Asian Journal of Statistical Sciences
Vol. 3, No-1, (2023), pp. 13-26

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

ROBUST PARTIAL CORRELATIONS

Rand R Wilcox[a] and Stanley F. Schilling-Friedemann[b]

[a] *Dept of Psychology, University of Southern California* [b] *Institute of Business Education and Management Training, University Leipzig*

ARTICLE HISTORY
Compiled August 28, 2023

Received 05 September 2022; Accepted 21 July 2023

ABSTRACT
The paper deals with four robust partial correlation coefficients. Two methods for testing the hypothesis of a zero correlation are studied via simulations. One of these methods deals with contamination bias, which refers to the fact that many robust regression estimators that have a high breakdown can be negatively impacted by few bad leverage points. The methods considered are illustrated using data that deals with playfulness and empathy.

KEYWORDS
skipped correlation; Winsorized correlation; Spearman's rho; Kendall's tau; MM-estimator

1. Introduction

A well established result is that Pearson's correlation, ρ, is not robust [1-2]. Roughly, this means that a very small change in a distribution can alter its value substantially. If, for example, $\rho = .8$, a slight change in the tails of the distributions can lower ρ to a value close to zero. The usual estimate of ρ, r, has a breakdown of only $1/n$, where the breakdown point refers to minimum proportion of points that must be altered to make the value of $|r|$ arbitrarily close to one or zero. Put another way, a single outlier can mask a strong association among the bulk of the participants, and it can result in a large value for r when in fact there is little or no association. It follows immediately that the partial correlation is not robust as well.

Motivated by personality research studies, this paper deals with robust analogs of the partial correlation coefficient. The approach used here is mentioned by Rao and Sievers [3] but not explored. To describe the basic strategy, let X, Y and Z denote three random variables and assume that

$$X = \beta_{11} Z + \beta_{01} + \epsilon_1 \tag{1}$$

CONTACT Rand R Wilcox. Email: rwilcox@usc.edu

and

$$Y = \beta_{12}Z + \beta_{02} + \epsilon_2, \tag{2}$$

where $(\beta_{11}, \beta_{01}, \beta_{12}, \beta_{01})$ are unknown parameters, and where ϵ_1 and ϵ_2 have some unknown bivariate distribution. Momentarily assume that the slopes and intercepts are estimated via the ordinary least squares (OLS) estimator based on a random sample of size n. Let r_{ij} ($i = 1, \ldots, n$; $j = 1, 2$) denote the corresponding residuals, where $j = 1$ refers to the residuals associated with (1) and $j = 2$ are the residuals associated with (2). Then the partial correlation is simply Pearson's correlation based on these residuals.

The strategy here is to replace the OLS estimator with some robust regression estimator that has a reasonably high breakdown point and to replace Pearson's correlation with some robust measure of association. Let $\tau_{xy.z}$ denote the resulting population version (the estimand) of some robust partial correlation estimator. One goal is to explore, via simulations, the properties of four robust estimators. Included are results when testing

$$H_0 : \tau_{xy.z} = 0. \tag{3}$$

A technical concern is that the residuals are dependent, but theoretical results derived by Randles [18] suggest that at least asymptotically, this is not a serious concern when testing (3). Here, both small and large sample sizes are considered to gain some sense about the practical importance of this issue.

2. Choosing a Robust Regression Estimator

First consider the issue of choosing some robust regression estimator. Many such estimators are available [2]several of which have a reasonably high breakdown point. In terms of efficiency, no single estimator dominates. An additional concern is that many robust estimators with a high breakdown point can be negatively impacted by what are called bad leverage points, which are type of outlier [2]. That is, a few outliers cannot result in an estimate of the slope that is arbitrarily large or small, but a small proportion of bad leverage points can substantially alter the estimate of the slope.

To elaborate on the notion of a bad leverage point, first focus on the linear model

$$Y = \beta_1 X + \beta_0 + \epsilon, \tag{4}$$

where ϵ has some unknown distribution and β_0 and β_1 are unknown parameters. Assume that for the bulk of the participants, (4) is the correct model. A bad leverage point is a point that satisfies two conditions. The first is that the value of the independent variable is an outlier. Such points are said to be leverage points. The second condition is that the residual associated with a leverage point is an outlier, based on the regression line given by (4).

Bad leverage points do not always result in a poor fit for the bulk of the data. But the reality is that they can have serious negative impact. Wilcox [2] illustrates this point when using the robust M-estimators derived by Markatou and Hettmansperger [4], Coakley and Hettmansperger [5] and Yohai [6] This problem persists when using the Theil [7] and Sen [8] estimator as well as the deepest regression line derived by

Rousseeuw and Hubert [9] The least trimmed squares estimator derived by Rousseeuw [10] relatively good at avoiding contamination bias but its asymptotic efficiency is relatively low [11]. The MGV and OP regression estimators in [2] avoid the negative impact of bad leverage points by eliminating all outliers. But this can result in a much higher standard error compared to eliminating only bad leverage points. S-estimators choose the slopes and intercept so as to minimize some robust measure of variation based on the residuals [12]. But Hössjer [13] shows that S-estimators cannot achieve simultaneously both a high breakdown point and high efficiency under the normal model. Davies [14] reports results on the inherent instability of S-estimators. Finally, the least absolute value (L_1) estimator chooses the values for the slope and intercept that minimize $\sum |r_i|$. This estimator can be impacted by bad leverage points as well.

In fairness, bad leverage points might have little or no impact on the MM-estimator and and the Theil-Sen estimator. But it is well established that this is not always the case. Consequently, it is prudent to use a method that deals with this issue. Because in general the MM-estimator has good efficiency, the approach here is to check for bad leverage points, remove any that are found, and use the MM-estimator on the remaining data. The details of how this is done are summarized in the next section of this paper.

3. Identifying Bad Leverage Points

In a major advance, Rousseeuw and van Zomeren [15] derived a method for identifying bad leverage points. Their method is based on the least median of squares estimator but at the time, concerns about the negative impact of bad leverage points on robust regression estimatores were not known. Here, a slight modification of their method is used that is aimed at dealing with this issue.

Consider the random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$. Here, the MAD-median rule is used to determine whether (X_i, Y_i) is a leverage point. The MAD-median rule is a special case of the multivariate outlier detection method derived by Rousseeuw and van Zomeren [15.]

Let M_x denote the usual sample median based on $X_1, \ldots, X_n$. The median absolute deviation (MAD) statistic is the median of $|X_1 - M_x|, \ldots, |X_n - M_x|$. Then X_i is labeled an outlier, and consequently (X_i, Y_i) labeled a leverage point, if

$$\frac{X_i - M_x}{\text{MADN}} > 2.24, \tag{5}$$

where MADN=MAD/.6745, and 2.24 is the square root of the .975 quantile of a chi-squared distribution with one degree of freedom. Under normality, MADN=MAD/.6745 estimates the standard deviation. This outlier detecting method has played a major role in a wide range of techniques [2]. Note that this outlier detection method has a breakdown point of .5.

Next, proceed as follows:

(1) Remove all leverage points, which of course means that any bad leverage points are eliminated.
(2) Estimate the slope and intercept using the MM-estimator based on the remaining data yielding say a_1 and a_0, respectively.

(3) Based on all of the data, compute the residuals

$$v_i = Y_i - a_0 - a_1 X_i$$

($i = 1, \ldots, n$).

(4) Declare (X_i, Y_i) a bad leverage point if it is a leverage point and simultaneously v_i is an outlier based on the MAD-median rule.

(5) Refit the regression line via the MM-estimator based on all points not declared a bad leverage point. The resulting estimates of the slope and intercept, say b_1 and b_0, are taken to be the final estimates of β_1 and β_0, respectively. This is called the AMM-estimator, an adjusted MM-estimator.

4. Methods

This section describes two methods for making inferences about four robust partial correlation coefficients.

Method M1

Consider (1) and (2) and let u_{ij} ($i = 1, \ldots, n$; $j = 1, 2$) denote the residuals based on the AMM-estimator, where again $j = 1$ refers to the residuals associated with (1) and $j = 2$ are the residuals associated with (2). Here, four robust measures of association are considered based on these residuals: A Winsorized correlation, Spearman's rho, Kendall's tau and a skipped correlation.

The Winsorized correlation begins by Winsorizing the marginal distributions. Here, 20% Winsorizing is used, which has been the focus of past studies [2]. Let $g = 0.2n$ rounded down to the nearest integer and let $u_{(1)j} \leq u_{(2)j} \leq \cdots \leq u_{(n)j}$ denote the residuals written in ascending order. Let

$$W_{ij} = \begin{cases} u_{(g+1),j}, & \text{if } u_{ij} \leq u_{(g+1),j} \\ u_{ij}, & \text{if } u_{(g+1,j)} < u_{ij} < u_{(n-g),j} \\ u_{(n-g),j}, & \text{if } u_{ij} \geq u_{(n-g),j}. \end{cases} \tag{6}$$

The Winsorized correlation, r_w, is simply Pearson's correlation based on the Winsorized values. Let ρ_w denote the population version of r_w. Then the hypothesis

$$H_0 : \rho_w = 0 \tag{7}$$

is rejected at the α level if $|T_w| > t_{1-\alpha/2}$, where $t_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of Student's t distribution with $\nu = h - 2$ degrees of freedom, $h = n - 2g$ and

$$T_w = r_w \sqrt{\frac{n-2}{1-r_w^2}}. \tag{8}$$

Both Kendall's Tau and Spearman's rho are covered in a basic statistics course, so for brevity the details are not reviewed here.

The Winsorized correlation, Spearman's rho and Kendall's tau guard against the deleterious impact of outliers among the marginal distributions, but they do not take into account the overall structure of the data cloud. That is, a point can be unusual relative to the cloud of points, but not an outlier based in the marginal distributions.

As a simple illustration, it is not unusual to be young, it is not unusual for someone to have heart disease, but it is unusual to be both young and have heart disease. A skipped correlation is aimed at dealing with this issue.

The basic idea is to use an outlier detection method that takes into account the overall structure of the data cloud, remove any outliers that are found and compute Pearson's correlation based on the remaining data. There are several such methods, none of which dominate. Donoho and Gasko [16] argued that a point is an outlier if it is an outlier based on any projection of the data. A version of this approach is used here. The somewhat involved computational details are described in [2]. Only an outline of the method is described here.

The method begins by computing a robust measure of location. Here, the marginal medians are used. Next, project all of the data onto the line connecting the center of the cloud to the first point. If the projection of a point is an outlier among the projected data, it is declared an outlier. This process is repeated for all n points. That is, n projections are used to determine whether a point is an outlier. The skipped correlation, r_p, is simply Pearson's correlation after any outliers are removed. This measure of association corresponds to the OP estimator [2]. The important point here is that this measure of association is easily applied via the R function scor in the R package WRS, which can be installed at https://github.com/nicebread/WRS. Alternatively, source the file Rallfun-v41, which can be downloaded from https://osf.io/xhe8u/.

Let

$$T_p = r_p \sqrt{\frac{n-2}{1-r_p^2}}. \tag{9}$$

The hypothesis $H_0 : \rho_p = 0$ is rejected at the $\alpha = 0.05$ level if $|T_p| \geq c$, where

$$c = \frac{6.947}{n} + 2.3197.$$

A percentile bootstrap method can be used for any α, which has the advantage of yielding both a confidence interval and a p-value at the expense of higher execution time. But the focus here is on T_p.

Method M2

Simulations reported in the next section indicated that method M1 performs very well, in terms of controlling the Type I error probability, when dealing with non-normal distributions. But a closer look revealed that there are situations where the Winsorized correlation, Spearman's rho and Kendall's tau perform poorly when there are bad leverage points among $(X_1, Y_1), \ldots, (X_n, Y_n)$. For this reason, a modification of M1 was considered. It consists of simply removing any bad leverage points among $(X_1, Y_1), \ldots, (X_n, Y_n)$. That is, rather than use all n residual as done by M1, use only the residuals based on the MM-estimator after bad leverage points are removed.

5. Simulation Results

A series of simulations was used to get some understanding of the methods considered here. First, attention is focused on testing the hypothesis of a zero correlation at the $\alpha = 0.05$ level using method M1. Data were generated based on (1) and (2) when all of the parameters are zero and the error terms have a bivariate g-and-h distribution, which contains a bivariate normal distribution as a special case. The correlation between X and Z was taken to $\rho_{xz} = 0$ and 0.5 when $\rho_{yz} = 0$. Data from a g-and-h distribution are generated by first generating from a standard normal distribution, V, and then transforming to

$$U = \begin{cases} \frac{\exp(gV)-1}{g}\exp(hV^2/2), & \text{if } g > 0 \\ V\exp(hV^2/2), & \text{if } g = 0, \end{cases}$$

where g and h are parameters that determine the first four moments. The standard normal corresponds to $g = h = 0$. Increasing g increases skewness and increasing h increases kurtosis. Based on a survey of papers dealing with the extent distributions are non-normal [2], four distributions are considered here: (g, h)= (0, 0), (0, 0.2), (1, 0) and (1, 0.2). These distributions are symmetric and light-tailed, symmetric and heavy-tailed, skewed and light-tailed, and skewed and heavy-tailed, respectively. Figure 1 shows a plot of these four distributions. The sample size was taken to be $n = 20$ and 500. A large sample size was included as a partial check on the extent results in Randles [17] apply to the situation at hand.

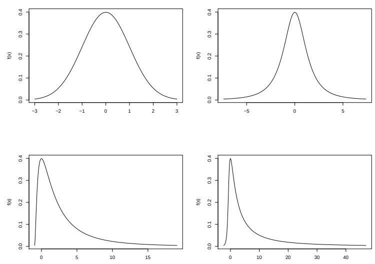

Figure 1. The four distributions used in the simulations. Upper left is (g, h)= (0, 0), upper right is (g, h)= (0, 0.2), lower left is (g, h)= (1, 0) and lower right is (g, h)= (1, 0.2)

Table 1 shows the results, based on 4000 replications, when X, Y and Z are independent and $\alpha = 0.05$. Bradley [18] suggested that as a general guide, when testing at the α level, a method is satisfactory if the actual level is between 0.5α and 1.5α. Presumably there are situations where this view is inappropriate, but for illustrative purposes, this view is adopted here. In particular, a method is viewed as satisfactory if the actual level is between 0.025 and 0.075 when testing at the 0.05 level. As can be seen, the estimates for all four methods satisfy Bradley's criterion. The results for $n = 500$ provide support for the theoretical results derived by Randles [17]. The simulations were repeated where Pearson's correlation for X and Z is 0.5. This altered the results by a few units in the third decimal place, so they are not reported.

An issue is how the four versions of M1 compare in terms of power. This depends on the nature of the association simply because the measures of association used here are sensitive to different features of the data. Nevertheless, some results are informative.

First consider the situation where slopes and intercepts in (1) and (2) are zero and $Y = \beta X + \epsilon$. Table 2 shows the estimates. The pattern is clear for the four distributions. Spearman's rho and Kendall's tau have the highest power, followed by the Winsorized correlation and then the skip correlation. For normal distributions, there is little separating the Winsorized and skipped correlations. But otherwise, there are situations where the power of the skipped correlation is substantially less than the power of the other methods, even with $n = 200$.

Table 1. Estimated probability of a Type I error, $\alpha = .05$ using method M1

n	g	h	WIN	SPEAR	KEN	SCOR
20	0	0	.048	.058	.056	.046
20	0	.2	.046	.055	.051	.049
20	1	0	.045	.056	.054	.053
20	1	.2	.041	.057	.058	.046
500	0	0	.048	.051	.051	.037
500	0	.2	.051	.048	.048	.044
500	1	0	.049	.050	.050	.039
500	1	.2	.045	.049	.048	.042

Table 2. Estimated power using method M1

n	g	h	WIN	SPEAR	KEN	SCOR
				$\beta = .5$		
20	0	0	.375	.451	.458	.353
20	0	.2	.389	.455	.451	.337
20	1	0	.437	.510	.496	.307
20	1	.2	.446	.524	.500	.316
				$\beta = .3$		
100	0	.0	.708	.781	.785	.684
100	0	.2	.819	.863	.860	.684
100	1	0	.879	.956	.944	.587
100	1	.2	.918	.967	.960	.625
				$\beta = .2$		
200	0	0	.677	.763	.760	.667
200	0	.2	.800	.848	.844	.650
200	1	0	.907	.981	.971	.588
200	1	.2	.947	.985	.980	.604

Table 3. Estimated probability of a Type I error when using M1 and there are two bad leverage points among data generated from a normal distribution

n	WIN	SPEAR	KEN	SCOR
20	.087	.188	.170	.050
30	.070	.150	.135	.049
40	.070	.143	.127	.049
50	.069	.119	.110	.051
75	.068	.108	.100	.050
100	.059	.088	.078	.048
150	.051	.065	.065	.044

Table 4. Estimated probability of a Type I error when using M2 and there are two bad leverage points among data generated from a normal distribution

n	WIN	SPEAR	KEN	SCOR
20	.087	.089	.110	.055
30	.060	.069	.082	.049
40	.059	.062	.073	.058
50	.059	.058	.066	.050
75	.058	.059	.059	.048
100	.054	.063	.063	.050
150	.055	.057	.057	.045

The failure rate of some hypothesis testing method is the minimum proportion of points needed to render control over the Type I error probability unsatisfactory. Consider the situation where X, Y and Z are independent. The next set of simulations are based on standard normal distributions with two bad leverage points. The issue is to what extent the failure rate is impacted by contamination bias. This was investigated by resetting $(X_1, Y_1) = (4, 4)$ and $(X_2, Y_2) = (5, 5)$.

Table 3 shows the estimated Type I error probability when using method M1. The method based on the skipped correlation performs well for all of the sample sizes considered. The Winsorized correlation is unsatisfactory for $n = 20$ based on Bradley's criterion but performs reasonably well for $n \geq 30$. Kendall's tau and Spearman's rho are unsatisfactory for $n \leq 100$ but they are satisfactory for $n = 150$.

Table 4 shows the results for the same situations in Table 3, only method M2 is used. M2 improves on M1 but when $n = 20$, only the skipped correlation performs reasonably well. For $n = 30$, the method based on Kendall's tau remains unsatisfactory but the other three methods perform well. For $n \geq 40$, all four methods are reasonably satisfactory.

Table 5 shows some results for both methods M1 and M2 when dealing with three bad leverage points: $(X_1, Y_1) = (3, 3)$, $(X_2, Y_2) = (4, 4)$ and $(X_3, Y_3) = (5, 5)$. First focus on M1. For $n = 30$, all four versions of method M1 are unsatisfactory. For $n = 40$, now the skipped correlation performs well, but the other three do not. Using M1, Kendall's tau and Spearman's rho are unsatisfactory even for $n = 200$. That is, the failure rate is less than 1.5 percent. The Type I error rate of the Winsorized correlation is unsatisfactory for $n = 75$ indicating that its failure rate is less than or equal 4 percent. The estimated Type I error probability, when $n = 100$, is .074 suggesting that it is satisfactory for $n \geq 100$. Increasing the number of replications to 10,000, again with $n = 100$, the estimated probability of a Type I error is .0733. The failure rate of the skipped correlation, for these 10,000 replications, is 10 percent.

Method M2 improves matters. Now the Winsorized correlation performed reason-

Table 5. Estimated probability of a Type I error, using methods M1 or M2, when there are three contamination points among data generated from a normal distribution

n	Method	WIN	SPEAR	KEN	SCOR
30	M1	.144	.287	.284	.083
	M2	.092	.112	.127	.068
40	M1	.115	.236	.217	.059
	M2	.076	.078	.092	.053
50	M1	.105	.202	.199	.055
	M2	.070	.081	.088	.061
75	M1	.080	.148	.143	.047
	M2	.051	.054	.060	.044
100	M1	.074	.130	.125	.047
	M2	.060	.060	.066	.051

Table 6. Estimated power using method M2, no bad leverage points

n	g	h	WIN	SPEAR	KEN	SCOR
				$\beta = 0.5$		
20	0	0	.372	.523	.523	.341
20	0	.2	.406	.490	.490	.326
20	1	0	.370	.348	.347	.278
20	1	.2	.347	.294	.294	.271

ably well for $n \geq 40$. However, $n \geq 75$ is needed when using Spearman's rho or Kendall's tau. Now the skipped correlation performs well for all of the sample sizes considered.

Boxplots of the partial correlations stemming from the situation in Table 5, based on M1 and $n = 100$, are shown in Figure 2. Note that the Winsorized, Spearman and Kendall measures are generally greater than zero, which explains why they are unsatisfactory in terms of controlling the Type I error probability. The median of the skipped correlation is very close to zero, but it clearly has a larger standard error compared to the other three methods.

If instead three bad leverage points occur for (1), again the Type I error probability, based on M1, can be inflated, but not as much as indicated in Table 5. Consider, for example, $(X_1, Z_1) = (3, 3)$, $(X_2, Z_2) = (4, 4)$ and $(X_3, Z_3) = (5, 5)$. For $n = 30$ the estimates are 0.062, 0.093, 0.091 and 0.045, respectively. As for M2, the estimates are 0.077, 0.076, 0.076 and 0.045.

Another issue is how method M2 compares to M1, in terms of power, when there are no bad leverage points. Table 6 shows some results for M2 when $n = 20$ and $\beta = 0.5$. Comparing these results to those in Table 2, generally, the power of M2 is similar to the power of M1, but the results for $g = 1$ and $h = 0$ are a notable exception. Now the power of M2 is substantially lower than the power of M1. A similar result was obtained with $n = 100$.

An issue is whether the test of independence derived by Hoeffding [19] is better at dealing with the contamination bias illustrated in Table 3. That is, does Hoeffding's method provide better control over the Type I error probability compared to the four measures of association considered here. Simulations indicate that the answer is no. For example, with $n = 50$, the estimated Type I error probability was 0.169. This is a bit better than Spearman or Kendall, but not by much. A method based on the percentage bend correlation [2] yields results similar to the Winsorized correlation. A

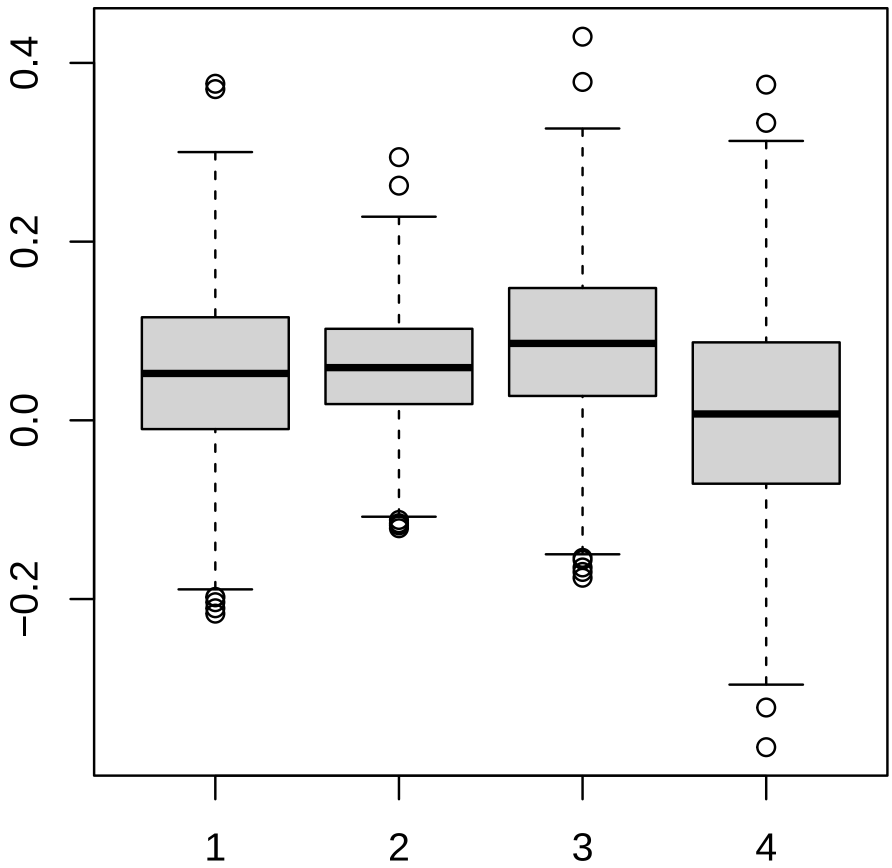

Figure 2. Boxplots of the partial correlations when there is bias contamination, $n = 100$. Left to right: Winsorized, Spearman, Kendall and skipped correlations

test of independence stemming from Stute et al. [20] failed as well.

The Rao and Sievers [3] method, which is based on the ranks of the residuals, worked reasonably well with $n = 20$ and no bad leverage points. But even with a single bad leverage point it performed poorly and generally performed much worse than the other techniques considered in Tables 3 and 4. Rao and Sievers [3] considered a single outlier among the dependent variable, but they did not consider situations where there are bad leverage points.

Some simulations were run using method M2 when Z has a p-variate distribution, $p > 1$. Table 7 shows the results for $p = 6$ and $n = 40$. For heavy-tailed distributions, the Winsorized correlation is unsatisfactory based on Bradley's criterion. Generally, the skipped correlation is best in terms of a Type I error probability close to the

Table 7. Estimated Type I error probabilities using method M2, no bad leverage points, $n = 40$, $p = 6$

g	h	WIN	SPEAR	KEN	SCOR
			$\beta = 0.5$		
0	0	.068	.070	.070	.062
0	.2	.069	.055	.055	.046
1	0	.086	.033	.033	.045
1	.2	.080	.025	.025	.049

nominal level. Increasing the sample size to $n = 60$, now the Winsorized correlation performs well. For $g = 1$ and $h = 0$ the estimate is .060. The estimate for the other three measures of association differed from 0.050 by a few units in the third decimal place. For $n = 60$ and $p = 8$, not shown here, again all four measures of association performed very well.

6. An Illustration

Method M2 is illustrated using unpublished data of the second author from research examining the relation of playfulness and empathy. Playfulness describes an individual differences variable "that allows people to frame or reframe everyday situations in a way such that they experience them as entertaining, and/or intellectually stimulating, and/or personally interesting. Those on the high end of this dimension seek and establish situations in which they can interact playfully with others (e.g., playful teasing, shared play activities) and they are capable of using their playfulness even under difficult situations to resolve tension (e.g., in social interactions, or in work-type settings [20] p. 114. Empathy can be understood as a competence to understand other people's feelings and is expressed by spontaneous emotional and volitional cognitive reactions to social situations [21] p. 549.

The sample comprised 254 participants, 78% women. Ages ranged between 18 to 65 with a mean equal to 27.8, and standard deviation equal 10.7. The participants completed measures of global playfulness (Short Measure of Playfulness, SMAP) [22], playfulness (OLIW: Other-directed, Lighthearted, Intellectual, and Whimsical playfulness, [20] and empathy (E-Scale) [20]. Further, 50 participants also provided peer-reports of their global playfulness and facets of playfulness. Prior research found mixed results on the relation of age with playfulness and empathy [20,22,24]. Therefore, age may be a relevant control variable when examining the relation between playfulness and empathy. Most psychometric measures had a slightly negative skewness (range: -0.48 to .14, MEAN=-0.14) and were strongly platykurtic (kurtosis range: -1.12 to 0.51, MEAN = -0.44). Age was positively skewed (skewness = 1.71) and slightly platykurtic (kurtosis = 2.00). Based on the skipped correlation, age was negatively associated with empathy (r_p = -.15, p = .017) and positively related to only one type of playfulness, namely the Lighthearted peer-report (r_p = .38, p = .013).

Table 8 presents the estimated partial correlations using Pearson's partial correlation and M2 methods. In sum, self-reported playfulness was positively related to empathy, except for Lighthearted playfulness (e.g., not worrying too much about future consequences of one's own behavior) and Whimsical playfulness (e.g., finding amusement in grotesque and strange situations) which were negatively or not found to be related to empathy, respectively. Neither peer-report of playfulness was found to be related to empathy. When comparing the coefficients of the methods, Winsorized

Table 8. Estimated partial correlation coefficients between playfulness and empathy controlling for age

	PEAR		WIN		SPEAR		KEN		SCOR	
Playfulness	r	p	r	p	r	p	r	p	r	p
Global	**.16**	.011	**.22**	.001	**.20**	.001	**.16**	.011	**.28**	<.001
Other-directed	**.30**	<.001	**.34**	<.001	**.33**	<.001	**.34**	<.001	**.34**	<.001
Lighthearted	-.09	.152	**-.17**	.007	**-.16**	.012	**-.15**	.020	**-.19**	.020
Intellectual	**.16**	.009	**.19**	.002	**.20**	.002	**.18**	.003	**.22**	<.001
Whimsical	.08	.181	.08	.198	.09	.133	.09	.177	.17	.080
Global	.26	.072	.15	.294	.17	.228	.26	.069	.13	.454
Other-directed	.11	.462	.21	.148	.19	.190	.21	.146	.17	.310
Lighthearted	-.13	.370	-.20	.168	-.20	.170	-.13	.362	-.20	.148
Intellectual	.14	.333	-.02	.914	-.04	.761	.08	607	.18	.736
Whimsical	.11	.462	.04	.799	.09	.535	.11	.459	.00	.902

Note: Significant coefficients are in bold
The last five lines refer to peer-reported data

and Spearman partial correlations were very similar. Kendall's tau and the skipped partial correlation were in most cases similar to the Winsorized and Spearman coefficients, in other cases they were either lower or higher in absolute value than those of the Winsorized and Spearman coefficients. In only two cases, Pearson coefficients were similar to most of the robust coefficients, namely for Other-directed and Whimsical playfulness. Note that for Lighthearted playfulness, Pearson's coefficient failed to reject the null hypothesis, whereas the null hypothesis was rejected using any version of method M2.

7. Concluding Remarks

In summary, all four versions of method M1 perform well in terms of Type I errors among the distributions considered in Table 1. Table 2 suggests that the Spearman and Kendall measures of association are best in terms of power. But as illustrated, the methods based on the Winsorized correlation, Kendall's tau and Spearman's rho run the risk of being negatively impacted by a few bad leverage points when using M1. Even with only two bad leverage points, Kendall's tau and Spearman's rho are unsatisfactory for $n \leq 100$. The skipped correlation is best in terms of dealing with contamination bias but at the expense of possibly lower power.

Method M2 deals with bad leverage points by checking for bad leverage points and removing any that are found. With two bad leverage points, the Type I error probability is controlled fairly well for $n \geq 30$, excluding Kendall's tau. All four measures of association performed reasonably well for $n \geq 40$. For three bad leverage points, the Winsorized correlation and Spearman's rho require $n \geq 40$, while Kendall's measures of association requires $n \geq 60$. The skipped correlation performs reasonably well even for $n = 30$.

The simulations based on M2 suggest that the skipped correlation tends to have the lowest power. However, the four measures of association used here are sensitive to different features of the data. Note, for example, that in Table 8, there are situations where the skipped correlation is larger in absolute value than the other measures that were used.

Finally, the R function part.cor can be used to apply all of the methods considered here and is stored in the file Rallfun-v41, which can be downloaded from https://osf.io/xhe8u/. It defaults to method M2. To use method M1, set the argument XOUT.blp = FALSE. To eliminate all leverage points, not just bad leverage points, set the argument XOUT.blp = FALSE and the argument xout=TRUE. The argument regfun controls which robust regression estimator is used. By default the MM-estimator is used assuming that the R package WRS has been installed. The argument corfun controls which measure of association is used. The default is the Winsorized correlation. Setting this argument equal to spear, tau or scor results in using Spearman, Kendall's tau or the skipped correlation, respectively. The R code used to analyze the data in the illustration is stored in the file analysis_PF_empathy.R and the data are stored in the file data_PF_empathy.Rds.

8. Reference

[1] Devlin, S.J., Gnanadesikan, R., Kettenring, J.R. (1981). Robust estimation of dispersion matrices and principal components. Journal of the American Statistical Association 76, 354–362.

[2] Wilcox, R. R. (2022). Introduction to Robust Estimation and Hypothesis Testing. 5th Edition. San Diego, CA: Academic Press.

[3] Rao, S. H. & Sievers, G.L. (1995) A robust partial correlation measure. Journal of Nonparametric Statistics, 5:1, 1–20, DOI: 10.1080/10485259508832630

[4] Markatou, M. & Hettmansperger, T. P. (1990). Robust bounded-influence tests in linear models. Journal of the American Statistical Association, 85, 187–190.

[5] Coakley, C. W. & Hettmansperger, T. P. (1993). A bounded influence, high breakdown, efficient regression estimator. Journal of the American Statistical Association, 88, 872–880. DOI: 10.2307/2290776

[6] Yohai, V. J. (1987). High breakdown point and high efficiency robust estimates for regression. Annals of Statistics, 15, 642–656. DOI: 10.1214/aos/1176350366

[7] Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. Indagationes Mathematicae, 12, 85–91.

[8] Sen, P. K. (1968). Estimate of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association, 63*, 1379–1389. DOI: 10.2307/2285891

[9] Rousseeuw, P. J., Hubert, M. (1999). Regression depth. *Journal of the American Statistical Association, 94*, 388–402.

[10] Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association, 79*, 871–880. DOI: 10.1080/01621459.1984.10477105

[11] Croux, C., Rousseeuw, P. J. & Hössjer, O. (1994). Generalized S-estimators. Journal of the American Statistical Association, 89, 1271–1281.

[12] Rousseeuw, P. J. & Yohai, V. (1984). Robust regression by means of S-estimators. Nonlinear Time Series Analysis. Lecture Notes in Statistics, 26, 256–272. New York: Springer.

[13] Hössjer, O. (1992). On the optimality of S-estimators. Statistics and Probability Letters, 14, 413–419.

[14] Davies, P. L. (1993). Aspects of robust linear regression. Annals of Statistics 21, 1843–1899.

[15] Rousseeuw, P. J. & van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association 85, 633–639. https://doi.org/10.2307/2289995.

[16] Donoho, D. L. & Gasko, M. (1992). Breakdown properties of the location estimates based on halfspace depth and projected outlyingness. Annals of Statistics, 20, 1803–1827.

[17] Randles, R. (1984). On tests applied to residuals. Journal of the American Statistical Association, 79, 349–354. doi.org/10.2307/2288275

[18] Bradley, J. V. (1978) Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144–152. doi.org/10.1111/j.2044-8317.1978.tb00581.x

[19] Hoeffding, W. (1948). A non-parametric test of independence. Annals of Mathematical Statistics, 19, 546–557. doi.org/10.1214/aoms/1177730150.

[20] Stute, W., Gonzalez Manteiga, W.G., Presedo Quindimil, M.P. (1998). Bootstrap approximations in model checks for regression. Journal of the American Statistical Association 93, 141–149. https://doi.org/10.2307/2669611

[21] Proyer, R. T. (2017). A new structural model for the study of adult playfulness: Assessment and exploration of an understudied individual differences variable. *Personality and Individual Differences, 108*, 113–122. doi.org/1.1016/j.paid.2016.12.011

[22]Leibetseder, M., Laireiter, A.-R., & Köller, T. (2007). Structural analysis of the E-scale. Personality and Individual *Differences, 42*, 547–561. doi.org/1.1016/j.paid.2006.08.002

[23] Proyer, R. T. (2012). Development and initial assessment of a short measure for adult playfulness: The SMAP. *Personality and Individual Differences, 53*, 989–994. doi.org/1.1016/j.paid.2012.07.018

[24] Sun, B., Luo, Z., Zhang, W., Li, W., & Li, X. (2018). Age-related differences in affective and cognitive empathy: Self-report and performance-based evidence. *Aging, Neuropsychology and Cognition, 25*, 655–672. doi.org/1.1080/13825585.2017.1360835