Journal of Modern Applied Statistical Methods

Volume 10 | Issue 1 Article 3

5-1-2011

Comparing the Strength of Association of Two Predictors via Smoothers or Robust Regression Estimators

Rand R. Wilcox
University of Southern California, rwilcox@usc.edu

Journal of Modern Applied Statistical Methods
May 2011, Vol. 10, No. 1, 8-18

1538 – 9472/11/$95.00

Comparing the Strength of Association of Two Predictors via Smoothers or Robust Regression Estimators

Rand R. Wilcox
University of Southern California,
Los Angeles, CA USA

Consider three random variables, Y, X_1 and X_2, having some unknown trivariate distribution and let η_j^2 (j = 1, 2) be some measure of the strength of association between Y and X_j. When η_j^2 is taken to be Pearson's correlation numerous methods for testing $H_o : \eta_1^2 = \eta_2^2$ have been proposed. However, Pearson's correlation is not robust and the methods for testing H_0 are not level robust in general. This article examines methods for testing H_0 based on a robust fit. The first approach assumes a linear model and the second approach uses a nonparametric regression estimator that provides a flexible way of dealing with curvature. The focus is on the Theil-Sen estimator and Cleveland's LOESS smoother. It is found that a basic percentile bootstrap method avoids Type I errors that exceed the nominal level. However, situations are identified where this approach results in Type I error probabilities well below the nominal level. Adjustments are suggested for dealing with this problem.

Key words: Explanatory power, Theil-Sen estimator, nonparametric regression, bootstrap methods, kernel density estimators.

Introduction

Consider three random variables, Y, X_1 and X_2 having some unknown trivariate distribution and let η_j^2 be some measure of association between Y and X_j (j=1, 2). This article considers the problem of testing

$$H_0 : \eta_1^2 = \eta_2^2 \qquad (1.1)$$

when η_j^2 is a robust version of explanatory power, which is estimated via the Theil (1950) and Sen (1968) regression estimator or the robust version of Cleveland's (1979) smoother

Rand R. Wilcox is a Professor of Psychology. He is the author of eight textbooks on statistics, including Modern Statistics for the Social and Behavioral Sciences (2012, New York, CRC Press). Email him at: rwilcox@usc.edu.

(LOESS). For the special case where η_j^2 is Pearson's correlation, ρ, numerous methods for testing $H_0 : \rho_1 = \rho_2$, as well as $H_0 : \rho_1^2 = \rho_2^2$, have been proposed by many authors (Hittner, May & Silver, 2003; Hotelling, 1940; Olkin, 1967; Dunn & Clark, 1971; Meng, Rosenthal & Rubin, 1992; Steiger, 1980; Wilcox & Tian, 2008; Wilcox, 2009; Williams, 1959; Zou, 2007). A general concern, however, is that r – the usual estimate of ρ – is not robust, roughly meaning that even a single outlier can result in a large value for r when there is little or no association among the bulk of the points. Similarly, a strong association among the bulk of the points can be masked by one or more outliers (Wilcox, 2005). Thus, ρ is not robust in the general sense as summarized by Huber (1981) and as illustrated by Wilcox (2005, p. 385).

Another concern is curvature. Experience with smoothers indicates that approximating the regression line with the usual parametric models can be unsatisfactory, which in turn raises concerns about how to measure the overall strength of association. A relatively simple strategy is to approximate the regression line with some type of nonparametric regression estimator or smoother (e.g., Efromovich, 1999; Eubank, 1999; Fan & Gijbels, 1996; Fox, 2001; Green & Silverman, 1993; Györfi, et al., 2002; Härdle, 1990; Hastie & Tibshirani, 1990) that can be used to estimate a robust measure of the strength of the association; this is the approach employed herein.

It is noted that there is a vast literature on identifying and ordering the importance of predictor variables; see for example Lafferty and Wasserman (2008) and the references they cite. It seems that none of these methods are based on a robust measure of association. Moreover, the precision of the resulting ordering is typically unclear. Thus, an additional goal of this research is to consider a formal hypothesis testing approach for determining which of two predictors has a stronger association with the outcome variable of interest, in contrast to merely estimating which has the stronger association.

Background

Basic results and methods used to measure and estimate the strength of an association are first reviewed. Consider the situation where the conditional mean of Y, given X, is assumed to be $Y = \beta_0 + \beta_1 X$ and ordinary least squares is used to estimate the unknown slope, β_1, and intercept, β_0. Let $\hat{Y} = b_0 + b_1 X$, where b_0 and b_1 are the least squares estimates of β_0 and β_1, respectively, based on the random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$. It is well known (and readily verified) that

$$r^2 = \frac{\hat{\sigma}^2(\hat{Y})}{\hat{\sigma}^2(Y)}, \tag{2.1}$$

where $\hat{\sigma}^2(\hat{Y})$ is the usual sample variance based on $\hat{Y}_i = b_0 + b_1 X_i, \quad i = 1, \ldots, n$. Slightly extending an approach to measuring the strength of an association used by Doksum and Samarov (1995), there is a simple and seemingly natural robust generalization of r^2. First, replace $\hat{Y}$ with $\tilde{Y}$, where $\tilde{Y}$ is any fit to the data, which might be obtained via a robust regression estimator (using a linear model) or some smoother that deals with curvature in a reasonably flexible manner. Next, let $\tau^2(Y)$ be some robust measure of variation associated with the marginal distribution of Y. It is assumed that $\tau^2(Y)$ has been chosen so that if there is no variation, $\tau^2(Y) = 0$. A general approach to measuring the strength of the association between Y and X is then

$$\eta^2 = \frac{\tau^2(\tilde{Y})}{\tau^2(Y)} \tag{2.2}$$

which Doksum and Samarov (1995) call explanatory power. To make η^2 practical, there are the issues of choosing τ and some method for computing $\tilde{Y}$. First consider τ. There are many robust alternatives to the usual variance (Wilcox, 2005). Lax (1985) compared the

efficiency of many scale estimators and concluded that two so-called A-estimators are best, one of which corresponds to what Shoemaker and Hettmansperger (1982) term the percentage bend midvariance. The other A-estimator found to perform well by Lax corresponds to what Shoemaker and Hettmansperger call the biweight midvariance.

Bickel and Lehmann (1976) argue that if both X and Y have symmetric distributions about zero, and if $|X|$ is stochastically larger than $|Y|$, then it should be the case that a measure of scale should be larger for X than it is for Y. Bickel and Lehmann define a measure of scale that satisfies this property to be a measure of dispersion. Shoemaker and Hettmansperger show that the percentage bend midvariance is a measure of dispersion but the biweight midvariance is not. A possible reason for preferring the biweight midvariance is that it has the highest possible breakdown point, namely .5. Here the focus is on the percentage bend midvariance, but this is not to suggest that all other measures of scale be eliminated from consideration.

Recently, Randal (2008) expanded on Lax's study and concluded that the two A-estimators recommended by Lax perform relatively well. However, Randal's study did not include Rocke's (1996) TBS (translated biweight S) estimator, and the tau measure of scale introduced by Yohai and Zamar (1988). As a partial check on the relative merits of these estimators, simulations based on 5,000 replications were used to estimate the standard error of the logarithm of these estimators when $n = 20$ for the same distributions used by Lax and Randal. (For this study the tau estimator was computed as described by Marrona & Zamar, 2002.) For a standard normal distribution, the results were 0.402, 0.388 and 0.530 for the percentage bend midvariance, tau and TBS, respectively.

For a 1-wild distribution (generate data from a normal distribution and multiply one value by 10), the results were 0.398, 0.420 and 0.516. For a slash distribution (Z / U, where Z has a standard normal distribution and U a uniform distribution), the results were 0.744, 0.631 and 0.670. No single estimator dominates. Although the focus here is on the percentage bend midvariance, it seems that the tau measure of scale deserves serious consideration based on these limited results.

For a random sample $Y_1,\ldots,Y_n$, the percentage bend midvariance is computed as follows. Let f be the value of $(1-\beta)n + .5$ rounded down to the nearest integer. The parameter β determines the finite breakdown point of the percentage bend midvariance, meaning the proportion of points that must be altered to make the estimate arbitrarily large. Roughly, β reflects the proportion of outliers that can be tolerated. Here β = 0.2 is used, which is motivated in part by the desire to obtain good efficiency under normality. Let $W_i = |Y_i - M|$, $i = 1,\ldots,n$, and let $W_{(1)} \leq \cdots \leq W_{(n)}$ be the W_i values written in ascending order. Let

$$\hat{\omega}_\beta = W_{(f)},$$

be the f[th] largest of the W_i values, let M be the usual sample median based on $Y_1,\ldots,Y_n$ and let

$$U_i = \frac{Y_i - M}{\hat{\omega}_\beta}$$

where $a_i = 1$ if $|U_i| < 1$; otherwise $a_i = 0$. The estimated percentage bend midvariance is

$$\hat{\zeta}^2 = \frac{n\hat{\omega}_\beta^2 \sum \Psi^2(U_i)}{(\sum a_i)^2}, \qquad (2.3)$$

where $\Psi(x) = \max[-1, \min(1, x)]$.

Henceforth, it is assumed that τ^2 is the percentage bend midvariance and that η^2 is estimated with

$$\hat{\eta}^2 = \frac{\hat{\tau}^2(\tilde{Y})}{\hat{\tau}^2(Y)}. \qquad (2.4)$$

There remains the problem of choosing a method for computing $\tilde{Y}$. First consider the situation where

$$Y = \beta_0 + \beta_1 X + \varepsilon, \qquad (2.5)$$

where X and ε are independent random variables. If b_0 and b_1 are estimates of β_0 and β_1, respectively, $\tilde{Y}_i = b_0 + b_1 X + \varepsilon$ $\left(i = 1,\ldots,n\right)$ can be used to compute $\hat{\tau}^2(\tilde{Y})$, which in turn can be used to compute $\hat{\eta}^2$. Wilcox (in press b) considered several robust regression estimates of β_0 and β_1 with the goal of estimating η^2 with $\hat{\eta}^2$. Based on mean squared error and bias, it was found that the Theil-Sen estimate of β_0 and β_1 performs relatively well, thus it is used here.

Consider all pairs of points for which the two predictor values differ. The Theil-Sen estimator computes the slope for all such pairs of points and the estimate of β_1, for example β_{1ts}, is taken to be the median of all these slopes. The intercept is taken to be $b_{0ts} = M_y - b_{1ts} M_x$, where M_y is the usual median of $Y_1,\ldots,Y_n$. The breakdown point of this estimator is approximately 0.29, where roughly, the breakdown point of an estimator is the proportion of points that must be altered to make it arbitrarily large or small. Moreover, the Theil-Sen estimator has excellent efficiency compared to many other robust estimators that have been proposed.

Next consider the more general case

$$Y = m(X) + \varepsilon \qquad (2.6)$$

where $m(X)$ is some unknown function of X and ε is some random variable that is independent of X. Wilcox (in press b) considered various nonparametric regression estimators with the goal of estimating η^2. In terms of mean squared error and bias, a so-called running interval smoother (Wilcox, 2005), as well as a method based on a cubic B-spline (Hastie & Tibshirani, 1990) were found to be relatively unsatisfactory. Bootstrap bagging combined with these estimators was considered, but was found to perform poorly. No method dominated, but a rough guideline is that , when a linear model holds , the Theil-Sen estimator is a good choice, otherwise use Cleveland's (1979) LOESS. A nonparametric estimator derived by Wood (2004) was found to perform relatively well when a linear model holds, but the Theil-Sen estimator seems preferable. Finally, when there is curvature LOESS was generally more satisfactory.

To briefly outline Cleveland's method, consider the random sample $(X_1, Y_1),\ldots,(X_n, Y_n)$. For any x, let

$$\delta_i = | X_i - x |.$$

Put the δ_i values in ascending order and retain the κn pairs of points that have the smallest δ_i values, where κ is some number between 0 and 1 and is called the span. Let

$$Q_i = \frac{| x - X_i |}{\delta_m},$$

where δ_m is the maximum of the retained δ_i values. If $0 \le Q_i < 1$, set

$$w_i = (1 - Q_i^3)^3,$$

otherwise $w_i = 0$. Let $m(X)$ be the estimate of Y given $X = x$ and use weighted least squares to estimate $m(X)$ using w_i as weights. Both R and S-PLUS provide access to a function, called lowess, that performs a robust version proposed by Cleveland, and the R version was used in the simulations reported here using the default value $\kappa = .75$. Cleveland's robust method in effect gives little or no weight to extreme Y values. (An outline of these additional computations also can be found in Härdle, 1990, p. 192.)

Methodology

Testing (1.1) when (2.5) Is True

Consider the case where (2.5) is true and η^2 is estimated via the Theil-Sen estimator. The initial strategy considered for testing (1.1) was a basic percentile bootstrap method (Efron & Tibshirani, 1993). Let (X_{i1}, X_{i2}, Y_i), $i = 1,\ldots,n$, be a random sample. A bootstrap sample is obtained by resampling with replacement n vectors from this random sample yielding, for example, $(X_{i1}^*, X_{i2}^*, Y_i^*)$. Let $\tilde{\eta}_j^2$ be the estimate of η_j^2 based on this bootstrap sample. Repeat this process B times yielding $\tilde{\eta}_{jb}^2$, $b = 1,\ldots,B$. Let

$$P = \frac{1}{B}\sum I_b$$

where the indicator function $I_b = 1$ if $\tilde{\eta}_{1b}^2 > \tilde{\eta}_{2b}^2$, otherwise $I_b = 0$. A (generalized) p-value is (Liu & Sing, 1997) is then:

$$p = 2\min(P, 1-P).$$

Let ρ_{12} be Pearson's correlation between X_1 and X_2. Initial simulations revealed that when testing at the $\alpha = .05$ level, the basic percentile bootstrap method described performs reasonably well when ρ_{12} = 0. However, as ρ_{12} increases, the estimate of the actual Type I error probability decreased. For ρ_{12} = 0.7 the estimates were less than 0.01.

The first attempt at reducing this problem was to use a bootstrap estimate of the squared standard error of $\eta_1^2 - \eta_2^2$, say, $\hat{\Upsilon}^2$, and then use the test statistic $(\eta_1^2 - \eta_2^2) / \hat{\Upsilon}$. However, in simulations with B = 100, Type I errors were found to be relatively sensitive to the distributions generating the data; increasing B to 400 reduced this problem somewhat but did not eliminate it, thus this approach was abandoned.

Consider the case ρ_{12} = 0 and let $D = \hat{\eta}_1^2 - \hat{\eta}_2^2$. It was found that Type I error control is improved if, rather than a single bootstrap sample, two independent bootstrap samples are used. That is, take a bootstrap sample from (X_{i1}, Y_i) and compute a bootstrap estimate of η_1^2, for example, $\tilde{\eta}_1^2$, take a new, independent bootstrap sample from (X_{i2}, Y_i) yielding $\tilde{\eta}_2^2$ and let $D = \tilde{\eta}_1^2 - \tilde{\eta}_2^2$. Repeating this process B times yields $D_1,\ldots,D_B$, which can be used to estimate $P = P(D < 0)$ in the manner already described. This in turn yields the generalized p value. Once again control over the probability of a Type I error was found to be unsatisfactory. However, it was found that control over the Type I error probability was improved if, instead of estimating $P = P(D < 0)$ with the bootstrap samples in the usual way, a kernel density estimate is used; this strategy was based on results from Racine and MacKinnon (2007).

Generally, kernel density estimates of the distribution of D take the form

$$\hat{f}(d) = \frac{1}{nh}\sum K(\frac{d - D_i}{\lambda}),$$

where K is some probability density function and λ is a constant to be determined called the span or smoothing parameter. Given h and a choice for K, which is assumed to be defined over some known interval (ℓ, u), an estimate of $P(D < 0)$ is

$$P(D < 0) = \frac{1}{nh}\sum_{i=1}^{n}\int_{\ell}^{0} K\left(\frac{t - D_i}{\lambda}\right)dt.$$

The focus here is on the Epanechnikov kernel where, for $| t |< \sqrt{5}$,

$$K(t) = \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}t^2\right),$$

otherwise $K(t) = 0$.

Following Silverman (1986, pp. 47-48), the span is taken to be

$$\lambda = 1.06\frac{A}{n^{1/5}},$$

where

$$A = \min(s, IQR / 1.34)$$

and s is the standard deviation, and IQR is the interquartile range.

From Silverman (1986), one possible way of improving on the basic kernel density estimator, is to use an adaptive method. Let $\tilde{f}(X_i)$ be an initial estimate of $f(X_i)$. Here, $\tilde{f}(X_i)$ is based on the so-called expected frequency curve (Wilcox, 2005, pp. 48-49). Let

$$\log(g) = \frac{1}{n}\sum \log(f(X_i))$$

and

$$\omega_i = \tilde{f}(X_i / g)^{-a}$$

where a is a sensitivity parameter satisfying $0 \le a \le 1$. Based on comments by Silverman (1986), if α = 0.5 is used, then the adaptive kernel estimate of the probability density function f is taken to be

$$\hat{f}(t) = K\{\lambda^{-1}\omega_i^{-1}(t - X_i)\}.$$

Henceforth, it is assumed that the adaptive method described is used to estimate P(D<0) based on $D_1^*,\ldots,D_B^*$, and the corresponding p-value is denoted by p.

There remains the problem of dealing with the general case $\rho_{12} \neq 0$. If it is assumed that there is normality and ρ_{12} is known, then simulations can be used to determine p_{adj} so that for some choice for α, $P(p \le p_{adj}) = \alpha$. In particular, imagine that simulations with N replications are performed resulting in the p-values, $p_1,\ldots,p_N$. Arranging these N values in ascending order yielding $p_{(1)} \le \cdots \le p_{(N)}$ and letting $C = \alpha N$ round to the nearest integer results in the adjusted p-value $p_{adj} = p_{(C)}$.

A simple approach when dealing with ρ_{12} unknown is to replace ρ_{12} with r_{12} in such a simulation. Execution time was found to be reasonably low, but to reduce it further, the following approach was considered when $\alpha = .05$. The value of p_{adj} was determined with n = 20 for ρ_{12} = 0, 0.2, 0.5 and 0.8. When ρ_{12} is known, it was found that p_{adj} is given approximately by $0.352|p_{adj}|+0.049$. But when $\alpha = .05$ the actual level can exceed 0.075 due to situations where $|r_{12}|$ exceeds $|\rho_{12}|$ resulting in over adjusting the critical p-value. In this situation, the additional concern is that r_{12} is not robust, and there is the issue of how to adjust the critical p-value when n > 20.

To deal with the lack of robustness associated with Pearson's correlation, r_{12} was replaced by Kendall's tau, resulting in r_{k12}. The population analog of r_{k12} is denoted by ρ_{k12}. Next, a 0.95 confidence interval for ρ_{k12} was computed using a basic percentile bootstrap method (Wilcox, 2005, p. 403), which has low execution time, even when the sample size is large. If this interval contains zero, let $\tilde{p}$ = 0.05, Otherwise, let $\tilde{p} = .352 \mid r_{k12} \mid +.049$. Rejecting (1) when the p-value is less than or equal to $\tilde{p}$ will be called method BTS.

This approximation depends on the sample size, n, but a convenient feature is that it was found to change slowly as n gets large. In particular, it continues to perform well when n = 100. For n = 200 this is no longer the case, but with $n \ge 100$ the adjustment makes little difference. So the suggestion is to use method BTS when $n \le 100$, otherwise reject if the p-value is less than or equal to α.

Testing (1.1) when (2.6) Is True

Consider now the more general case where the regression line is given by (2.6). Method BTS can be extended in an obvious way. In particular, again the strategy is to use independent bootstrap samples to estimate η_1^2 and η_2^2 and the adaptive kernel density

estimation method for computing a p-value is used. However, now the actual level of the test is more sensitive to ρ_{12} and for the case $\alpha = .05$, a modification of $\tilde{p}$ is required. As was the case when (2.5) is assumed, simulations indicate that if (1.1) is rejected when the p-value is less than or equal to α, the actual level will be less than or equal to α; avoiding actual Type I error probabilities substantially less than the nominal level is more difficult in this case. Based on preliminary simulations, under normality, when testing at the $\alpha = .05$ level, the following approach performed best among the methods considered. Let

$$\breve{p} = .25 \mid r_{k12} \mid + .05 + (100 - n)/10000,$$

$\hat{p} = \max(.05, \breve{p})$, and reject (1.1) if $p \leq \hat{p}$. For n>200, $\hat{p}$ is taken to be 0.05 and this will be called method SM. Note that in contrast to method BTS, a confidence interval for ρ_{k12} is not used.

Results

Simulations were used as a partial check on the actual level of methods SM and BTS when testing at the 0.05 level. Values for X_1 and X_2 were generated from a bivariate distribution for which the marginal distributions belong to the family of g-and-h distributions, which contains the standard normal as a special case. The R function rmul was used, in conjunction with the function ghdist, which are part of the library of R functions described in Wilcox (2005).

The R function rmul generates data from an m-variate distribution having a population correlation matrix $\mathbf{R}$ by first forming the Cholesky decomposition $\mathbf{U}\mathbf{U}' = \mathbf{R}$, where $\mathbf{U}$ is the matrix of factor loadings of the principal components of the square-root method of factoring a correlation matrix, and $\mathbf{U}'$ is the transpose of $\mathbf{U}$. Next, an $n \times m$ matrix of data, $\mathbf{X}$, for which the marginal distributions are independent, is generated, then $\mathbf{XU}$ produces an $n \times m$ matrix of data that has population correlation matrix $\mathbf{R}$.

To elaborate, let Z be a standard normal distribution. For $g > 0$, let

$$X = \frac{\exp(gZ) - 1}{h} \exp(hZ^2 / 2),$$

and for $g = 0$, let

$$X = \exp(hZ^2 / 2),$$

in which case $\mathbf{X}$ has a g-and-h distribution where g and h are parameters that determine the first four moments. When g = h = 0, $\mathbf{X}$ has a standard normal distribution. With g = 0 this distribution is symmetric and it becomes increasingly skewed as g gets large. As h gets large, the g-and-h distribution becomes more heavy-tailed. Table 1 shows the skewness (κ_1) and kurtosis (κ_2) for each distribution considered in the simulations used herein. They correspond to a standard normal (g = h = 0), a symmetric heavy-tailed distribution (h = 0.2, g = 0.0), an asymmetric distribution with relatively light tails (g = 0.2, h = 0) and an asymmetric distribution with relatively heavy tails (g = h = 0.2).

Table 1: Some Properties of the g-and-h Distribution

g	h	κ_1	κ_2
0.0	0.0	0.00	3.0
0.0	0.2	0.00	21.46
0.2	0.0	0.61	3.68
0.2	0.2	2.81	155.98

Simulations were run with ρ_{12} = 0 and 0.7, where ε has the same distribution as X_1 and X_2. Additional simulations were run where X_1 is normal but X_2 has one of the non-normal g-and-h distributions previously described.

Table 2 shows the estimated probability of a Type I error based on 1,000 replications when using method BTS, n = 50 and $Y = X_1 + X_2 + \varepsilon$. The columns headed by $X_1 \sim X_2$ indicate that X_1 and X_2 have

identical distributions, while $X_1 \sim N(0,1)$ means that X_1 has a standard normal distribution and X_2 has the g-and-h distribution indicated. Table 3 shows the results when using method SM when $Y = X_1 + X_2 + \varepsilon$ and $Y = X_1^2 + X_2^2 + \varepsilon$ with n = 60. As is evident, method BTS performs reasonably well in terms of avoiding a Type I error well above the nominal level, at least for the situations considered. A deficiency of the method is that the estimates drop below 0.025 in some situations. Method SM also performs reasonably well, but the actual level drops well below the nominal level in some situations.

Table 2: Estimated Type I Error Rates, n=50, Method BTS, α=0.05

			$X_1 \sim X_2$	$X_1 \sim N(0,1)$
g	h	ρ_{12}	$\hat{\alpha}$	$\hat{\alpha}$
0.0	0.0	0.0	0.039	-
-	-	0.7	0.048	-
0.0	0.2	0.0	0.026	0.054
-	-	0.7	0.023	0.059
0.2	0.0	0.0	0.041	0.027
-	-	0.7	0.063	0.048
0.2	0.2	0.0	0.064	0.056
-	-	0.7	0.044	0.071

Table 3: Estimated Type I Error Rates, n=60, α=0.05, Method SM

			$X_1 \sim X_2$	$X_1 \sim X_2$	$X_1 \sim N(0,1)$	$X_1 \sim N(0,1)$
g	h	ρ_{12}	$Y = X_1 + X_2 + \varepsilon$	$Y = X_1^2 + X_2^2 + \varepsilon$	$Y = X_1 + X_2 + \varepsilon$	$Y = X_1^2 + X_2^2 + \varepsilon$
0.0	0.0	0.0	0.036	0.026	-	-
		0.5	0.020	0.034	-	-
		0.7	0.008	0.014	-	-
0.0	0.2	0.0	0.036	0.012	0.048	0.049
		0.5	0.014	0.020	0.022	0.054
		0.7	0.008	0.012	0.014	0.021
0.2	0.0	0.0	0.032	0.024	0.036	0.022
		0.5	0.014	0.024	0.023	0.036
		0.7	0.014	0.016	0.008	0.022
0.2	0.2	0.0	0.026	0.014	0.040	0.024
		0.5	0.004	0.016	0.026	0.042
		0.7	0.008	0.008	0.010	0.023

Power

There is the issue of how much power is sacrificed if method SM is used rather than BTS when the regression line is straight. Table 4 shows the probability of rejecting when $Y = X_1 + \varepsilon$. As is evident, both methods have fairly high power for this special case and BTS can offer a substantial gain in power when the regression line is straight.

Table 4: Power Comparisons, n=60, α=0.05, $Y = X_1 + \varepsilon$

g	h	ρ_{12}	BTS	SM
0.0	0.0	0.0	0.960	0.833
-	-	0.5	0.861	0.659
0.0	0.2	0.0	0.930	0.668
-	-	0.5	0.777	0.467
0.2	0.0	0.0	0.968	0.807
-	-	0.5	0.836	0.710
0.2	0.2	0.0	0.942	0.672
-	-	0.5	0.772	0.460

An Illustration

In an unpublished study by Doi, a general goal was to identify good predictors of reading ability in children. Two of the predictors were a measure of letter naming speed and the speed at which lowercase letters could be identified. The outcome of interest was a measure of reading comprehension. A scatterplot of the data and the LOESS estimate of the regression strongly suggests that there is curvature, and a test of the hypothesis that the regression line is straight (using the method in Wilcox, 2005, section 11.5.1) is rejected at the 0.05 level; thus method SM is used and it rejects at the 0.05 level. The estimated explanatory power for the plot in the left panel is 0.444, and in the right panel it is 0.171. These results suggest that naming speed has a stronger association with comprehension.

If the apparent curvature is ignored, BTS also rejects at the 0.05 level, but now the estimated explanatory is 0.351 for the left panel and 0.142 for the right. That is, the estimated difference in explanatory power is substantially smaller compared to using a smoother. If instead Pearson correlations are compared using the method in Zou (2007), the 0.95 confidence interval for the difference is (–0.490, 0.024). Therefore, fail to reject at the 0.05 level.

Conclusion

In summary, numerous methods for comparing two predictors were considered based on a robust measure of the strength of the association. Two methods were found that perform reasonably well in simulations, one of which is based on a smoother and so provides a flexible approach to curvature. All indications are that Type I errors that exceed the nominal level are avoided using a basic percentile bootstrap method; however, there is a practical problem that the actual level can drop well below the nominal level, particularly when the sample size is small. Adjustments were suggested that substantially reduce this problem among the situations considered. The adjustment used by method BTS performed reasonably well in simulations, but when using method SM, situations occurred where the actual level drops well below the nominal level even with n = 60. In principle, if there are p predictors and the goal is to compare subsets of k predictors, a strategy similar to those used here could be used, but it remains to be determined whether reasonable control over the probability of a Type I error can be achieved.

Regarding the use of a bootstrap method, Hall and Wilson (1991) argue in favor of using a pivotal test statistic, which is not done here. When working with means, more recent results, summarized in Wilcox (2005), also support the conclusion that a pivotal test statistic be used. When working with robust estimators, however, there are general situations where a percentile bootstrap method has a substantial advantage. In addition, when using a percentile bootstrap method, there is no need to approximate the null distribution of some test statistic (Liu & Singh, 1997). Roughly, the percentile bootstrap method is based on determining how deeply the null value is nested within the sampling distribution of some estimator. Finally, R functions for applying the methods considered are available from the

author; download the file Rallfun-v17 from www-rcf.usc.edu~rwilcox. The function sm2strv7 performs method SM, and the function ts2str performs methods BTS.

References

Bickel, P. J., & Lehmann, E. L. (1976). Descriptive statistics for nonparametric models III dispersion. *Annals of Statistics*, *4*, 1139-1158.

Doksum, K. A., & Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *Annals of Statistics*, *23*, 1443-1473.

Dunn, O. J., & Clark, V. A. (1971). Comparison of tests of the equality of dependent correlation coefficients. *Journal of the American Statistical Association*, *66*, 904-908.

Efromovich, S. (1999). *Nonparametric curve estimation: methods, theory and applications*. New York: Springer-Verlag.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap* . New York: Chapman & Hall.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. New York: Marcel Dekker.

Fan, J., & Gijbels, I. (1996). *Local polynomial modeling and its applications*. Boca Raton, FL: CRC Press.

Fox, J. (2001). *Multiple and generalized nonparametric regression*. Thousand Oaks, CA: Sage.

Godfrey, L. G. (2006). Tests for regression models with heteroskedaticity of unknown form. *Computational Statistics & Data Analysis*, *50*, 2715-2733.

Green, P. J., & Silverman, B. W. (1993*). Nonparametric regression and generalized linear models: A roughness penalty approach.* Boca Raton, FL: CRC Press.

Györfi, L., Kohler, M., Krzyzk, A., & Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. New York: Springer-Verlag.

Hall, P., & Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, *47*, 757-762.

Härdle, W. (1990). *Applied nonparametric regression*. Econometric Society Monographs, No. 19, Cambridge, UK: Cambridge University Press.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. New York: Chapman and Hall.

Hendrickson, G. F., Stanley, J. C., & Hills, J. R. (1970). Olkin's new formula for significance levels of r_{13} vs. r_{23} compared with Hotelling's method. *American Educational Research Journal*, *7*, 189-195.

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *Journal of General Psychology*, *130*, 149-168.

Hoaglin, D. C. (1985) Summarizing shape numerically: The g-and-h distributions. In *Exploring data tables, trends, and shapes*, D. Hoaglin, F. Mosteller, & J. Tukey (Eds.), 461-515. New York: Wiley.

Hotelling, H. (1940). The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. *Annals of Statistics*, *11*, 271-283.

Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

Lafferty, J., & Wasserman, L. (2008). RODEO: Sparse, greedy nonparametric regression. *Annals of Statistics*, *3*, 28-63.

Lax, D. A. (1985). Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association*, *80*, 736-741.

Liu, R. G., & Singh, K. (1997). Notions of limiting P values based on data depth and bootstrap. *Journal of the American Statistical Association*, *92*, 266-277.

Maronna, R. A., & Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, *44*, 307-317.

Meng, X.-L., Rosenthal, R. & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172-175.

Olkin, I. (1967). Correlations revisited. In *Improving Experimental Design and Statistical Analysis*, J. C. Stanley (Ed.), 102-128. Chicago: Rand McNally

Racine, J. S., & MacKinnon, J. G. (2007). Inference via kernel smoothing of bootstrap p values. *Computational Statistics & Data Analysis*, *51*, 5940-5957.

Randal, J. A. (2008). A reinvestigation of robust scale estimation in finite samples. *Computational Statistics & Data Analysis*, *52*, 5014-5021.

Rocke, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimensions. *Annals of Statistics*, *24*, 1327-1345.

Sen, P. K. (1968). Estimate of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, *63*, 1379-1389.

Shoemaker, L. H., & Hettmansperger, T. P. (1982) Robust estimates and tests for the one- and two-sample scale models. *Biometrika*, *69*, 47-54.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman and Hall.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245-251.

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

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing*. New York: Academic Press.

Wilcox, R. R. (2009). Comparing Pearson correlations: Dealing with heteroscedasticity and nonnormality. *Communications in Statistics-Simulations and Computations*, *38*, 2220-2234.

Wilcox, R. R. (in press). Measuring and detection associations: Methods based on robust regression estimators or smoothers that allow curvature. *British Journal Mathematical and Statistical Psychology*.

Wilcox, R. R., & Tian, T. (2008). Comparing dependent correlations. *Journal of General Psychology*, *135*, 105-112.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society*, *Series B*, *21*, 396-399.

Yohai, V. J., & Zamar, R. H. (1988). High breakdown point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, *83*, 406-414.

Yu, M. C., & Dunn, O. J. (1982). Robust tests for the equality of two correlation coefficients: A Monte Carlo study. *Educational and Psychological Measurement*, *42*, 987-1004.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*, 399-413.