Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Robust Permutation Tests For Correlation And Regression Coefficients

Cyrus J. DiCiccio & Joseph P. Romano

To cite this article: Cyrus J. DiCiccio & Joseph P. Romano (2017) Robust Permutation Tests For Correlation And Regression Coefficients, Journal of the American Statistical Association, 112:519, 1211-1220, DOI: 10.1080/01621459.2016.1202117

To link to this article: https://doi.org/10.1080/01621459.2016.1202117

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2017, VOL. 112, NO. 519, 1211–1220
http://dx.doi.org/10.1080/01621459.2016.1202117

Taylor & Francis
Taylor & Francis Group

Robust Permutation Tests For Correlation And Regression Coefficients

Cyrus J. DiCiccio[a] and Joseph P. Romano[b]

[a]Department of Statistics, Stanford University, Stanford, CA; [b]Departments of Statistics and Economics, Stanford University, Stanford, CA

ABSTRACT
Given a sample from a bivariate distribution, consider the problem of testing independence. A permutation test based on 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 a 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. Simulations and empirical applications are included. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received August 2015
Revised June 2016

KEYWORDS
Least squares; Partial correlation; Randomization tests; Studentization; Testing independence

1. Introduction

Assume $(X_1, Y_1), \dots, (X_n, Y_n)$ are iid according to a joint distribution P with (nondegenerate) marginal distributions P_X and P_Y. Define $X^n = (X_1, \dots, X_n)$ and $Y^n = (Y_1, \dots, Y_n)$. Let $\rho = \rho(P) = \mathrm{corr}(X_1, Y_1)$ 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 $\{1, \dots, n\}$. The permutation distribution of any given test statistic $T_n(X^n, Y^n)$ is defined as

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

where we write Y^n_π for $(Y_{\pi(1)}, \dots, Y_{\pi(n)})$. 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 kth largest ordered value of $\{T_n(X^n, Y^n_\pi) : \pi \in \mathbf{G}_n\}$, $m_1 = n! - \lfloor (1 - \alpha/2)n! \rfloor + 1$, $m_2 = n! - \lfloor \alpha/2n! \rfloor$, and γ_1, γ_2 are chosen so that

$$\frac{1}{n!} \sum_{\pi : T(X^n, Y^n_\pi) \le T_n^{(m_1)}} \phi(X^n, Y^n_\pi) = \alpha_1,$$

and

$$\frac{1}{n!} \sum_{\pi : T(X^n, Y^n_\pi) \ge T_n^{(m_2)}} \phi(X^n, Y^n_\pi) = \alpha_2$$

which ensures

$$\frac{1}{n!} \sum_{\pi \in \mathbf{G}_n} \phi(X^n, Y^n_\pi) = \alpha.$$

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^n_π) 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^n_π) 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 can have large error rates.

CONTACT Joseph Romano romano@stanford.edu Departments of Statistics and Economics, Stanford University, Stanford, CA 94305.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

References

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 $E(X_1)^2 < \infty$, $E(Y_1)^2 < \infty$ and $E(X_1 Y_1)^2 < \infty$, 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

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

and μ_X and μ_Y are the means of the X_i and Y_i, respectively. 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. Section S.1 of the supplement gives results analogous to Section 2 using Fisher's z-transformation to the sample correlation, as well as corresponding simulations. Finally, Section S.2 gives empirical applications comparing the studentized and unstudentized permutation tests, as well as plots comparing the resulting permutation distributions. Proofs of the results in Sections 2, 3, and 4 are given in the Appendix, and auxiliary results are stated in Section S.3 of the supplement.

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 (1999a), and more generally by Chung and Romano (2013). Applications of permutation tests can be found in Good (2005). For general results on permutation test, see Janssen and Pauls (2003). Also see Janssen (2005) for general central limit theorems for resampling procedures which are useful in showing the asymptotic validity of tests. The goal of this article 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 are 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. applications

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) discussed 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 heteroscedastic 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 heteroscedasticity-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 $(X_1, Y_1), \dots, (X_n, Y_n)$ are iid according to P such that X_1 and Y_1 are uncorrelated but not necessarily independent. Also assume that $E(X_1^2) < \infty$ and $E(Y_1^2) < \infty$. 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 where Φ is the distribution function of a standard normal random variable. However, when in addition to finite second moments, $E(X_1^2 Y_1^2) < \infty$,

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

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

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

almost surely, 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, as is seen in Lemma S.1 of the supplement), but the sampling distribution has limiting variance $\tau^2(P)$, which can be arbitrarily large or small.

Remark 2.1. If X_i and Y_i are uncorrelated (but not necessarily independent), then there exists a joint distribution P of (X_1, Y_1) such that $Cov(X_1, Y_1) = 0$, but $\tau^2(P)$ can take any value in $[0, \infty]$. Details on finding such a joint distribution are provided in the Appendix.

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. Because S_n is a pivotal statistic, the true sampling distribution of S_n under P has the same asymptotic behavior as the true sampling distribution of S_n under $P_X \times P_Y$. When the randomization hypothesis is satisfied, the permutation test using the statistic S_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 permutation distribution under P should behave asymptotically like the true sampling distribution under $P_X \times P_Y$. If the test statistic is not pivotal, this may lead to a test that is not asymptotically valid. However, it is reasonable to expect that the quantiles of the permutation distribution of a studentized statistic will approximate those of the true sampling distribution under P because the permutation distribution under P approximates the sampling distribution under $P_X \times P_Y$ which is asymptotically the same as the distribution under P.

Theorem 2.2. Assume $(X_1, Y_1), \dots, (X_n, Y_n)$ are iid 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

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

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 almost surely 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 2.2 (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, the $1-\alpha$ quantile of the permutation distribution of S_n converges to $z_{1-\alpha}$, the $1-\alpha$ quantile of the standard normal distribution. By contiguity, the $1-\alpha$ quantile of the permutation distribution of S_n also converges to $z_{1-\alpha}$ under P_n. Therefore, 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 S_n under P_n. In the case where P_n is "are just uncorrelated (but possibly dependent)"

a bivariate normal distribution with correlation $\rho_n = h/\sqrt{n}$,

$$S_n \xrightarrow{\mathcal{L}} N(h, 1)$$

under P_n and the limiting power against this sequence of alternatives is $1 - \Phi(z_{1-\alpha} - h)$. For a bivariate normal distribution with means zero and variances one, 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 as the score test, 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 nonnormality.

Remark 2.3 (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, that is, the limiting probability that the test rejects and the sample correlation is negative, is 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, it is common to use Fisher's z-transformation, $\sqrt{n}\tanh^{-1}(\hat{\rho}_n)$ when testing that the data are uncorrelated (for a detailed discussion of Fisher's z-transformation, see DasGupta 2008). As with using the sample correlation, a permutation test using Fisher's z-transformation is not guaranteed to be level α. In fact, the permutation distribution of the transformed correlation is asymptotically standard normal, but it is readily seen that the sampling distribution has asymptotic variance $\tau^2(P)$ when the correlation is zero. Consequently, using a studentized version of the test statistic, $\sqrt{n}\tanh^{-1}(\hat{\rho}_n)/\hat{\tau}^2$, gives an asymptotically level α permutation test. Details are given in Section S.1 of the supplement.

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 iid (X_i, Y_i) pairs following a simple univariate linear regression model why call them iid? doesn't this imply e ind. X?

$$Y_i = \alpha + \beta X_i + \epsilon_i, \quad i = 1, \ldots, n$$

why?

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_i and ϵ_i 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

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

or the least-squares estimator $\hat{\beta}_n$. If the X_i's are independent of the ϵ_i's (and therefore independent of the Y_i's under $H : \beta = 0$), 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 fit when computing the ordinary least-squares coefficient. The permutation approach may not work when using the least-squares statistic for the reduced model $Y = \beta X + \epsilon$ if $\alpha = 0$ is known in advance. Further details on why the constant should be included are provided in Section S.3 of the supplement.

Consider the regression model specified by the following assumptions:

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

Throughout, we will also use the usual matrix notation $Y = X\beta + \epsilon$, where X denotes the matrix whose rows are the $X_i^\top$, and Y and ϵ are the vectors with entries Y_i and ϵ_i, respectively.

Whenever iid (X_i, Y_i) pairs are observed, the least squares estimator estimates $\beta = E((X_i - \mu_X)(X_i - \mu_X)^\top)^{-1} E((X_i - \mu_X)(Y_i - \mu_Y))$ and $\alpha = E(Y_i) - E(X_i^\top)\beta$. For these α and β (which can be interpreted as the intercept and slope of the best fitting line), the data follows the model $Y_i = \alpha + X_i^\top\beta + \epsilon_i$ with $\epsilon_i = Y_i - \alpha - X_i^\top\beta$. It is easily verified that this error term has mean zero, and is uncorrelated with X_i. Therefore, the assumption in (2) that the predictors are uncorrelated with the errors is not strictly necessary. A commonly used sub-family of the regression models specified by assumptions (1)–(3) is heteroscedastic 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 nonconstant scedastic function $\sigma^2(X_i) = E(\epsilon_i^2 | X_i)$. (Note that we are not assuming $E(\epsilon_i | X_i) = 0$, so the scedastic 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,

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

need (1/n) (X' X)^-1 -> limit.

Tests beta=0

In the case of heteroscedasticity, 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 almost surely asymptotically normal with mean zero and variance $E(\epsilon_i^2) \cdot \Sigma_{XX}^{-1}$ when $\beta = 0$. Unless $E(\epsilon_i^2 X_i X_i^\top) = E(\epsilon_i^2)E(X_iX_i^\top)$, 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.

need indep. e, x

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

$$U_n(X, Y) = n\hat\beta_n^\top \left(\hat\Sigma_{XX}^{-1}\hat\Omega\hat\Sigma_{XX}^{-1}\right)\hat\beta_n,$$

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.

don't we assume Y,X iid?

Theorem 3.1. 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 $E(X_{1j}^4) < \infty$, $j = 1, \ldots, p$. If $\beta = 0$, then the permutation distribution $\hat R_n^{U_n}(t)$ of U_n obtained by permuting the Y_i's satisfies

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

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

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

almost surely, where $J^{U_n}(t, P)$ is the sampling distribution of U_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 $g_\delta(y_1, \ldots, y_n) = (\delta_1 y_1, \ldots, \delta_n y_n)$ for any $y \in \mathbb{R}^n$. In particular, if the errors are symmetric and $\beta = 0$, then $U_n(X, g_\delta(Y))$ is distributed as $U_n(X, Y)$ for any uniformly chosen transformation g_δ and the test is exact because the randomization hypothesis is satisfied. This permutation test, which changes the signs of the error terms in the model is closely related to the wild bootstrap, which is often performed by changing signs of the fitted residuals. Studentized permutation tests with respect to this group of transformations have been shown to be asymptotically valid in Janssen (1999b). The next theorem (which follows from Theorem 2.4 of Janssen 1999b) studies the asymptotic behavior of the permutation distribution

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

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

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 $E(X_{1j}^4) < \infty$, $j = 1, \ldots, p$. If $\beta = 0$, then the permutation distribution $\hat R_n^{U_n,\delta}(t)$ of U_n obtained changing the sign of the Y_i's satisfies

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

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

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

almost surely, where $J^{U_n}(t, P)$ is the sampling distribution of U_n.

Under the sequence of local alternatives $\beta = h/\sqrt{n}$, the permutation distribution of U_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 U_n under these alternatives is asymptotically chi squared with p degrees of freedom and noncentrality parameter

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

and so the local power of either of the randomization tests is $P(C > \chi_{p,1-\alpha}^2)$ where $C \sim \chi_p^2(\lambda)$ and $\chi_{p,1-\alpha}^2$ is the $1-\alpha$ 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 nonzero, 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 permuting the regressors corresponding to the coefficients of interest.

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

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

2'. $\{(Y_i, X_i, Z_i)\}$ are iid 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:

$$\sqrt{n}\left(\left(\hat\beta_n^\top, \hat\gamma_n^\top\right)^\top - \left(\beta^\top, \gamma^\top\right)^\top\right) \xrightarrow{\mathcal{L}} N(0, \Sigma_{\tilde X\tilde X}^{-1}\Omega\Sigma_{\tilde X\tilde X}^{-1}).$$

To test the hypothesis

$$H : \beta = 0,$$

one could use a Wald statistic

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

with $R = \text{diag}(1, \dots, 1, 0, \dots, 0)$, $\hat{\Sigma}_{\tilde{X}\tilde{X}} = \frac{1}{n} \sum_i \tilde{X}_i^\top \tilde{X}_i$ 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 under H.

If the X_i's are independent of the (Z_i, Y_i)'s, then for any permutation π of $\{1, \dots, n\}$, $(X_{\pi(i)}, Z_i, Y_i)$ is distributed as (X_i, Z_i, Y_i). Moreover, $W_n(X_\pi, Z, Y)$ is distributed as $W_n(X, Z, Y)$. 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

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

is almost surely asymptotically χ_k^2 and the test is asymptotically level α.

Theorem 3.3. Suppose that $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ satisfies the regression model specified by conditions (1')–(3') and also assume that $E(Y_1^4) < \infty$, $E(X_{1j}^4) < \infty$, $j = 1, \dots, k$, and $E(Z_{1j}^4) < \infty$, $j = 1, \dots, p - k$. If $\beta = 0$, the permutation distribution $\hat{R}_n^{W_n}(t)$ of W_n obtained by permuting the X_i's satisfies

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

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

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

almost surely, 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

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

  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

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

Freedman and Lane assumed that one of the columns of Z is constant. 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 normal with mean zero, and covariance given by the upper left $k \times k$ submatrix of $\Sigma_{XZ}^{-1} \Sigma_{\epsilon XZ} \Sigma_{XZ}^{-1}$ where

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

and

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

Asymptotically, the F-statistic is distributed as $N^\top (E(\epsilon_i^2))^{-1} R \Sigma_{XZ} R^\top N$, where $N \sim N(0, R\Sigma_{XZ}^{-1} \Sigma_{\epsilon XZ} \Sigma_{XZ}^{-1} R^\top)$. 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 3.4. Under the conditions of Theorem 3.3, if $\beta = 0$, the permutation distribution $\hat{P}_n^{W_n}(t)$ of W_n obtained by the Freedman–Lane procedure satisfies

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

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

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

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

4. Partial Correlation

Suppose we have univariate variables X and Y, and a multivariate variable Z which satisfy model assumptions 1'.–3'. for $k = 1$. The partial correlation between X and Y given Z, denoted $\rho_{X,Y|Z}$ is defined as 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 $Y = X\beta + Z^\top \gamma + \epsilon$, 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 $(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n)$ are iid according to P such that X_1 and Y_1 are uncorrelated but not necessarily independent conditionally on the Z_i. Also assume that $E(X_1^4) < \infty$, $E(Y_1^4) < \infty$, and $E(Z_{1j}^4) < \infty$ for $j = 1, \ldots, p - 1$. The permutation distribution $\hat{R}_n^{V_n}(t)$ of $V_n = \sqrt{n}\hat{\rho}_{X,Y|Z}/\hat{\sigma}_n$ satisfies

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

almost surely.

As with the Freedman–Lane procedure, this test relies on permuting residuals, and is never exact.

5. Simulations

In this section, simulations will be presented investigating the claims made in Sections 2 and 3.

The simulations presented were conducted using MATLAB. Each reported result used 10,000 simulations. For bootstrap or permutation tests, 9999 bootstrap iterations or permutations of the data were computed. Finally, all hypothesis tests presented were conducted at the nominal level $\alpha = 0.05$.

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

Distribution	n:	10	25	50	100	200
$N(0, 1)$	Bootstrap	0.0903	0.0735	0.0683	0.0586	0.0582
	Normal Approx	0.0346	0.0439	0.0506	0.0489	0.0532
	Unstudentized Perm	0.0510	0.0466	0.0537	0.0520	0.0534
	Studentized Perm	0.0470	0.0525	0.0525	0.0511	0.0561
Multivariate t_5	Bootstrap	0.1056	0.0883	0.0778	0.0705	0.0669
	Normal Approx	0.0253	0.0359	0.0394	0.0412	0.0443
	Unstudentized Perm	0.0880	0.1176	0.1440	0.1616	0.1844
	Studentized Perm	0.0507	0.0462	0.0460	0.0456	0.0471
Exponential	Bootstrap	0.0647	0.0540	0.0499	0.0504	0.0484
	Normal Approx	0.0639	0.0528	0.0477	0.0495	0.0482
	Unstudentized Perm	0.0150	0.0082	0.0077	0.0053	0.0074
	Studentized Perm	0.0679	0.0508	0.0476	0.0502	0.0485
Circular	Bootstrap	0.0646	0.0497	0.0491	0.0509	0.0543
	Normal Approx	0.0628	0.0482	0.0479	0.0482	0.0508
	Unstudentized Perm	0.0151	0.0085	0.0067	0.0067	0.0052
	Studentized Perm	0.0674	0.0468	0.0488	0.0484	0.0521
$t_{4.1}$	Bootstrap	0.1090	0.0922	0.0809	0.0768	0.0659
	Normal Approx	0.0211	0.0336	0.0362	0.0377	0.0358
	Unstudentized Perm	0.1187	0.1904	0.2414	0.2820	0.3138.
	Studentized Perm	0.0444	0.0428	0.0426	0.0442	0.0391

Table 1 compares the rejection probabilities of the permutation tests for the hypothesis $H : \rho = 0$ based on the unstudentized correlation $\sqrt{n}\hat{\rho}_n$ with the studentized correlation $\sqrt{n}\hat{\rho}_n/\hat{\tau}$. Also included in the table are the rejection probabilities of the test based on the normal approximation (comparing $\sqrt{n}\hat{\rho}_n$ with the appropriate quantiles of the $N(0, \hat{\tau}_n^2)$ distribution) and the pairs bootstrap (which computes $\sqrt{n}\hat{\rho}_n$ on resampled (X_i, Y_i) pairs).

For the purposes of this table, we are interested in sampling from bivariate distributions which have correlation zero, but not independence. One source of such distributions are elliptical distributions, for which correlation zero implies dependence between the variables only when the data are normally distributed.

One such distribution is the d-dimensional multivariate t-distribution with ν degrees of freedom, and covariance $\nu/(1-\nu)\Sigma$, denoted by $t_\nu(\Sigma)$, which 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 (which has finite fourth moments).

Other elliptically distributed, p-dimensional random vectors 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 $r \sim \exp(1)$, as well as the circular uniform distribution, which gives (X, Y) uniformly distributed on the unit circle.

Finally, we consider an example where the data used are of the form $X = W + Z$ and $Y = W - Z$, where X and Y are independent and identically distributed. It is seen in the Appendix that τ^2 can be made arbitrarily large by choosing W and Z to be independent t_υ random variables with υ close to four. As an example, we will consider $\upsilon = 4.1$.

To summarize, the distributions used are

- $(N(0,1))$ multivariate normal distribution with mean zero and identity covariance.
- (Multivariate t_5) multivariate t-distribution with identity covariance and 5 degrees of freedom
- (Exponential) $(X_i, Y_i) = rS'u$ where $S = \text{diag}(\sqrt{2}, 1)$, u is uniformly distributed on the two-dimensional unit circle, and $r \sim \exp(1)$
- (Circular) The uniform distribution on the two dimensional unit circle.
- $(t_{4.1})$ $X_i = W_i + Z_i$ and $X_i = W_i - Z_i$ where the W's and Z' are iid $t_{4.1}$ variables

The simulations for testing sample correlation indicate that an unstudentized permutation test can indeed have rejection probability that is far from the nominal level, even in large sample. On the other hand, the studentized permutation test has the desired rejection probability for large n in all of the simulations, while still retaining the exactness property of the unstudentized test under independence (as seen in the normal example). Even in examples where the permutation test is not exact, it appears to have rejection probability much closer to the nominal level when compared to the normal approximation, especially in small-sample sizes. The studentized permutation test also outperforms the bootstrap in the normal example and both t examples with the performance being similar in the remaining examples.

Table 2. Type 3 error rates for normal approximation and permutation tests of $\rho = 0$ using the sample correlation statistic.

Distribution	n:	10	25	50	100	200
t_5	Normal Approx	0.0062	0.0070	0.0056	0.0029	0.0007
	Not Studentized	0.0233	0.0258	0.0277	0.0202	0.0144
	Studentized	0.0134	0.0094	0.0065	0.0031	0.0008
Normal	Normal Approx	0.0086	0.0071	0.0039	0.0019	0.0005
	Not Studentized	0.0149	0.0078	0.0047	0.0017	0.0004
	Studentized	0.0129	0.0080	0.0041	0.0023	0.0005

Further in the supplement, analogous simulations to those in Table 1 are given for Fisher's z-transformation. Using Fisher's z-transformation appears to perform similarly to using the untransformed correlation.

Table 2 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. As an example with nonzero, but small correlation, we will use

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

as the covariance matrix for both simulation settings.

It is seen in Table 2 that the two permutation tests as well as the normal approximation behave similarly when the underlying distribution is normal. When the underlying distribution is multivariate t, the studentized permutation test and the normal approximation have similar error rates, and both are lower

Table 3. Rejection probabilities for tests of $\beta = 0$.

Dist	$\sigma^2(x)$	n:	10	25	50	100	200
$N(0,1)$	1	Permutation	0.0518	0.0501	0.0509	0.0505	0.0515
		Sign Change	0.0513	0.0492	0.0512	0.0501	0.0510
		Wild Bootstrap	0.0650	0.0560	0.0523	0.0509	0.0533
		Normal Approx	0.1673	0.0896	0.0668	0.0585	0.0560
	$\|x\|$	Permutation	0.0520	0.0525	0.0543	0.0501	0.0505
		Sign Change	0.0509	0.0518	0.0535	0.0492	0.0502
		Wild Bootstrap	0.0635	0.0591	0.0585	0.0513	0.0514
		Normal Approx	0.1631	0.0931	0.0724	0.0586	0.0546
	$\|\log(x)\|$	Permutation	0.0517	0.0534	0.0504	0.0509	0.0507
		Sign Change	0.0484	0.0517	0.0493	0.0503	0.0500
		Wild Bootstrap	0.0718	0.0665	0.0548	0.0528	0.0512
		Normal Approx	0.1578	0.0932	0.0671	0.0585	0.0523
t_5	1	Permutation	0.0486	0.0492	0.0515	0.0508	0.516
		Sign Change	0.0496	0.0504	0.0514	0.0527	0.0520
		Wild Bootstrap	0.0601	0.0543	0.0537	0.0529	0.0517
		Normal Approx	0.1529	0.0812	0.0670	0.0597	0.0559
	$\|x\|$	Permutation	0.0542	0.0494	0.0507	0.0486	0.0504
		Sign Change	0.0495	0.0490	0.0481	0.0475	0.0504
		Wild Bootstrap	0.0641	0.0572	0.0512	0.0507	0.0520
		Normal Approx	0.1599	0.0830	0.0634	0.0546	0.0537
	$\|\log(x)\|$	Permutation	0.0489	0.0531	0.0541	0.0491	0.0489
		Sign Change	0.0496	0.0527	0.0526	0.0490	0.0495
		Wild Bootstrap	0.0693	0.0642	0.0584	0.0525	0.0518
		Normal Approx	0.1482	0.0880	0.0704	0.0555	0.0517
Exponential	1	Permutation	0.0523	0.0507	0.0498	0.0513	0.0496
		Sign Change	0.0431	0.0460	0.0471	0.0521	0.0490
		Wild Bootstrap	0.0574	0.0545	0.0514	0.0529	0.0492
		Normal Approx	0.1468	0.0769	0.0627	0.0596	0.0532
	$\|x\|$	Permutation	0.0620	0.0630	0.0589	0.0616	0.0526
		Sign Change	0.0438	0.0553	0.0575	0.0594	0.0505
		Wild Bootstrap	0.0670	0.0667	0.0622	0.0628	0.0514
		Normal Approx	0.1659	0.0925	0.0731	0.0675	0.0534
	$\|\log(x)\|$	Permutation	0.0699	0.0766	0.0636	0.0572	0.0562
		Sign Change	0.0490	0.0667	0.0610	0.0555	0.0564
		Wild Bootstrap	0.0826	0.0809	0.0665	0.0593	0.0565
		Normal Approx	0.1769	0.1043	0.0761	0.0628	0.0604

Replicate.

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

Dist	$\sigma^2(x)$	n:	10	25	50	100	200
$U(1,4), N(0,1)$	1	Permutation	0.0488	0.0487	0.0505	0.0532	0.0502
		Normal Approx	0.2457	0.1133	0.0781	0.0641	0.0553
	$\lvert x\rvert$	Permutation	0.0529	0.0542	0.0520	0.0546	0.0518
		Normal Approx	0.2512	0.1183	0.0807	0.0686	0.0575
	$\lvert\log(x)\rvert$	Permutation	0.0525	0.0529	0.0512	0.0505	0.0479
		Normal Approx	0.2536	0.1138	0.0771	0.0642	0.0552
$N(0, I_3)$	1	Permutation	0.0502	0.0489	0.0528	0.0473	0.0502
		Normal Approx	0.2597	0.1214	0.0881	0.0640	0.0596
	$\lvert x\rvert$	Permutation	0.0915	0.0798	0.0684	0.0573	0.0507
		Normal Approx	0.3325	0.1620	0.1062	0.0754	0.0600
t_5	1	Permutation	0.0803	0.0880	0.0791	0.0740	0.0718
		Normal Approx	0.3414	0.1993	0.1383	0.1063	0.0930
	$\lvert x\rvert$	Permutation	0.1360	0.1286	0.1170	0.1024	0.0913
		Normal Approx	0.4212	0.2433	0.1835	0.1393	0.1222
t_{15}	1	Permutation	0.0588	0.0578	0.0550	0.0540	0.0521
		Normal Approx	0.2912	0.1404	0.0973	0.0733	0.0618
	$\lvert x\rvert$	Permutation	0.1052	0.0880	0.0776	0.0651	0.0597
		Normal Approx	0.3620	0.1830	0.1213	0.0869	0.0697

than the unstudentized test. It is seen in this example that using the unstudentized correlation statistic can lead to excessively large Type 3 error rates, even in large samples. The normal approximation appears to substantially outperform the studentized permutation tests when $n = 10$, but this likely explained by the results in Table 1 showing that the normal approximation is very conservative in small samples under the null. But for $n \geq 25$, the studentized permutation test and the normal approximation have very similar Type 3 error rates in each of the examples.

Finally, we present simulation results showing the performance of the permutation tests described in Section 3 for linear regression. Table 3 compares the rejection probabilities of the two methods of permutation test (permuting the predictors and changing the signs) with the normal approximation, which compares the Wald statistic with the appropriate quantiles of the chi-squared distribution, when the nominal level is $\alpha = 0.05$.

For the simulations in Table 3, the regression model used is

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

where $X_i \sim U(1, 4)$ and $\epsilon_i = \sigma(X_i) \cdot N_i$ where N_i is a random variable, and $\sigma(\cdot)$ is a scedastic function specified in the table. The distributions of N_i used are $N(0, 1)$, t_5, and Exponential(1) (centered to have mean zero). The simulations are performed with $\beta = 0$ so that the rejection probabilities reported are the Type 1 error rates.

In the cases of normal and t errors given in Table 3, the permutation and sign change methods perform similarly. Although the permutation method is only guaranteed to be exact under homoscedasticity, the rejection probabilities are very close to the nominal level in the presence of heteroscedasticity. In the exponential example, the permutation method is exact under homoscedasticity, while the sign change method is conservative. Under heteroscedasticity, the sign change method has rejection probabilities that are closer to the nominal level than the permutation method. Had it not been known at the outset that $\alpha = 0$, the sign change method would no longer be valid. On the other hand, the permutation test is not affected by changes in α. In such cases, the wild bootstrap is an appealing alternative to the sign change method. In each of the simulation settings, the permutation method has rejection probabilities that are closer to the nominal level than the wild bootstrap.

The next simulation is a comparison of the rejection probability of the normal approximation which compares the Wald statistic with the appropriate quantiles of the chi-squared distribution, and permutation tests of the hypothesis $H_0 : \beta = 0$ in the regression model

compare with χ_1^2

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

Table 4 gives rejection probabilities for simulations under this model using $\alpha = 0$ and $\gamma = 1$, and several joint distributions of X, Z, and ϵ. The first joint distribution used is $X_i \sim U(1, 4)$, $Z_i \sim N(0, 1)$ and $\epsilon_i = \sigma(X_i) \cdot N_i$, independently, where N_i is a standard normal random variable, and $\sigma(\cdot)$ is a scedastic function specified in the table. The remaining simulations take $\epsilon_i = \sigma(X_i) \cdot e_i$ with (X_i, Z_i, e_i) following a multivariate normal distribution with mean zero and identity covariance, a multivariate t-distribution with 5 degrees of freedom and identity covariance, and a multivariate t-distribution with 5 degrees of freedom and identity covariance. Again, we simulate under the null hypothesis and report the Type 1 error rates. Table 5 gives the Type 3 error rates for the multivariate normal and multivariate t (with 15 d.f.) settings when $\beta = 0.1$.

In heteroscedastic examples where the predictors and errors are independent, the permutation test is exact. In examples where the permutation test is not exact, we see that the permutation test has the correct rejection probability asymptotically. Similarly to what was seen in the simulations for correlation, the

Table 5. Type 3 error rates for tests of $\beta = 0$.

Dist	$\sigma^2(x)$	n:	10	25	50	100	200
$N(0, I_3)$	1	Permutation	0.0172	0.0088	0.0042	0.0027	0.0003
		Normal Approx	0.0905	0.0246	0.0081	0.0036	0.0005
	$\lvert x\rvert$	Permutation	0.0281	0.0153	0.0088	0.0036	0.0010
		Normal Approx	0.1158	0.0381	0.0141	0.0060	0.0013
t_{15}	1	Permutation	0.0181	0.0106	0.0056	0.0025	0.0010
		Normal Approx	0.1030	0.0303	0.0110	0.0038	0.0012
	$\lvert x\rvert$	Permutation	0.0412	0.0310	0.0239	0.0146	0.0103
		Normal Approx	0.1518	0.0662	0.0378	0.0207	0.0121

permutation test for regression coefficients has a rejection probability that is far closer to the nominal level than using a normal approximation, which can give rejections probabilities that are considerably higher than the nominal level. Furthermore, the Type 3 error rates of the permutation test are much lower than those of the normal approximation.

6. 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 applied to 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 rejection probabilities (under the null) of the permutation test are closer to the nominal level than using the usual F-statistic, which agrees with simulation results on Wald-type statistics for regression problems provided in Pauly et al. (2015), and that asymptotically the tests are indeed of the correct level. Moreover, simulation results suggest that permutation tests based on the Wald statistic can have much better small sample properties than using an asymptotic approximation. Even in cases where the permutation test is not exact, the rejection probability can be much closer to the nominal level.

Supplementary Materials

The online supplementary materials contain additional examples and results and the appendix for the article.

Acknowledgment

We would like to thank the referees whose comments greatly improved the article.

Funding

DiCiccio's research supported by NSF Grants DMS-1307973, and DMS-1007732. Romano's research supported by NSF Grants DMS-1307973, and DMS-1007732.

References

Anderson, M., and Robinson, J. (2002), "Permutation Tests for Linear Models," *Australian & New Zealand Journal of Statistics*, 43, 75–88. [1212]
Chung, E., and Romano, J. P. (2013), "Exact and Asymptotically Robust Permutation Tests," *Annals of Statistics*, 41, 458–507. [1212]
DasGupta, A. (2008), *Asymptotic Theory of Statistics and Probability*, New York: Springer. [1214]
Freedman, D., and Lane, D. (1983), "A Nonstochastic Interpretation of Reported Significance Levels," *Journal of Business and Economic Statistics*, 1, 292–298. [1216]
Good, P. I. (2005), *Permutation, Parametric and Bootstrap Tests of Hypotheses* (3rd ed.), Springer Series in Statistics, New York: Springer. [1212]
Janssen, A. (1997), "Studentized Permutation Tests for Non i.i.d. Hypotheses and the Generalized Behrens Fisher problem," *Statistics and Probability Letters*, 36, 9–21. [1212]
——— (1999b), "Nonparametric Symmetry Tests for Statistical Functionals," *Mathematical Methods of Statistics*, 8, 320–343. [1212,1215]
——— (2005), "Resampling Student's T-type Statistics," *Annals of the Institute of Statistical Mathematics*, 57, 507–529. [1212]
Janssen, A. and Pauls, T. (2003), "How do Bootstrap and Permutation Tests Work," *Annals of Statistics*, 31, 768–806. [1212]
Neuhaus, G. (1993), "Conditional Rank Tests for the Two-sample Problem Under Random Censorship," *Annals of Statistics*, 21, 1760–1779. [1212]
Omelka, M. and Pauly, M. (2012), "Testing Equality of Correlation Coefficients in a Potentially Unbalanced Two-sample Problem Via Permutation Methods," *Journal of Statistical Planning and Inference*, 142, 1396–1406. [1212]
Pauly, M. (2011), "Discussion About the Quality of F-ratio Resampling Tests for Comparing Variances," *TEST*, 20, 163–179. [1212]
Pauly, M., Brunner, E., and Konietschke, F. (2015), "Asymptotic Permutation Tests in General Factorial Designs," *Journal of the Royal Statistical Society*, Series B, 77, 461–473. [1220]
Romano, J. P. (1990), "On the Behavior of Randomization Tests Without a Group Symmetry Assumption," *Journal of the American Statistical Association*, 85, 686–692. [1212]
White, H. (1980), "A Heteroscedasticity-Consistent Covariance Matrix and a Direct Test for Heteroscedasticity," *Econometrica*, 48, 817–838. [1212,1214]
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, 381–397. [1212]
Wu, C. F. J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *Annals of Statistics*, 14, 1261–1350. [1212]