ROBUST PERMUTATION TESTS FOR CORRELATION AND REGRESSION COEFFICIENTS

By

Cyrus J. DiCiccio
Joseph P. Romano

Technical Report No. 2015-15
August 2015

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

ROBUST PERMUTATION TESTS FOR CORRELATION AND REGRESSION COEFFICIENTS

By

Cyrus J. DiCiccio
Joseph P. Romano
Stanford University

Technical Report No. 2015-15
August 2015

This research was supported in part by National Science Foundation grants DMS 1007732 and DMS 1307973.

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://statistics.stanford.edu

Robust Permutation Tests For Correlation And Regression Coefficients

Cyrus J. DiCiccio [1]
Department of Statistics
Stanford University

Joseph P. Romano [2]
Departments of Statistics and Economics
Stanford University

August 6, 2015

Abstract

Given a sample from a biviariate distribution, consider the problem of testing independence. A permutation test using the sample correlation is known to be an exact level α test. However, when used to test the null hypothesis that the samples are uncorrelated, the permutation test can have rejection probability that is far from the nominal level. Further, the permutation test can have large Type 3 (directional) error rate, whereby there can be a large probability that the permutation test rejects because the sample correlation is a large positive value, when in fact the true correlation is negative. It will be shown that studentizing the sample correlation leads to a permutation test which is exact under independence and asymptotically controls the probability of Type 1 (or Type 3) errors. These conclusions are based on our results describing the almost sure limiting behavior of the randomization distribution. We will also present asymptotically robust randomization tests for regression coefficients, including a result based on a modified procedure of Freedman and Lane (1983). Simulations and empirical applications are included.

Keywords: Testing independence; Randomization tests; Least squares; Partial correlation; Studentization

1 Introduction

Assume $(X_1, Y_1), ..., (X_n, Y_n)$ are i.i.d. according to a joint distribution P with marginal distributions P_X and P_Y. Define $X^n = (X_1, ..., X_n)$ and $Y^n = (Y_1, ..., Y_n)$. Let $\rho = \rho(P) = \text{corr}(X_1, Y_1)$ and first consider the problem of testing the null hypothesis of independence,

$$H_0 : P = P_X \times P_Y.$$

A permutation test can be constructed as follows. Define $\mathbf{G_n}$ to be the set of all permutations π of $\{1, ..., n\}$. The permutation distribution of any given test statistic $T_n(X^n, Y^n)$ is defined as

$$\hat{R}_n^{T_n}(t) = \frac{1}{n!} \sum_{\pi \in \mathbf{G_n}} I\{T_n(X^n, Y^n_\pi) \leq t\}$$

[1,2] Research supported by NSF Grants DMS-1307973 and DMS-1007732

where we write Y_π^n for $(Y_{\pi(1)}, ..., Y_{\pi(n)})$. A level α permutation test rejects if $T_n(X^n, Y^n)$ is smaller than the $\alpha_1/2$ quantile or larger than the $1-\alpha_2/2$ quantile of the permutation distribution (where α_1 and α_2 are chosen so that $\alpha = \alpha_1 + \alpha_2$). More precisely, the permutation test is given by

$$\phi(X^n, Y^n) = \begin{cases} 1 & T_n(X^n, Y^n) < T_n^{(m_1)} \text{ or } T_n(X^n, Y^n) > T_n^{(m_2)} \\ \gamma_1 & T_n(X^n, Y^n) = T_n^{(m_1)} \\ \gamma_2 & T_n(X^n, Y^n) = T_n^{(m_2)} \\ 0 & T_n^{(m_1)} < T_n(X^n, Y^n) < T_n^{(m_2)} \end{cases}$$

where $T_n^{(k)}$ denotes the k^{th} largest ordered value of $\{T_n(X^n, Y_\pi^n) : \pi \in \mathbf{G_n}\}$, $m_1 = n! - \lfloor (1 - \alpha/2)n! \rfloor + 1$, $m_2 = n! - \lfloor \alpha/2 n! \rfloor$, and γ_1, γ_2 are chosen so that

$$\frac{1}{n!} \sum_{\pi : T(X^n, Y_\pi^n) \le T_n^{(m_1)}} \phi(X^n, Y_\pi^n) = \alpha_1 \quad \text{and} \quad \frac{1}{n!} \sum_{\pi : T(X^n, Y_\pi^n) \ge T_n^{(m_2)}} \phi(X^n, Y_\pi^n) = \alpha_2$$

which ensures

$$\frac{1}{n!} \sum_{\pi \in \mathbf{G_n}} \phi(X^n, Y_\pi^n) = \alpha.$$

Usually for a two sided test, $\alpha_1 = \alpha_2 = \alpha/2$, and for a one sided test, $\alpha_1 = 0$. The *randomization hypothesis* is said to hold if the distribution of (X^n, Y^n) is invariant under the group of transformations, $\mathbf{G_n}$ (i.e. (X^n, Y_π^n) is distributed as (X^n, Y^n)). If the randomization hypothesis holds, then the permutation test is exact level α (See Definition 15.2.1 and Theorem 15.2.2 of Lehmann and Romano).

To test the null hypothesis of independence, the normalized sample correlation

$$\sqrt{n}\hat{\rho}_n\left(X^n, Y^n\right) = \sqrt{n}\frac{\sum X_i Y_i - n\bar{X}_n\bar{Y}_n}{\sqrt{\sum (X_i - \bar{X}_n)^2 \sum (Y_i - \bar{Y}_n)^2}}.$$

can be used as the test statistic. Under the null hypothesis of independence, the distribution of (X^n, Y^n) is the same as that of (X^n, Y_π^n) for any permutation π. The randomization hypothesis is satisfied and this test is exact level α. However, even in this case, if rejection of independence is accompanied by the claim that ρ is positive (negative) when $\hat{\rho}_n$ is large positive (negative), then such claims have large error rates.

Suppose instead we are interested in testing the null hypothesis

$$H_0 : \rho(P) = 0,$$

against two sided alternatives. When testing H_0, the X_i and Y_i may be dependent under the null, and the distribution of the test statistic may not be the same under all permutations of the data. Therefore, the randomization hypothesis is violated and the test is not guaranteed to be level α, even asymptotically. Under the null hypothesis that $\rho = 0$, if $\text{var}(X_1) > 0$, $\text{var}(Y_1) > 0$, $E(X_1)^2 < \infty$, $E(Y_1)^2 < \infty$ and $E(X_1Y_1)^2 < \infty$, then the sampling distribution of $\sqrt{n}\hat{\rho}_n(X^n, Y^n)$ is asymptotically normal with mean zero and variance

$$\tau^2 = \tau^2(P) = \frac{\mu_{2,2}}{\mu_{2,0}\mu_{0,2}} \tag{1}$$

where

$$\mu_{r,s} = \mu_{r,s}(P) = E\left[(X_1 - \mu_X)^r (Y_1 - \mu_Y)^s\right].$$

It will be shown in the next section that the permutation distribution of the sample correlation is not guaranteed to asymptotically approximate this distribution, and the permutation test obtained by comparing the sample correlation with the quantiles of the permutation distribution will not have the desired level asymptotically. Even more troubling, this discrepancy can lead to large Type 3 (directional) error rate if one is interested in deciding the sign of the correlation based on the sample correlation. A Type 3 error occurs when declaring $\rho < 0$ when in fact $\rho > 0$, or the other way around. For example, a researcher who rejects H_0 due to a large positive value of $\hat{\rho}_n$ would like to claim $\rho > 0$. The results of Section 2 will show that the permutation distribution does not approximate the true sampling distribution of the sample correlation; however, appropriate studentization of the sample correlation yields a permutation test which is asymptotically level α for testing correlation zero, but is also exact in finite samples under independence. Randomization tests for regression coefficients will be presented in Section 3 and for partial correlations will be given in Section 4. It will be shown that appropriate studentization of the test statistic in the regression setting leads to a permutation test that is exact when the error terms are independent of the predictor variables, and asymptotically valid when they are only assumed to be uncorrelated. In Section 5, simulation results will be presented showing the true rejection probability of the studentized and unstudentized permutation tests. Finally, Section 6 gives empirical applications comparing the studentized and unstudentized permutation tests, as well as plots comparing the resulting permutation distributions. Proofs of the results in Section 2, 3 and 4 are given in the appendix.

Robust asymptotic inference for parameters based on permutation tests has been much studied in the two sample problem. While the context of observing two independent samples is distinct from the context of paired samples studied here, we will now provide a short review. Typically, two sample permutation tests are exact when the underlying distributions are the same, however the exactness property may fail if the parameters of interest are equal but the distributions differ; see Romano (1990). Extensive work has been done to show how studentizing two sample permutation tests can lead to an asymptotically valid test whenever the parameters of interest are equal which also retains the exactness property in finite samples when the underlying distributions are the same. This was first discovered by Neuhaus (1993), in the context of random censoring models. Further work on studentizing two sample permutation tests has been done on comparing means by Janssen (1997), comparing variances by Pauly (2011), and comparing correlations by Omelka and Pauly (2012). Results on studentizing linear statistics are given by Janssen (1999), and more generally by Chung and Romano (2013). Applications of permutation tests can be found in Good (2005). The goal of this paper is to show how studentizing the sample correlation coefficient calculated from one sample of a bivariate density can lead to an asymptotically robust test. In the two sample case, the permutation test is exact when the marginal distributions are equal, but in the case of testing for independence, the permutation test is exact when the joint distribution is the product of the marginal distributions. When permuting pairs of uncorrelatated variables, the difficulty with the permutation test arises when the data is not independent as opposed to the two sample test where the two samples are always assumed independent, but the permutation test may fail when the underlying distributions are unequal. As with the two sample problem, the "fix" of using a studentized or asymptotically distribution free statistic works here, although the heuristics and proofs are distinct.

Permutation tests are often used to test significance of one or more regression coefficients in multiple regression, especially in biological or ecological studies. Winkler et al. (2014) discusses the applications of permutation methods for multiple linear regression to neuroimaging data. Depending on the regression model assumed, permutation methods often fail to be exact in regression (see Anderson and Robinson (2002) for a comparison of existing methods). A common concern is that when the errors and predictor variables are uncorrelated, but not independent (which includes heteroskedastic regression), permutation methods are not guaranteed to be exact or even asymptotically valid. In these situations, asymptotically valid tests based on a normal approximation using White's heteroskedasticity-consistent covariance estimators (White (1980)), or bootstrap methods such as the pairs bootstrap or the Wild bootstrap, proposed by Wu (1986), may be used. However, the results on robust tests for correlation extend naturally to testing for regression coefficients, and asymptotically valid permutation tests can be constructed by using appropriately studentized test statistics.

2 Main Results

The first theorem will show that the permutation distribution does not asymptotically approximate the true sampling distribution of the correlation statistic. Instead, the permutation distribution behaves asymptotically like the true sampling distribution of the correlation of a sample from $P_X \times P_Y$ (instead of from P). As a result, comparing the sample correlation with the quantiles of the permutation distribution will not give an asymptotically level α test.

Theorem 2.1 *Assume $(X_1, Y_1), ..., (X_n, Y_n)$ are i.i.d. according to P such that X_1 and Y_1 are uncorrelated but not necessarily independent. Also assume that $E(X_1^4) < \infty$ and $E(Y_1^4) < \infty$. Then, the permutation distribution of $T_n = \sqrt{n}\hat{\rho}_n$ satisfies*

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| \hat{R}_n^{T_n}(t) - \Phi(t) \right| = 0$$

almost surely. However,

$$\sqrt{n}\hat{\rho}_n(X^n, Y^n) \xrightarrow{\mathcal{L}} N\left(0, \tau^2(P)\right)$$

where $\tau^2(P)$ is defined by equation 1. Therefore,

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| \hat{R}_n^{T_n}(t) - F_n(t) \right| > 0$$

where F_n is the law of $\sqrt{n}\hat{\rho}_n(X^n, Y^n)$, unless $\tau^2(P) = 1$.

When the X_i and Y_i are independent, the asymptotic variance of both the sampling distribution and the permutation distribution of $\sqrt{n}\hat{\rho}_n(X^n, Y^n)$ are one. In this case, the quantiles of the permutation distribution and the true sampling distribution will converge to the corresponding quantiles of the standard normal distribution (with probability one), and the test is asymptotically level α. However, this is not necessarily the case when the (X_i, Y_i) are uncorrelated but dependent. In fact, when the (X_i, Y_i) are uncorrelated but dependent, the permutation distribution has variance one (asymptotically and $n/(n-1)$ in finite samples), but the sampling distribution has limiting variance $\tau^2(P)$, which can take any value in the interval $[0, \infty]$ as is summarized by the next theorem.

Theorem 2.2 *If X_i and Y_i are uncorrelated (but not necessarily independent),*

$$0 \leq \frac{E\left[(X_i - \mu_X)^2 (Y_i - \mu_Y)^2\right]}{\sigma_X^2 \sigma_Y^2} \leq \infty$$

and these bounds can be attained in the sense that there exists a joint distribution of (X_1, Y_1) where $cov(X_1, Y_1) = 0$, but this ratio is 0, and likewise for which it is ∞.

When $\tau^2(P) < 1$, the permutation test can have asymptotic null rejection probability much smaller than the nominal level α, and when $\tau^2(P) > 1$, the permutation test can have rejection probability much larger than α. Further, this discrepancy can cause the permutation test to have large Type 3 (directional) error when concluding that the true sign of the correlation is equal to that of the sample correlation after rejecting H_0. For example, when the X_i and Y_i are uncorrelated, and $\tau^2(P)$ is much larger than one, the permutation test will have rejection probability α' much larger than the nominal level α. By continuity, the rejection probability when X_i and Y_i have some small positive correlation can be made very near to α'. In this case, the true sampling distribution of the sample correlation will be almost symmetric about 0, and the Type 3 error rate will be close to $\alpha'/2$. Since α' can be arbitrarily close to one, the Type 3 error rate can be unacceptably large in this situation. Moreover, in cases when $\tau^2(P)$ is small, the test can have large Type 2 error.

To remedy these problems, the test statistic can be studentized by

$$\hat{\tau}_n = \sqrt{\frac{\hat{\mu}_{2,2}}{\hat{\mu}_{2,0}\hat{\mu}_{0,2}}}$$

where

$$\hat{\mu}_{r,s} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^r (Y_i - \bar{Y}_n)^s$$

are the sample central moments. The studentized correlation statistic defined by $S_n = \sqrt{n}\hat{\rho}_n / \hat{\tau}_n$ is asymptotically pivotal in the sense that it is asymptotically distribution free whenever the underlying distribution satisfies H_0. If T_n is a pivotal statistic, then the true sampling distribution of T_n under P has the same asymptotic behavior as the true sampling distribution of T_n under $P_X \times P_Y$. But, if the randomization hypothesis is satisfied, the permutation test using the statistic T_n is exact under $P_X \times P_Y$, and therefore, the permutation distribution should asymptotically approximate the true sampling distribution under $P_X \times P_Y$. Applying a permutation to data sampled under P effectively destroys the dependence between the pairs, and the distribution of the permutation distribution under P should behave asymptotically like the permutation distribution under $P_X \times P_Y$. It is then reasonable to expect that the quantiles of the permutation distribution of the studentized statistic will approximate those of the true sampling distribution under P. Although $\hat{\tau}_n^2(X^n, Y^n)$ converges in probability to $\tau^2(P)$, if a permutation π is chosen uniformly, then $\hat{\tau}_n^2(X^n, Y_\pi^n)$ converges in probability to 1, which is the same as the limit when the X_i and Y_i are independent. By

Slutsky's theorem, the sampling distribution of $\sqrt{n}\hat{\rho}_n/\hat{\tau}_n$ is asymptotically standard normal and the next theorem gives that the permutation distribution of $\sqrt{n}\hat{\rho}_n/\hat{\tau}_n$ is also standard normal in probability (or almost surely under stronger moment assumptions).

Theorem 2.3 *Assume $(X_1, Y_1), ..., (X_n, Y_n)$ are i.i.d. according to P such that X_1 and Y_1 are uncorrelated but not necessarily independent. Also assume that $E(X_1^4) < \infty$ and $E(Y_1^4) < \infty$. The permutation distribution $\hat{R}_n^{S_n}(t)$ of $S_n = \sqrt{n}\hat{\rho}_n/\hat{\tau}_n$ satisfies*

$$\sup_{t \in \mathbb{R}} \left| \hat{R}_n^{S_n}(t) - \Phi(t) \right| \to 0$$

in probability. Under the stronger assumption that $E(X_1^8) < \infty$ and $E(Y_1^8) < \infty$,

$$\lim_{n \to \infty} \sup_{t \in \mathbb{R}} \left| \hat{R}_n^{S_n}(t) - \Phi(t) \right| = 0$$

almost surely.

Consequently, if $\sqrt{n}\hat{\rho}_n$ is studentized by $\hat{\tau}_n$, then the quantiles of the permutation distribution and the true sampling distribution converge in probability to the corresponding quantiles of the standard normal distribution. The permutation test using the studentized statistic is appealing because it retains the exactness property under $P_X \times P_Y$ but is also asymptotically level α under P.

Remark 2.4 (Limiting Local Power) *To study the limiting local power of the studentized permutation test, suppose that P_0 satisfies H_0 and consider a sequence $\{P_n\}$ of contiguous alternatives to P_0. Under P_0,*

$$\hat{r}_n(1-\alpha) \xrightarrow{P} z_{1-\alpha}$$

where $\hat{r}_n(1-\alpha)$ is the $1-\alpha$ quantile of the permutation distribution of $\sqrt{n}\hat{\rho}_n$ and $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. By contiguity, it follows that

$$\hat{r}_n(1-\alpha) \xrightarrow{P} z_{1-\alpha}$$

under P_n. Using Slutsky's theorem, this gives that the probability under P_n that the permutation test rejects H_0 converges to $P(Y > z_{1-\alpha})$ where Y is distributed as the limiting distribution of $T_n = \sqrt{n}\hat{\rho}_n$ under P_n. In the case where P_n is a bivariate normal with correlation $\rho_n = h/\sqrt{n}$,

$$\sqrt{n}\hat{\rho}_n \xrightarrow{\mathcal{L}} N(h, 1)$$

under P_n and the limiting power against this sequence of alternatives is $1 - \Phi(z_{1-\alpha} - h)$. For a bivariate normal distribution, the Rao Score test which rejects for large values of $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i Y_i$ is known to be locally asymptotically uniformly most powerful (see Lehmann and Romano Theorem 13.3.2 and Example 13.3.5). Since the studentized permutation test has the same limiting local power, the limiting local power of the studentized permutation test is optimal. Thus, the loss of power of the permutation test is negligible in large samples. However, the permutation test is widely robust against non-normality.

Remark 2.5 (Limiting Type 3 Error Rate) *Because this test controls the Type 1 error rate, the Type 3 error rate is also controlled. Under contiguous alternatives, the Type 3 error rate is asymptotically bounded above by* $\alpha/2$. *For instance, let* P_n *be a sequence of normal contiguous alternatives with correlations* $\rho_n = h/\sqrt{n}$, *where* $h > 0$. *Then the probability that a Type 3 error occurs, i.e. the limiting probability that the test rejects and the sample correlation is negative, is bounded bounded by* $(1 - \Phi(z_{1-\alpha} - h))/2 < \alpha/2$.

The same techniques can be used to correct for permutation tests based on functions of the sample correlation. For instance, a permutation test using Fisher's z-transformation is not guaranteed to be level α, but a studentized version is asymptotically level α. If the (X_i, Y_i)'s follow a bivariate normal distribution with correlation ρ, then

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} N(0, (1-\rho^2)^2).$$

Fisher proposed using the variance stabilizing transformation $\tanh^{-1}(\rho) = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$, which has derivative $\frac{1}{1-\rho^2}$ so that

$$\sqrt{n}(\tanh^{-1}(\hat{\rho}) - \tanh^{-1}(\rho)) \xrightarrow{\mathcal{L}} N(0,1).$$

When the data is not normally distributed but has finite fourth moments, this transformation is no longer variance stabilizing (it is readily seen using the delta method that $\mathrm{var}(\tanh^{-1}(\hat{\rho})) \approx \mathrm{var}(\hat{\rho})/(1-\rho^2)^2$ which is not constant when the data is not normally distributed). When $\rho = 0$,

$$\sqrt{n}(\tanh^{-1}(\hat{\rho}) - \tanh^{-1}(\rho)) \xrightarrow{\mathcal{L}} N\left(0, \tau^2(P)\right),$$

but the permutation distribution of $\sqrt{n}\tanh^{-1}(\hat{\rho})$ has asymptotic variance 1. This problem can again be fixed asymptotically by studentizing by $\hat{\tau}_n$ (as defined above).

Theorem 2.6 *Assume* $(X_1, Y_1), ..., (X_n, Y_n)$ *are i.i.d. such that* X_1 *and* Y_1 *are uncorrelated but not necessarily independent. If* $E(X_1^4) < \infty$ *and* $E(Y_1^4) < \infty$, *then the permutation distribution* $R_n^{U_n}(t)$ *of* $U_n = \sqrt{n}\tanh^{-1}(\hat{\rho}_n)/\hat{\tau}_n$ *satisfies*

$$\sup_{t \in \mathbb{R}} \left|\hat{R}_n^{U_n}(t) - \Phi(t)\right| \to 0$$

in probability. Under the stronger assumption that $E(X_1^8) < \infty$ *and* $E(Y_1^8) < \infty$,

$$\lim_{n\to\infty} \sup_{t \in \mathbb{R}} \left|\hat{R}_n^{U_n}(t) - \Phi(t)\right| = 0$$

almost surely.

Of course, if bivariate normality holds, then $\rho = 0$ implies independence and each of the permutation tests discussed control Type 1 (and Type 3) errors.

3 Applications to Linear Regression

The techniques of using studentizing test statistics for asymptotically valid permutation tests seen in the previous section can be extended to regression problems. As motivation, consider a simple univariate linear regression model

$$Y = \alpha + \beta X + \epsilon,$$

where $X_i \in \mathbb{R}$ and ϵ_i are errors with mean zero and variance σ^2. (For the moment, assumptions on the joint distribution of X and ϵ are not specified, but will be described below). To test the hypothesis

$$H : \beta = 0,$$

it is natural to base a test on the sample correlation

$$\sqrt{n}\hat{\rho}_n\left(X, Y\right) = \sqrt{n}\frac{\sum X_i Y_i - n\bar{X}_n\bar{Y}_n}{\sqrt{\sum(X_i - \bar{X}_n)^2 \sum(Y_i - \bar{Y}_n)^2}}$$

or the least squares estimator $\hat{\beta}_n$. If the X_i's are independent of the ϵ_i's, then an exact permutation test can be performed by permuting the X_i's. A permutation test may not be exact if the predictors and errors are uncorrelated, however, following the results of the previous section, studentizing the correlation coefficient leads to an asymptotically level α test. Using the same techniques, asymptotically valid permutation tests can be performed in multiple linear regression as well.

Remark 3.1 *Throughout this section, we will assume that the constant is included in the regression model. Otherwise, it is possible that the permutation distribution of* $\sqrt{n}\hat{\beta}_n$ *will not approximate the sampling distribution asymptotically when* $\beta = 0$*. For example, in the simple linear regression case discussed above, if the constant is not fit, basing a permutation test on* $\sqrt{n}\hat{\beta}_n$ *is equivalent to basing a test on* $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i Y_i$*. The distribution of* $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i Y_{\pi(i)}$ *conditional on the* X_i*'s and* Y_i*'s has mean* $n\bar{X}_n\bar{Y}_n$*, which may not converge to zero. But the sampling distribution has mean zero asymptotically, so it will not be approximated by the permutation distribution. A more rigorous justification of this claim is given in the appendix.*

Consider the regression model specified by the following assumptions:

(1) $Y_i = \alpha + X_i^\top\beta + \epsilon_i$, $i = 1, ..., n$ where $X_i \in \mathbb{R}^p$ is a vector of predictor variables, and ϵ_i is a mean zero error term.

(2) $\{(Y_i, X_i)\}$ are i.i.d. according to a distribution P with $E(\epsilon_i \cdot X_i) = 0$.

(3) $\Sigma_{XX} := E((X_i - \mu_X)(X_i - \mu_X)^\top)$ and $\Omega := E(\epsilon_i^2(X_i - \mu_X)(X_i - \mu_X)^\top)$ are nonsingular. Furthermore, $\sum_{i=1}^n X_i X_i^\top$ is almost surely invertible.

When $\{X_i, Y_i\}$ pairs are observed, one can define $\epsilon_i = Y_i - X_i^\top\beta$, where β is chosen so that X_i and ϵ_i are uncorrelated. β can therefore be interpreted as the slope of the best fitting line, and assumption (2) is unnecessary. A commonly used sub-family of the regression models specified by assumptions (1)-(3) is heteroskedastic models, which assume that the conditional variance of Y_i given X_i changes with X_i, satisfy assumptions (1)-(3). For these models, there exists a non-constant skedastic function $\sigma^2(X_i) = E\left(\epsilon_i^2|X_i\right)$. (Note that we are not assuming $E(\epsilon_i|X_i) = 0$, so the skedastic function may not be the conditional variance.)

Since the constant is included in the model, we may assume, without loss of generality, that the X_i's and Y_i's have been standardized to have sample mean zero (i.e. consider the regression of $Y - \bar{Y}$ on $X - \bar{X}$). Under these model assumptions, White (1980) showed that the ordinary least squares estimator $\hat{\beta}_n = (X^\top X)^{-1}X^\top Y$ is asymptotically normal,

$$\sqrt{n}(\hat{\beta}_n - \beta) \to N(0, \Sigma_{XX}^{-1}\Omega\Sigma_{XX}^{-1}).$$

In the case of heteroskedasticity, the center matrix in the covariance can be written as $\Omega = E(\sigma^2(X_i)X_iX_i^\top)$. As will be seen in the proof of Theorem 3.1, permutation distribution of $\sqrt{n}(\hat{\beta}_n - \beta)$ is asymptotically normal with mean zero and variance $E(\epsilon_i^2) \times \Sigma_{XX}^{-1}$ when $\beta = 0$. Unless $E(\epsilon_i^2 X_iX_i^\top) = E(\epsilon_i^2)E(X_iX_i^\top)$, the covariance of the permutation distribution is not equal to that of the sampling distribution. Consequently, a permutation test of the hypothesis $H_0 : \beta = 0$ using the usual F-statistic will not be exact, and will not even be asymptotically valid when the predictor and error terms are dependent.

Our aim is to test the hypothesis

$$H_0 : \beta = 0.$$

To test this hypothesis, two randomization tests will be considered, each using the studentized test statistic

$$S_n(X, Y) = n\hat{\beta}_n^\top \left(\hat{\Sigma}_{XX}^{-1}\hat{\Omega}\hat{\Sigma}_{XX}^{-1}\right)\hat{\beta}_n$$

where $\hat{\Sigma}_{XX} = \frac{1}{n}\sum_i X_i^\top X_i$ and $\hat{\Omega} = \frac{1}{n}\sum_i Y_i^2 X_iX_i^\top$.

If the ϵ_i are independent of the X_i, we can use this statistic for a permutation test done by permuting the $Y_i's$. This test will not be exact if there is dependence, but since the statistic is asymptotically pivotal, we expect that the permutation test will be asymptotically level α.

Theorem 3.2 *Suppose that $\{(X_i, Y_i)\}_{i=1}^n$ satisfies the regression model described by conditions (1)-(3) above, and also assume that $E(Y_1^4) < \infty$ and $E(X_{1j}^4) < \infty$, $j = 1, ..., p$. If $\beta = 0$, then the permutation distribution $\hat{R}_n^{T_n,\pi}(t)$ of S_n obtained by permuting the Y_i's satisfies*

$$\lim_{n\to\infty}\sup_{t\in\mathbb{R}}\left|\hat{R}_n^{S_n,\pi}(t) - R(t)\right| \xrightarrow{P} 0$$

where $R(\cdot)$ is the law of a χ_p^2 random variable. Therefore,

$$\lim_{n\to\infty}\sup_{t\in\mathbb{R}}\left|\hat{R}_n^{S_n,\pi}(t) - J_n^S(t, P)\right| = 0.$$

where $J^S(t, P)$ is the sampling distribution of S_n.

Alternatively, if it is assumed that the X_i are independent of the ϵ_i, and the errors are symmetric, then an exact randomization test can be obtained using the group of transformations $\mathbf{G}_n^\delta = \{g_\delta : \delta \in \{1, -1\}^n\}$ such that $g_\delta(y_1, ..., y_n) = (\delta_1 y_1, ..., \delta_n y_n)$ for any $y \in \mathbb{R}^n$. In particular, if the errors are symmetric and $\beta = 0$, then $S_n(X, g_\delta(Y))$ is distributed as $S_n(X, Y)$

for any uniformly chosen transformation g_δ and the test is exact because the randomization hypothesis is satisfied. The next theorem studies the asymptotic behavior of the permutation distribution

$$\hat{R}_n^{S_n,\delta}(t) = \frac{1}{2^n} \sum_{g_\delta \in \mathbf{G}_n^\delta} I\left\{S_n(X, g_\delta(Y)) \leq t\right\}$$

when the errors are not assumed to be symmetric or independent of the X_i, but instead satisfy $E\left(\epsilon_i \cdot X_i\right) = 0$.

Theorem 3.3 *Suppose that $\{(X_i, Y_i)\}_{i=1}^n$ satisfies the regression model described by conditions (1)-(3) above. If $\beta = 0$, then the permutation distribution $\hat{R}_n^{S_n,\delta}(t)$ of S_n obtained changing the sign of the Y_i's satisfies*

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left|\hat{R}_n^{S_n,\delta}(t) - R(t)\right| \xrightarrow{P} 0$$

where $R(\cdot)$ is the law of a χ_p^2 random variable. Therefore,

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left|\hat{R}_n^{S_n,\delta}(t) - J_n^S(t, P)\right| = 0,$$

where $J^S(t, P)$ is the sampling distribution of S_n.

Under the sequence of local alternatives $\beta = h/\sqrt{n}$, the permutation distribution of S_n remains asymptotically χ_p^2 under either of the methods of permuting the data described above. On the other hand, the sampling distribution of S_n under these alternatives is asymptotically chi squared with p degrees of freedom and non-centrality parameter

$$\lambda = \left\|\Omega^{-1/2}\Sigma_{XX}h\right\|_2^2.$$

and so the local power of either of the randomization tests is $P\left(C > \chi_{p,1-\alpha}^2\right)$ where $C \sim \chi_p^2\left(\lambda\right)$ and $\chi_{p,1-\alpha}^2$ is the $1-\alpha^{th}$ quantile of a χ_p^2 distribution. Therefore, the two tests have the same limiting local power functions under such alternative sequences. An advantage of the randomization test conducted by permuting the Y_i's is that it can be extended to test if only a subset of the regression coefficients are zero.

Suppose instead that we are interested in testing only a subset of the coefficients. Then, a subset of the regression coefficients may be non-zero, and the Y_i's may no longer have mean zero (conditionally). Even if the errors are symmetric, the randomization test using sign changes will no longer work since Y_i and $-Y_i$ will have different means. However, an asymptotically valid permutation test can still be conducted by permtuing the regressors corresponding to the coefficients of interest.

Consider the multiple linear regression model specified by the following assumptions:

(1') $Y_i = \alpha + X_i^\top \beta + Z_i^\top \gamma + \epsilon_i$, $i = 1, ..., n$ where $X_i \in \mathbb{R}^k$ and $Y_i \in \mathbb{R}^{p-k}$ are vectors of predictor variables, and ϵ_i is a mean zero error term.

(2') $\{(Y_i, X_i, Z_i)\}$ are i.i.d. according to a distribution P with $E(\epsilon_i|X_i, Z_i) = 0$.

(3') $\Sigma_{\tilde{X}\tilde{X}} := E((\tilde{X}_i - \mu_{\tilde{X}})(\tilde{X}_i - \mu_{\tilde{X}})^\top)$ and $\Omega := E(\epsilon_i^2(\tilde{X}_i - \mu_{\tilde{X}})(\tilde{X}_i - \mu_{\tilde{X}})^\top)$ are nonsingular where $\tilde{X}_i^\top = (X_i^\top, Z_i^\top)$. Furthermore, $\sum_{i=1}^n \tilde{X}_i\tilde{X}_i^\top$ is almost surely invertible.

Without loss of generality, assume that the X_i's, Y_i's and Z_i's have been standardized to have sample mean zero. In this setting, the ordinary least squares estimator, $(\hat{\beta}_n^\top, \hat{\gamma}_n^\top)^\top = (\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top Y$ is asymptotically normal:

$$\sqrt{n}\left(\begin{pmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{pmatrix} - \begin{pmatrix} \beta \\ \gamma \end{pmatrix}\right) \xrightarrow{\mathcal{L}} N(0, \Sigma_{\tilde{X}\tilde{X}}^{-1}\Omega\Sigma_{\tilde{X}\tilde{X}}^{-1}).$$

To test the hypothesis

$$H : \beta = 0,$$

one could use a Wald statistic

$$W_n(X, Z, Y) = n \cdot \left(R\begin{pmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{pmatrix}\right)^\top \left[R\hat{\Sigma}_{\tilde{X}\tilde{X}}^{-1}\hat{\Omega}\hat{\Sigma}_{\tilde{X}\tilde{X}}^{-1}R^\top\right]^{-1}\left(R\begin{pmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{pmatrix}\right) \tag{2}$$

with $R = \text{diag}(1, ..., 1, 0, ..., 0)$, $\hat{\Sigma}_{\tilde{X}\tilde{X}} = \frac{1}{n}\sum_i \tilde{X}_i^\top\tilde{X}_i$ and $\hat{\Omega} = \frac{1}{n}\sum_i \hat{\epsilon}_i^2\tilde{X}_i\tilde{X}_i^\top$ (which are consistent estimators of Σ and Ω respectively). This Wald statistic is asymptotically χ_k^2.

If the X_i's are independent of the (Z_i, Y_i)'s, then for any permutation π of $\{1, ..., n\}$, $(X_{\pi(i)}, Z_i, Y_i)$ is distributed as (X_i, Z_i, Y_i). Moreover, $W(X_\pi, Z, Y)$ is distributed as $W(X, Z, Y)$. The randomization hypothesis is satisfied, and a permutation test conducted by permuting the X_i's while keeping the (Y_i, Z_i) pairs together is exact. When there is dependence, the test is no longer exact, but the permutation distribution

$$\hat{R}_n^{W_n}(t) = \frac{1}{n!}\sum_{\pi \in \mathbf{G}_n} I\left\{W(X_{\pi(i)}, Z_i, Y_i) \le t\right\}$$

is asymptotically χ_p^2 and the test is asymptotically level α.

Theorem 3.4 *Suppose that $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ satisfies the regression model specified by conditions (1')-(3'). If $\beta = 0$, the permutation distribution $\hat{R}_n^W(t)$ of W_n obtained by permuting the X_i's satisfies*

$$\lim_{n\to\infty}\sup_{t\in\mathbb{R}}\left|\hat{R}_n^{W_n}(t) - R(t)\right| \xrightarrow{P} 0$$

where $R(\cdot)$ is the law of a χ_k^2 random variable. Therefore,

$$\lim_{n\to\infty}\sup_{t\in\mathbb{R}}\left|\hat{R}_n^{W_n}(t) - J_n^{W_n}(t, P)\right| = 0.$$

where $J_n^{W_n}(t, P)$ is the sampling distribution of W_n.

Another approach to testing $H_0 : \beta = 0$ using permutations of the data was suggested by Freedman and Lane (1983). They proposed the following procedure:

- Fit the model $Y = X\hat{\beta}_n + Z\hat{\gamma}_n + \hat{\epsilon}$

- Compute

$$F_n(Y,X,Z) = n \cdot \left(R \begin{pmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{pmatrix} \right)^\top \left[R s^2 (\tilde{X}^\top \tilde{X})^{-1} R^\top \right]^{-1} \left(R \begin{pmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{pmatrix} \right),$$

  the F-statistic to test $H : \beta = 0$ with $s^2 = \frac{1}{n-p} \sum_{i=1}^n \hat{\epsilon}_i^2$

- Fit $Y = Z\hat{\gamma}_n + \hat{\epsilon}$
- Generate new samples $Y_i^* = Z_i^\top \hat{\gamma}_n + \hat{\epsilon}_{\pi(i)}$
- Regress Y^* on X and Z, get a new F statistic, say F^π
- Compare the original F-statistic with the appropriate quantiles of the permutation distribution

$$\hat{P}_n^{F_n}(t) = \frac{1}{n!} \sum_\pi I\left\{ F^\pi \le t \right\}$$

Their procedure does not appear to be exact under any circumstance, but is asymptotically valid when the regressors are independent of the error terms. This procedure may fail to be asymptotically valid under the relaxed model assumptions (1')-(3'). When $\beta = 0$, the sampling distribution of $\sqrt{n}\hat{\beta}_n$ is asymptotically $N(0, R\Sigma_{\epsilon XZ}R^\top)$ where

$$\Sigma_{\epsilon XZ} = \begin{pmatrix} E(\epsilon_i^2 X_i X_i^\top) & E(\epsilon_i^2 Z_i X_i^\top) \\ E(\epsilon_i^2 Z_i X_i^\top) & E(\epsilon_i^2 Z_i Z_i^\top) \end{pmatrix}.$$

Asymptotically, the F-statistic is distributed as $N^\top E(\epsilon_i^2) R\Sigma_{XZ}R^\top N$, where $N \sim N(0, R\Sigma_{\epsilon XZ}R^\top)$ and

$$\Sigma_{XZ} = \begin{pmatrix} E(X_i X_i^\top) & E(Z_i X_i^\top) \\ E(Z_i X_i^\top) & E(Z_i Z_i^\top) \end{pmatrix}.$$

Therefore, the sampling distribution of the F-statistic is not always asymptotically chi-squared. Nevertheless, the permutation distribution, which behaves as though the regressors are independent of the error terms, will always be asymptotically chi-squared. This procedure is asymptotically correct if instead of the F-statistic, the Wald statistic defined in Equation 2 is used.

Theorem 3.5 *Suppose that $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ satisfies the regression model specified by conditions (1')-(3'). If $\beta = 0$, the permutation distribution $\hat{P}_n^W(t)$ of W_n obtained by the Freedman-Lane procedure satisfies*

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| \hat{P}_n^{W_n}(t) - R(t) \right| \xrightarrow{P} 0$$

where $R(\cdot)$ is the law of a χ_k^2 random variable. Therefore,

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| \hat{P}_n^{W_n}(t) - J_n^{W_n}(t, P) \right| = 0.$$

where $J_n^{W_n}(t, P)$ is the sampling distribution of W_n.

While the Freedman-Lane procedure does not have any of the usual exactness properties of a permutation test, it may be favorable to the previous permutation test when there is dependence between the regressors of interest and nuisance regressors.

4 Partial Correlation

Suppose we have univariate variables X and Y, and a multivariate variable Z which satisfy model assumptions (1')-(3'). The partial correlation between X and Y given Z, denoted $\rho_{X,Y|Z}$ is the correlation between the residual $R_X = X_i - Z_i^\top E(Z_i Z_i^\top)^{-1} E(Z_i X_i)$ of regressing X on Z and the residual $R_Y = Y_i - Z_i^\top E(Z_i Z_i^\top)^{-1} E(Z_i Y_i)$ of regressing Y on Z.

The problem of testing partial correlation is related to the problem of inference for a single regression coefficient in the presence of nuisance regressors. The sample partial correlation is proportional to the ordinary least squares estimate of the coefficient β in the model $Y = X\beta + Z\gamma + \epsilon$, and testing that the sample correlation is zero is equivalent to testing $H_0 : \beta = 0$. Consequently, either of the randomization tests for this hypothesis proposed in the previous section are appropriate for testing partial correlation. Alternatively, a randomization test can be based on permuting residuals and recomputing the partial correlation on permuted residuals.

Write

$$r_X = X - (Z^\top Z)^{-1} Z^\top X = X - \hat{X}$$

and

$$r_Y = Y - (Z^\top Z)^{-1} Z^\top Y = Y - \hat{Y}.$$

The sample partial correlation is the sample correlation between r_X and r_Y

$$\hat{\rho}_{X,Y|Z} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)(X_i - \hat{X}_i)}{\sqrt{\sum_{i=1}^n (X_i - \hat{X}_i)^2 \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}}$$

It is easily seen that the sample correlation is related to the ordinary least squares estimate of β:

$$\hat{\rho}_{X,Y|Z} = \sqrt{\frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}} \hat{\beta}_n.$$

Define

$$\check{X}_i = X_i - Z_i^\top E(Z^\top Z) E(Z^\top X)$$

and

$$\check{Y}_i = Y_i - Z_i^\top E(Z^\top Z) E(Z^\top Y).$$

When the partial correlation is zero, the asymptotic distribution of $\hat{\rho}_{X,Y|Z}$ is normal with mean zero and variance $\sigma^2 = \frac{E(\check{X}^2 \check{Y}^2)}{E(\check{X}^2) E(\check{Y}^2)}$. However, the permutation distribution of the sample correlation computed on permuted residuals is asymptotically standard normal. If we instead use the studentized statistic $\hat{\rho}_{X,Y|Z} / \hat{\sigma}_n$ where

$$\hat{\sigma}_n = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 (X_i - \hat{X}_i)^2}{\sum_{i=1}^n (X_i - \hat{X}_i)^2 \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}},$$

then both the sampling distribution and the permutation distribution are asymptotically standard normal when the partial correlation is zero.

Theorem 4.1 *Assume* $(X_1, Y_1, Z_1), ..., (X_n, Y_n, Z_n)$ *are i.i.d. according to* P *such that* X_1 *and* Y_1 *are uncorrelated but not necessarily independent conditionally on the* Z_i*. Also assume that* $E(X_1^4) < \infty$ *and* $E(Y_1^4) < \infty$*. The permutation distribution* $\hat{R}_n^{T_n}(t)$ *of* $T_n = \sqrt{n}\hat{\rho}_{X,Y|Z}/\hat{\sigma}_n$ *satisfies*

$$\sup_{t \in \mathbb{R}} \left| \hat{R}_n^{T_n}(t) - \Phi(t) \right| \to 0$$

in probability.

Although asymptotically valid, the permutation test using permutations of the X observations discussed in the previous section may not be appropriate when X and Y are not independent of Z. However, the Freedman-Lane procedure, and the test permuting the residuals preserve the dependence of X and Y on Z and are a better choice for testing partial correlation.

5 Simulations

Recall that a random vector X is said to have an elliptical distribution with parameters μ (a vector) and Σ (a positive definite symmetric matrix) if it has a characteristic function of the form $\phi(t) = \exp(it'\mu)\psi(t'\Sigma t)$, where ψ is a scalar function. If X has an elliptical distribution, then $E(X) = \mu$ and $\text{cov}(X) = -\frac{\partial \psi(0)}{\partial t}\Sigma$, i.e. the covariance is proportional to Σ. The elliptical distributions provide many joint distributions having correlation zero but dependence between the variables since correlation zero implies independence only when the data is normally distributed.

The density functions of elliptical distributions have the form $f(x) = cg((x-\mu)'\Sigma^{-1}(x-\mu))$, where g is a univariate density, and c is the appropriate normalizing constant. For instance, if we take g to be the density function of a t distribution with ν degrees of freedom, we get the distribution function of a d-dimensional multivariate t-distribution, denoted by $t_\nu(\mu, \Sigma)$:

$$f(x) = \frac{1}{|\Sigma|^{1/2}(\nu\pi)^{d/2}} \cdot \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)} \cdot \left(1 + \frac{x'\Sigma^{-1}x}{v}\right)^{-(\nu+d)/2}.$$

When $\mu = 0$, samples from this distribution can be generated by dividing a d-dimensional normal vector with mean 0 and covariance Σ by an independent chi-squared random variable with ν degrees of freedom. We will consider a bivariate t distribution with covariance the identity and 5 degrees of freedom (so that there are finite fourth moments).

Also, an elliptically distributed, p-dimensional random vector X can be obtained by taking

$$X = \mu + rS'u^{(k)}$$

where $r \geq 0$ is a random variable, S is a $k \times p$ matrix satisfying $S'S = \Sigma$, and $u^{(k)}$ is a random variable that is uniformly distributed on the k-dimensional unit sphere and independent of r. For example, we will consider the case where $k = p = 2$, μ is zero, $\Sigma = \text{diag}(2, 1)$, and $r \sim \exp(1)$.

Another common example of a distribution is the circular uniform distribution, which gives (X, Y) uniformly distributed on the unit circle.

Distribution	n:	10	25	50	100	200
$N(0,1)$	Bootstrap	0.0800	0.0700	0.0700	0.0640	0.0580
	Normal Approx	0.2240	0.1220	0.0840	0.0820	0.0600
	Not Studentized	0.0480	0.0480	0.0440	0.0500	0.0480
	Studentized	0.0480	0.0540	0.0480	0.0600	0.0500
Multivariate t_5	Bootstrap	0.0920	0.0820	0.0940	0.0620	0.0640
	Normal Approx	0.2960	0.1640	0.1260	0.1000	0.0760
	Not Studentized	0.0920	0.0980	0.1580	0.1680	0.1640
	Studentized	0.0760	0.0580	0.0700	0.0560	0.0570
Exponential	Bootstrap	0.0700	0.0620	0.0380	0.0500	0.0700
	Normal Approx	0.1320	0.0680	0.0580	0.580	0.0660
	Not Studentized	0.0240	0.0120	0.0100	0.0060	0.0020
	Studentized	0.0360	0.0400	0.0460	0.0560	0.0580
Circular	Bootstrap	0.0500	0.0540	0.0600	0.0620	0.0480
	Normal Approx	0.1600	0.0800	0.0680	0.0600	0.0440
	Not Studentized	0.0080	0.0020	0.0060	0.0040	0.0200
	Studentized	0.0320	0.0300	0.0420	0.0510	0.0480
$t_{4.1}$	Bootstrap	0.0960	0.0880	0.0920	0.0800	0.0660
	Normal Approx	0.3100	0.2180	0.1520	0.1180	0.0780
	Not Studentized	0.1040	0.1920	0.2440	0.2960	0.2900
	Studentized	0.0820	0.1020	0.1140	0.0980	0.0700

Table 1: Rejection probabilities for bootstrap, normal approximation and permutation tests for $\rho = 0$ using the sample correlation statistic.

It is also seen in the appendix that $\frac{\mu_{22}}{\mu_{20}\mu_{02}}$ can be made arbitrarily large by taking $X = W+Z$ and $Y = W - Z$, where W and Z are independent t_v random variables with v close to four. As an example, we will consider $v = 4.1$.

Finally, we consider two examples where the data used is of the form $X = W + Z$ and $Y = W - Z$, where X and Y are independent and identically distributed. In the first example, we will choose $W, X \sim t_{4.1}$ so that the ratio $\frac{\mu_{22}}{\mu_{20}\mu_{02}}$ is large (and the quantiles of the limiting sampling distribution are much larger than those of the permutation distribution of the unstudentized test statistic).

Table 1 presents rejection probabilities estimated by 1000 simulations at nominal level $\alpha = .05$. Results are for tests using the bootstrap distribution, the normal approximation, the unstudentized permutation, and the studentized permutation distribution. For the permutation tests, 800 permutations were used.

Table 2 below gives the same simulations as Table 1, but instead uses Fisher's z-transformation. Because Fisher's z-transformation is monotonic, the unstudentized test rejects exactly when the unstudentized permutation test based on the untransformed correlation rejects.

The simulations for testing sample correlation indicate that an unstudentized permutation test can indeed have rejection probability , even in large sample. The studentized permutation test has the desired rejection probability for large n, and appears to have rejection probability much closer to the nominal level when compared to the normal approximation, especially in

Distribution	n:	10	25	50	100	200
$N(0,1)$	Normal Approx	0.0540	0.0500	0.0520	0.0520	0.0480
	Normal Approx Studentized	0.1560	0.0900	0.0760	0.0720	0.0540
	Not Studentized	0.0480	0.0480	0.0440	0.0500	0.0480
	Studentized	0.0460	0.0540	0.0520	0.0600	0.0500
Multivariate t_5	Normal Approx	0.1080	0.1580	0.2120	0.2220	0.2080
	Normal Approx Studentized	0.2220	0.1360	0.1220	0.1060	0.0740
	Not Studentized	0.0920	0.0980	0.1580	0.1680	0.1640
	Studentized	0.0620	0.0580	0.0660	0.0800	0.0640
Exponential	Normal Approx	0.0160	0.0120	0.0080	0.0040	0.0060
	Normal Approx Studentized	0.0960	0.0740	0.0460	0.0540	0.0480
	Not Studentized	0.0240	0.0120	0.0100	0.0060	0.0020
	Studentized	0.0360	0.0540	0.0380	0.0480	0.0480
Circular	Normal Approx	0.0140	0.0160	0.0100	0.0060	0.0040
	Normal Approx Studentized	0.1180	0.0540	0.0400	0.0560	0.0600
	Not Studentized	0.0080	0.0020	0.0060	0.0040	0.0200
	Studentized	0.0360	0.0540	0.0340	0.0580	0.0560
$t_{4.1}$	Normal Approx	0.1700	0.2060	0.2520	0.2860	0.3080
	Normal Approx Studentized	0.2260	0.1500	0.1320	0.0940	0.0700
	Not Studentized	0.1040	0.1920	0.2240	0.2960	0.2900
	Studentized	0.0640	0.0760	0.0920	0.0760	0.0620

Table 2: Rejection probabilities for bootstrap, normal approximation and permutation tests for $\rho = 0$ using Fisher's z-transformation.

Distribution	n:	10	25	50	100	200
t_5	Normal Approx	0.0880	0.0280	0.0160	0.0040	0.0060
	Not Studentized	0.0220	0.0320	0.0220	0.0180	0.0220
	Studentized	0.0140	0.0140	0.0080	0.0020	0.0040
Normal	Normal Approx	0.0740	0.0500	0.0080	0.0040	0.0000
	Not Studentized	0.0140	0.0240	0.0060	0.0000	0.0020
	Studentized	0.0160	0.0280	0.0080	0.0020	0.0000

Table 3: Type 3 error rates for normal approximation and permutation tests of $\rho = 0$ using the sample correlation statistic.

small sample sizes. Further, using Fisher's z-transformation appears to perform similarly to using the untransformed correlation.

Table 3 reports the Type 3 error rate when the data is generated according to a multivariate t-distribution (as described above) with 5 degrees of freedom, as well as a multivariate normal distribution. For both examples, we will use

$$\Sigma = \begin{pmatrix} 1 & .1 \\ .1 & 1 \end{pmatrix}$$

as the covariance matrix.

The studentized permutation test exhibits a lower Type 3 error rate than the unstudentized test, or the normal approximation, especially in small samples.

Finally, we present simulation results showing the performance of the permutation tests described in Section 3 for linear regression. Table 4 compares the rejection probabilities of the two methods of permutation test with the normal approximation when the nominal level is $\alpha = 0.05$.

For the simulations in Table 4, the regression model used is

$$Y = \alpha + \beta X + \epsilon$$

where $X_i \sim U(1, 4)$ and $\epsilon_i = \sigma(X_i) \cdot N_i$ where N_i is a standard normal random variable, and $\sigma(\cdot)$ is a skedastic function specified in the table. The simulations are performed with $\beta = 0$ so that the rejection probabilities reported are the Type 1 error rates at nominal level 0.05.

The next simulation is a comparison of the rejection probability of the normal approximation, and permutation tests of the hypothesis $H_0 : \beta = 0$ in the regression model

$$Y = \alpha + \beta X + \gamma Z + \epsilon.$$

Table 5 gives rejection probabilities for simulations under this model using $\gamma = 1$, $X_i \sim U(1, 4)$, $Z \sim N(0, 1)$ and $\epsilon_i = \sigma(X_i) \cdot N_i$ where N_i is a standard normal random variable, and $\sigma(\cdot)$ is a skedastic function specified in the table. Again, we simulate under the null hypothesis and report the Type 1 error rates at nominal level 0.05.

Similarly to what was seen in the simulations for correlation, a studentized permutation test for regression coefficients has a rejection probability that is far closer to the nominal level

	n:	10	25	50	100	200
$\sigma(x) = 1$	Permutation	0.0440	0.0560	0.0520	0.0530	0.0490
	Sign Change	0.0490	0.0540	0.0530	0.0480	0.0530
	Normal Approx	0.1270	0.0860	0.0650	0.0570	0.0540
$\sigma(x) = \|x\|$	Permutation	0.0690	0.0660	0.0560	0.0530	0.0510
	Sign Change	0.0910	0.0670	0.0580	0.0560	0.0510
	Normal Approx	0.1790	0.1080	0.0640	0.0590	0.0550
$\sigma(x) = \|\log(x)\|$	Permutation	0.0650	0.0570	0.0630	0.0580	0.0520
	Sign Change	0.0990	0.0720	0.0670	0.0660	0.0570
	Normal Approx	0.1940	0.0940	0.0740	0.0670	0.0580

Table 4: Rejection probabilities for tests of $\beta = 0$.

	n:	10	25	50	100	200
$\sigma(x) = 1$	Permutation	0.0610	0.0570	0.0600	0.0540	0.0560
	Normal Approx	0.2480	0.1240	0.0750	0.0590	0.0600
$\sigma(x) = \|x\|$	Permutation	0.0680	0.0670	0.0600	0.0620	0.0540
	Normal Approx	0.2780	0.1330	0.0820	0.0810	0.0610
$\sigma(x) = \|\log(x)\|$	Permutation	0.0770	0.0730	0.0570	0.0550	0.0540
	Normal Approx	0.2810	0.1430	0.0850	0.0640	0.0620

Table 5: Rejection probabilities for tests of $\beta = 0$.

than using a normal approximation. The randomization test using sign changes outperforms the normal approximation, but appears to have a rejection probability that is farther above the nominal level than the permutation test.

6 Empirical Applications

Bardsley and Chambers (1984) collected data on the number of cattle and sheep on large farms in Australia. This data was used for an exercise in Sprent and Smeeton (2007) asking if there is evidence of correlation.

Sheep	Cattle
4716	41
4605	0
4951	42
2745	15
6592	47
8934	0
9165	0
5917	0
2618	56
1105	67
150	707
2005	368
3222	231
7150	104
8658	132
6304	200
1800	172
5270	146
1537	0

The p-values found by the normal approximation, the permutation test, and the studentized permutation tests are 0.1682, 0.0390, and 0.1670 respectively. Figure 1 shows the unstudentized permutation distribution compared with a $N(0, \hat{\tau}_n^2)$ density, and figure 2 shows the studentized permutation distribution compared with a $N(0, 1)$ density.

The following is a randomly chosen subset of the measurements of blood pressure and obesity taken from a South African heart disease data set used in The Elements of Statistical Learning (Hastie, Tibshirani, and Friedman).

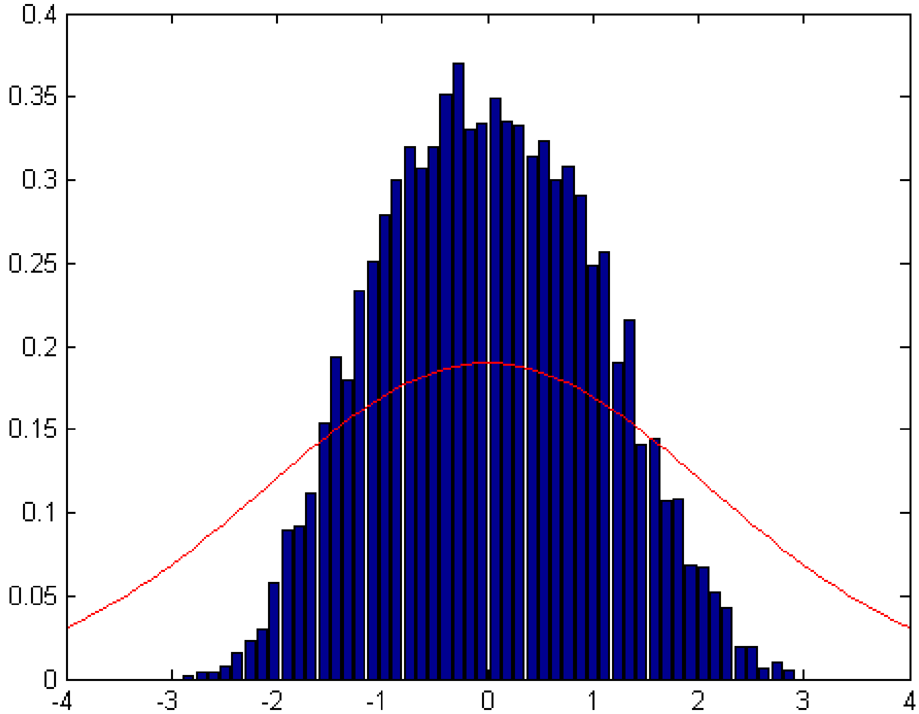

Figure 1: Unstudentized permutation distribution

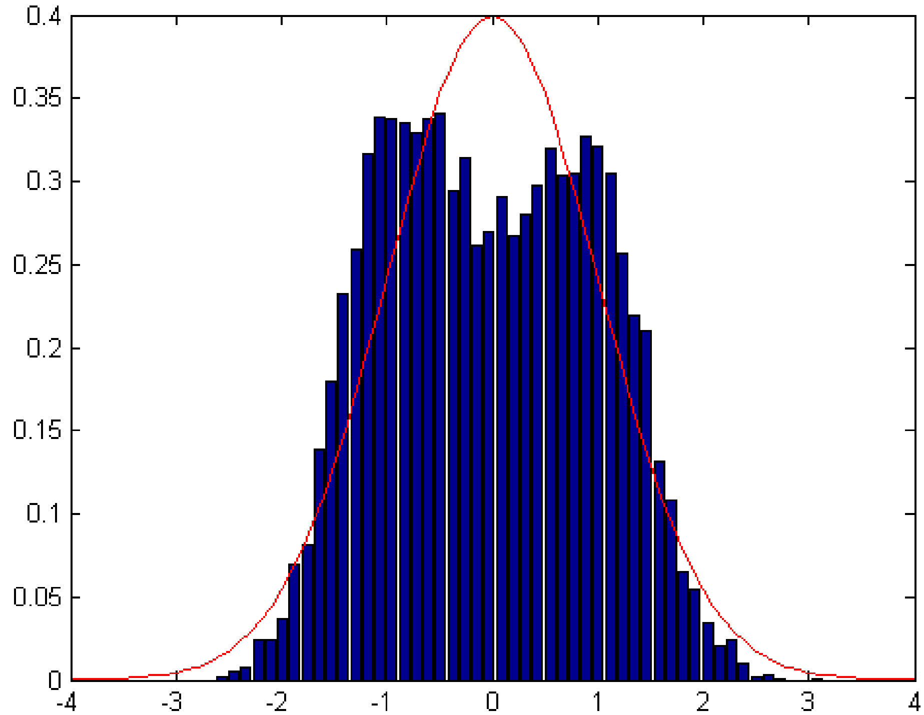

Figure 2: Studentized permutation distribution

Blood Pressure	Obesity
142.0000	24.9800
134.0000	26.0900
114.0000	17.8900
126.0000	27.6200
116.0000	30.0100
118.0000	30.9800
122.0000	28.0700
180.0000	26.7000
138.0000	26.0600
132.0000	30.7700
134.0000	29.1600
148.0000	31.8300
114.0000	20.1700
130.0000	24.3400
136.0000	25.0000
118.0000	25.1500
146.0000	36.4600
108.0000	22.6100
118.0000	29.1400
126.0000	19.3700
206.0000	27.3600
117.0000	25.8900
127.0000	22.0300
138.0000	22.1600
156.0000	28.4000
128.0000	23.8100
136.0000	27.6800
128.0000	29.3800
114.0000	20.3100
146.0000	32.7300

On the entire data set (which consists of 462 observations), the sample correlation is 0.23. For the subsampled data set shown above, the p-values computed using the normal approximation, the permutation test, and the studentized permutation test are 0.0212, 0.0907, and 0.0088 respectively. If we were to do hypothesis test at .05 level, the permutation test and the studentized permutation test would have different decisions, with the studentized permutation test correctly rejecting. Of 1000 randomly chosen subsets of size 30, in 93 cases the unstudentized test rejects, but the studentized test does not, and 35 cases the studentized test rejects but the unstudentized test does not, and in 186 cases, both tests reject.

7 Conclusion

The permutation test using the sample correlation as the test statistic is exact when testing that two variables are independent. However, this test fails to be exact, or even asymptotically valid, for testing the null hypothesis that the variables are uncorrelated. When used to test the correlation, the permutation test may have a rejection probability which is far from the nominal level or have an excessively large Type 3 error rate.

This problem can be resolved by studentizing the sample correlation so that the test statistic is asymptotically pivotal (or distribution free). The permutation test asymptotically behaves as if the variables are independent. When using a studentized sample correlation statistic, the sampling distribution has the same asymptotic behavior regardless of whether the variables are independent or only uncorrelated. The permutation distribution is exact under independence, and has the same asymptotic behavior as the sampling distribution. But under the weaker assumption of the variables being uncorrelated, the permutation distribution has the same asymptotic behavior as the permutation distribution under dependence, and thus should approximate the sampling distribution. A permutation test based on a studentized test statistic retains the exactness property under independence, but also has the desired asymptotic level.

Simulations results confirm that the unstudentized permutation test can have rejection probability that is far from the nominal level in large samples. Using a studentized statistic not only leads to a test with the correct rejection probability, but that simulations suggest has rejection probability much closer to the nominal level.

The techniques used to find the limiting behavior of the permutation test can also be used to describe permutation tests for regression coefficients. When testing that several of many regression coefficients are zero, using a Wald type statistic (which is inherently studentized) leads to a permutation test that is exact when the predictors of interest are independent of the predictors with non-zero coefficients and the error, and that is asymptotically correct when there is dependence. Simulation results show that the level of the permutation test is closer to the nominal level than using the usual chi-squared statistics.

8 Appendix

The next two lemmas are useful for proving the theorems stated in Section 2.

Lemma 8.1 *Let $Z_1, Z_2, \ldots$ be an iid sequence of random variables. If $E\left|Z_1\right|^2 < \infty$, then $\lim_{n\to\infty} \frac{\max_{1\le i\le n}|Z_i|}{\sqrt{n}} = 0$ almost surely.*

Proof. By assumption,

$$E\left(\left|Z_1\right|/\epsilon\right)^2 = \int_0^\infty P\left(\left|Z_1\right|^2 > \epsilon^2 z\right) dz < \infty$$

for any $\epsilon > 0$. Thus,

$$\sum_{n=1}^{\infty} P\left(\left|Z_n\right| \ge \epsilon\sqrt{n}\right) < E\left(Z^2/\epsilon^2\right) < \infty.$$

and it follows from the Borel-Cantelli lemma that $\lim_{n\to\infty} Z_n/\sqrt{n} = 0$ almost surely. Now let $\Lambda = \left\{\omega : \left|\frac{Z_n(\omega)}{\sqrt{n}}\right| \to 0\right\}$. For any $\omega \in \Lambda$, the sequence $\left|\frac{Z_n(\omega)}{\sqrt{n}}\right|$ is bounded, say by K. Then given any $\epsilon > 0$, there exists a constant N_1 such that $\left|\frac{Z_n(\omega)}{\sqrt{n}}\right| \leq \epsilon/2$ for all $n > N_1$. We can write

$$\left|\frac{\max_{1\leq i\leq n} Z_i(\omega)}{\sqrt{n}}\right| \leq \left|\frac{\max_{1\leq i\leq N_1} Z_i(\omega)}{\sqrt{n}}\right| + \left|\frac{\max_{N_1+1\leq i\leq n} Z_i(\omega)}{\sqrt{n}}\right|.$$

The second term on the right hand side is bounded by $\epsilon/2$. The first term is bounded by $\max_{1\leq i\leq N_1} K/\sqrt{n}$ which can be made smaller than $\epsilon/2$ for all n greater than a constant N_2. So,

$$\left|\frac{\max_{1\leq i\leq n} Z_i(\omega)}{\sqrt{n}}\right| \leq \epsilon$$

for all $n > \max\{N_1, N_2\}$. ■

The second lemma is due to McLeish (1974, Lemma 2.11) which is stated next.

Lemma 8.2 (McLeish) *If $\{X_n\}$ and X are positive, integrable random variables such that $E(X_n) \to E(X)$ and $P(X - X_n > \epsilon) \to 0$ for all $\epsilon > 0$, then X_n converges in L_1 to X.*

The next theorem is a condition due to Hoeffding that characterizes convergence of randomization distributions. (sufficiency given in Lehmann and Romano (2005), and necessity given in Chung and Romano (2013)).

Theorem 8.3 (Hoeffding's condition) *Suppose that X^n has distribution P_n in $\mathcal{X}_n$, and that $\mathbf{G_n}$ is a finite group of transformations from $\mathcal{X}_n$ to $\mathcal{X}_n$. Let $\hat{R}_n(\cdot)$ denote the permutation distribution of a statistic T_n. For any G_n and G'_n chosen independently and uniformly from $\mathbf{G_n}$,*

$$\big(T_n(G_nX^n), T_n(G'_nX^n)\big) \xrightarrow{\mathcal{L}} (T, T')$$

under P_n where T and T' are independent with common c.d.f. $R(\cdot)$ if and only if

$$\hat{R}_n(t) \xrightarrow{P} R(t)$$

for any continuity point t of $R(\cdot)$.

The following combinatorial central limit theorem due to Hoeffding (1951, Theorem 4) is helpful for finding the asymptotic behavior of the randomization distribution of sample correlations.

Theorem 8.4 (Hoeffding's Combinatorial Central Limit Theorem) *Let $a = (a_1, ..., a_n)$ and $b = (b_1, ..., b_n)$ be two vectors whose components a_i and b_i are real numbers, possibly depending on n. Let π be a uniformly chosen permutation of $\{1, ..., n\}$. Define*

$$F_n(y, a, b) = P\left(\frac{\sqrt{n-1}\sum_{i=1}^n (a_i - \bar{a})b_{\pi(i)}}{\left(\sum_{i=1}^n (a_i - \bar{a})^2 \sum_{i=1}^n (b_i - \bar{b})^2\right)^{\frac{1}{2}}} \leq y\right)$$

where $\bar{a} = \frac{1}{n}\sum_{i=1}^n a_i$ and $\bar{b} = \frac{1}{n}\sum_{i=1}^n b_i$. A sufficient condition for

$$F_n(y, a, b) \to \Phi(y)$$

as $n \to \infty$ is that

$$n^{\frac{1}{2}p-1} \frac{\sum_{i=1}^n (a_i - \bar{a})^p}{\left(\sum_{i=1}^n (a_i - \bar{a})^2\right)^{\frac{1}{2}p}} \frac{\sum_{i=1}^n (b_i - \bar{b})^p}{\left(\sum_{i=1}^n (b_i - \bar{b})^2\right)^{\frac{1}{2}p}} \to 0, \quad p = 3, 4, ...$$

This condition is satisfied if

$$n \frac{\max_{1 \le i \le n} (a_i - \bar{a})^2}{\sum_{i=1}^n (a_i - \bar{a})^2} \frac{\max_{1 \le i \le n} (b_i - \bar{b})^2}{\sum_{i=1}^n (b_i - \bar{b})^2} \to 0.$$

The last result needed is a multivariate extension of Hoeffding's combinatorial central limit theorem which is used to prove the results in Section 3 on regression coefficients.

Theorem 8.5 *Suppose that $(a_1^{(r)}, ..., a_n^{(r)})$ and $(b_1^{(r)}, ..., b_n^{(r)})$, $r = 1, ..., k$ are sequences of constants satisfying $\sum_i a_i^{(r)} = \sum_i b_i^{(r)} = 0$,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \left(a^{(r)}\right)^2 < \infty$$

and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \left(b^{(r)}\right)^2 < \infty$$

for all $r = 1, ..., k$. If π is a uniformly chosen permutation of $\{1, ..., n\}$, then

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^n a_i^{(1)} b_{\pi(i)}^{(1)}, ..., \sum_{i=1}^n a_i^{(k)} b_{\pi(i)}^{(k)} \right) \xrightarrow{\mathcal{L}} N(0, \Sigma)$$

where $\Sigma_{rs} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n a_i^{(r)} b_i^{(s)} \frac{1}{n} \sum_{j=1}^\infty a_j^{(s)} b_j^{(r)}$, provided these limits are finite and

$$\lim_{n \to \infty} \frac{\max_{i \le n} \left(a_i^{(r)}\right)^2}{\sqrt{n}} \times \frac{\max_{i \le n} \left(b_i^{(r)}\right)^2}{\sqrt{n}} = 0$$

for $r = 1, ..., k$.

Proof. Define $S_n = \sum_{i=1}^n c_n(i, \pi(i))$ where $c_n(i, j)$ are real numbers. Hoeffding's combinatorial central limit theorem states that if

$$\frac{\max_{g,h} d^2(g, h)}{\frac{1}{n} \sum_{i,j} d^2(i, j)} \to 0,$$

then

$$\frac{S_n - ES_n}{\sqrt{\text{var} S_n}} \xrightarrow{\mathcal{L}} N(0, 1)$$

where $ES_n = \frac{1}{n} \sum_{i,j} c(i, j)$, and $\text{var} S_n = \frac{1}{n-1} \sum_{i,j} d^2(i, j)$ with

$$d(i, j) = c(i, j) - \frac{1}{n} \sum_g c(g, j) - \frac{1}{n} \sum_h c(i, h) + \frac{1}{n} \sum_{g,h} c(g, h).$$

To prove the theorem using the Cramér-Wold device, it is enough to show that for any constants $\phi_1, ..., \phi_k$, $\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{r=1}^k \phi_r a_i^{(r)} b_{\pi(i)}^{(r)}$ is asymptotically normal with mean zero and variance $\phi' \Sigma \phi$. To show this, we will use Hoeffding's combinatorial central limit theorem with

$$c(i,j) = \frac{1}{\sqrt{n}} \sum_{r=1}^k \phi_r a_i^{(r)} b_j^{(r)}$$

where the ϕ_r are arbitrary constants. Then

$$d(i,j) = \frac{1}{\sqrt{n}} \sum_{r=1}^k \phi_r (a_i^{(r)} - \bar{a}^{(r)})(b_j^{(r)} - \bar{b}^{(r)}) = \frac{1}{\sqrt{n}} \sum_{r=1}^k \phi_r a_i^{(r)} b_j^{(r)}$$

and

$$d^2(i,j) = \frac{1}{n} \sum_{r=1}^k \left(\phi_r a_i^{(r)} b_j^{(r)}\right)^2 + \frac{1}{n} \sum_{r \neq s} \phi_r \phi_s a_i^{(r)} b_j^{(s)}.$$

Thus,

$$\text{var} S_n = \frac{1}{n-1} \left[\sum_r \phi_r \frac{1}{n} \sum_{i=1}^n \left(a_i^{(r)}\right)^2 \sum_{j=1}^n \left(b_j^{(s)}\right)^2 + \sum_{r \neq s} \phi_r \phi_s \frac{1}{n} \sum_{i=1}^n a_i^{(r)} b_i^{(r)} \sum_{j=1}^n a_j^{(s)} b_j^{(s)} \right]$$

which converges to $\phi' \Sigma \phi$, and the result of the theorem follows if we can check the conditions of Hoeffding's combinatorial central limit theorem. Indeed,

$$\frac{\max_{g,h} d^2(g,h)}{\frac{1}{n} \sum_{i,j} d^2(i,j)} = \frac{\frac{1}{n} \max_{g,h} \sum_{r=1}^k \left(\phi_r a_g^{(r)} b_h^{(r)}\right)^2 + \sum_{r \neq s} \phi_r \phi_s a_g^{(r)} b_h^{(s)}}{\frac{1}{n} \left[\sum_r \phi_r \sum_{i=1}^n \left(a_i^{(r)}\right)^2 \sum_{j=1}^n \left(b_j^{(s)}\right)^2 + \sum_{r \neq s} \phi_r \phi_s \sum_{i=1}^n a_i^{(r)} b_i^{(r)} \sum_{j=1}^n a_j^{(s)} b_j^{(s)} \right]}$$

$$\leq \sum_{r=1}^k \frac{\left(\sum_{q,s} |\phi_q \phi_s|\right) \max_g \left(a_g^{(r)}\right)^2 \frac{1}{\sqrt{n}} \max_h \left(b_h^{(r)}\right)^2 \frac{1}{\sqrt{n}}}{\frac{1}{n} \left[\sum_r \phi_r \sum_{i=1}^n \left(a_i^{(r)}\right)^2 \sum_{j=1}^n \left(b_j^{(s)}\right)^2 + \sum_{r \neq s} \phi_r \phi_s \sum_{i=1}^n a_i^{(r)} b_i^{(r)} \sum_{j=1}^n a_j^{(s)} b_j^{(s)} \right]}$$

which converges to zero by assumption since the denominator converges to $\phi' \Sigma \phi < \infty$. ∎

Justification of Remark 3.1. If both the mean of X_i and Y_i are non-zero, then the ordinary least squares coefficient tends to infinity in probability. We will assume that the mean of the Y_i is zero, and the mean of the X_i is non-zero. In this case, the sampling distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i Y_i$ is asymptotically normal with mean zero, and covariance $E(X_i^2 Y_i^2)$. Using the necessity part of Hoeffding's condition (Theorem 8.3), if the permutation distribution approximated the sampling distribution, then for any randomly chosen permutations π and π',

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i Y_{\pi(i)} \right) \xrightarrow{d} (T, T')$$

where T and T' are independent with a common distribution. However, writing

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i Y_{\pi(i)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_{\pi(i)} - \bar{Y}_n) - n \bar{X}_n \bar{Y}_n$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i Y_{\pi'(i)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_{\pi'(i)} - \bar{Y}_n) - n\bar{X}_n\bar{Y}_n,$$

and noting that $n\bar{X}_n\bar{Y}_n$ is asymptotically normal, we see that these statistics cannot converge jointly to independent random variables. ■

Proof of Theorem 2.1. To prove this theorem, we will use Hoeffding's combinatorial central limit theorem (Theorem 8.4), replacing a_i by X_i and b_i by Y_i where (X_i, Y_i), $i = 1, ..., n$, are independent and identically distributed. If it can be shown that

$$n\frac{\max_{1\le i\le n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\frac{\max_{1\le i\le n}(Y_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \xrightarrow{a.s.} 0,$$

then the convergence in Hoeffding's central limit theorem holds almost surely and our first theorem is proved. Indeed, this holds using Lemma 8.1 since we can write

$$n\frac{\max_{1\le i\le n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\frac{\max_{1\le i\le n}(Y_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = \frac{1}{\sqrt{n}}\frac{\max_{1\le i\le n}(X_i - \bar{X})^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}\frac{1}{\sqrt{n}}\frac{\max_{1\le i\le n}(Y_i - \bar{Y})^2}{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}.$$

■

Proof of Theorem 2.2. Clearly $\frac{E\left[(X-\mu_X)^2(Y-\mu_Y)^2\right]}{\sigma_X^2\sigma_Y^2} \ge 0$. We will now show that this quantity can be made arbitrarily large. Let W and Z be independent and identically distributed random variables with mean zero. Set $X = W - Z$, and $Y = W + Z$ so that $\mathrm{cov}(X, Y) = \mathrm{var}(W) - \mathrm{var}(Y) = 0$. Therefore, X and Y are uncorrelated, but not necessarily independent. Also, $E(X) = E(Y) = 0$ and $\mathrm{var}(X) = \mathrm{var}(Y) = 2\sigma_W^2$.

Because

$$\begin{aligned}
E(X^2Y^2) &= E\left((W - Z)^2(W + Z)^2\right) \\
&= E\left((W^2 - 2WZ + Z^2)(W^2 + 2WZ + Z^2)\right) \\
&= E\left(W^4 + 2W^3Z + W^2Z^2 - 2W^3Z - 4W^2Z^2 - 2WZ^3 + W^2Z^2 + 2WZ^3 + Z^4\right) \\
&= E\left(W^4\right) + E\left(Z^4\right) - 2E\left(W^2Z^2\right) \\
&= E\left(W^4\right) + E\left(Z^4\right) - 2\sigma_W^4,
\end{aligned}$$

we have

$$\frac{E\left((X-\mu_X)^2(Y-\mu_Y)^2\right)}{\sigma_X^2\sigma_Y^2} = \frac{E\left(W^4\right) + E\left(Z^4\right) - 2\sigma_W^4}{4\sigma_W^4} = \frac{E\left(W^4\right) + E\left(Z^4\right)}{4\sigma_W^4} - \frac{1}{2}$$

which we claim can be made arbitrarily large. For instance, if W and Z follow a t distribution with $\nu > 4$ degrees of freedom, $E(W^2) = \frac{\nu}{\nu-2}$ and $E(W^4) = \nu^2\left(\frac{1}{\nu-2}\right)\left(\frac{3}{\nu-4}\right)$. Consequently,

$$\frac{E(W^4)}{4\sigma_W^2} = \frac{3(\nu-2)}{4(\nu-4)}$$

which can be made arbitrarily large by choosing ν sufficiently close to 4. Moreover, this quantity is infinite if $\nu = 4$. Note, however, that if X and Y have finite fourth moments, that this quantity is always finite.

To show that this quantity can be made exactly equal to 0, choose $X = W - Z$, and $Y = W + Z$ where W and Z are independent with $P(W = \pm 1/2) = P(Z = \pm 1/2) = 1/2$. It is easily seen that always either $X = 0$ or $Y = 0$ and consequently $E\left[(X-\mu_X)^2(Y-\mu_Y)^2\right] = 0$. ■

Proof of Theorem 2.3. We will assume without loss of generality that the X_i and Y_i have mean zero since the sample correlation only depends on the data through $X_i - \bar{X} = (X_i - \mu_X) - (\bar{X} - \mu_X)$ and $Y_i - \bar{Y} = (Y_i - \mu_Y) - (\bar{Y} - \mu_Y)$. To prove the first part of the theorem we first show that show $\hat{\mu}_{2,2} \xrightarrow{P} \mu_{2,0}\mu_{0,2}$ under the assumption of finite fourth moments. The result of the theorem will then follow from Slutsky's theorem as well as Hoeffding's condition.

By Hoeffding's condition (Theorem 8.3), and Theorem 2.1, if π and π' are independent, uniformly chosen permutations of $\{1, ..., n\}$, then

$$\left(\sqrt{n}\hat{\rho}_n(X^n, Y^n_\pi), \sqrt{n}\hat{\rho}_n(X^n, Y^n_{\pi'})\right) \xrightarrow{\mathcal{L}} N(0, I_2).$$

If we can show that for any uniformly chosen permutation π, $\hat{\tau}_n(X^n, Y^n_\pi) \xrightarrow{P} 1$, then Slutsky's Theorem implies

$$\left(\sqrt{n}\hat{\rho}_n(X^n, Y^n_\pi)/\hat{\tau}_n(X^n, Y^n_\pi), \sqrt{n}\hat{\rho}_n(X^n, Y^n_{\pi'})/\hat{\tau}_n(X^n, Y^n_{\pi'})\right) \xrightarrow{\mathcal{L}} N(0, I_2).$$

and by Hoeffding's condition, it follows that the permutation distribution of $\sqrt{n}\hat{\rho}_n/\hat{\tau}_n$ is asymptotically standard normal in probability. To show $\hat{\tau}_n(X^n, Y^n_\pi) \xrightarrow{P} 1$, it is enough to show that $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 (Y_{\pi(i)} - \bar{Y})^2 \xrightarrow{P} \mu_{2,0}\mu_{0,2}$ since $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} \mu_{2,0}$ and $\frac{1}{n}\sum_{i=1}^n Y_{\pi(i)}^2 \xrightarrow{P} \mu_{0,2}$. Because $\sqrt{n}\bar{X}\bar{Y} \xrightarrow{a.s.} 0$, we need only show $\frac{1}{n}\sum_{i=1}^n (X_i Y_{\pi(i)})^2 \xrightarrow{P} \mu_{2,0}\mu_{0,2}$. Define

$$\mu_n = E\left(\frac{1}{n}\sum_{i=1}^n (X_i Y_{\pi(i)})^2\right).$$

Then by conditioning on the number of fixed points,

$$\mu_n = \mu_{2,0}\mu_{0,2} + \frac{1}{n}(E(X_1 Y_1) - \mu_{20}\mu_{02})E(\#\{i : \pi(i) = i\}).$$

Because

$$\begin{aligned} E(\#\{i : \pi(i) = i\}) &= E\sum_{k=1}^n I_{\{\pi(k)=k\}} \\ &= \sum_{k=1}^n P(\pi(k) = k) \\ &= \sum_{k=1}^n \frac{1}{k} \\ &= 1, \end{aligned}$$

$\frac{1}{n}E(\#\{i : \pi(i) = i\}) \to 0$ and it follows that $\mu_n \to \mu_{2,0}\mu_{0,2}$. Also define

$$\begin{aligned} \sigma_n^2 &= \text{var}\left(\frac{1}{n}\sum_{i=1}^n (X_i Y_{\pi(i)})^2\right) \\ &\frac{1}{n^2}\sum_{i=1}^n \text{var}\left((X_i Y_{\pi(i)})^2\right) + \frac{1}{n^2}\sum_{i \neq j} \text{cov}\left((X_i Y_{\pi(i)})^2, (X_j Y_{\pi(j)})^2\right). \end{aligned}$$

The first term trivially converges to zero, and the second term converges to zero since $(X_i Y_{\pi(i)})^2$ is independent of all but at most two of the other terms $(X_j Y_{\pi(j)})^2$. Thus $\sigma_n^2 \to 0$.

By Chebychev's inequality,

$$\begin{aligned}
&\lim_{n\to\infty} P\left(\frac{1}{n}\sum_{i=1}^{n}(X_i Y_{\pi(i)})^2 < \mu_{2,0}\mu_{0,2} - \epsilon\right) \\
&= \lim_{n\to\infty} P\left(\frac{1}{n}\sum_{i=1}^{n}(X_i Y_{\pi(i)})^2 - \mu_n < \mu_{2,0}\mu_{0,2} - \mu_n - \epsilon\right) \\
&\leq \lim_{n\to\infty} \frac{\sigma_n^2}{(\mu_{2,0}\mu_{0,2} - \mu_n - \epsilon)^2} \\
&= 0.
\end{aligned}$$

and we get the desired convergence by appealing to Lemma 8.2.

Under the stronger assumption of finite eighth moments, the convergence can be strengthened to almost sure convergence by showing

$$P\left(\frac{1}{n}\sum_{i=1}^{n}(X_i Y_{\pi(i)})^2 > \epsilon\right) \leq \frac{1}{\epsilon^4} E\left(\frac{1}{n^4}\left(\sum_{i=1}^{n}(X_i Y_{\pi(i)})^2\right)^4\right) = O(1/n^2).$$

The desired convergence would then follow from the Borel-Cantelli lemma. Indeed,

$$\begin{aligned}
E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i Y_{\pi(i)}\right)^4\right] =& \frac{1}{n^4} E\left[\left(\sum_{i=1}^{n} X_i^2 Y_{\pi(i)}^2 + \sum_{i\neq j} X_i Y_{\pi(i)} X_j Y_{\pi(j)}\right)^2\right] \\
=& \frac{1}{n^4} E\left[\left(\sum_{i=1}^{n} X_i^2 Y_{\pi(i)}^2\right)^2 + 2\sum_{i\neq j} X_i^2 Y_{\pi(i)}^2 X_i Y_{\pi(i)} X_j Y_{\pi(j)}\right. \\
&+ 3\sum_{i\neq j\neq k} X_i^2 Y_{\pi(i)}^2 X_j Y_{\pi(j)} X_k Y_{\pi(k)} + \sum_{i\neq j} X_i^2 Y_{\pi(i)}^2 X_j^2 Y_{\pi(j)}^2 \\
&\left. + \sum_{i\neq j\neq k} X_i Y_{\pi(i)} X_j Y_{\pi(j)} X_k Y_{\pi(k)} X_l Y_{\pi(l)}\right] \\
=& \frac{1}{n^4} E\left(3\sum_{i\neq j\neq k} X_i^2 Y_{\pi(i)}^2 X_j Y_{\pi(j)} X_k Y_{\pi(k)}\right. \\
&\left. + \sum_{i\neq j\neq k} X_i Y_{\pi(i)} X_j Y_{\pi(j)} X_k Y_{\pi(k)} X_l Y_{\pi(l)}\right) + O(1/n^2)
\end{aligned}$$

Now $E(X_i^2 Y_{\pi(i)}^2 X_j Y_{\pi(j)} X_k Y_{\pi(k)}|\pi)$ is zero, except perhaps when $\{j,k\} \subset \{\pi_i, \pi_j, \pi_k\}$ which happens with probability less than $1/n$. Consequently, the first term in the expectation above is $O(1/n^2)$. Similarly, $E(X_i Y_{\pi(i)} X_j Y_{\pi(j)} X_k Y_{\pi(k)} X_l Y_{\pi(l)}|\pi)$ is zero except when $\{i,j,k,l\} \subset \{\pi_i, \pi_j, \pi_k, \pi_l\}$ which happens with probability less than $1/n^2$. So the second term is also $O(1/n^2)$.

Applying this result to $\frac{1}{n}\sum_{i=1}^{n}(X_i^2-\sigma_X^2)(Y_{\pi(i)}^2-\sigma_Y^2)$, we have that $\frac{1}{n}\sum_{i=1}^{n} X_i^2 Y_{\pi(i)}^2 \xrightarrow{a.s.} \sigma_X^2\sigma_Y^2$ if the X_i's and Y_i's have finite eight moments. ■

Proof of Theorem 2.6. It was shown above that the permutation distribution of $\sqrt{n}\hat{\rho}_n$ also converges weakly to $N(0,1)$. By Hoeffding's condition (Theorem 8.3), this implies that if π and π' are two independent and uniformly chosen permutations of $\{1, ..., n\}$, then

$$\left(\sqrt{n}(\hat{\rho}_n(X^n, Y_\pi^n)), \sqrt{n}(\hat{\rho}_n(X^n, Y_{\pi'}^n))\right) \xrightarrow{\mathcal{L}} N(0, I_2).$$

Using Slutsky's theorem gives

$$\left(\sqrt{n}(\tanh^{-1}(\hat{\rho}_n(X^n, Y_\pi^n))/\hat{\tau}_n(X^n, Y_\pi^n)), \sqrt{n}(\tanh^{-1}(\hat{\rho}_n(X^n, Y_{\pi'}^n))/\hat{\tau}_n(X^n, Y_\pi^n))\right) \xrightarrow{\mathcal{L}} N(0, I_2).$$

Hence, by Hoeffding's condition, the permutation distribution of $\sqrt{n}(\tanh^{-1}(\hat{\rho}_n))/\hat{\tau}_n$ converges weakly to $N(0,1)$ in probability. Convergence to $N(0,1)$ almost surely follows exactly as in the proof of Theorem 2.3. ■

Proof of Theorem 3.2. By the multivariate extension of Hoeffdings combinatorial central limit theorem (Theorem 8.5), if π is a uniformly chosen permutation of $\{1, ..., n\}$, then

$$\frac{1}{\sqrt{n}}X^\top Y_\pi = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{i1}Y_{\pi(i)}, ..., \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{ip}Y_{\pi(i)}\right) \xrightarrow{\mathcal{L}} N(0, \Omega^*), \quad \text{a.s.}$$

where $\Omega^* = E(\epsilon_i^2)E(X_iX_i^\top)$. But also, $\hat{\Sigma}_{XX}(X, Y_\pi) \xrightarrow{P} \Sigma_{XX}$ and $\hat{\Omega}(X, Y_\pi) \xrightarrow{P} \Omega^*$. Because

$$E\left(\frac{1}{n}\sum_{i=1}^{n} Y_{\pi(i)}^2 X_{ij}X_{ik}\right) = \Omega_{jk}^* + \frac{1}{n}E(\#\{i : \pi(i) = i\})\left(\Omega_{jk}^* - \Omega_{jk}^*\right) \to \Omega_{jk}^*$$

and using the law of total variance,

$$\begin{aligned}\text{var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_{\pi(i)}^2 X_{ij}X_{ik}\right) =& \text{var}\left(\frac{1}{n}\#\{i : \pi(i) = i\}\left(\Omega_{jk}^* - \Omega_{jk}^*\right)\right) \\ &+ E\left(\frac{\left(\text{var}(Y_1^2 X_{1j}X_{1k}) - \text{var}(Y_1^2 X_{2j}X_{2k})\right)^2}{n^2}\#\{i : \pi(i) = i\}\right) \\ &\to 0\end{aligned}$$

second convergence is a consequence of Chebychev's inequality. Using Slutsky's theorem for permutation distributions (see Chung and Romano (2014)), the permutation distribution of $\Omega^{-1/2}\hat{\Sigma}_{XX}\sqrt{n}\hat{\beta}_n$ is asymptotically multivariate normal with mean zero and the identity covariance matrix (in probability), and we have that the permutation distribution of $S_n(X, Y)$ is asymptotically χ_p^2 in probability. ■

Proof of Theorem 3.3. Appealing to Hoeffding's condition (Theorem 8.3), it is enough to show that for any independent and uniformly chosen transformations g_δ and g'_δ,

$$(S_n(X, g_\delta(Y)), S_n(X, g_{\delta'}(Y))) \xrightarrow{\mathcal{L}} (T, T')$$

where T and T' are independent χ_p^2 random variables. By the multivariate central limit theorem, if $\delta_1, ..., \delta_n, \delta_1', ..., \delta_n'$ are independent random variables taking values +1 or -1, each with probability 1/2, then

$$\frac{1}{\sqrt{n}}(X^\top g_\delta(Y), X^\top g_{\delta'}(Y)) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{i1}Y_i\delta_i, ..., \frac{1}{\sqrt{n}}\sum_{i=1}^n X_{ip}Y_i\delta_i, \frac{1}{\sqrt{n}}\sum_{i=1}^n X_{i1}Y_i\delta_i', ..., \frac{1}{\sqrt{n}}\sum_{i=1}^n X_{ip}Y_i\delta_i'\right)$$
$$\xrightarrow{\mathcal{L}} N\left(0, I_2 \otimes \Omega\right)$$

Further, $\hat{\Sigma}_{XX} \xrightarrow{p} \Sigma_{XX}$ because this estimator only depends on the data through X, and

$$\hat{\Omega}(X, g_\delta(Y)) = \hat{\Omega}(X, g_{w'}(Y)) = \frac{1}{n}\sum_i Y_i^2 X_i X_i^\top \xrightarrow{P} \Omega.$$

By the multivariate delta method, this yields,

$$(S_n(X, g_\delta(Y)), S_n(X, g_{\delta'}(Y))) \xrightarrow{\mathcal{L}} (T, T')$$

as required. ■

Proof of Theorem 3.4. The Wald statistic W_n is a quadratic form of

$$\begin{pmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{pmatrix} = \begin{bmatrix} X^\top X & X^\top Z \\ X^\top Z & Z^\top Z \end{bmatrix} \begin{pmatrix} X^\top \\ Z^\top \end{pmatrix} \epsilon.$$

Using Hoeffding's combinatorial central limit theorem, $\frac{1}{\sqrt{n}} X_\pi^\top \epsilon$ is asymptotically $N(0, E(\epsilon_i^2)E(X_i X_i^\top))$ almost surely. Moreover,

$$\frac{1}{n}\begin{bmatrix} X_\pi^\top X_\pi & X_\pi^\top Z \\ X_\pi^\top Z & Z^\top Z \end{bmatrix} \xrightarrow{P} \begin{bmatrix} E(X_i X_i^\top) & 0 \\ 0 & E(Z_i Z_i^\top) \end{bmatrix}$$

almost surely and we have that

$$\sqrt{n}R\hat{\beta}(X_\pi, Z, Y) \xrightarrow{\mathcal{L}} \begin{pmatrix} N \\ 0 \end{pmatrix}$$

almost surely where $N \sim N(0, E(\epsilon_i^2)E(X_i X_i^\top)^{-1})$. Also,

$$n\hat{\Sigma}_{\tilde{X}\tilde{X}}(X_\pi, Z, Y)^{-1}\hat{\Omega}(X_\pi, Z, Y)\hat{\Sigma}_{\tilde{X}\tilde{X}}(X_\pi, Z, Y)^{-1}$$
$$\xrightarrow{P} \begin{bmatrix} E(\epsilon_i^2)E(X_i X_i^\top)^{-1} & 0 \\ 0 & E(Z_i Z_i^\top)^{-1}E(\epsilon_i^2 Z_i Z_i^\top)E(Z_i Z_i^\top)^{-1} \end{bmatrix}$$

so it follows from the continuous mapping theorem that the permutation distribution has the required chi-squared distribution asymptotically. ■

Proof of Theorem 3.5. We will first look at the randomization distribution of $\hat{\beta}_n$ under the Freedman-Lane procedure. For any permutation π, we can write

$$\begin{aligned}
\sqrt{n}R\begin{pmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{pmatrix} &= \sqrt{n}R\left[(X, Z)^\top(X, Z)\right]^{-1}(X, Z)^\top Y^* \\
&= \sqrt{n}R\left[(X, Z)^\top(X, Z)\right]^{-1}(X, Z)^\top\left(Z\hat{\gamma}_n + \hat{\epsilon}_\pi\right) \\
&= \sqrt{n}R\left[(X, Z)^\top(X, Z)\right]^{-1}\begin{pmatrix} X^\top Z \\ Z^\top Z \end{pmatrix}\hat{\gamma}_n + \sqrt{n}R\left[(X, Z)^\top(X, Z)\right]^{-1}(X, Z)^\top\hat{\epsilon}_\pi
\end{aligned}$$

The first term is zero, so it is enough to find the limiting behavior of

$$\sqrt{n}\begin{pmatrix} X^\top \\ Z^\top \end{pmatrix}\epsilon_\pi.$$

Appealing to the multivariate extension of Hoeffding's combinatorial central limit theorem (Theorem 8.5), this is asymptotically $N(0, E(\epsilon_i^2)\Sigma_{X,Z})$, in probability. It follows that the permutation distribution of $\sqrt{n}R\hat{\beta}$ is asymptotically $N(0, E(\epsilon_i^2)R\Sigma_{XZ}^{-1}R^\top)$, in probability. Now, under permutations, $\hat{\Omega} = \frac{1}{n}\sum_i \hat{\epsilon}_i^2 \tilde{X}_i\tilde{X}_i^\top$ converges in probability to $E(\epsilon_i^2)\Sigma_{XZ}$, and the result follows from Slutsky's theorem. ∎

Proof of Theorem 4.1. This follows immediately from Hoeffding's combinatorial central limit theorem (Theorem 8.4), and Slutsky's theorem for permutation distributions. ∎

References

[1] Anderson, M., Robinson, J. (2002). Permutation Tests for Linear Models. Australian & New Zealand Journal of Statistics, 43: 75-88.

[2] Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. Annals of Statistics, **41**, 458-507.

[3] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability.* Springer, New York.

[4] Freedman, D., and Lane, D. (1983). A nonstochastic interpretation of reported significance levels. Journal of Business and Economic Statistics, Vol. 1, No. 4, 292-298.

[5] Good, P. I., (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses.* 3rd ed. Springer Series in Statistics, New York.

[6] Hall, P. and Heyde C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.

[7] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning.* 2nd ed. Springer, New York.

[8] Hoeffding, W. (1951). A combinatorial central limit theorem. The Annals of Mathematical Statistics, Vol. 22, pp. 558-566.

[9] Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. The Annals of Mathematical Statistics, 23, 169-192.

[10] Janssen, A. (1997). Studentized permutation tests for non i.i.d. hypotheses and the generalized BehrensFisher problem. Statist. Probab. Lett. **36** 921.

[11] Janssen, A. (1999). Testing nonparametric statistical functionals with applications to rank tests. Journal of Statistical Planning and Inference **81** 7193.

[12] Lehmann, E.L. and Romano, J. P. (2005). *Testing Statistical Hypotheses.* 3rd ed. Springer, New York.

[13] McLeish, D.L. (1974). Dependent central limit theorems and invariance principles. Annals of Probability **2**, 620-628.

[14] Omelka, M. and Pauly, M. (2012). Testing equality of correlation coefficients in a potentially unbalanced two-sample problem via permutation methods. J. Statist. Plann. Inference **142** 13961406.

[15] Romano, J.P. (1990). On the behavior of randomization tests without a group symmetry assumption. Journal of the American Statistical Association, Vol. **85**, 686-692.

[16] Pauly, M. (2011). Discussion about the quality of F-ratio resampling tests for comparing variances. TEST **20** 163179.

[17] Sprent, P. and Smeeton, N. C. (2007). *Applied Nonparametric Statistical Methods*. Chapman & Hall, Florida.

[18] White, H. (1980). A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity, Econometrica, 48, 817-838.

[19] Winkler A.M., Ridgway G.R., Webster M.A., Smith S.M., and Nichols T.E. (2014). Permutation inference for the general linear model. Neuroimage. 92(100):381-397.

[20] Wu, C.F.J. (1986). Jacknife, bootstrap and other resampling methods in regression analysis. Ann. Statist., 14, 1261-1350.