A Powerful Bootstrap Test of Independence in High Dimensions[†]

Mauricio Olivares
Department of Statistics
LMU Munich
m.olivares@lmu.de

Tomasz Olma
Department of Statistics
LMU Munich
t.olma@lmu.de

Daniel Wilhelm
Departments of Statistics and Economics
LMU Munich
d.wilhelm@lmu.de

March 26, 2025

Abstract

This paper proposes a nonparametric test of independence of one random variable from a large pool of other random variables. The test statistic is the maximum of several Chatterjee's rank correlations and critical values are computed via a block multiplier bootstrap. The test is shown to asymptotically control size uniformly over a large class of data-generating processes, even when the number of variables is much larger than sample size. The test is consistent against any fixed alternative. It can be combined with a stepwise procedure for selecting those variables from the pool that violate independence, while controlling the family-wise error rate. All formal results leave the dependence among variables in the pool completely unrestricted. In simulations, we find that our test is very powerful, outperforming existing tests in most scenarios considered, particularly in high dimensions and/or when the variables in the pool are dependent.

Keywords: independence test, high-dimensional data, Chatterjee's rank correlation, block multiplier bootstrap, family-wise error rate.

[†]The authors gratefully acknowledge financial support from the European Research Council (Starting Grant No. 852332).

1 Introduction

This paper is concerned with nonparametric testing of independence between a random variable X and many other random variables $Y_1, \dots, Y_p$,

$$H_0 \colon Y_j \perp X \text{ for all } j \in \{1, \dots, p\},$$

against the alternative H_1, which is the negation of H_0. The goal is to propose a powerful test of H_0 allowing for p to be much larger than the sample size while at the same time not restricting the dependence among $Y_1, \dots, Y_p$ in any way. In a second step, we want to combine the new test with a stepwise procedure for screening out variables from $Y_1, \dots, Y_p$ that violate independence so as to control the family-wise error rate.

There are many applied examples in which testing H_0 and, in particular, screening out variables that violate independence is of interest. For instance, in causal inference, one might want to test whether a treatment indicator has an effect on various outcomes and then select those outcomes on which there is an effect. Such a test could also be applied to "placebo" outcomes, i.e. pre-treatment outcomes that the researcher knows cannot have been affected by the treatment, to validate unconfoundedness assumptions. Another example concerns testing fairness of machine learning predictions, where one might want to test independence of a prediction from a set of protected characteristics. As a final example, one might want to test independence of a measure of environmental exposure (e.g., whether or not a person smokes) from a vector of genetic markers. In our empirical application, we use the proposed test for identifying genes whose transcript levels oscillate during the cell cycle.

As test statistic we consider the maximum of p rank correlation coefficients by Chatterjee (2021) for testing independence between Y_j and X, $j = 1, \dots, p$. Critical values for the test are computed via a block multiplier bootstrap that perturbs an asymptotically linear representation of the rank correlations. We then show that, as the sample size grows, the proposed test controls size uniformly over a large class of data-generating processes. This result allows for the dimension p to grow exponentially in sample size and the dependence among $Y_1, \dots, Y_p$ is left completely unrestricted.

While it has been shown that the nonparametric bootstrap does not consistently estimate the distribution of a single Chatterjee's rank correlation (Lin and Han, 2024), the proposed block multiplier bootstrap achieves this by accounting for the dependence of the summands in the asymptotically linear representation. The existence of an asymptotic linear representation for Chatterjee's rank correlation together with suitably strong control of the remainder terms allows us to employ recent results for the approximation of maxima of high-dimensional vectors of sums (Chernozhukov, Chetverikov, and Kato, 2013, 2015, 2019) to establish the validity of the block multiplier bootstrap also for the maximum of many Chatterjee's rank correlation coefficients.

The block multiplier bootstrap requires a tuning parameter choice, namely the block size that should tend to infinity with sample size. We are able to provide a simple choice that enjoys a certain optimality property: it minimizes the distance of the bootstrap variance and the asymptotic variance for each individual test statistic. In simulations, however, we find that the size and power of our test are almost insensitive to the particular value of the block size.

The test is consistent against all fixed alternatives under which $Y_1, \dots, Y_p, X$ are continuously distributed. Finally, our test is computationally attractive even for very large sample sizes and high dimensions p.

The test can be combined with a stepwise multiple testing procedure (as in Romano and Wolf (2005)) for screening out variables from $Y_1, \dots, Y_p$ that violate independence so as to asymptotically control the family-wise error rate uniformly over a large set of data-generating processes. This result allows for the dimension p to grow exponentially in sample size and the dependence among $Y_1, \dots, Y_p$ is left completely unrestricted.

Related Literature. Our paper is most closely related to the literature on testing independence of two random vectors, $H_0^{joint}: Y \perp X$, where $Y \in \mathbb{R}^p$ and $X \in \mathbb{R}^q$. This is a different hypothesis from the one we consider; for $q = 1$, it implies, but is not implied by, H_0. Therefore, tests that control the rejection probability under H_0^{joint} are not guaranteed to control it under our hypothesis H_0. Sinha and Wieand (1977), Taskinen, Oja, and Randles (2005), Bakirov, Rizzo, and Székely (2006), Székely, Rizzo, and Bakirov (2007), Heller, Gorfine, and Heller (2012), Heller, Heller, and Gorfine (2012), Shi, Hallin, Drton, and Han (2022) propose nonparametric tests of H_0^{joint}, where p and q are of arbitrary, but fixed (with sample size) dimensions. Székely and Rizzo (2013) show that the test statistic of Székely, Rizzo, and Bakirov (2007) is biased in high dimensions and an independence test based on it therefore does not control size in high dimensions. They also propose a bias-corrected test statistic and derive its asymptotic distribution under the null of independence when the dimensions of both vectors grow with the sample size. The asymptotic regime under which their test is valid requires the dimensions of both vectors to grow, so it is not clear (at least to us) whether it is also valid when one of the two dimensions remains constant as the sample size grows. In addition, their derivation of the test statistic's limiting distribution requires the elements of the two vectors to be exchangeable, a condition we do not require for $Y_1, \dots, Y_p$. Ramdas, Jakkam Reddi, Poczos, Singh, and Wasserman (2015) show that both independence tests, Székely, Rizzo, and Bakirov (2007) and Székely and Rizzo (2013) have low power against "fair alternatives" in high dimensions. More recently, Zhou, Xu, Zhu, and Li (2024) and Wang, Liu, and Feng (2024) propose other rank-based tests, e.g. based on Hoeffding's D, Blum-Kiefer-Rosenblatt's R and Bergsma-Dassios-Yanagimoto's τ among others, of H_0^{joint} in high dimensions. The validity of these tests relies on at least one of the dimensions p and/or q diverging so that a central limit theorem across the elements of, say, Y can be invoked. This approach necessarily restricts the dependence of $Y_1, \dots, Y_p$, while our validity results leave the dependence completely unrestricted. In simulations, we compare our test to the distance covariance tests by Székely, Rizzo, and Bakirov (2007), Székely and Rizzo (2013), Zhu, Zhang, Yao, and Shao (2020), and various tests proposed in Zhou, Xu, Zhu, and Li (2024). Even though we are not aware of theoretical guarantees that these tests of H_0^{joint} also control the rejection probability under our null, H_0, in the simulations, we find that they do. However, in most scenarios considered, our test is substantially more powerful than all the other tests, particularly in high dimensions and/or when the variables $Y_1, \dots, Y_p$ are dependent. An additional contribution of our paper to this literature is our proposal of a stepwise procedure for selecting hypotheses that are not rejected so as to control the family-wise error rate.

There is a large literature on nonparametric tests of mutual independence among the elements of a random vector. Some examples are Leung and Drton (2018), Yao, Zhang, and Shao (2018), Drton, Han, and Shi (2020), Wang, Liu, Feng, and Ma (2024), Bastian, Dette, and Heiny (2024); see also references therein. Xia, Cao, Du, and Dai (2024) propose such a test based on Chatterjee's rank correlation. While the hypothesis considered in our paper also involves many nonparametric independence tests, it substantially differs from the null of mutual independence. This is because, in our testing problem, X occurs in every independence hypothesis and the dependence of $Y_1, \ldots, Y_p$ is left unrestricted.

Finally, since our proposed test is based on the rank correlation for two random variables proposed by Chatterjee (2021), our paper is also related to a recent and fast-growing literature that examines the rank correlation coefficient's properties. Chatterjee (2021) shows asymptotic normality of the correlation coefficient under independence of the two random variables. Lin and Han (2022a) and Kroll (2024) show that it is also asymptotically normal under dependence. Shi, Drton, and Han (2021) and Lin and Han (2022b) examine and propose improvements of the power of tests of independence based on the rank correlation coefficient. For a recent review of this literature, see Chatterjee (2024).

2 The Test

This section first introduces the new test, establishes asymptotic size control uniformly over a large class of data-generating processes, and then consistency against all fixed alternatives under which $Y_1, \ldots, Y_p$ and X have continuous distributions. The section concludes with the development of an optimal tuning parameter choice.

Let $\mathbb{D} := \{(X_i, Y_{1,i}, \ldots, Y_{p,i})\}_{i=1}^n$ be an i.i.d. sample drawn from the distribution of $(X, Y_1, \ldots, Y_p)$. For each individual hypothesis $H_{0,j} \colon Y_j \perp X$ there are many available tests in the literature. In this paper, we focus on the test statistic by Chatterjee (2021). The motivation for this choice will become clear later in this section. To define the test statistic let $X_{(k)}$ be the k-th order statistic of $X_1, \ldots, X_n$, i.e. $X_{(1)} \leq \cdots \leq X_{(n)}$, and $Y_{j,(k)}$ be the concomitant of $X_{(k)}$, i.e. if $X_{(k)} = X_l$, then $Y_{j,(k)} = Y_{j,l}$. Denote by F_{Y_j} the cumulative distribution function (cdf) of Y_j and by $\hat{F}_{Y_j}$ the empirical cdf. Then, Chatterjee's rank correlation for testing an individual hypothesis $H_{0,j}$ is

$$\hat{\xi}_j := 1 - \frac{3n}{n^2 - 1} \sum_{i=1}^{n-1} \left| \hat{F}_{Y_j}(Y_{j,(i+1)}) - \hat{F}_{Y_j}(Y_{j,(i)}) \right| . \tag{1}$$

Chatterjee (2021) shows that $\hat{\xi}_j$ is a consistent estimator of

$$\xi_j := \frac{\int Var(\mathbb{E}[\mathbb{1}\{Y_j \geq t\} | X]) f_{Y_j}(t) dt}{\int Var(\mathbb{1}\{Y_j \geq t\}) f_{Y_j}(t) dt},$$

a measure of dependence introduced by Dette, Siburg, and Stoimenov (2013) in the case in which Y_j has a continuous distribution with density f_{Y_j}. This measure has several attractive features (Chatterjee, 2021). Two features that are particularly important for the test proposed in this paper are that (i) ξ_j is equal to zero if, and only if, X and Y_j are independent, and (ii) the estimator $\hat{\xi}_j$ admits an asymptotic linear representation (shown in (3) below) with a remainder that we can show to be sufficiently small. Our arguments for validity of the proposed test in high dimensions crucially depend on property (ii).

The proposed test statistic for H_0 is the maximum of the individual Chatterjee's rank correlations:

$$\hat{T} \coloneqq \sqrt{n} \max_{1 \leq j \leq p} \hat{\xi}_j. \tag{2}$$

We propose to compute critical values for the test statistic via a block multiplier bootstrap. To describe the procedure consider first an individual hypothesis $H_{0,j}$. If the null $H_{0,j}$ holds and both random variables are continuously distributed, then the arguments in Angus (1995) imply that Chatterjee's rank correlation has an asymptotically linear representation of the form

$$\sqrt{n}\,\hat{\xi}_j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} W_{j,i} + r_{j,n} \ , \tag{3}$$

where $r_{j,n}$ is a negligible remainder term and $W_{j,i}$ is defined as

$$W_{j,i} \coloneqq 2 - 3|U_{j,i+1} - U_{j,i}| - 6U_{j,i}(1 - U_{j,i})$$

with $U_{j,i} \coloneqq F_{Y_j}(Y_{j,(i)})$. A naive application of the multiplier bootstrap idea would be to repeatedly draw bootstrap multipliers $\varepsilon_1, \ldots, \varepsilon_n$ as independent standard normal random variables that are independent of the data $\mathbb{D}$ and then compute a critical value from the distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} \varepsilon_i W_{j,i}$ given the data. However, there are two problems with this approach. First, $\{W_{j,i}\}_{i=1}^{n}$ is not an i.i.d. sequence but a 1-dependent process. Second, $\{W_{j,i}\}_{i=1}^{n}$ are not directly observed because F_{Y_j} is unknown.

To address the first challenge, we decompose the sum $\sum_{i=1}^{n-1} W_{j,i}$ into the sums over "big" and "small" blocks formed of $\{W_{j,i}\}_{i=1}^{n}$ with the property that the big blocks are independent of each other. Formally, let $q \geq 1$ denote the size of big blocks. It is a tuning parameter to be chosen by the researcher; in Section 2.3 we develop an optimal choice of q. Since $\{W_{j,i}\}_{i=1}^{n}$ is a 1-dependent sequence, we let the small blocks be of length one. Further, let $m \coloneqq \lfloor (n-1)/(q+1) \rfloor$, where $\lfloor \nu \rfloor$ is the integer part of ν, denote the number of big blocks. Lastly, $r \coloneqq n - 1 - m(q+1)$, $0 \leq r < q+1$, is the number of remaining summands that are not allocated to any of the small or big blocks. With this notation, we can thus write

$$\sum_{i=1}^{n-1} W_{j,i} = \sum_{k=1}^{m} A_{j,k} + \sum_{k=1}^{m} B_{j,k} + R_j,$$

where

$$A_{j,k} \coloneqq \sum_{l=(k-1)(q+1)+1}^{kq+(k-1)} W_{j,l}, \quad B_{j,k} \coloneqq W_{j,k(q+1)}, \quad \text{and} \quad R_j \coloneqq \sum_{k=1}^{r} W_{j,m(q+1)+k}.$$

The big block sums $A_{j,1}, \ldots, A_{j,m}$ are independent of each other, and we show that the terms $B_{j,1}, \ldots, B_{j,m}$ and R_j are asymptotically negligible. It then follows that $\sqrt{n}\,\hat{\xi}_j$ can be approximated by $\frac{1}{\sqrt{mq}} \sum_{k=1}^{m} A_{j,k}$, the sum of independent components.

We now need to address the second challenge, which is that $W_{j,i}$, and thus also $A_{j,k}$, are not observed. $W_{j,i}$ can be estimated by

$$\hat{W}_{j,i} := 2 - 3\left|\hat{U}_{j,i+1} - \hat{U}_{j,i}\right| - 6\hat{U}_{j,i}(1 - \hat{U}_{j,i}),$$

where $\hat{U}_{j,i} := \hat{F}_{Y_j}(Y_{j,(i)})$, and $A_{j,k}$ by

$$\hat{A}_{j,k} := \sum_{l=(k-1)(q+1)+1}^{kq+(k-1)} \hat{W}_{j,l}.$$

While $A_{j,1}, \ldots, A_{j,m}$ are independent, $\hat{A}_{j,1}, \ldots, \hat{A}_{j,m}$ are only asymptotically independent, i.e. in the limit as $n, m \to \infty$.

Finally, bootstrap multipliers $\varepsilon_1, \ldots, \varepsilon_m$ are drawn as independent standard normal random variables that are independent of the data $\mathbb{D}$. The bootstrap statistic is then defined as

$$\hat{T}^B := \max_{1 \le j \le p} \frac{1}{\sqrt{mq}} \sum_{k=1}^{m} \varepsilon_k \hat{A}_{j,k}. \tag{4}$$

For a nominal level $\alpha \in (0, 1)$, the proposed critical value $\hat{c}(\alpha)$ for our test is the conditional $(1-\alpha)$-quantile of $\hat{T}^B$ given the data $\mathbb{D}$. The test rejects H_0 if, and only if, the test statistic $\hat{T}$ exceeds the critical value $\hat{c}(\alpha)$.

2.1 Size Control

In this subsection, we show that our proposed test asymptotically controls size uniformly over a large class of data-generating processes.

Assumption 1. $X, Y_1, \ldots, Y_p$ *are continuously distributed.*

Assuming all random variables have a continuous distribution simplifies the presentation, but is not essential. Remark 1 below discusses extensions to the case with discrete distributions.

Assumption 2. *Suppose* $p \ge 2$*. There exist constants* $C_1 > 0$ *and* $0 < \gamma < 1/4$ *such that* $(1/q)\log^2 p \le C_1 n^{-\gamma}$ *and* $\max\left\{q \log^{5/2} p,\ \sqrt{q}\log^{7/2}(pn)\right\} \le C_1 n^{1/2-\gamma}$*.*

This assumption requires q to diverge as the sample size grows, but restricts its rate to be neither too slow nor too fast. The assumption also restricts the rate at which the dimension p is allowed to grow with sample size. However, the upper bound on the rate is very large: p may be an exponential function of sample size and thus is allowed to be much larger than sample size. For instance, there are positive constants δ_1, δ_2 so that $p = \mathrm{e}^{n^{\delta_1}}$ and $q = n^{\delta_2}$ satisfy Assumption 2.

Theorem 1. *Suppose that Assumptions 1–2 hold. Then, under the null hypothesis* H_0*, there exist positive constants* c*,* C *depending only on* γ *and* C_1 *such that*

$$\left|\mathrm{P}\left(\hat{T} > \hat{c}(\alpha)\right) - \alpha\right| \le Cn^{-c}. \tag{5}$$

The result in (5) implies that the proposed test asymptotically controls size. In fact, the asymptotic size is equal to the nominal level α and, in this sense, the test is not conservative. Furthermore, the probability of rejecting H_0 when H_0 is satisfied can deviate from the nominal level only by a term that is polynomially small in n. Importantly, the constants c and C depend on the data-generating process only through the constants γ and C_1 from Assumption 2. Therefore, under H_0, the inequality in (5) holds uniformly over all data-generating processes that satisfy the assumption with the same constants, denoted by $\mathbf{P}_{\gamma,C_1}$:

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_{\gamma,C_1}} \left| \mathrm{P}\left(\hat{T} > \hat{c}(\alpha)\right) - \alpha \right| = 0,$$

i.e. our test has asymptotic size equal to α uniformly over those data-generating processes.

It is worth noting that the validity of our test is guaranteed in high dimensions without restricting the dependence of $Y_1, \ldots, Y_p$ in any way.

To establish this result we need to show that the distribution of the bootstrap statistic $\hat{T}^B$ given the data is close to the distribution of the test statistic $\hat{T}$. This is achieved in several steps: (i) show that $\hat{T}$ is close to

$$T_0 := \max_{1\le j\le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} W_{j,i} \ ,$$

(ii) show that $\hat{T}^B$ is close to

$$T_0^B := \max_{1\le j\le p} \frac{1}{\sqrt{mq}} \sum_{k=1}^{m} \varepsilon_k A_{j,k},$$

and then (iii) show that both T_0 and T_0^B are close to the maximum of Gaussian random variables,

$$Z_0 := \max_{1\le j\le p} V_j,$$

where $V := (V_1, \ldots, V_p)'$ is a Gaussian vector such that $\mathbb{E}(V) = 0$ and $\mathbb{E}(V V') = \frac{1}{mq}\sum_{k=1}^m \mathbb{E}(A_k A_k')$, $A_k := (A_{1,k}, \ldots, A_{p,k})'$. Steps (i) and (ii) are developed in the proof of the theorem and step (iii) is delegated to Lemmas 3 and 4. These three steps establish that the distribution of the test statistic is close to that of the bootstrap statistic. However, this result does not yet imply the statement in (5) because $\hat{c}(\alpha)$ is random and may be correlated with $\hat{T}$. The final step of the proof therefore consists of passing from a deterministic to a random critical value.

Step (iii) of the derivation combines several results on high-dimensional Gaussian approximations from Chernozhukov, Chetverikov, and Kato (2013, 2015, 2019). The main challenge of the proof is contained in steps (i) and (ii), where we need to establishing that all remainder terms $r_{1,n}, \ldots, r_{p,n}$ from the representation (3) vanish at a suitably fast rate. Angus (1995) shows that, for each j, $r_{j,n} = o_P(n^{-1/2})$, but in our high-dimensional setting, we need stronger control of these remainder terms, and these results are developed in Lemma 2.

Remark 1 (discrete distributions)**.** If any of $X, Y_1, \ldots, Y_p$ have a noncontinuous distribution, then they can be replaced by new random variables that are equal to the original ones plus a sufficiently small amount of noise (which is independent of the data). Our proposed test can then be applied to the new random variables. One can show that the test continues to be valid in this case because the null hypothesis H_0 (in terms of the original variables) implies that the probability limits of Chatterjee's rank correlations in terms of the new variables are all equal to zero. ■

Remark 2 (bootstrapping Chatterjee's rank correlation)**.** Lin and Han (2024) show that, under the null of independence of two random variables, the nonparametric bootstrap does not consistently estimate the limiting distribution of Chatterjee's rank correlation. In particular, they show that the bootstrap yields a variance estimate whose expectation is below 2/5, the correct limiting variance of Chatterjee's rank correlation under the null of independence and Assumption 1.

Assumption 2 requires $p \geq 2$, but an inspection of the proof of Theorem 1 reveals that the result can also be proven for $p = 1$ under slightly simplified rate conditions. Therefore, the block multiplier bootstrap correctly approximates the limiting distribution of Chatterjee's rank correlation under the null of independence. The reason for this is that it correctly accounts for the 1-dependence in the asymptotic linear representation, and thus correctly estimates the variance of Chatterjee's rank correlation, while the nonparametric bootstrap ignores this dependence.[1] ■

Remark 3 (studentization)**.** The test proposed in this section does not studentize the test statistics. In simulations in Section 4, we find that studentization may improve size and power of the test. We consider studentizing the individual test statistics by their standard deviation under the null, i.e.,

$$\hat{T}^{stud} := \sqrt{n} \max_{1\leq j\leq p} \frac{\hat{\xi}_j}{\sqrt{v_n}}.$$

where

$$v_n := \frac{n(n-2)(4n-7)}{10(n-1)^2(n+1)} = \mathrm{Var}(\sqrt{n}\hat{\xi}_j)$$

under the null (Zhang, 2023, Lemma 2). The sequence v_n is monotonically increasing and, as $n \to \infty$, it converges to 2/5, the asymptotic variance derived by Chatterjee (2021).

For the bootstrap statistics there are at least two different possibilities for studentization. First, one could studentize it by the square root of $\mathbb{E}[A_{j,k}^2]/q = 0.4 + 0.1/q$, i.e.

$$\hat{T}^{B,stud\,1} := \max_{1\leq j\leq p} \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \varepsilon_k \frac{\hat{A}_{j,k}}{\sqrt{0.4q + 0.1}}.$$

This standardization ensures that the diagonal elements of the bootstrap covariance matrix of individual tests are all approximately centered at one for any q, and hence approximately match the variance of $\sqrt{n}\hat{\xi}_j/\sqrt{v_n}$. The second possibility is to employ a bootstrap test statistic in which the big blocks sums are demeaned and standardized by their sample standard deviation, i.e.

$$\hat{T}^{B,stud\,2} := \max_{1\leq j\leq p} \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \varepsilon_k \frac{\hat{A}_{j,k} - \frac{1}{m}\sum_{k=1}^{m} \hat{A}_{j,k}}{\sqrt{\frac{1}{m}\sum_{k=1}^{m} \hat{A}_{j,k}^2}}.$$

This standardization ensures that the diagonal elements of the bootstrap covariance matrix of individual test statistics are all equal to one for any q. ■

[1] Dette and Kroll (2024) show that the m-out-of-n bootstrap is also valid.

2.2 Consistency

The following theorem shows that the proposed test is consistent against any (fixed) violation of the null:

Theorem 2. *Suppose that Assumptions 1–2 hold. Then, under the alternative hypothesis H_1,*

$$\mathrm{P}\left(\hat{T} > \hat{c}(\alpha)\right) \to 1 \qquad \textit{as } n \to \infty.$$

In light of recent discussions related to the local asymptotic power of Chatterjee's test of independence (e.g., Shi, Drton, and Han (2021) and Lin and Han (2022b)), it would be interesting to study the power of our proposed test to local alternatives. We view this analysis as beyond the scope of the paper and relegate it to future research.

2.3 Choice of q

Theorem 1 shows that our proposed block multiplier bootstrap test asymptotically controls size under rate conditions on q, the size of blocks used to construct the critical value $\hat{c}(\alpha)$. These rate conditions specify that $q \to \infty$ as $n \to \infty$ at a rate that is neither too fast nor too slow, but they do not provide any guidance on how to choose q in finite samples. In this section, we develop a choice of q that is optimal in a certain finite-sample sense.

To describe the optimality criterion, recall from the discussion following Theorem 1 that the distribution of the bootstrap test statistic $\hat{T}^B$ is, to first order, determined by its infeasible counterpart $T_0^B = \max_{1 \le j \le p} T_{0,j}^B$, where

$$T_{0,j}^B := \frac{1}{\sqrt{mq}} \sum_{k=1}^{m} \varepsilon_k A_{j,k}$$

with bootstrap multipliers $\varepsilon_1, \ldots, \varepsilon_m$ that are i.i.d. standard normal and independent of the data. By construction, $T_{0,j}^B$ follows a normal distribution with mean zero and variance that depends on the data and, implicitly, on q,

$$T_{0,j}^B | \mathbb{D} \sim \mathcal{N}\left(0, V_j^B\right), \quad V_j^B := \mathrm{Var}\left(T_{0,j}^B | \mathbb{D}\right).$$

In our bootstrap procedure, $T_{0,j}^B$ mimics the behavior of the individual test statistic $\sqrt{n}\hat{\xi}_j$, which is asymptotically normally distributed and, under the null, has variance v_n given in Remark 3. Since q affects the conditional distribution of $T_{0,j}^B$ only through the bootstrap variance V_j^B and it does not affect the test statistic itself, we aim to choose q such that V_j^B is close to v_n.

The following lemma characterizes the expectation and the variance of the bootstrap variance V_j^B under the null $H_{0,j}$.

Lemma 1. *Suppose that Assumption 1 and the null $H_{0,j}$ hold. For any $j = 1, \dots, p$, it holds that*

$$\mathbb{E}[V_j^B] = \frac{2}{5} + \frac{1}{10q}, \qquad \textit{for any } q \geq 1,$$

$$\mathrm{Var}(V_j^B) = \frac{1}{m} \begin{cases} \frac{7}{20} & \textit{if } q = 1, \\ \frac{1353}{2800} & \textit{if } q = 2, \\ \frac{8}{25} + \frac{88}{175q} - \frac{229}{700q^2} & \textit{if } q \geq 3, \end{cases}$$

where $m := \lfloor (n-1)/(q+1) \rfloor$.

Lemma 1 shows that the expectation of V_j^B is above $2/5$ for any fixed q and it monotonically converges to $2/5$ as q grows large. Since the target variance v_n approaches $2/5$ from below (see Remark 3), this means that V_j^B is biased upwards with a bias that vanishes asymptotically when $q, n \to \infty$. The variance of V_j^B, in turn, is generally increasing in q. Figure 1 illustrates these relationships for samples of size $n = 500$ and $n = 1000$. We resolve this bias-variance trade-off by considering the mean squared error of the bootstrap variance V_j^B:

$$MSE_{j,n}(q) := \mathbb{E}\left[\left(V_j^B - v_n\right)^2\right] = \left\lfloor \frac{n-1}{q+1} \right\rfloor^{-1} \left\{ \begin{matrix} \frac{7}{20} & \text{if } q = 1 \\ \frac{1353}{2800} & \text{if } q = 2 \\ \frac{8}{25} + \frac{88}{175q} - \frac{229}{700q^2} & \text{if } q \geq 3 \end{matrix} \right\} + \left(\frac{2}{5} + \frac{1}{10q} - v_n\right)^2.$$

Minimizing $MSE_{j,n}(q)$ over $q \in \mathbb{N}^+$ yields our proposed optimal choice of q:

$$q^*(n) := \arg\min_{q \in \mathbb{N}^+} MSE_{j,n}(q).$$

Since $MSE_{j,n}(q)$ is a known function of the sample size, the optimal choice $q^*(n)$ does not depend on the data (beyond n). The reason for that is that the individual bootstrap statistic $T_{0,j}^B$ depends on the data only through the ranks of the concomitants, $U_{j,i} := F_{Y_j}(Y_{j,(i)})$, and these are independent and uniformly distributed under the null. In consequence, the optimal choice $q^*(n)$ is independent of j and thus the same for each independence hypothesis $H_{0,j}$.

The minimizer $q^*(n)$ can be computed for each n by evaluating the function $MSE_{j,n}(q)$ over a grid of values $q \in \{1, 2, \dots, \lfloor (n-1)/2 \rfloor\}$. On most of its domain, $q^*(n)$ is a step function with large flat regions, but it can oscillate in small transition regions. For example,

$$q^*(n) = \begin{cases} 1 & \text{if } 3 \leq n \leq 87, \\ 2 & \text{if } 88 \leq n \leq 224, \\ 2 \text{ or } 3 & \text{if } 225 \leq n \leq 244, \\ 3 & \text{if } 245 \leq n \leq 615, \\ 3 \text{ or } 4 & \text{if } 616 \leq n \leq 645, \\ 4 & \text{if } 646 \leq n \leq 1344, \\ \dots & \end{cases}$$

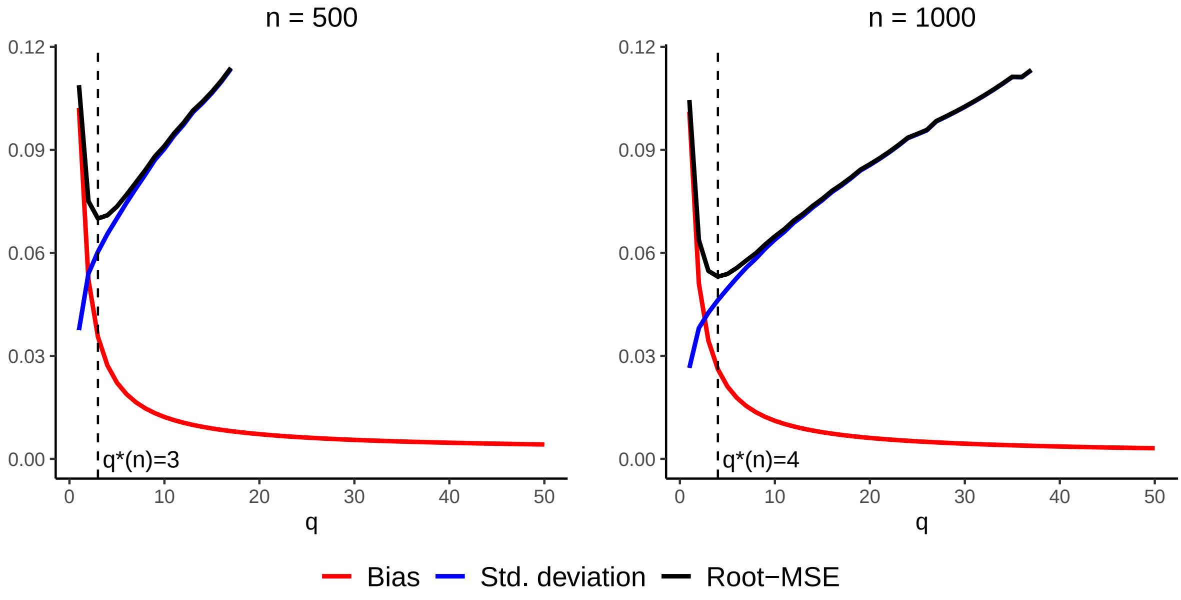


Figure 1: Bias, standard deviation, and the root mean squared error of V_j^B based on the formulas in Lemma 1 together with the optimal choice $q^*(n)$.

where $q^*(225) = 3$, $q^*(n) = 2$ for $n \in \{226, \ldots, 232\}$, $q^*(233) = 3$, etc. The non-monotone ranges make it impractical to tabularize $q^*(n)$. Instead, we approximate $q^*(n)$ using a smooth function. To motivate this approximation, note that in a setting where n and q are large but q is much smaller than n, which is consistent with the asymptotic regime of Assumption 2, $MSE_{j,n}(q)$ is close to $0.32q/n + 0.01/q^2$, which is minimized at

$$\tilde{q}(n) := (n/16)^{1/3}.$$

This expression turns out to provide a good approximation to $q^*(n)$ even for small n in the sense that $|\tilde{q}(n) - q^*(n)| < 1$ for any $n \in \mathbb{N}^+$. This property implies that

$$q^*(n) = \begin{cases} \lceil \tilde{q}(n) \rceil \text{ if } MSE_{j,n}(\lceil \tilde{q}(n) \rceil) \leq MSE_{j,n}(\lfloor \tilde{q}(n) \rfloor), \\ \lfloor \tilde{q}(n) \rfloor \text{ otherwise.} \end{cases}$$

The above approximating property of $\tilde{q}(n)$ is illustrated in Figure 2.

Remark 4 (compatibility with rate conditions)**.** The optimal choice of q diverges at the rate $n^{1/3}$. This rate is compatible with Assumption 2 as long as $p = O(e^{n^a})$ for some $a < 1/15$. ■

We note that the goal of this paper is to test the hypothesis H_0 that all individual hypotheses $H_{0,j}$, $j = 1, \ldots, p$, hold simultaneously, while we derived the optimal $q^*(n)$ for an individual test statistic. Therefore, this choice does not necessarily minimize the distance between the distributions of the max-test statistic $\hat{T} := \max_{1 \leq j \leq p} \sqrt{n}\hat{\xi}_j$ and the bootstrap statistic $\hat{T}^B := \max_{1 \leq j \leq p} \hat{T}_j^B$ in any sense. However, minimizing the distance between these two distributions is considerably more difficult because the individual statistics may be arbitrarily dependent and the optimal q would then depend on their (unknown) dependence structure. Developing a feasible version of this approach would require estimation of the copula of a high-dimensional random vector, which our proposal above avoids.

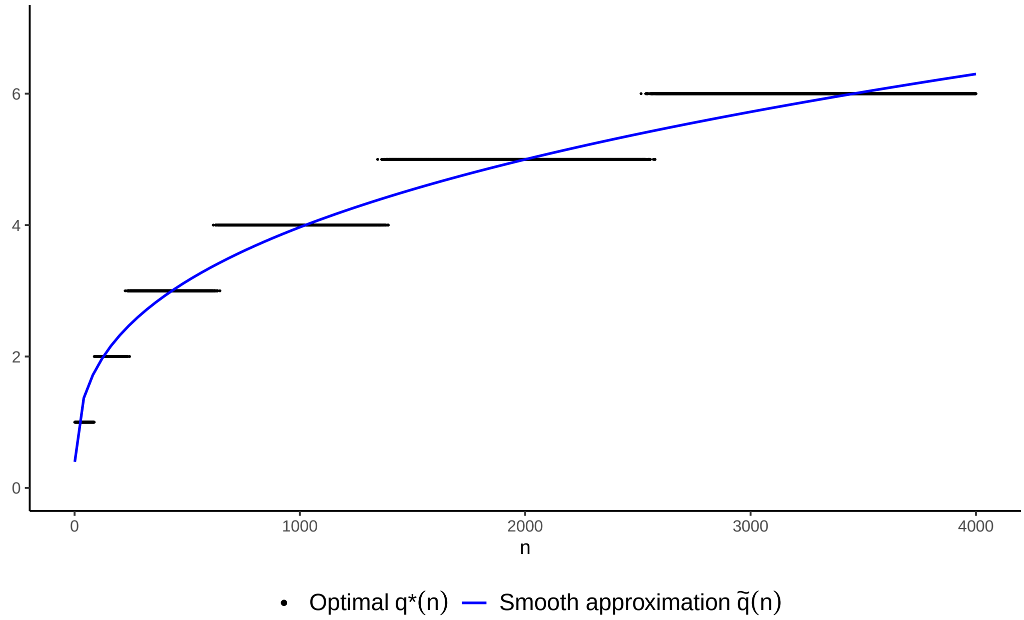


Figure 2: Optimal choice $q^*(n)$ and the smooth approximation $\tilde{q}(n) := (n/16)^{1/3}$.

In simulations in Section 4, we find that our proposed choice of q not only optimizes the bootstrap approximation of the individual statistics, but also yields a good bootstrap approximation for the max-statistic.

3 Stepwise Procedure

In the previous section, we considered testing the null H_0 that all variables $Y_1, \ldots, Y_p$ are independent of X. Now, consider the problem of selecting individual hypotheses $H_{0,j} \colon Y_j \perp X$ that are violated. The previous section already yields a ("single-step") method of selection: simply select all hypotheses $H_{0,j}$ for which $\sqrt{n}\hat{\xi}_j > \hat{c}(\alpha)$. This section introduces a stepdown procedure that improves upon the single-step procedure by possibly rejecting more hypotheses in finite samples. In addition, we show that the stepdown procedure (and, thus, by extension also the single-step procedure) guarantees asymptotic control of the family-wise error rate.

Let $J(P) \subseteq \{1, \ldots, p\}$ denote the set of hypotheses $H_{0,j}$ that are true under P. The family-wise error rate is defined as the probability of rejecting at least one true hypothesis,

$$FWER_P := \mathrm{P}(\text{reject at least one } H_{0,j} \colon j \in J(P)).$$

For any $I \subseteq \{1, \ldots, p\}$, let

$$\hat{T}(I) := \sqrt{n} \max_{j \in I} \hat{\xi}_j \qquad \text{and} \qquad \hat{T}^B(I) := \max_{j \in I} \frac{1}{\sqrt{mq}} \sum_{k=1}^{m} \varepsilon_k \hat{A}_{j,k},$$

where ε_k and $\hat{A}_{j,k}$ are the multipliers and estimated blocks as introduced in the previous section. Finally, define $\hat{c}(\alpha; I)$ as the $(1-\alpha)$-quantile of $\hat{T}(I)$ given the data $\mathbb{D}$.

The following algorithm introduces the stepdown procedure.

Algorithm 1 (Stepdown Procedure)
Initialize $I_0 = \{1, \ldots, p\}$ *and* $s = 0$.

1. *Compute* $\hat{T}(I_s)$ *and* $\hat{c}(\alpha; I_s)$.
2. *Is* $\hat{T}(I_s) > \hat{c}(\alpha; I_s)$ *satisfied?*
   (a) **yes:** *Reject any hypothesis* $H_{0,j}$ *with* $j \in I_s$ *for which* $\sqrt{n}\hat{\xi}_j > c(\alpha; I_s)$, *then let* $I_{s+1} \subset I_s$ *denote the set of hypotheses that have not previously been rejected, set* $s \to s+1$, *and return to Step 1.*
   (b) **no:** *Stop.*

Theorem 3. *Suppose that Assumptions 1–2 hold. Then, the procedure for rejecting individual hypotheses defined in Algorithm 1 satisfies*

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_{\gamma,C_1}} FWER_P \leq \alpha$$

The theorem shows that the stepdown procedure in Algorithm 1 asymptotically controls the family-wise error rate uniformly over data-generating processes in $\mathbf{P}_{\gamma,C_1}$.

4 Simulations

In this section, the finite-sample performance of our proposed test is investigated in simulations. We consider three variants of the test that employ different types of studentization. The procedure introduced in Section 2, constructs the test statistic $\hat{T}$ based on the individual Chatterjee's rank correlations $\hat{\xi}_j$ as in (2) and the critical value $\hat{c}(\alpha)$ from the distribution of the bootstrap statistic $\hat{T}^B$ as in (4). This test does not use studentization and is denoted by BMB0. We compare this test to the two studentized variants described in Remark 3. Both variants use the test statistic $\hat{T}^{stud}$, but different critical values: BMB1 refers to the test with critical value from the bootstrap statistic $\hat{T}^{B,stud\,1}$ and BMB2 to the one based on $\hat{T}^{B,stud\,2}$.

In the main text, we report results for two different designs (Models 1 and 2) in which deviations from the null are such that only one of the $Y_1, \ldots, Y_p$ violates independence. In Appendix C.1, we consider a third design (Model 3) in which all $Y_1, \ldots, Y_p$ violate independence. The qualitative conclusions therein are similar to those in the main text.

Model 1: Let $(X, Y_1, \ldots, Y_p)' \sim \mathcal{N}(0, \Sigma_{\rho,\tau})$, where

$$\Sigma_{\rho,\tau} = \begin{bmatrix} 1 & \rho & 0 & \ldots & 0 \\ \rho & 1 & \tau & \ldots & \tau \\ 0 & \tau & 1 & \ldots & \tau \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \tau & \tau & \ldots & 1 \end{bmatrix} \in \mathbb{R}^{(1+p)\times(1+p)}.$$

In this model, H_0 is satisfied when $\rho = 0$. If $\rho \neq 0$, then ρ determines the magnitude of the violation of the first hypothesis $H_{0,1}\colon Y_1 \perp X$, while all other hypotheses $H_{0,j}\colon Y_j \perp X$, $j \geq 2$ hold. τ determines the correlation between $Y_1, \ldots, Y_p$.[2]

Model 2: Let $X \sim Unif[-1,1]$ and $(\epsilon_1, \ldots, \epsilon_p) \sim \mathcal{N}(0, \Sigma_\tau)$, where Σ_τ has diagonal elements equal to one and off-diagonal elements equal to τ, and $(\epsilon_1, \ldots, \epsilon_p)$ is independent of X. Further, let

$$Y_j = \begin{cases} 2\rho\cos(8\pi X) + \epsilon_1, & \text{for } j = 1 \\ \epsilon_j & \text{for } 2 \leq j \leq p \end{cases}.$$

As in Model 1, H_0 is satisfied when $\rho = 0$, and if $\rho \neq 0$, then ρ determines the magnitude of the violation of the first hypothesis of independence. τ is the correlation between $Y_1, \ldots, Y_p$.[3]

All the results in this section are for sample size $n = 500$ and significance level $\alpha = 0.05$.

4.1 Simulation I: Size and power for different block sizes

The first set of simulations concerns the rejection rates for different choices of the block size q. Figure 3 and 4 present the results for Models 1 and 2, respectively, based on $B = 999$ bootstrap replications and $10,000$ Monte Carlo draws. First, we note that our baseline procedure BMB0 controls the size across different scenarios for all considered values of q, and its performance is very similar regardless of whether there is dependence between individual tests or not. It is, however, conservative in higher dimensions. The fact that the rejection rate is particularly low for $q = 1$ is generally consistent with the upward bias in the bootstrap variance characterized in Lemma 1. The optimal rule from Section 2.3 yields $q^*(n) = 3$ for $n = 500$, marked by the vertical dashed lines in the graphs. This choice proves reasonable in all the considered scenarios.

The issue of conservativeness is alleviated by the studentization. Both BMB1 and BMB2 effectively shrink the bootstrap distribution and yield rejection probabilities very close to the nominal level of 5% for small values of q. Since the correction in BMB1 is negligible for large q, the blue and black lines approach each other as q increases. BMB2 maintains rejection rates closer to 5% as q grows, but it generally slightly overrejects.

Panels C and D of Figure 3 indicate that all tests have very high power against the considered alternative in low dimensions under Model 1. The power generally deteriorates, albeit very slowly, as p grows in this model. Apart from very small q in BMB0, the power is not very sensitive to the choice of q. In Panels C and D of Figure 4, the simulated rejection rates are all equal to one, which is consistent with the observations that Chatterjee's rank correlation is particularly well-suited for detecting oscillatory behavior (Chatterjee, 2021). As in the case of size, the power is not sensitive to correlation between $Y_1, \ldots, Y_p$.

[2] This model is similar to Example 5.3 of Shi, Hallin, Drton, and Han (2022), except that we allow for all correlations between $Y_1, \ldots, Y_p$ to be nonzero and consider one-dimensional X.

[3] For $p = 1$, this model is a reparameterized version of DGP (4) in Chatterjee (2021).

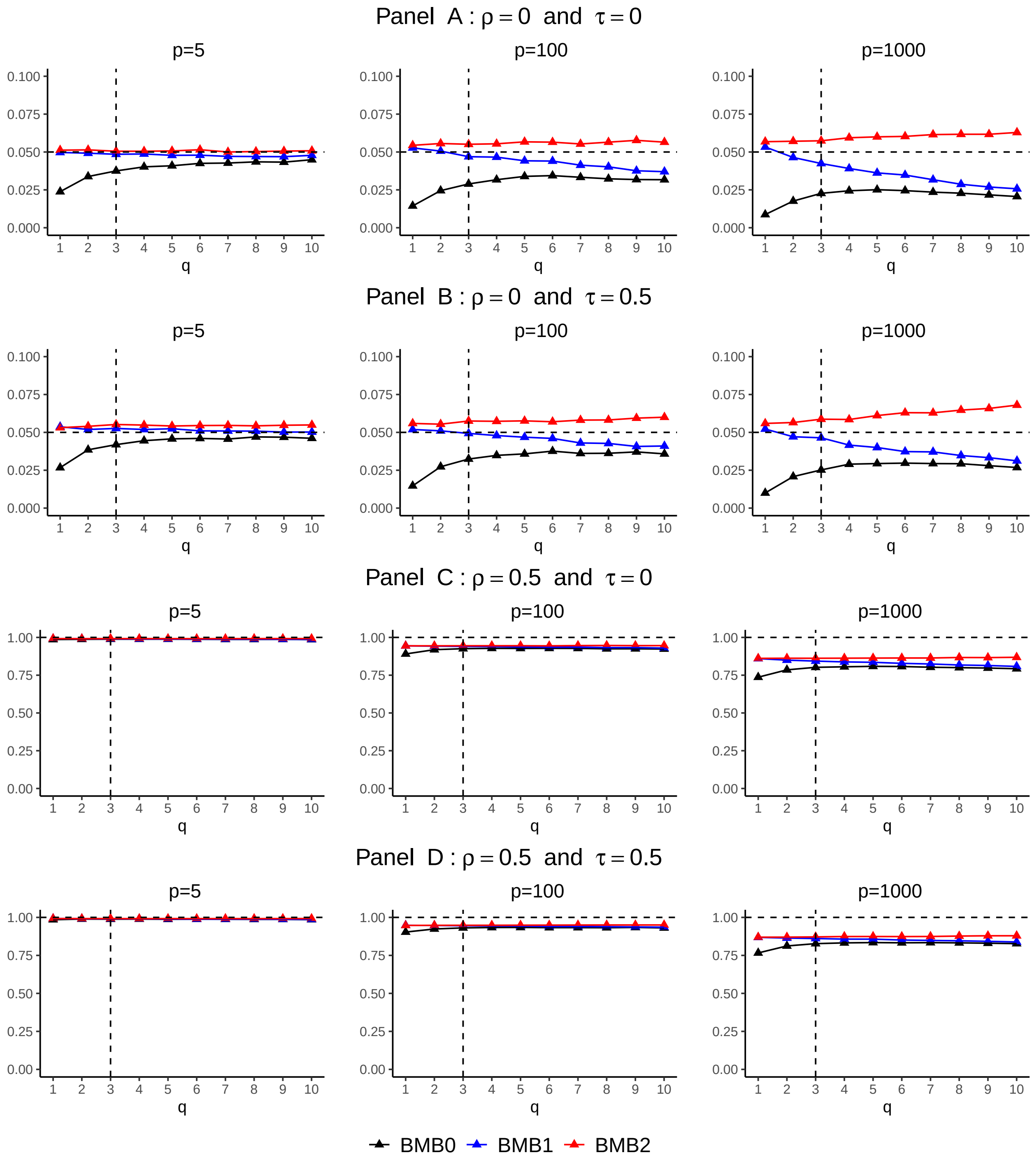


Figure 3: Rejection rates for Model 1 under different choices of the big block size q.
Notes: The tests have nominal level of 5%. The null of independence holds if $\rho = 0$. τ denotes the correlation between components of Y. Results for sample size $n = 500$, $B = 999$ bootstrap replications, and $S = 10,000$ Monte Carlo draws.

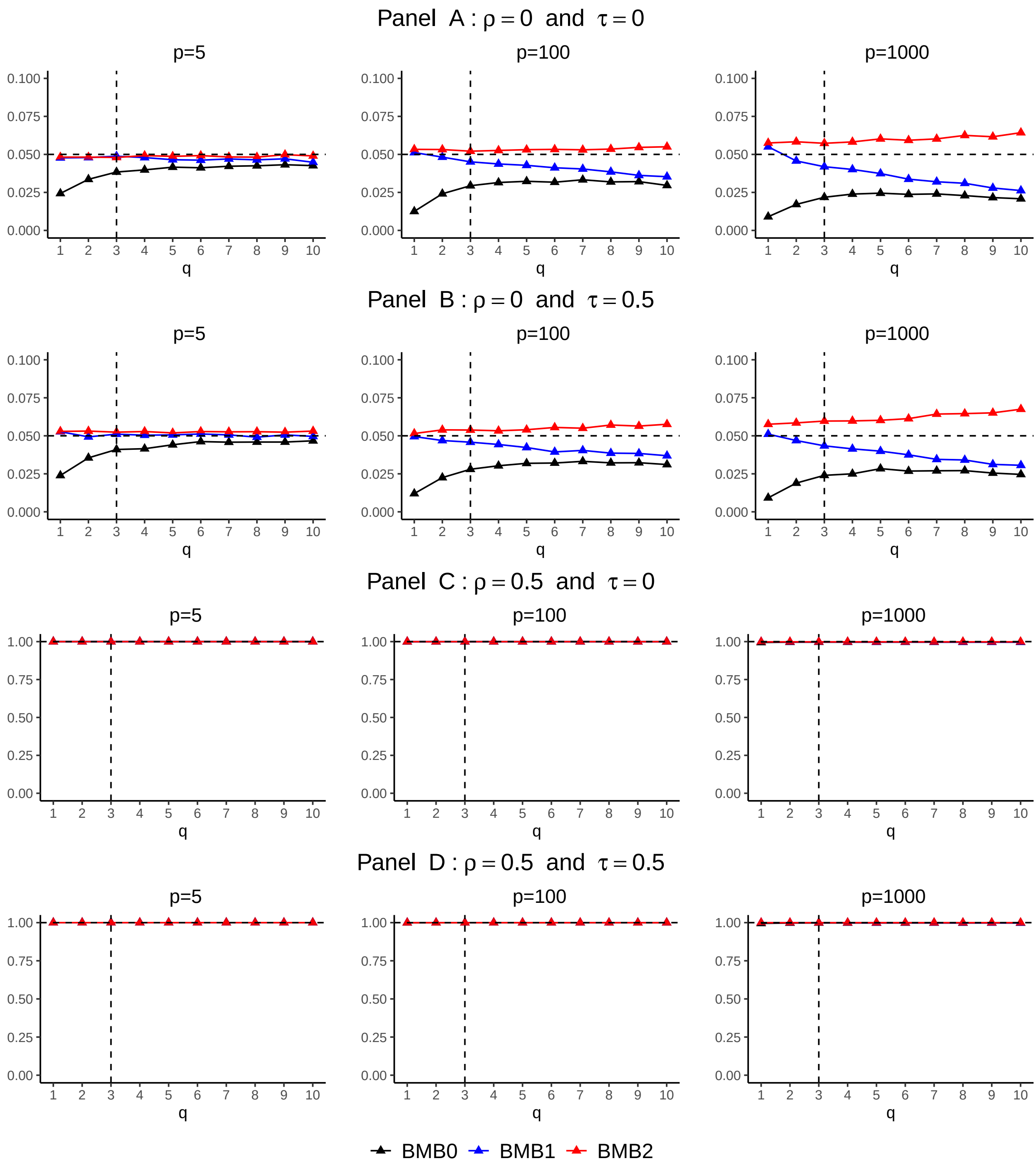


Figure 4: Rejection rates for Model 2 under different choices of the big block size q.
Notes: The tests have nominal level of 5%. The null of independence holds if $\rho = 0$. τ denotes the correlation between components of Y under the null. Results for sample size $n = 500$, $B = 999$ bootstrap replications, and $S = 10,000$ Monte Carlo draws.

4.2 Simulation II: Power curves

Since our test was seen in the previous subsection to be fairly insensitive to the particular choice of block size q, we now analyze power only for the optimal choice $q^*(n) = 3$. We use $B = 499$ bootstrap replications and $S = 5,000$ Monte Carlo draws in this subsection.

We compare our test to the distance covariance test of Székely, Rizzo, and Bakirov (2007) and its adaption to high-dimensional settings proposed by Székely and Rizzo (2013). Specifically, we use the commands `dcov.test` and `dcorT.test` from the `R` package `energy`. In Appendix C.2, we also compare our test to tests proposed in Zhu, Zhang, Yao, and Shao (2020) and Zhou, Xu, Zhu, and Li (2024), but since these are computationally demanding, we are able to only run a small simulation experiment. In these, the additional tests perform very similarly to the distance covariance tests reported in the this section. Remember from the discussion in Section 1 that all of these alternative tests were developed as tests of the hypothesis $H_0^{joint}\colon (Y_1, \ldots, Y_p) \perp X$, which implies, but is not implied by, our hypothesis H_0. We are not aware of theoretical guarantees that these tests also control the rejection probability under H_0. The results in Figure 5 show, however, that at least in these simulations, the alternative tests approximately control the rejection frequency under H_0. Remarkably, the power curves for our test are very similar for both values of τ and the power decreases only slowly with p.

The power of any test of independence depends on the type of alternatives under consideration. Under Model 1 and when the dimension p is not too large, the distance covariance tests exhibit higher power than ours. In high dimensions, however, our test dominates. One can also see that the distance covariance tests' power decreases when the $Y_1, \ldots, Y_p$ are dependent, while our tests' power is not affected by this dependence. Under Model 2, the distance covariance tests have no power while our test remains powerful with power curves that are almost insensitive to the dimension p and the degree of dependence among $Y_1, \ldots, Y_p$.

5 Empirical Application

We illustrate the practical usefulness of our proposed test by revisiting the study conducted by Hughes, DiTacchio, Hayes, Vollmers, Pulivarthy, Baggs, Panda, and Hogenesch (2009), which investigates transcriptional oscillations from mouse liver, NIH3T3, and U2OS cells. The goal is to identify genes whose transcript levels oscillate during the cell cycle. These cycles are crucial, as they play a key role in regulating metabolism and liver function, e.g., many liver genes operate on daily cycles, influencing vital processes such as detoxification, energy metabolism, and hormone regulation.

Hughes, DiTacchio, Hayes, Vollmers, Pulivarthy, Baggs, Panda, and Hogenesch (2009) collected liver tissue samples from mice at hourly intervals over a 48-hour period, pooling samples from 3-5 mice at each point. Their gene-level expression data set is available from Gene Expression Omnibus (GEO). In this section, we focus on the liver data set (accession GSE11923). We extracted the data using the `GEOquery` and `BiocManager` R Packages. The final data set contains $p = 45,101$ genes. For each gene j, we observe $n = 48$ transcript level measurements $Y_{j,1}, \ldots, Y_{j,n}$ at different points in time, recorded in $X_1, \ldots, X_n$.

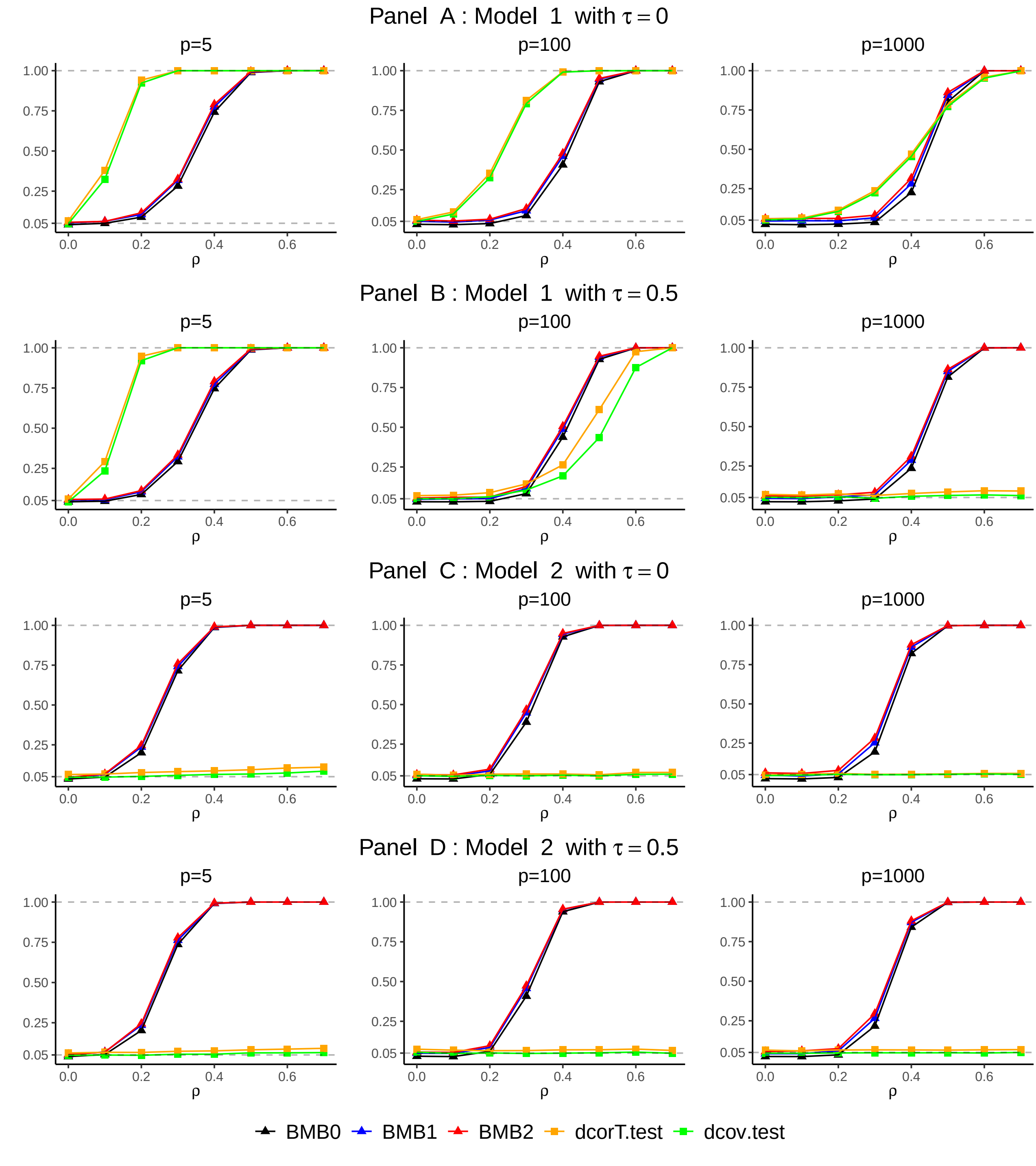


Figure 5: Power curves of the BMB for Chatterjee's rank correlation and distance covariance tests. Notes: The tests have nominal level of 5%. Results for sample size $n = 500$, $B = 499$ bootstrap replications, and $S = 5,000$ Monte Carlo draws.

The goal is to test whether gene transcriptions Y_j are independent of the time of measurement X and, in particular, to identify genes that violate independence.

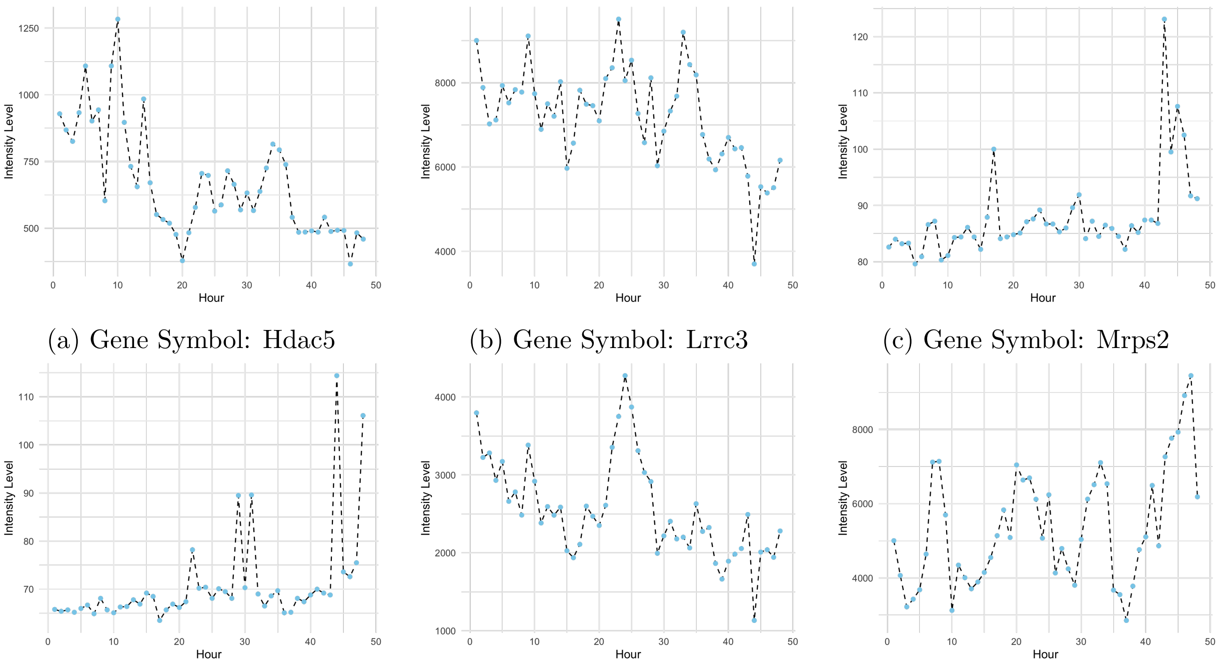
We apply our proposed bootstrap test based on the studentized test statistic BMB1 described in Remark 3, combined with the stepwise procedure developed in Section 3. We set $\alpha = 0.05$, the nominal level at which the family-wise error rate is to be controlled, and the number of bootstrap samples to $B = 1,000$. We choose the optimal block size developed in Section 2.3, which in this application is $q^*(n) = 1$.

The stepdown procedure identifies $4,554$ genes that violate the independence hypothesis. These correspond to approximately 10.1% of all gene transcripts. The stepdown procedure rejects $4,052$ hypotheses in the first and 502 in the second step, demonstrating the power gains from employing multiple steps.

The original study by Hughes, DiTacchio, Hayes, Vollmers, Pulivarthy, Baggs, Panda, and Hogenesch (2009) identified $3,667$ gene transcripts showing oscillatory behavior, but based on a different methodology for testing a different hypothesis than ours.[4] Their procedure is guaranteed to control the false discovery rate at $\alpha = 0.05$. Interestingly, our procedure finds more genes than the original study while at the same time providing stronger guarantees in the form of family-wise error rate control.

Among the genes identified by our method, $1,919$ were also identified in the original study, thus demonstrating substantial overlap. On the other hand, it singled out $2,635$ genes not identified in the original study (about 5.8% of all gene transcripts), thus highlighting its ability to detect novel rhythmic gene activity. Figure 6 shows transcript level measurements of a random sample of six genes identified by our method, but not in the original study. A complete list of identified gene transcripts is available in our replication code.

[4] The authors identify oscillatory patterns by testing for hidden periodicities in gene transcriptions using Fisher's G and Straume (2004) tests. These tests are designed to test the null that the data are generated by a Gaussian white noise process against the alternative that the data is generated by a Gaussian white noise with a deterministic sinusoidal component. The authors then obtain so-called q-values following Storey and Tibshirani (2003) so as to select gene transcripts whose q-values are less than $\alpha = 0.05$.



(a) Gene Symbol: Hdac5 (b) Gene Symbol: Lrrc3 (c) Gene Symbol: Mrps2

(d) Gene Symbol: Uap1 (e) Gene Symbol: Mir5136 (f) Gene Symbol: Impg1

Figure 6: Visualization transcript levels of 6 genes uniquely identified by our proposed stepdown method in Section 3, with $B = 1000$ Gaussian multiplier bootstrap samples, $\alpha = 0.05$, and optimal $q^*(n) = 1$.

A Proofs

A.1 Proofs for Section 2.1

The proof of Theorem 1 relies on a few auxiliary lemmas. Define

$$\Delta_1 \coloneqq \max_{1\le j\le p} |r_{j,n}| \quad \text{and} \quad \Delta_2 \coloneqq \max_{1\le j\le p} \frac{1}{mq}\sum_{k=1}^{m}(\hat{A}_{j,k} - A_{j,k})^2. \tag{6}$$

Lemma 2. *Suppose that Assumption 1 and the null hypothesis H_0 hold, and*

$$\max\left\{\log^{7/2} p, \sqrt{q}\log^{3/2} p\right\} \le C_3 n^{1/2-c_3}$$

for some positive constants c_3 and C_3. Then, there exist positive constants c, c_1, c_2, C, C_1, and C_2 depending only on c_3 and C_3 such that

(i) $\mathrm{P}\left(\Delta_1 > C_1 n^{-c_1}/\sqrt{\log p}\right) \le Cn^{-c}$,

(ii) $\mathrm{P}\left(\Delta_2 > C_2 n^{-c_2}/\log^2 p\right) \le Cn^{-c}$.

Proof. The proof is divided into two parts, one for showing (i) and one for showing (ii). In the subsequent derivations, c, c_1, c_2, C, C_1, C_2 denote generic positive constants depending only on c_3 and C_3. Their values may change from place to place.

Part (i). Note that $\hat{F}_{Y_j}(Y_{j,(i)}) = \hat{F}_{U_j}(U_{j,i})$, where $U_{j,i} \coloneqq F_{Y_j}(Y_{j,(i)})$ and $\hat{F}_{U_j}(\cdot)$ is the empirical CDF of $\{U_{j,i}\}_{i=1}^n$. Recall that under the null hypothesis of independence and Assumption 1, for each j, $\{U_{j,i}\}_{i=1}^n$ are independent random variables distributed uniformly on $[0,1]$. Following Angus (1995), we note that

$$\begin{aligned}
\hat{\xi}_j &= 1 - \frac{3n}{n^2-1}\sum_{i=1}^{n-1}\left|\hat{F}_{U_j}(U_{j,i+1}) - \hat{F}_{U_j}(U_{j,i})\right| \\
&= \frac{n}{n^2-1}\sum_{i=1}^{n-1}\left(1 - 3|U_{j,i+1}-U_{j,i}| - 3\left|\hat{F}_{U_j}(U_{j,i+1}) - \hat{F}_{U_j}(U_{j,i})\right| + 3|U_{j,i+1}-U_{j,i}|\right) + \frac{1}{n+1} \\
&= \frac{n}{n^2-1}\sum_{i=1}^{n-1}(1 - 3|U_{j,i+1}-U_{j,i}|) + \frac{1}{n+1} \\
&\quad - \frac{n}{n+1}\left(3\int_0^1\int_0^1 \left(\left|\hat{F}_{U_j}(y) - \hat{F}_{U_j}(x)\right| - |y-x|\right) H(dx,dy) + 3\tilde{I}_{j,n}\right),
\end{aligned}$$

where $H(x,y) \coloneqq xy$ and

$$\begin{aligned}
\tilde{I}_{j,n} &\coloneqq \int_0^1\int_0^1 \left(|\hat{F}_{U_j}(y) - \hat{F}_{U_j}(x)| - |y-x|\right)(\hat{H}_j(dx,dy) - H(dx,dy)), \\
\hat{H}_j(x,y) &\coloneqq \frac{1}{n-1}\sum_{i=1}^{n-1}\mathbb{1}\{U_{j,i}\le x, U_{j,i+1}\le y\}.
\end{aligned}$$

Furthermore, Angus (1995) showed that

$$\int_0^1 \int_0^1 \left(\left| \hat{F}_{U_j}(y) - \hat{F}_{U_j}(x) \right| - |y - x| \right) dydx = \frac{1}{n} \sum_{i=1}^{n} 2U_{j,i}(U_{j,i} - 1) - \frac{1}{3}.$$

It follows that

$$\hat{\xi}_j = \frac{1}{n} \sum_{i=1}^{n-1} W_{j,i} - 3\tilde{I}_{j,n} + u_{j,n},$$

where $W_{j,i} := 2 - 3|U_{j,i+1} - U_{j,i}| - 6U_{j,i}(U_{j,i} - 1)$ and $|u_{j,n}| < C/n$.

To prove part (i), it therefore suffices to show that

$$\mathrm{P}\left(\max_{1 \le j \le p} |I_{j,n}| > C_1 n^{-c_1}/\sqrt{\log p} \right) \le Cn^{-c}, \quad I_{j,n} := \sqrt{n}\tilde{I}_{j,n}. \tag{7}$$

To show that (7) holds, first note that by the union bound, for any $\varepsilon > 0$,

$$\mathrm{P}\left(\max_{1 \le j \le p} |I_{j,n}| > \varepsilon \right) \le p \max_{1 \le j \le p} \mathrm{P}\left(|I_{j,n}| > \varepsilon \right). \tag{8}$$

Next, we derive a bound on $\mathrm{P}\left(|I_{j,n}| > \varepsilon \right)$ that holds uniformly in j. Let $\hat{B}_j(x) = \sqrt{n}(\hat{F}_{U_j}(x) - x)$ and observe that

$$\begin{aligned} I_{j,n} = &\int_{0 \le y \le x \le 1} (\hat{B}_j(x) - \hat{B}_j(y))(\hat{H}_j(dx, dy) - H(dx, dy)) \\ &- \int_{0 \le x < y \le 1} (\hat{B}_j(x) - \hat{B}_j(y))(\hat{H}_j(dx, dy) - H(dx, dy)). \end{aligned} \tag{9}$$

By the Komlós-Major-Tusnády approximation, there exists a sequence of Brownian bridges B_j (Shorack and Wellner, 2009, Theorem 3, Ch. 12) such that

$$\mathrm{P}\left(\left\| \hat{B}_j(x) - B_j(x) \right\|_\infty > \varepsilon \right) \le b^{KMT} \exp\left(-c^{KMT}(\sqrt{n}\varepsilon - a^{KMT} \log n) \right) \tag{10}$$

for some positive constants a^{KMT}, b^{KMT}, and c^{KMT} and all $\varepsilon > a^{KMT} \log n/\sqrt{n}$.

By suitably adding and subtracting $B_j(x)$ and $B_j(y)$ in (9), we obtain that

$$|I_{j,n}| \le 8\left\| \hat{B}_j(x) - B_j(x) \right\|_\infty + \left| \int_0^1 \int_0^1 B_j(x) d\tilde{h}(x, y) \right| + \left| \int_0^1 \int_0^1 B_j(y) d\tilde{h}(x, y) \right|, \tag{11}$$

where

$$\tilde{h}(x, y) = \begin{cases} \hat{H}_j(x, y) - H(x, y) \text{ if } y \le x, \\ H(x, y) - \hat{H}_j(x, y) \text{ if } x < y. \end{cases}$$

We now derive a bound on the second and third term on the right-hand side in (11). Fix a $\delta > 0$ and consider a partition $\mathcal{S}$ of the interval $[0, 1]$ into $\lceil 1/\delta \rceil$ intervals of length at most δ. Let B_j^δ be a process with piecewise-constant paths on the subintervals in $\mathcal{S}$ such that for any $S \in \mathcal{S}$ and $x \in S$, $B_j^\delta(x) = \inf_{y \in S} B_j(y)$. Recall that all B_j are standard Brownian bridges on $[0, 1]$, and therefore their distribution is the same for all j. By Levy's modulus of continuity theorem for the Brownian bridge (Shorack and Wellner, 2009, Theorem 1, Ch. 14), with probability one, it holds that

$$\left\|B_j^\delta(x) - B_j(x)\right\|_\infty \leq 2\sqrt{\delta \log(1/\delta)}.$$

It follows that

$$\begin{aligned}\left|\int_0^1\int_0^1 B_j(x) d\tilde{h}(x,y)\right| &= \Big|\int_0^1\int_0^1 (B_j(x) - B_j^\delta(x)) d\tilde{h}(x,y) + \int_0^1\int_0^1 B_j^\delta(x) d\tilde{h}(x,y)\Big| \\ &\leq 4\sqrt{\delta\log(1/\delta)} + \frac{2}{\delta}\|B_j\|_\infty \left\|\hat{H}_j - H\right\|_\infty.\end{aligned}$$

Using the union bound, we obtain that

$$\begin{aligned}\mathrm{P}\left(\frac{1}{\delta}\|B_j\|_\infty\left\|\hat{H}_j - H\right\|_\infty \geq \varepsilon\right) &\leq \mathrm{P}\left(\left\{\|B_j\|_\infty \geq \varepsilon^{1/2}\delta^{1/2}n^{1/4}\right\} \cup \left\{\left\|\hat{H}_j - H\right\|_\infty \geq \varepsilon^{1/2}\delta^{1/2}n^{-1/4}\right\}\right) \\ &\leq \mathrm{P}\left(\|B_j\|_\infty \geq \varepsilon^{1/2}\delta^{1/2}n^{1/4}\right) + \mathrm{P}\left(\left\|\hat{H}_j - H\right\|_\infty \geq \varepsilon^{1/2}\delta^{1/2}n^{-1/4}\right).\end{aligned}$$

By classical results for the Brownian bridge (e.g. Adler and Brown, 1986, Theorem 4.1),

$$\mathrm{P}\left(\|B_j\|_\infty \geq \lambda\right) \leq C\exp(-c\lambda^2)$$

for all $\lambda > 0$. By Lemma 8, which is a simple implication of the two-dimensional version Dvoretsky-Kiefer-Wolfowitz inequality (Kiefer and Wolfowitz, 1958),

$$\mathrm{P}\left(\sqrt{n}\left\|\hat{H}_j - H\right\|_\infty \geq \lambda\right) \leq C\exp(-c\lambda^2).$$

for all $\lambda > 0$. In total, we obtain that

$$\mathrm{P}\left(\frac{1}{\delta}\|B_j\|_\infty\left\|\hat{H}_j - H\right\|_\infty \geq \varepsilon\right) \leq C^{BH}\exp(-c^{BH}\varepsilon\delta\sqrt{n}). \tag{12}$$

for some positive constants c^{BH} and C^{BH}.

Using (11) and the subsequent steps, we obtain that

$$|I_{j,n}| \leq 8\left\|\hat{B}_j(x) - B_j(x)\right\|_\infty + 8\sqrt{\delta\log(1/\delta)} + \frac{4}{\delta}\|B_j\|_\infty\left\|\hat{H}_j - H\right\|_\infty.$$

Using the union bound, we have that

$$\mathrm{P}\left(|I_{j,n}| > \varepsilon\right) \leq \mathrm{P}\left(8\left\|\hat{B}_j(x) - B_j(x)\right\|_\infty > \frac{\varepsilon}{3}\right) + \mathrm{P}\left(8\sqrt{\delta\log(1/\delta)} > \frac{\varepsilon}{3}\right) + \mathrm{P}\left(\frac{4}{\delta}\|B_j\|_\infty\left\|\hat{H}_j - H\right\|_\infty > \frac{\varepsilon}{3}\right).$$

Now, let $\varepsilon = C_1 n^{-c_1}/\sqrt{\log p}$ and take $\delta = C_\delta n^{-r}$ for some $r > 1/7$. Recall that, by assumption, $\log p \leq C_3 n^a$ for some $a < 1/7$. With this choice of r, we can choose C_δ sufficiently large and c_1

sufficiently small such that $8\sqrt{\delta\log(1/\delta)} < \varepsilon/3$. Using (10), (12), we have that

$$\begin{aligned}
\mathrm{P}\left(|I_{j,n}| > C_1 n^{-c_1}/\sqrt{\log p}\right) &\le \mathrm{P}\left(|I_{j,n}| > C_1 C_3^{-1/2} n^{-c_1-a/2}\right) \\
&\le b^{KMT}\exp(-c^{KMT}(C_1C_3^{-1/2}n^{1/2-c_1-a/2}/24 - a^{KMT}\log n)) \\
&\quad + C^{BH}\exp(-c^{BH}C_1C_3^{-1/2}C_\delta n^{1/2-c_1-r-a/2}/12) \\
&\le C\exp(-cn^{1/2-c_1-r-a/2}).
\end{aligned}$$

Combining this result with the union bound in (8), we obtain that

$$\begin{aligned}
\mathrm{P}\left(\max_{1\le j\le p}|I_{j,n}| > C_1 n^{-c_1}/\sqrt{\log p}\right) &\le p\,C\exp(-cn^{1/2-c_1-r-a/2}) \\
&\le C\exp(n^a - cn^{1/2-c_1-r-a/2}).
\end{aligned}$$

Since $a < 1/7$, we can choose c_1 sufficiently small and r sufficiently close to $1/7$ such that $a < 1/2 - c_1 - r - a/2$. Hence, we can choose positive constants c, c_1, C, and C_1 such that part (i) holds.

Part (ii). First, note that

$$|\hat{W}_{j,l} - W_{j,l}| \le CD_j, \quad D_j := \sup_{t\in\mathbb{R}}\left|\hat{F}_j(t) - F_j(t)\right|.$$

Further,

$$(A_{j,k} - \hat{A}_{j,k})^2 = \left(\sum_{l=(k-1)(q+1)+1}^{kq+(k-1)}\left(\hat{W}_{j,l} - W_{j,l}\right)\right)^2 \le Cq^2D_j^2.$$

It follows that

$$\Delta_2 = \max_{1\le j\le p}\frac{1}{mq}\sum_{k=1}^{m}(A_{j,k} - \hat{A}_{j,k})^2 \le Cq\max_{1\le j\le p}D_j^2.$$

Thus, using the union bound and the DKW inequality, we have that for any $c_2, C_2 > 0$,

$$\begin{aligned}
\mathrm{P}(\Delta_2 > C_2n^{-c_2}/\log^2 p) &\le p\max_{1\le j\le p}\mathrm{P}\left(CqD_j^2 > C_2n^{-c_2}/\log^2 p\right) \\
&\le 2p\exp\left(-c\frac{n^{1-c_2}}{q\log^2 p}\right) \\
&\le 2\exp\left(\log p - c\frac{n^{1-c_2}}{q\log^2 p}\right).
\end{aligned}$$

Note that

$$\log p - c\frac{n^{1-c_2}}{q\log^2 p} = \left(1 - c\frac{n^{1-c_2}}{q\log^3 p}\right)\log p \le -Cn^c,$$

provided that c_2 is chosen small enough. It then follows that $\mathrm{P}(\Delta_2 > C_2n^{-c_2}/\log^2 p) \le Cn^{-c}$. Q.E.D.

Lemma 3. *Suppose that Assumption 1 holds and there exist constants $C_1 > 0$ and $0 < \gamma < 1/4$ such that $(1/q)\log^2 p \le C_1n^{-\gamma}$ and $\max\{q\log^{1/2}p, \log^{3/2}p, \sqrt{q}\log^{7/2}(pn)\} \le C_1n^{1/2-\gamma}$. Then there exist* *constants $c, C > 0$ depending only on γ and C_1 such that, under the null hypothesis H_0,*

$$\sup_{t \in \mathbb{R}} |\mathrm{P}(T_0 \leq t) - \mathrm{P}(Z_0 \leq t)| \leq C n^{-c} . \tag{13}$$

Remark 5. Lemma 3 is similar to Theorem E.1 in Chernozhukov, Chetverikov, and Kato (2019), but it differs from the latter in two ways: the mixing rate does not appear in our bound and we do not need to assume that the small blocks grow with the sample size. This difference stems from our derivations directly imposing the 1-dependence of $\{W_{j,i}\}_{i=1}^n$ rather than a general mixing condition. ■

Proof. In this proof, c, C denote generic positive constants depending only on γ and C_1; their values may change from place to place. We divide the proof into several steps. In the first three steps, we show that

$$-Cn^{-c} \leq \mathrm{P}(T_0 \leq t) - \mathrm{P}\left(\max_{1 \leq j \leq p} \sqrt{\frac{mq}{n}}\, V_j \leq t\right) \leq Cn^{-c} \tag{14}$$

for all $t \in \mathbb{R}$. For $\star \in \{+, -\}$, consider the following decomposition

$$\begin{aligned}
&\mathrm{P}(T_0 \leq t) - \mathrm{P}\left(\max_{1 \leq j \leq p} \sqrt{\frac{mq}{n}}\, V_j \leq t\right) \\
&\quad = \mathrm{P}(T_0 \leq t) - \mathrm{P}\left(\max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{k=1}^{m} A_{j,k} \leq t \star \frac{C}{n^c \sqrt{\log p}}\right) && \text{(i)} \\
&\qquad + \mathrm{P}\left(\max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{k=1}^{m} A_{j,k} \leq t \star \frac{C}{n^c \sqrt{\log p}}\right) - \mathrm{P}\left(\max_{1 \leq j \leq p} \sqrt{\frac{mq}{n}} V_j \leq t \star \frac{C}{n^c \sqrt{\log p}}\right) && \text{(ii)} \\
&\qquad + \mathrm{P}\left(\max_{1 \leq j \leq p} \sqrt{\frac{mq}{n}} V_j \leq t \star \frac{C}{n^c \sqrt{\log p}}\right) - \mathrm{P}\left(\max_{1 \leq j \leq p} \sqrt{\frac{mq}{n}} V_j \leq t\right). && \text{(iii)}
\end{aligned}$$

In steps 1–3, we show that each of the terms in (i)–(iii) is bounded in the supremum norm (w.r.t. t) by Cn^{-c}. To show the first and second inequality in (14), we use $\star = -$ and $\star = +$, respectively.

Step 1. (Reduction to independence). We wish to show that

$$\begin{aligned}
\mathrm{P}\left(\max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{k=1}^{m} A_{j,k} \leq t - Cn^{-c}/\sqrt{\log p}\right) - n^{-c} &\leq \mathrm{P}(T_0 \leq t) \\
&\leq \mathrm{P}\left(\max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{k=1}^{m} A_{j,k} \leq t + Cn^{-c}/\sqrt{\log p}\right) + n^{-c}.
\end{aligned}$$

We only prove the second inequality; the first inequality follows from an analogous argument. Recall that $\sum_{i=1}^{n-1} W_{j,i} = \sum_{k=1}^{m} A_{j,k} + \sum_{k=1}^{m} B_{j,k} + R_j$, so that

$$\left|\max_{1 \leq j \leq p} \sum_{i=1}^{n-1} W_{j,i} - \max_{1 \leq j \leq p} \sum_{k=1}^{m} A_{j,k}\right| \leq \max_{1 \leq j \leq p} \left|\sum_{k=1}^{m} B_{j,k}\right| + \max_{1 \leq j \leq p} |R_j|.$$

Hence for every $\delta_1, \delta_2 > 0$,

$$\mathrm{P}(T_0 \le t) \le \mathrm{P}\left(\max_{1\le j\le p} \frac{1}{\sqrt{n}}\sum_{k=1}^{m} A_{j,k} \le t+\delta_1+\delta_2\right) + \mathrm{P}\left(\frac{1}{\sqrt{n}}\max_{1\le j\le p}\left|\sum_{k=1}^{m} B_{j,k}\right| > \delta_1\right)$$
$$+\mathrm{P}\left(\max_{1\le j\le p}|R_j| > \sqrt{n}\delta_2\right)$$
$$=: I + II + III.$$

This inequality holds because the event on the LHS implies that at least one of the events on the RHS holds.

Since $|W_{j,i}| \le 2$ for all j and i, we have $|R_j| \le 2(q+1)$ a.s., uniformly in j. Thus, take $\delta_2 = 4qn^{-1/2}$ ($\le Cn^{-c}/\sqrt{\log p}$) and conclude $III = 0$. Moreover, for every $\varepsilon > 0$, Markov's inequality and $\delta_1 = \varepsilon^{-1}\,\mathbb{E}(n^{-1/2}\max_{1\le j\le p}|\sum_{k=1}^{m} B_{j,k}|)$ imply $II \le \varepsilon$. This way, it remains to bound the magnitude of $\mathbb{E}(n^{-1/2}\max_{1\le j\le p}|\sum_{k=1}^{m} B_{j,k}|)$. To this end, we note that $|B_{j,k}| \le 2$ a.s. uniformly in j and k, and that $\max_j \sum_{k=1}^{m}\mathrm{Var}(B_{j,k})/m$ is bounded above since $\mathrm{Var}\,(B_{j,k}) = 1/2$, independent of j. Thus, independence of $\{B_{j,k}\}_{k=1}^{m}$ and Lemma 6 imply

$$\mathbb{E}\left(\max_{1\le j\le p}\left|\frac{1}{\sqrt{n}}\sum_{l=1}^{m} B_{j,k}\right|\right) \le K\left(\sqrt{\frac{\log p}{q}} + \frac{\log p}{\sqrt{n}}\right),$$

where K is universal (here we have used the simple fact that $m/n \le 1/q$), so that the left side is bounded by $Cn^{-2c}/\sqrt{\log p}$ (by taking c sufficiently small). The conclusion of this step follows from taking $\varepsilon = n^{-c}$ so that $\delta_1 \le Cn^{-c}/\sqrt{\log p}$.

Step 2. (Normal approximation to the sum of independent blocks). We wish to show that

$$\sup_{t\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{k=1}^{m} A_{j,k} \le t\right) - \mathrm{P}\left(\max_{1\le j\le p}\sqrt{\frac{mq}{n}}V_j \le t\right)\right| \le Cn^{-c}.$$

Since $\{A_{j,k}\}_{k=1}^{m}$ are independent, we may apply Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2013) (note that the covariance matrix of $\sqrt{mq/n}\,V$ is the same as that of $n^{-1/2}\sum_{k=1}^{m} A_{k,j}$). To this end, we seek to verify the conditions of the corollary applied to our case. Observe that

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{m} A_{j,k} = \frac{1}{\sqrt{m}}\sum_{k=1}^{m}\frac{A_{j,k}}{\sqrt{n/m}},$$

and $\sqrt{q} \le \sqrt{n/m} \le 2\sqrt{q}$ (recall that $q+1 \le (n-1)/2$). For any $1 \le q \le n$, let

$$\bar{\sigma}^2(q) = \max_{1\le j\le p}\max_{I}\mathrm{Var}\left(\frac{1}{\sqrt{q}}\sum_{i\in I} W_{ij}\right) \quad \text{and} \quad \underline{\sigma}^2(q) = \min_{1\le j\le p}\min_{I}\mathrm{Var}\left(\frac{1}{\sqrt{q}}\sum_{i\in I} W_{ij}\right),$$

where the $\max_I$ and $\min_I$ are taken over all $I \subset \{1,\dots,n\}$ of the form $I = \{i+1,\dots,i+q\}$. Angus (1995) shows that $\mathbb{E}(W_{j,i}) = 0$, $\mathrm{Var}(W_{j,i}) = 1/2$, and $Cov(W_{j,i}, W_{j,i+k}) = -1/20$ for $k = 1$ (0 otherwise). Therefore,

$$\operatorname{Var}\left(\frac{1}{\sqrt{q}}\sum_{i\in I}W_{i,j}\right)=\frac{1}{q}\left(\sum_{i\in I}\operatorname{Var}(W_{i,j})+2\sum_{i=1}^{q-1}\operatorname{Cov}(W_{j,i},W_{j,i+1})\right)=\frac{2}{5}+\frac{1}{10q}\ ,$$

so that the variance in the previous display is a monotone decreasing function in q, independent of j with maximal value $1/2$ at $q=1$ and bounded from below by $2/5$. Therefore $\bar{\sigma}^2(q)=\underline{\sigma}^2(q)=2/5+1/(10q)$. Collect these facts to conclude that

$$\frac{2}{5}\frac{1}{4}\leq\frac{\underline{\sigma}^2(q)}{4}\leq\operatorname{Var}\left(\frac{A_{j,k}}{\sqrt{n/m}}\right)\leq\bar{\sigma}^2(q)\leq\frac{1}{2},\quad 1\leq j\leq p\ ,$$

and $\left|A_{j,k}/\sqrt{n/m}\right|\leq 2\sqrt{q}$ a.s. so the conditions of Corollary 2.1 (i) in Chernozhukov, Chetverikov, and Kato (2013) are verified with $B_n=2\sqrt{q}$, which leads to the assertion of this step (note that $C^{-1}n^c\leq n/(4q)\leq m$).

Step 3. (Anti-concentration). We wish to verify that, for every $\varepsilon>0$,

$$\sup_{t\in\mathbb{R}}\mathrm{P}\left(\left|\max_{1\leq j\leq p}V_j-t\right|\leq\varepsilon\right)\leq C\varepsilon\sqrt{\max\{1,\log(p/\varepsilon)\}}.\tag{15}$$

Indeed, since V is a normal random vector with

$$\frac{2}{5}\leq\underline{\sigma}^2(q)\leq\operatorname{Var}(V_j)\leq\bar{\sigma}^2(q)\leq\frac{1}{2},\quad 1\leq j\leq p\ ,$$

the desired result follows by Corollary 1 in Chernozhukov, Chetverikov, and Kato (2015) restated in Lemma 5. We apply this result with $\varepsilon=C/(2n^c\sqrt{\log p})(mq/n)^{-1/2}$ and conclude that

$$\sup_{t\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\leq j\leq p}\sqrt{\frac{mq}{n}}V_j\leq t\star\frac{C}{n^c\sqrt{\log p}}\right)-\mathrm{P}\left(\max_{1\leq j\leq p}\sqrt{\frac{mq}{n}}V_j\leq t\right)\right|\leq Cn^{-c}\sqrt{\log n}\leq Cn^{-c'}$$

for some c'.

Step 4. (Conclusion). By Steps 1–3, we have

$$\sup_{t\in\mathbb{R}}\left|\mathrm{P}(T_0\leq t)-\mathrm{P}\left(\max_{1\leq j\leq p}\sqrt{\frac{mq}{n}}V_j\leq t\right)\right|\leq Cn^{-c}.$$

It remains to replace $\sqrt{(mq)/n}$ by 1 on the left side. Observe that

$$1-\sqrt{\frac{mq}{n}}\leq 1-\frac{mq}{n}\leq 1-\left(\frac{n}{q+1}-1\right)\left(\frac{1}{n}\right)=\frac{1}{q+1}+\frac{1}{n},$$

and the right side is bounded by $Cn^{-c}/\sqrt{\log p}$. With this c, by Markov's inequality,

$$\mathrm{P}\left(\left|\max_{1\leq j\leq p}V_j\right|>n^{c/2}\sqrt{\log p}\right)\leq Cn^{-c/2},$$

as $\mathbb{E}(|\max_{1\le j\le p} V_j|) \le C\sqrt{\log p}$, so that with probability larger than $1 - Cn^{-c/2}$,

$$\left(1-\sqrt{\frac{mq}{n}}\right)\left|\max_{1\le j\le p} V_j\right| \le C'n^{-c/2}\log^{-1/2} p.$$

Apply the anti-concentration property of $\max_{1\le j\le p} V_j$ (see Step 3) to conclude

$$\sup_{t\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\le j\le p}\sqrt{\frac{mq}{n}}\,V_j \le t\right) - \mathrm{P}\left(\max_{1\le j\le p} V_j \le t\right)\right| \le Cn^{-c}.$$

This finishes the proof of the lemma. Q.E.D.

Lemma 4. *Let $\varepsilon_1,\dots,\varepsilon_m$ be independent standard normal random variables, independent of the data $\mathbb{D}$. Suppose that there exist constants $C_1 > 0$ and $0 < \gamma < 1/2$ such that $q\log^{5/2} p \le C_1 n^{1/2-\gamma}$. Then, under Assumption 1 and the null hypothesis H_0, there exist constants $c, c', C, C' > 0$ depending only on γ and C_1 such that, with probability larger than $1 - Cn^{-c}$,*

$$\sup_{t\in\mathbb{R}}\left|\mathrm{P}\left(T_0^B \le t\,|\mathbb{D}\right) - \mathrm{P}\left(Z_0 \le t\right)\right| \le C'n^{-c'}.$$

Proof. Here c, c', C, C' denote generic positive constants depending only on γ; their values may change from place to place. By Theorem 2 in Chernozhukov, Chetverikov, and Kato (2015), the left side in the conclusion of the lemma is bounded by $CD^{1/3}\max\{1, \log(p/D)\}^{2/3}$, where

$$D = \max_{1\le j,j'\le p}\left|\frac{1}{mq}\sum_{k=1}^{m}(A_{j,k}A_{j',k} - \mathbb{E}[A_{j,k}A_{j',k}])\right|.$$

Hence, it suffices to prove that $\mathrm{P}(D > C'n^{-c'}\log^{-2} p) \le Cn^{-c}$ with suitable c, c', C, C'. Observe that $|A_{j,k}A_{j',k}| \le 4q^2$ a.s. and $\mathbb{E}[(A_{j,k}A_{j',k})^2] \le 4q^3\,\bar{\sigma}^2(q)$, where $\bar{\sigma}^2(q)$ is given as in Step 2 in the proof of Lemma 3. Hence, by Lemma 6, we have

$$\begin{aligned}\mathbb{E}[D] &\le C\left(\sqrt{4mq^3\bar{\sigma}^2(q)}\frac{\sqrt{\log p}}{mq} + 4q^2\frac{\log p}{mq}\right),\\ &\le C\left(q\sqrt{\frac{\log p}{mq}} + q^2\frac{\log p}{mq}\right)\end{aligned}$$

Since $mq > n/2$, it follows that

$$\mathbb{E}[D] \le C\left(q\sqrt{\frac{\log p}{n}} + q^2\frac{\log p}{n}\right).$$

Since $q\log^{5/2} p \le C_1 n^{1/2-\gamma}$, the right side is bounded by $C'n^{-\gamma}\log^{-2} p$. The conclusion of the lemma follows from application of Markov's inequality. Q.E.D.

Proof of Theorem 1. Let c and C denote generic positive constants depending only on γ and C_1. Their values may change from place to place. The proof consists of several steps that rely on Lemmas 2, 3, and 4.

Step 1: We will show that

$$\mathrm{P}\left(\left|\hat{T} - T_0\right| > \zeta_n\right) \le Cn^{-c} \tag{16}$$
$$\mathrm{P}\left(\mathrm{P}\left(\left|\hat{T}^B - T_0^B\right| > \zeta_n | \mathbb{D}\right) > Cn^{-c}\right) \le Cn^{-c} \ . \tag{17}$$

for some sequence $\zeta_n \le Cn^{-c}/\sqrt{\log p}$ for sufficiently small $c > 0$ and large $C > 0$ depending only on the constants c_1, c_2, C_1, and C_2 from Lemma 2. Note the assumptions of Lemma 2 are satisfied under the conditions of Theorem 1 because $\max\{\log^{7/2} p, \sqrt{q}\log^{3/2} p\} \le C_3 n^{1/2-c_3}$ follows from $\sqrt{q}\log^{7/2}(pn) \le C_1 n^{1/2-\gamma}$ with $C_3 = C_1$ and $c_3 = \gamma$.

To show (16), observe that

$$\left|\hat{T} - T_0\right| \le \max_{1\le j\le p}\left|\hat{\xi}_j - \frac{1}{\sqrt{n}}\sum_{i=1}^{n-1} W_{j,i}\right| \le \max_{1\le j\le p}|r_{j,n}| = \Delta_1 \ .$$

The claim follows from Lemma 2 by taking $\zeta_n \ge C_1 n^{-c_1}/\sqrt{\log p}$.

To show (17), note first that

$$\left|\hat{T}^B - T_0^B\right| \le \max_{1\le j\le p}\left|\frac{1}{\sqrt{mq}}\sum_{k=1}^{m}\varepsilon_k\left(\hat{A}_{j,k} - A_{j,k}\right)\right|.$$

Conditional on the data $\mathbb{D}$, the vector $\left(\frac{1}{\sqrt{mq}}\sum_{k=1}^{m}\varepsilon_k(\hat{A}_{j,k} - A_{j,k})\right)_{1\le j\le p}$ is normal with mean zero and all diagonal elements of the covariance matrix bounded by

$$\max_{1\le j\le p}\frac{1}{mq}\sum_{k=1}^{m}\left(\hat{A}_{j,k} - A_{j,k}\right)^2 . \tag{18}$$

The last expression is bounded by $C_2 n^{-c_2}/\log^2 p$ with probability larger than $1 - Cn^{-c}$ by Lemma 2. Conditional on this event,

$$\mathbb{E}\left[\left|\hat{T}^B - T_0^B\right| | \mathbb{D}\right] \le \sqrt{\frac{C_2 n^{-c_2}}{\log^2 p}} \cdot 2\sqrt{2\log p} \le \frac{\zeta_n}{2}.$$

Further, by Borell's inequality,

$$\mathrm{P}\left(\left|\hat{T}^B - T_0^B\right| > \zeta_n | \mathbb{D}\right) \le \exp\left(-\frac{\zeta_n^2 \log^2 p}{8C_2 n^{-c_2}}\right) \le \exp(-Cn^c) \le Cn^{-c}.$$

This yields (17). In the following derivations, ζ_n is as selected in this step.

Step 2: Let $c(\alpha)$ denote the $(1-\alpha)$-quantile of Z_0. In this step, we show that

$$\mathrm{P}\left\{\hat{c}(\alpha) \ge c(\alpha + Cn^{-c} + 20\zeta_n\sqrt{\log p})\right\} \ge 1 - Cn^{-c}, \tag{19}$$
$$\mathrm{P}\left\{\hat{c}(\alpha) \le c(\alpha - Cn^{-c} - 20\zeta_n\sqrt{\log p})\right\} \ge 1 - Cn^{-c}, \tag{20}$$

and that, for any $\gamma \in (0, 1 - 20\zeta_n\sqrt{\log p})$,

$$c(\gamma + 20\zeta_n\sqrt{\log p}) + \zeta_n \le c(\gamma). \tag{21}$$

We first show (21). By Theorem 3 in Chernozhukov, Chetverikov, and Kato (2015), for any $t \in \mathbb{R}$ and any $\epsilon > 0$,

$$\mathrm{P}(|Z_0 - t| \le \epsilon) \le 4\epsilon(\sqrt{2\log p} + 1)/\sigma_V,$$

where $\sigma_V^2 \coloneqq \mathbb{E}[V_j^2]$. Since $\{A_{j,k} \colon j = 1, \dots, p, k = 1, \dots, m\}$ are i.i.d., $V_1, \dots, V_p$ are also i.i.d., and

$$\sigma_V^2 = \frac{1}{mq}\sum_{k=1}^{m}\mathbb{E}[A_{j,k}^2] = \frac{1}{q}\mathbb{E}[A_{j,k}^2]$$

with $2/5 \le E[A_{j,k}^2]/q \le 1/2$, as shown in the proof of Lemma 3. Therefore,

$$\begin{aligned}
&\mathrm{P}(Z_0 \le c(\gamma + 20\zeta_n\sqrt{\log p}) + \zeta_n) - \mathrm{P}(Z_0 \le c(\gamma + 20\zeta_n\sqrt{\log p})) \\
&\quad = \mathrm{P}(c(\gamma + 20\zeta_n\sqrt{\log p}) \le Z_0 \le c(\gamma + 20\zeta_n\sqrt{\log p}) + \zeta_n) \\
&\quad = \mathrm{P}(-\zeta_n/2 \le Z_0 - [c(\gamma + 20\zeta_n\sqrt{\log p}) + \zeta_n/2] \le \zeta_n/2) \\
&\quad \le 2\zeta_n(\sqrt{2\log p} + 1)/\sigma_V
\end{aligned}$$

This implies

$$\begin{aligned}
\mathrm{P}(Z_0 \le c(\gamma + 20\zeta_n\sqrt{\log p}) + \zeta_n) &\le \mathrm{P}(Z_0 \le c(\gamma + 20\zeta_n\sqrt{\log p})) + 2\zeta_n(\sqrt{2\log p} + 1)/\sigma_V \\
&\le \mathrm{P}(Z_0 \le c(\gamma + 20\zeta_n\sqrt{\log p})) + 8\zeta_n\sqrt{\log p}\cdot\frac{5}{2} \\
&\le 1 - \gamma - 20\zeta_n\sqrt{\log p} + 20\zeta_n\sqrt{\log p}) \\
&= 1 - \gamma
\end{aligned}$$

and the desired result in (21) follows.

Next, we show (19). For any $t \in \mathbb{R}$, by the union bound,

$$\begin{aligned}
\mathrm{P}(\hat{T}^B \le t|\mathbb{D}) &\le \mathrm{P}(T_0^B \le t + \zeta_n|\mathbb{D}) + \mathrm{P}(|\hat{T}^B - T_0^B| > \zeta_n|\mathbb{D}) \\
&\le \mathrm{P}(Z_0 \le t + \zeta_n) + \rho_n^B + \mathrm{P}(|\hat{T}^B - T_0^B| > \zeta_n|\mathbb{D})
\end{aligned}$$

where

$$\rho_n^B \coloneqq \sup_{t\in\mathbb{R}}\left|\mathrm{P}\left(T_0^B \le t\,|\mathbb{D}\right) - \mathrm{P}\left(Z_0 \le t\right)\right|.$$

Therefore, with probability at least $1 - Cn^{-c}$,

$$\begin{aligned}
\mathrm{P}(\hat{T}^B \le c(\alpha + Cn^{-c} + 20\zeta_n\sqrt{\log p})|\mathbb{D}) &\le \mathrm{P}(Z_0 \le c(\alpha + Cn^{-c} + 20\zeta_n\sqrt{\log p}) + \zeta_n) + Cn^{-c} \\
&\le \mathrm{P}(Z_0 \le c(\alpha + Cn^{-c}) + Cn^{-c} \\
&\le 1 - \alpha - Cn^{-c} + Cn^{-c} \\
&= 1 - \alpha
\end{aligned}$$

where the first inequality follows from Lemma 4 and (17), and the second follows from (21).

Since $\mathrm{P}(\hat{T}^B \leq \hat{c}(\alpha)|\mathbb{D}) = 1-\alpha$, the inequality implies that with probability at least $1-Cn^{-c}$, $\hat{c}(\alpha) \geq c(\alpha + Cn^{-c} + 20\zeta_n\sqrt{\log p})$, which establishes (19). The second inequality (20) can be shown in an analogous fashion.

Step 3: This step combines results from Step 1, Step 2, and Lemmas 3-4 to prove the statement of the theorem. Define

$$\rho_n := \sup_{t\in\mathbb{R}} |\mathrm{P}\left(T_0 \leq t\right) - \mathrm{P}\left(Z_0 \leq t\right)|.$$

Then:

$$\begin{aligned}
\mathrm{P}\left(\hat{T} > \hat{c}(\alpha)\right) &\leq \mathrm{P}\left(T_0 + |\hat{T} - T_0| > \hat{c}(\alpha)\right) \\
&\leq \mathrm{P}\left(\{T_0 + \zeta_n > \hat{c}(\alpha)\} \cup \left\{|\hat{T} - T_0| > \zeta_n\right\}\right) \\
&\leq \mathrm{P}\left(T_0 + \zeta_n > \hat{c}(\alpha)\right) + \mathrm{P}\left(|\hat{T} - T_0| > \zeta_n\right) \\
&= \mathrm{P}\left(T_0 + \zeta_n > \hat{c}(\alpha), \hat{c}(\alpha) \geq c(\alpha + Cn^{-c} + 20\zeta_n\sqrt{\log p})\right) \\
&\quad + \mathrm{P}\left(T_0 + \zeta_n > \hat{c}(\alpha)|\hat{c}(\alpha) < c(\alpha + Cn^{-c} + 20\zeta_n\sqrt{\log p})\right) \times \\
&\qquad \times \mathrm{P}\left(\hat{c}(\alpha) < c(\alpha + Cn^{-c} + 20\zeta_n\sqrt{\log p})\right) \\
&\quad + \mathrm{P}\left(|\hat{T} - T_0| > \zeta_n\right) \\
&\leq \mathrm{P}\left(T_0 + \zeta_n > c(\alpha + Cn^{-c} + 20\zeta_n\sqrt{\log p})\right) + Cn^{-c} \\
&\leq \mathrm{P}\left(T_0 > c(\alpha + Cn^{-c} + 40\zeta_n\sqrt{\log p})\right) + Cn^{-c} \\
&\leq \mathrm{P}\left(Z_0 > c(\alpha + Cn^{-c} + 40\zeta_n\sqrt{\log p})\right) + \rho_n + Cn^{-c} \\
&\leq \alpha + Cn^{-c} + 40\zeta_n\sqrt{\log p} + \rho_n + Cn^{-c} \\
&\leq \alpha + Cn^{-c},
\end{aligned}$$

where the third inequality follows from the union bound, the fourth inequality follows from (16) and (19), the fifth inequality follows from (21), and the last inequality follows from Lemma 3 and $40\zeta_n\sqrt{\log p} \leq Cn^{-c}$.

Similarly,

$$\begin{aligned}
\mathrm{P}\left(\hat{T} > \hat{c}(\alpha)\right) &\geq \mathrm{P}\left(T_0 > \hat{c}(\alpha) + |\hat{T} - T_0|\right) \\
&\geq \mathrm{P}\left(T_0 > \hat{c}(\alpha) + \zeta_n\right) - \mathrm{P}\left(|\hat{T} - T_0| > \zeta_n\right) \\
&\geq \mathrm{P}\left(T_0 > c(\alpha - Cn^{-c} - 20\zeta_n\sqrt{\log p}) + \zeta_n\right) - Cn^{-c} \\
&\geq \mathrm{P}\left(T_0 > c(\alpha - Cn^{-c} - 40\zeta_n\sqrt{\log p})\right) - Cn^{-c} \\
&\geq \mathrm{P}\left(Z_0 > c(\alpha - Cn^{-c} - 40\zeta_n\sqrt{\log p})\right) - \rho_n - Cn^{-c} \\
&\geq \alpha - Cn^{-c} - 40\zeta_n\sqrt{\log p} - Cn^{-c} \\
&\geq \alpha - Cn^{-c},
\end{aligned}$$

where the second inequality follows from the union bound, the third from (16) and (20), the fourth from (21), the sixth from Lemma 3, and the last from $40\zeta_n\sqrt{\log p} \leq Cn^{-c}$. This concludes the proof. Q.E.D.

A.2 Proofs for Section 2.2

Proof of Theorem 2. Let v_n and r_n be two deterministic sequences such that $v_n < r_n$. Using the union bound, we obtain that

$$\begin{aligned}
\mathrm{P}\left(\hat{T} \leq \hat{c}(\alpha)\right) &= \mathrm{P}\left(\hat{T} - \hat{c}(\alpha) \leq r_n - v_n + v_n - r_n\right) \\
&\leq \mathrm{P}\left(\hat{T} - \hat{c}(\alpha) \leq r_n - v_n\right) + \mathrm{P}\left(v_n \geq r_n\right) \\
&\leq \mathrm{P}\left(\hat{T} \leq r_n\right) + \mathrm{P}\left(\hat{c}(\alpha) \geq v_n\right) + \mathrm{P}\left(v_n \geq r_n\right) \ .
\end{aligned}$$

To establish consistency, it suffices to show that

$$\mathrm{P}\left(\hat{c}(\alpha) \geq v_n\right) = o(1), \tag{22}$$

$$\mathrm{P}\left(\hat{T} \leq r_n\right) = o(1) \ . \tag{23}$$

We first establish (22). To ease notation, define, for $j = 1, \ldots, p$,

$$\hat{T}_j^B = \frac{1}{\sqrt{mq}} \sum_{k=1}^{m} \varepsilon_k \hat{A}_{j,k}.$$

Observe that the variance of $\hat{T}_j^B$ conditional on the data $\mathbb{D}$ is bounded from above by

$$\begin{aligned}
\frac{1}{mq} \sum_{k=1}^{m} \hat{A}_{j,k}^2 &= \frac{1}{mq} \sum_{k=1}^{m} \left(\sum_{l=(k-1)(q+1)+1}^{kq+(k-1)} \hat{W}_{j,l} \right)^2 \\
&\leq \frac{1}{mq} \sum_{k=1}^{m} (2q)^2 \\
&= 4q,
\end{aligned}$$

where the inequality follows from $|\hat{W}_{j,l}| \leq 2$. Then, by Proposition 1.1.3 in Talagrand (2003),

$$\mathbb{E}\left[\hat{T}^B | \mathbb{D}\right] \leq 2\sqrt{2q \log p} \ .$$

A modification of Borell's Lemma (see Lemma 9) with $\sigma^2 = 4q$ and $\lambda := \sqrt{8q \log(1/\alpha)}$ implies that

$$\mathrm{P}\left\{ \max_{1 \leq j \leq p} \hat{T}_j^B > \mathbb{E}\left[\max_{1 \leq j \leq p} \hat{T}_j^B \;\middle|\; \mathbb{D} \right] + \lambda \;\middle|\; \mathbb{D} \right\} \leq \alpha \ .$$

By the definition of $\hat{c}(\alpha)$, together with the anti-concentration inequality (Chernozhukov, Chetverikov, and Kato, 2015, Corollary 1), we have that

$$\mathrm{P}\left(\max_{1 \leq j \leq p} \hat{T}_j^B > \hat{c}(\alpha) \;\middle|\; \mathbb{D} \right) = \alpha \ .$$

Collecting these facts, we conclude that

$$\begin{aligned}\hat{c}(\alpha) &\le \mathbb{E}\left[\max_{1\le j\le p} \hat{T}_j^B \,\middle|\, \mathbb{D}\right] + \sqrt{8q\log(1/\alpha)}\\ &\le 2\sqrt{2q\log p} + 2\sqrt{2q\log(1/\alpha)}\end{aligned}$$

with probability one. This implies (22) with $v_n := 2\sqrt{2q\log p} + 2\sqrt{2q\log(1/\alpha)}$.

We now establish (23). Note that under the alternative, at least one of the individual hypotheses $H_{0,j} : Y_j \perp X$, $1 \le j \le p$, fails. Let j^* be the index of one of these hypotheses, so that $\xi_{j^*} > 0$ (since ξ_j is nonnegative for any j). Define $Z_n := \sqrt{n}(\hat{\xi}_{j^*} - \xi_{j^*})$. It follows from Kroll (2024) that $Z_n = O_P(1)$. Further, observe that

$$\hat{T} = \sqrt{n}\max_{1\le j\le p} \hat{\xi}_j \ge \sqrt{n}\hat{\xi}_{j^*} = Z_n + \sqrt{n}\xi_{j^*}\ .$$

For any sequence $r_n = o(n^{1/2})$, we have that

$$\begin{aligned}\mathrm{P}\left(\hat{T} > r_n\right) &\ge \mathrm{P}\left(Z_n + \sqrt{n}\xi_{j^*} > r_n\right)\\ &= \mathrm{P}\left(Z_n + \sqrt{n}\xi_{j^*} > r_n \mid |Z_n| \le \sqrt{n}\xi_{j^*}/2\right)\mathrm{P}\left(|Z_n| \le \sqrt{n}\xi_{j^*}/2\right)\\ &\quad + \mathrm{P}\left(Z_n + \sqrt{n}\xi_{j^*} > r_n \mid |Z_n| > \sqrt{n}\xi_{j^*}/2\right)\mathrm{P}\left(|Z_n| > \sqrt{n}\xi_{j^*}/2\right)\\ &\ge \mathrm{P}\left(\sqrt{n}\xi_{j^*}/2 > r_n\right)(1 - o(1)) + o(1)\\ &= 1 - o(1).\end{aligned}$$

It remains to show that we can choose a sequence $r_n = o(n^{1/2})$ larger than v_n defined above. Indeed, Assumption 2 implies that $\sqrt{q\log p} \le C_1 n^{1/2-\gamma}$. Therefore, we can choose positive constants c and C, such that $r_n := Cn^{1/2-c}$ satisfies $r_n = o(n^{1/2})$ and $r_n > v_n$, as required. This proves the claim in (23) and concludes the proof as a whole. Q.E.D.

A.3 Proofs for Section 2.3

Proof of Lemma 1. Suppose that Assumption 1 holds. In the subsequent derivations, we use the following values of moments of $W_{j,i}$ under the null $H_{0,j}$:

(i) $\mathbb{E}[W_{j,1}] = 0$,

(ii) $\mathbb{E}[W_{j,1}^2] = 1/2$,

(iii) $\mathbb{E}[W_{j,1}W_{j,2}] = -1/20$,

(iv) $\mathbb{E}[W_{j,1}^4] = 3/5$,

(v) $\mathbb{E}[W_{j,1}^2 W_{j,2}^2] = 23/70$,

(vi) $\mathbb{E}[W_{j,1}W_{j,2}(W_{j,1}^2 + W_{j,2}^2)] = -3/28$,

(vii) $\mathbb{E}[W_{j,1}W_{j,2}W_{j,3}(W_{j,1} + W_{j,2} + W_{j,3})] = -37/700$,

(viii) $\mathbb{E}[W_{j,1}W_{j,2}W_{j,3}W_{j,4}] = 1/700$.

Parts (i)–(iii) were shown by Angus (1995). Parts (iv)–(viii) were calculated using Mathematica.

Note that

$$\mathbb{E}[V_j^B] = \frac{1}{q}\,\mathbb{E}[A_{j,1}^2],$$
$$\mathbb{E}[A_{j,1}^2] = q\mathrm{Var}(W_{j,1}) + 2(q-1)Cov(W_{j,1}, W_{j,2}) = \frac{2}{5}q + 1/10,$$

which yields the first part of the lemma.

Furthermore, by the independence of blocks, we have that

$$\mathrm{Var}(V_j^B) = \frac{1}{mq^2}\mathrm{Var}(A_{j,1}^2) = \frac{1}{mq^2}\left(\mathbb{E}[A_{j,1}^4] - \mathbb{E}[A_{j,1}^2]^2\right).$$

For $q = 1$, $\mathbb{E}[A_{j,1}^4] = \mathbb{E}[W_{j,1}^4] = 3/5$, and hence

$$\mathbb{E}[A_{j,1}^4] - \mathbb{E}[A_{j,1}^2]^2 = 3/5 - 1/2^2 = 7/20.$$

For $q = 2$, $\mathbb{E}[A_{j,1}^4] = 2\,\mathbb{E}[W_{j,1}^4] + 4\,\mathbb{E}[W_{j,1}W_{j,2}(W_{j,1}^2 + W_{j,2}^2)] + 6\,\mathbb{E}[W_{j,1}^2 W_{j,2}^2] = 96/35$, and hence

$$\mathbb{E}[A_{j,1}^4] - \mathbb{E}[A_{j,1}^2]^2 = 96/35 - (9/10)^2.$$

For $q \geq 3$, a tedious calculation shows that $\mathbb{E}[A_{j,1}^4] = \frac{48}{100}q^2 + \frac{102}{175}q - \frac{111}{350}$, which combined with the expression for $\mathbb{E}[A_{j,1}^2]$ then leads to the desired result. Q.E.D.

A.4 Proofs for Section 3

Proof of Theorem 3. The result follows from Romano and Wolf (2005) if

$$\hat{c}(\alpha; I') \leq \hat{c}(\alpha; I'') \text{ whenever } I' \subset I'', \tag{24}$$
$$\sup_{P \in \mathbf{P}_{\gamma, C_1}} P\left(\hat{T}(J(P)) > \hat{c}(\alpha; J(P))\right) \leq \alpha + o(1). \tag{25}$$

Condition (24) holds by construction. By inspecting the proof of Theorem 1, we see that (25) holds. Q.E.D.

B Auxiliary Results

Lemma 5 (Corollary 1 in Chernozhukov, Chetverikov, and Kato (2015))**.** *Let $(X_1, \ldots, X_p)^T$ be a centered Gaussian random vector in $\mathbb{R}^p$ with $\sigma_j^2 := \mathbb{E}[X_j^2] > 0$ for all $1 \leq j \leq p$. Let $\underline{\sigma} := \min_{1 \leq j \leq p} \sigma_j$ and $\bar{\sigma} := \max_{1 \leq j \leq p} \sigma_j$. Then for every $\epsilon > 0$,*

$$\sup_{x \in \mathbb{R}} \mathrm{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \epsilon\right) \leq C\epsilon\sqrt{\max\{1, \log(p/\epsilon)\}},$$

where $C > 0$ depends only on $\underline{\sigma}$ and $\bar{\sigma}$. When σ_j are all equal, $\log(p/\epsilon)$ on the right side can be replaced by $\log p$.

Lemma 6 (Lemma 8 in Chernozhukov, Chetverikov, and Kato (2015))**.** *Let $X_1, \dots, X_n$ be independent random vectors in $\mathbb{R}^p$ with $p \geq 2$. Define $M := \max_{1\leq i\leq n} \max_{1\leq j\leq p} |X_{ij}|$ and $\sigma^2 := \max_{1\leq j\leq p} \sum_{i=1}^n \mathbb{E}[X_{ij}^2]$. Then*

$$\mathbb{E}\left(\max_{1\leq j\leq p}\left|\sum_{i=1}^n (X_{ij} - \mathbb{E}[X_{ij}])\right|\right) \leq K\left(\sigma\sqrt{\log p} + \sqrt{\mathbb{E}[M^2]}\log p\right),$$

where K is a universal constant.

Lemma 7 (Multivariate DKW inequality, Kiefer and Wolfowitz (1958))**.** *Let $X_1, \dots, X_n$ be independent random variables in $\mathbb{R}^k$ with CDF $F(\cdot)$, and let $\hat{F}$ be their empirical CDF. There exist constants c_k and c_k' such that for all $\varepsilon > 0$ and $n \in \mathbb{N}$ it holds that*

$$\mathrm{P}\left(\sup_{t\in\mathbb{R}^k} |F_n(t) - F(t)| \geq \varepsilon\right) \leq c_k \exp\left(-c_k' n\varepsilon^2\right) \qquad \text{for any } \varepsilon > 0.$$

The following DKW-type inequality for 1-dependent data is a straightforward corollary of Lemma 7.

Lemma 8. *Let $X_1, \dots, X_n$ be a sequence of 1-dependent, identically distributed random variables in $\mathbb{R}^k$. Then the conclusion of Lemma 7 holds.*

Proof. Suppose that $n = 2N$, and let F_n' be the empirical CDF calculated based on observations with odd indices i, and F_n'' based on the even indices. Then for any $\varepsilon > 0$, using the union bound, we obtain that

$$\begin{aligned}
\mathrm{P}\left(\sup_{t\in\mathbb{R}^k} |F_n(t) - F(t)| > \varepsilon\right) &= \mathrm{P}\left(\sup_{t\in\mathbb{R}^k} |(F_n'(t) - F(t)) + (F_n''(t) - F(t))| > 2\varepsilon\right) \\
&\leq \mathrm{P}\left(\sup_{t\in\mathbb{R}^k} |F_n'(t) - F(t)| > \varepsilon\right) + \mathrm{P}\left(\sup_{t\in\mathbb{R}^k} |F_n''(t) - F(t)| > \varepsilon\right) \\
&\leq 2c_k \mathrm{e}^{-c_k' n\varepsilon^2/2}.
\end{aligned}$$

An analogous argument holds for odd n. Q.E.D.

Lemma 9 (Modified Borell's Inequality)**.** *Let $X_1, \dots, X_p$ be mean-zero normal random variables with $\sigma^2 := \max_{1\leq i\leq p} Var(X_i)$ finite and nonzero. Then, for any $\lambda > 0$,*

$$\mathrm{P}\left(\max_{1\leq i\leq p} X_i > \mathbb{E}\left[\max_{1\leq i\leq p} X_i\right] + \lambda\right) \leq \mathrm{e}^{-\lambda^2/(2\sigma^2)}.$$

Proof. Let $Z := (Z_1, \dots, Z_p)$ be a vector independent standard normal random variables and let A be the symmetric square-root of $Var(X)$, $X := (X_1, \dots, X_p)$. For any $z \in \mathbb{R}^p$, let

$$f(z) := \frac{1}{\sigma} \max_{i=1,\dots,p} (Az)_i$$

and note that f is a Lipschitz function with Lipschitz constant 1: for any $z_1, z_2 \in \mathbb{R}^p$

$$\begin{aligned}|f(z_1) - f(z_2)| &= \frac{1}{\sigma}\left|\max_{1\le i\le p}(Az_1)_i - \max_{i=1,\ldots,p}(Az_2)_i\right| \\ &\le \frac{1}{\sigma}\max_{1\le i\le p}|(A(z_1-z_2))_i| \\ &\le \frac{1}{\sigma}\max_{1\le i\le p}\|A_{i\cdot}\|_\infty\,\|z_1-z_2\|_1 \\ &\le \frac{1}{\sigma}\left(\max_{1\le i\le p}A_{ii}^2\right)^{1/2}\|z_1-z_2\|_1 \\ &= \|z_1-z_2\|_1\end{aligned}$$

where the second inequality follows from Hölder's inequality. We can therefore apply van der Vaart and Wellner (1996, Lemma A.2.2) to obtain

$$\mathrm{P}\left(\max_{1\le i\le p}X_i > \mathbb{E}\left[\max_{1\le i\le p}X_i\right]+\lambda\right) = \mathrm{P}\left(f(Z)-\mathbb{E}[f(Z)] > \frac{\lambda}{\sigma}\right) \le \mathrm{e}^{-\lambda^2/(2\sigma^2)} ,$$

as desired. Q.E.D.

C Additional Simulation Results

C.1 Additional Results for Simulation II

In this section, we extend the results of Simulation II from the main text by including samples of size $n = 200$ and $n = 1000$ together with intermediate values of p. A third model is also considered. It is a version of Model 2 where all individual hypotheses are violated when $\rho \neq 0$.

Model 3: Let $X \sim Unif[-1,1]$ and $\varepsilon \sim \mathcal{N}(0,\Sigma_\tau)$, where Σ_τ has diagonal elements equal to one and off-diagonal elements equal to τ, and ε is independent of X. Further, let

$$Y_j = 2\rho\cos(8\pi X) + \varepsilon_j \quad \text{for } 1 \le j \le p.$$

The results are presented in Figures 7–12. As expected, the rejection rates (weakly) increase with the sample size for any given combination of p and ρ, but the qualitative patterns in Models 1 and 2 remain similar to those discussed in the main text. In Model 3, when there is no dependence between the components of Y, the power of the distance correlation test of Székely and Rizzo (2013) substantially improves relative to Model 2, achieving similar rejection rates to our test when $n = 1000$ and $p = 1000$. Its performance, however, deteriorates dramatically in the presence of dependence between the components of Y, while our procedure remains insensitive to such dependence in the data.

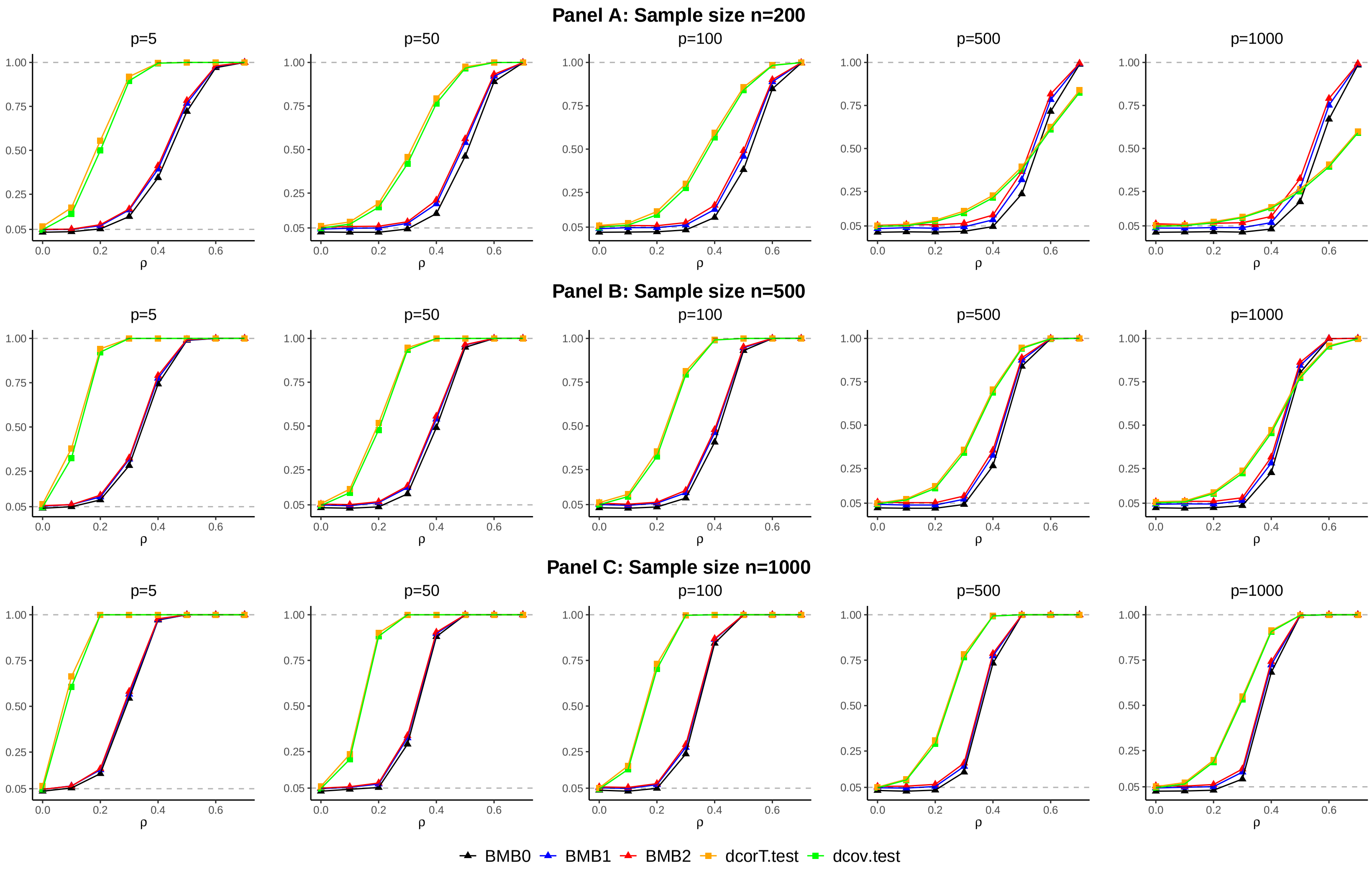


Figure 7: Simulation II for Model 1 with $\tau = 0$.

Notes: The tests have nominal level of 5%. Results are based on $S = 5,000$ Monte Carlo draws. The BMB tests and the distance covariance test used $B = 499$ bootstrap samples/replicates.

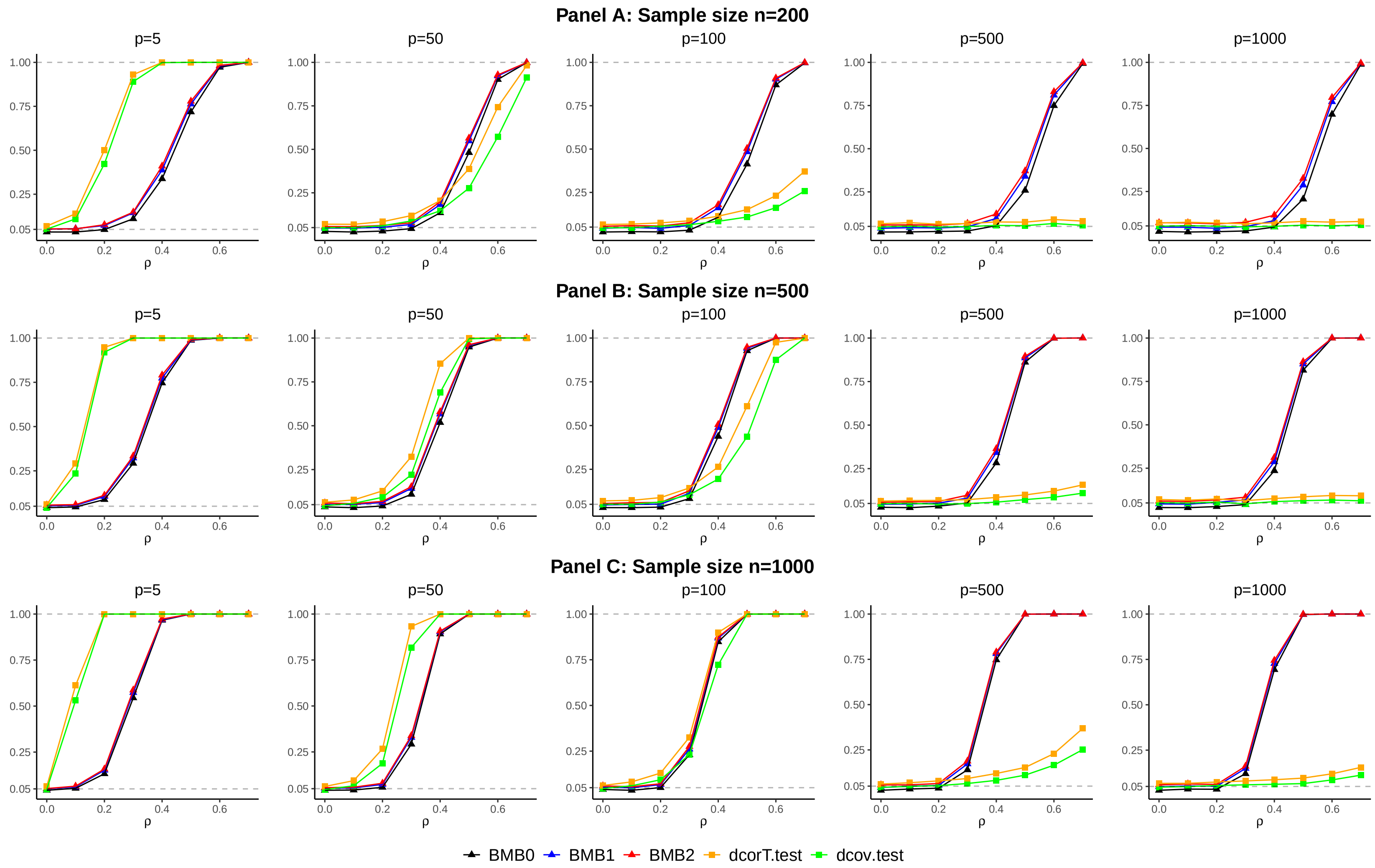


Figure 8: Simulation II for Model 1 with $\tau = 0.5$.

Notes: The tests have nominal level of 5%. Results are based on $S = 5,000$ Monte Carlo draws. The BMB tests and the distance covariance test used $B = 499$ bootstrap samples/replicates.

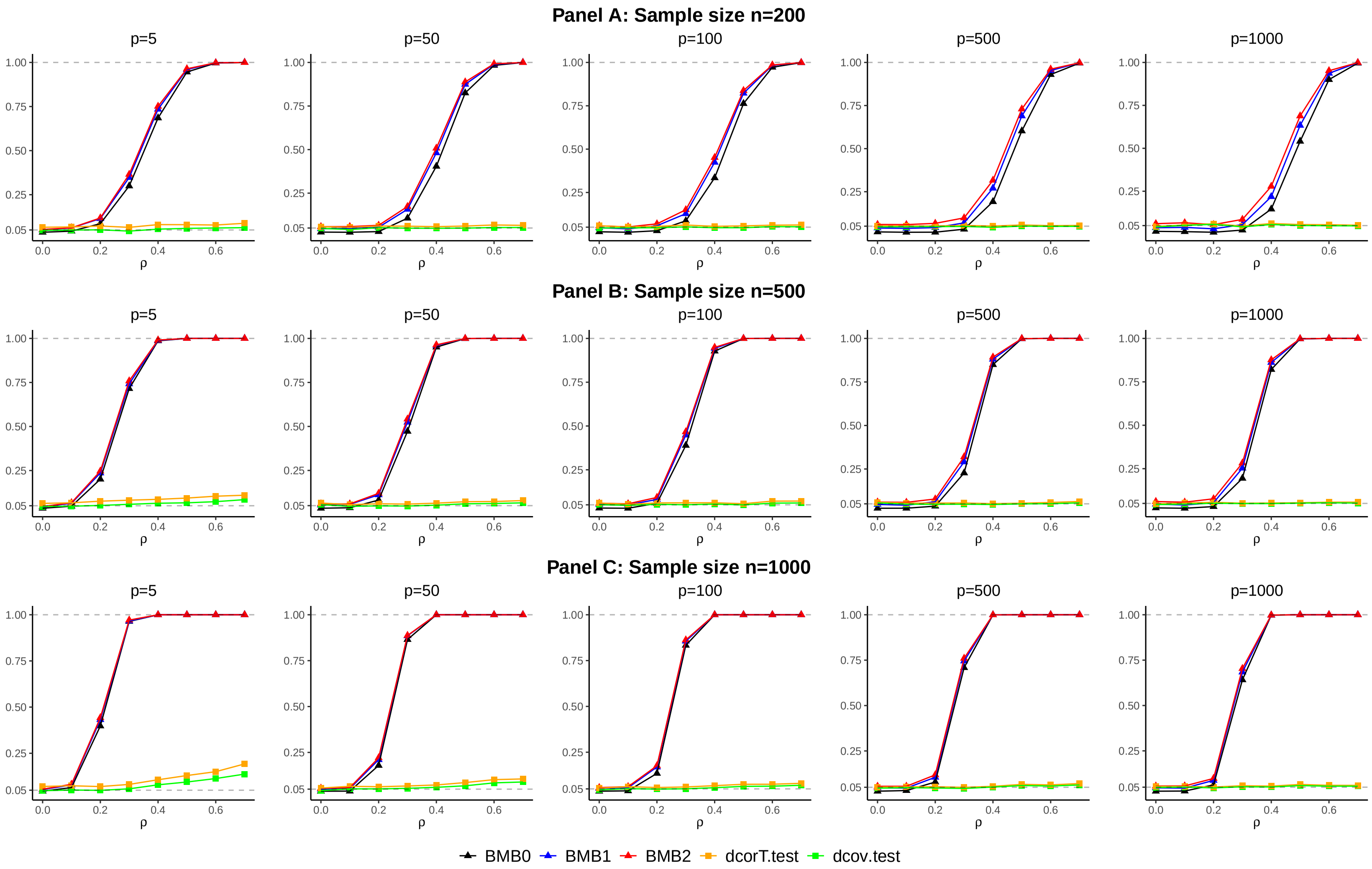


Figure 9: Simulation for Model 2 with $\tau = 0$.

Notes: The tests have nominal level of 5%. Results are based on $S = 5,000$ Monte Carlo draws. The BMB tests and the distance covariance test used $B = 499$ bootstrap samples/replicates.

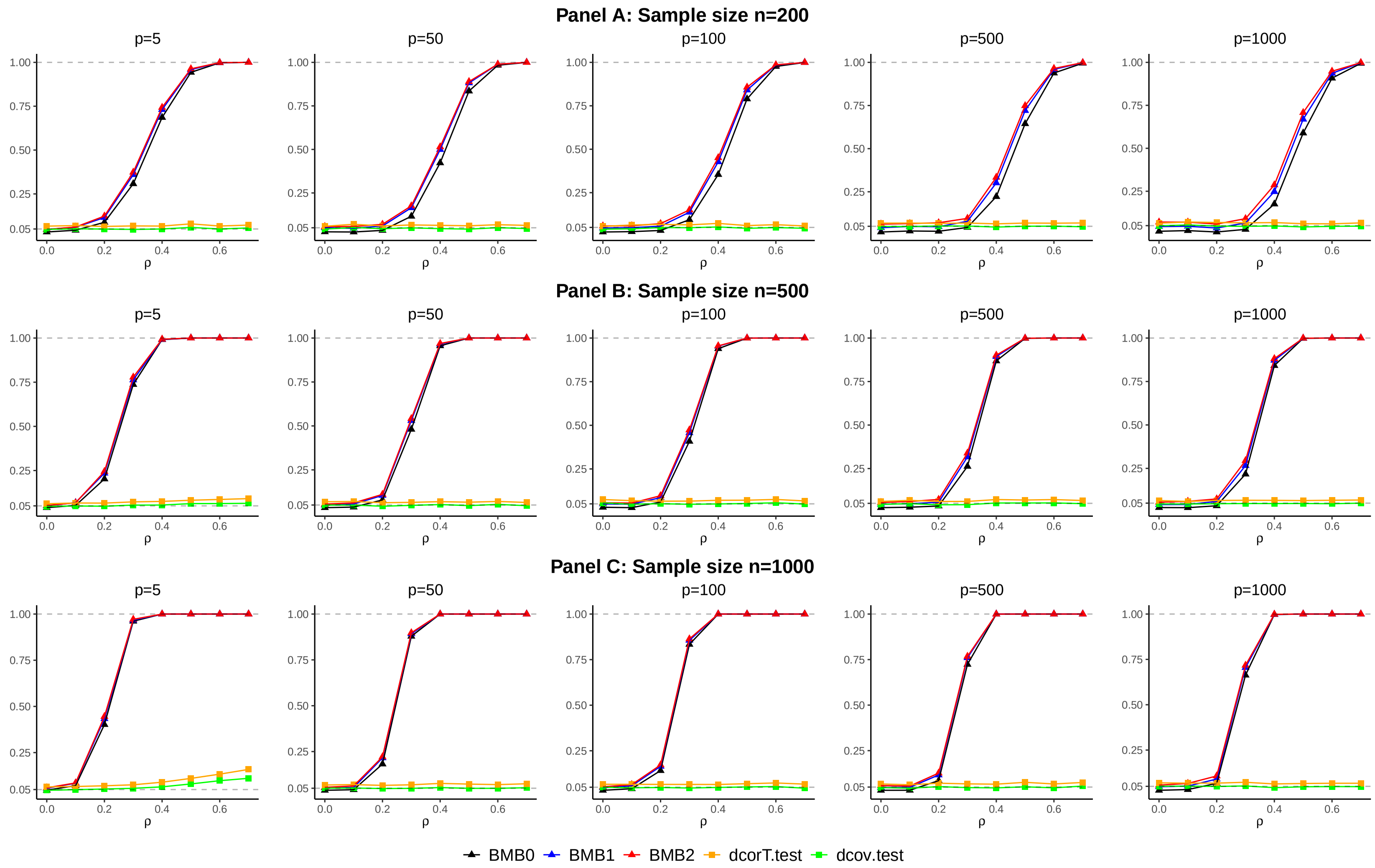


Figure 10: Simulation II for Model 2 with $\tau = 0.5$.

Notes: The tests have nominal level of 5%. Results are based on $S = 5,000$ Monte Carlo draws. The BMB tests and the distance covariance test used $B = 499$ bootstrap samples/replicates.

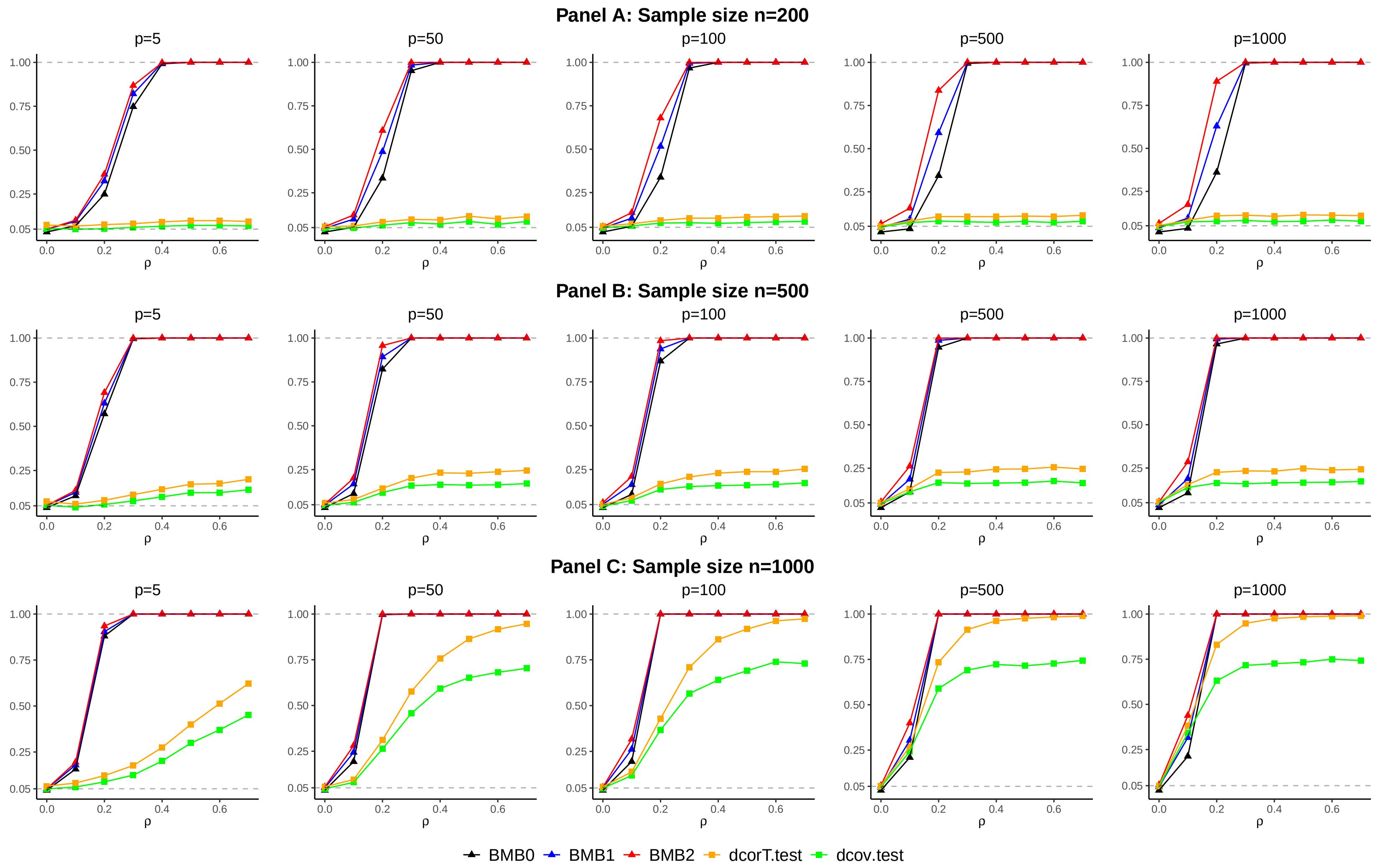


Figure 11: Simulation II for Model 3 with $\tau = 0$.

Notes: The tests have nominal level of 5%. Results are based on $S = 5,000$ Monte Carlo draws. The BMB tests and the distance covariance test used $B = 499$ bootstrap samples/replicates.

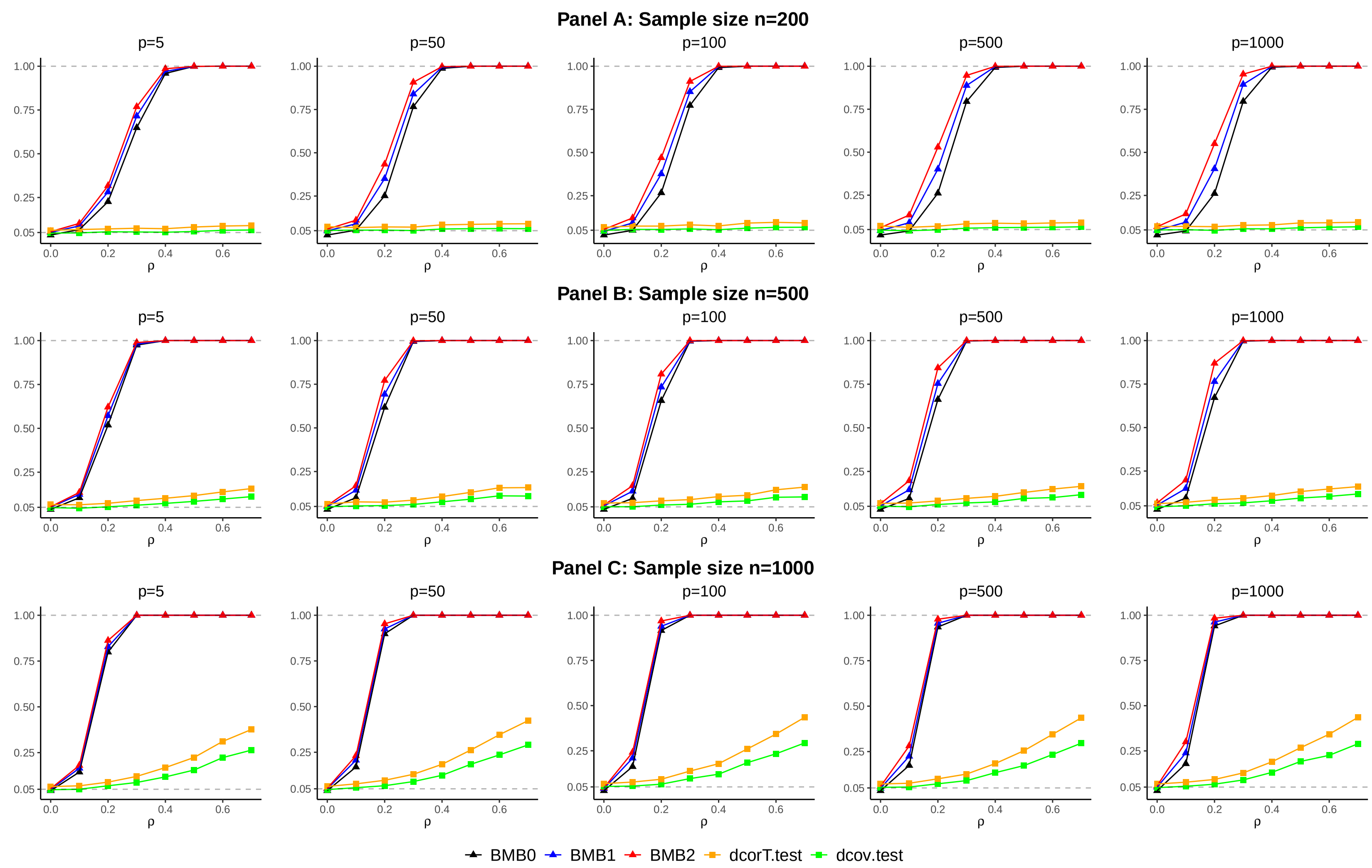


Figure 12: Simulation II for Model 3 with $\tau = 0.5$.

Notes: The tests have a nominal level of 5%. Results are based on $S = 5,000$ Monte Carlo draws. The BMB tests and the distance covariance test used $B = 499$ bootstrap samples/replicates.

C.2 Simulation III

In this section, we compare our method to the independence tests of Zhou, Xu, Zhu, and Li (2024) based on Hoeffding's D (ZXZL-D), Blum-Kiefer-Rosenblatt's R (ZXZL-R) and Bergsma-Dassios-Yanagimoto's τ^* (ZXZL-τ^*), and to the test of Zhu, Zhang, Yao, and Shao (2020) based on an aggregated distance covariance measure (ZZYS-agg.dcov).[5] Due to the high computational complexity of these additional tests, we limit the simulation setup to samples of size $n = 200$ with the maximum of $p = 200$ individual hypotheses.[6] Figure 13 shows that, in the considered designs, the tests of Zhou, Xu, Zhu, and Li (2024) and Zhu, Zhang, Yao, and Shao (2020) perform similarly to the distance-covariance-based tests considered in Simulation II. In particular, they exhibit higher power than our test in Model 1 when there is no dependence between the components of Y, but they performance deteriorates significantly in the presence of such dependence. They also have virtually no power against the cosine alternatives in Models 2 and 3.

[5]The tests ZXZL-D, ZXZL-R, ZXZL-τ^*, and ZZYS-agg.dcov were run using the implementation of Zhou, Xu, Zhu, and Li (2024) provided at `https://github.com/Yeqing-TJ/Rank-based-test-in-high-dimension` [Accessed on March 10, 2025].

[6]Estimation of the variance of the rank-based indices of Zhou, Xu, Zhu, and Li (2024) in one sample with $p = 500$ and $n = 500$ takes over 40 minutes using an Intel i7-1185G7 @ 3.00 GHz processor, which renders simulations for such settings infeasible. For comparison, our bootstrap test with $B = 499$ takes less than a second to compute in this setting.

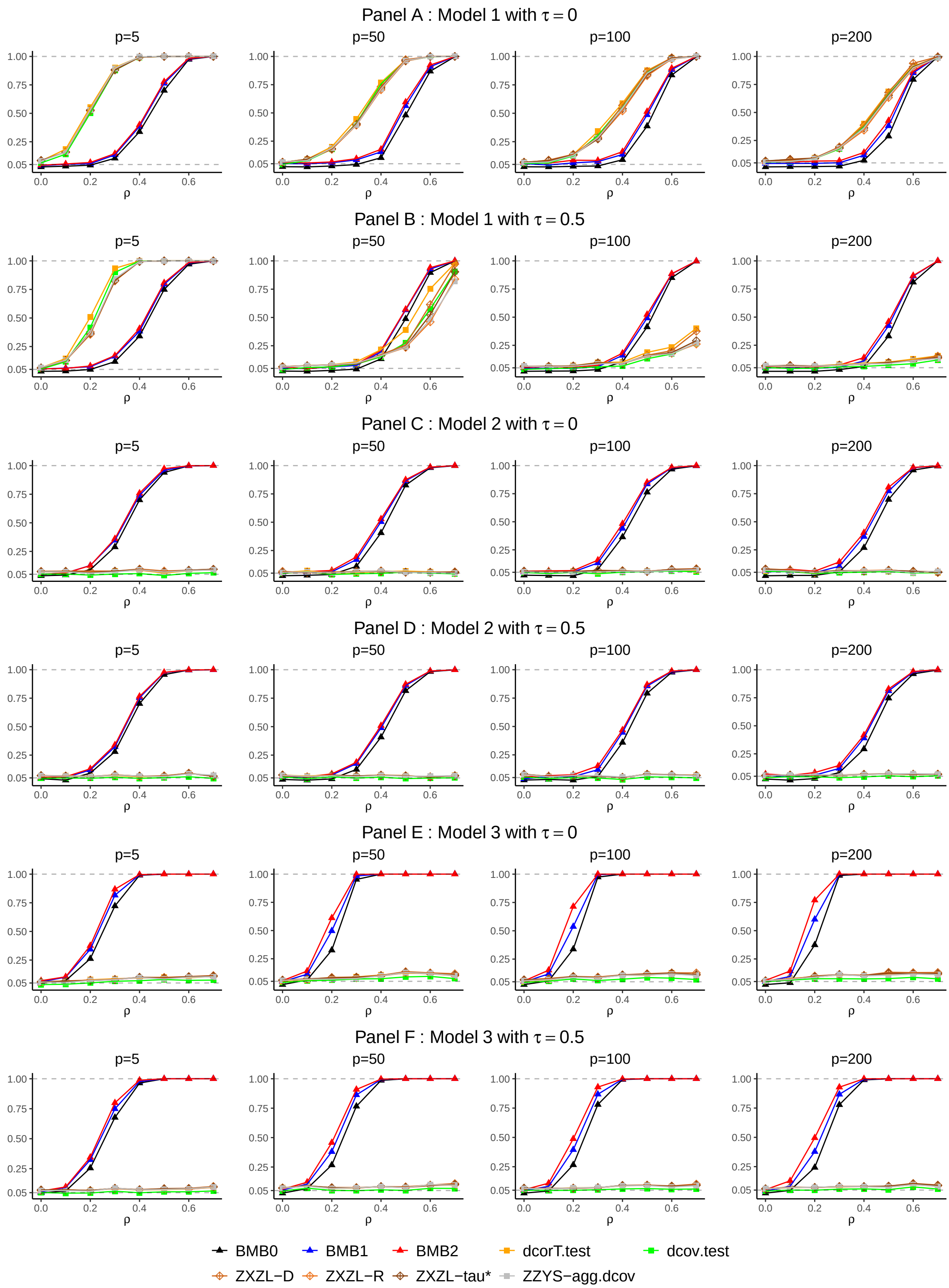


Figure 13: Simulation III.

Notes: The tests have nominal level of 5%. Results for $n = 200$ and $S = 1,000$ Monte Carlo draws. The BMB tests and the distance covariance used $B = 499$ bootstrap samples/replicates. Where not distinguishable, the power curves of the tests dcorT.test, dcov.test, ZXZL-D, ZXZL-R, ZXZL-τ^*, and ZZYS-agg.cov overlap.

References

Adler, R. J., and L. D. Brown (1986): "Tail behaviour for suprema of empirical processes," *The Annals of Probability*, pp. 1–30.

Angus, J. E. (1995): "A coupling proof of the asymptotic normality of the permutation oscillation," *Probability in the Engineering and Informational Sciences*, 9(4), 615–621.

Bakirov, N. K., M. L. Rizzo, and G. J. Székely (2006): "A multivariate nonparametric test of independence," *Journal of Multivariate Analysis*, 97(8), 1742–1756.

Bastian, P., H. Dette, and J. Heiny (2024): "Testing for practically significant dependencies in high dimensions via bootstrapping maxima of U-statistics," *The Annals of Statistics*, 52(2), 628 – 653.

Chatterjee, S. (2021): "A new coefficient of correlation," *Journal of the American Statistical Association*, 116(536), 2009–2022.

——— (2024): "A Survey of Some Recent Developments in Measures of Association," in *Probability and Stochastic Processes: A Volume in Honour of Rajeeva L. Karandikar*, ed. by S. Athreya, A. G. Bhatt, and B. V. Rao, pp. 109–128. Springer Nature Singapore, Singapore.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors," *The Annals of Statistics*, 41(6), 2786–2819.

——— (2015): "Comparison and anti-concentration bounds for maxima of Gaussian random vectors," *Probability Theory and Related Fields*, 162(1), 47–70.

——— (2019): "Inference on causal and structural parameters using many moment inequalities," *The Review of Economic Studies*, 86(5), 1867–1900.

Dette, H., and M. Kroll (2024): "A simple bootstrap for Chatterjee's rank correlation," *Biometrika*, forthcoming.

Dette, H., K. F. Siburg, and P. A. Stoimenov (2013): "A Copula-Based Non-parametric Measure of Regression Dependence," *Scandinavian Journal of Statistics*, 40(1), 21–41.

Drton, M., F. Han, and H. Shi (2020): "High-dimensional consistent independence testing with maxima of rank correlations," *The Annals of Statistics*, 48(6), 3206 – 3227.

Heller, R., M. Gorfine, and Y. Heller (2012): "A class of multivariate distribution-free tests of independence based on graphs," *Journal of Statistical Planning and Inference*, 142(12), 3097–3106.

Heller, R., Y. Heller, and M. Gorfine (2012): "A consistent multivariate test of association based on ranks of distances," *Biometrika*, 100(2), 503–510.

Hughes, M. E., L. DiTacchio, K. R. Hayes, C. Vollmers, S. Pulivarthy, J. E. Baggs, S. Panda, and J. B. Hogenesch (2009): "Harmonics of circadian gene transcription in mammals," *PLoS genetics*, 5(4), e1000442.

Kiefer, J., and J. Wolfowitz (1958): "On the deviations of the empiric distribution function of vector chance variables," *Transactions of the American Mathematical Society*, 87(1), 173–186.

Kroll, M. (2024): "Asymptotic Normality of Chatterjee's Rank Correlation," *arXiv preprint arXiv:2408.11547*.

Leung, D., and M. Drton (2018): "Testing independence in high dimensions with sums of rank correlations," *The Annals of Statistics*, 46(1), 280 – 307.

Lin, Z., and F. Han (2022a): "Limit theorems of Chatterjee's rank correlation," Discussion paper.

Lin, Z., and F. Han (2022b): "On boosting the power of Chatterjee's rank correlation," *Biometrika*, 110(2), 283–299.

Lin, Z., and F. Han (2024): "On the failure of the bootstrap for Chatterjee's rank correlation," *Biometrika*, 111(3), 1063–1070.

Ramdas, A., S. Jakkam Reddi, B. Poczos, A. Singh, and L. Wasserman (2015): "On the Decreasing Power of Kernel and Distance Based Nonparametric Hypothesis Tests in High Dimensions," *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1).

Romano, J. P., and M. Wolf (2005): "Exact and approximate stepdown methods for multiple hypothesis testing," *Journal of the American Statistical Association*, 100(469), 94–108.

Shi, H., M. Drton, and F. Han (2021): "On the power of Chatterjee's rank correlation," *Biometrika*, 109(2), 317–333.

Shi, H., M. Hallin, M. Drton, and F. Han (2022): "On universally consistent and fully distribution-free rank tests of vector independence," *The Annals of Statistics*, 50(4), 1933 – 1959.

Shorack, G. R., and J. A. Wellner (2009): *Empirical processes with applications to statistics*. SIAM.

Sinha, B. K., and H. Wieand (1977): "Multivariate nonparametric tests for independence," *Journal of Multivariate Analysis*, 7(4), 572–583.

Storey, J. D., and R. Tibshirani (2003): "Statistical significance for genomewide studies," *Proceedings of the National Academy of Sciences*, 100(16), 9440–9445.

Straume, M. (2004): "DNA microarray time series analysis: automated statistical assessment of circadian rhythms in gene expression patterning," *Methods in enzymology*, 383, 149–166.

Székely, G. J., and M. L. Rizzo (2013): "The distance correlation t-test of independence in high dimension," *Journal of Multivariate Analysis*, 117, 193–213.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007): "Measuring and testing dependence by correlation of distances," *The Annals of Statistics*, 35(6), 2769 – 2794.

Talagrand, M. (2003): *Spin Glasses: A Challenge for Mathematicians*, vol. 46. Springer.

Taskinen, S., H. Oja, and R. H. Randles (2005): "Multivariate Nonparametric Tests of Independence," *Journal of the American Statistical Association*, 100(471), 916–925.

van der Vaart, A. W., and J. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer, New York.

Wang, H., B. Liu, and L. Feng (2024): "Testing Independence Between High-Dimensional Random Vectors Using Rank-Based Max-Sum Tests," Discussion paper.

Wang, H., B. Liu, L. Feng, and Y. Ma (2024): "Rank-based max-sum tests for mutual independence of high-dimensional random vectors," *Journal of Econometrics*, 238(1), 105578.

Xia, L., R. Cao, J. Du, and J. Dai (2024): "Consistent complete independence test in high dimensions based on Chatterjee correlation coefficient," *Statistical Papers*, 66(1), 3.

Yao, S., X. Zhang, and X. Shao (2018): "Testing Mutual Independence in High Dimension via Distance Covariance," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(3), 455–480.

Zhang, Q. (2023): "On the asymptotic null distribution of the symmetrized Chatterjee's correlation coefficient," *Statistics & Probability Letters*, 194, 109759.

Zhou, Y., K. Xu, L. Zhu, and R. Li (2024): "Rank-based indices for testing independence between two high-dimensional vectors," *The Annals of Statistics*, 52(1), 184–206.

Zhu, C., X. Zhang, S. Yao, and X. Shao (2020): "Distance-based and RKHS-based dependence metrics in high dimension," *The Annals of Statistics*, 48(6), 3366–3394.