 조건하에서 제안된 방법의 검정력이 점근적으로 최적의 신호 탐지 경계를 달성할 수 있음을 보인다. 대규모의 모의실험 결과를 통해 제안된 방법의 검정력이 실제로 변수들 사이의 상관 구조에 강건하며, 신호가 희박한 상황에서 다른 방법들에 비해 검정력이 높다는 사실을 제시한다. 마지막으로 사례연구로서, 제안된 방법을 염증성 장질환 (Inflammatory bowel disease, IBD) 전체유전체 상관분석 연구에 적용한다.

주요어: 유의확률 결합, 코쉬 분포, 전역 가설 검정, 전체유전체 상관분석연구

학번: 2016-30092