Comparing the Performance of Non-monotone Dependence Measures in Psychology

A simulation study on the power of mutual information and distance correlation

Sarah M.Y. Makonnen

Abstract

The standard methods for quantifying the dependence between two continuous variables in psychological research are the Pearson correlation and the rank correlation coefficients Spearman's ρ and Kendall's τ. A severe limitation of these correlation measures is that they fail to capture non-monotone bivariate relationships that are ubiquitous in psychology. Two well-established and promising non-monotone dependence measures used in other fields are mutual information (MI) and distance correlation (Dcor). Two questions that remain unexamined are how well these measures can detect dependencies typically found in psychological research and how they compare to conventional correlation measures. Thus, in this paper, we compared the statistical power of MI and Dcor to the traditional correlation measures via a simulation study. Our results suggest that Dcor can detect any functional relationship in principle, as long as sample sizes are large enough. The MI fails to detect the linear relationship when the data is noisier, but it outperforms all other methods on the sine wave relationship for small sample sizes. Consequently, we recommend using Dcor as the standard measure for detecting any relationship in psychological research when sample sizes are large enough. For small and noisy datasets, and when there is a strong belief that the underlying relationship is periodic, MI should be used instead.

Keywords: mutual information, distance correlation, non-monotone dependence, statistical power

Table of Content

Introduction

In psychological research, the dependence structure between two variables is usually described by the product-moment correlation coefficient, also known as Pearson's r, which measures the degree of linear association between two variables (Colman, 2015; Pearson, 1895). Other commonly used correlation measures include Spearman's ρ and Kendall's τ, which are both nonparametric measures of rank correlation and assess how well the relationship between two variables can be described using a monotonic function (Kendall, 1938; Spearman, 1987).

At the same time, it has long been recognized that functional relationships in psychology are not limited to linear or monotone associations. Different functional forms have been studied both theoretically and empirically in many subfields, including psychiatry, cognitive, educational, as well as social and organizational psychology (Bohon, Hembacher, Moller, Moody, & Feusner, 2012; Grijalva, Harms, Newman, Gaddis, & Fraley, 2015; Klassen & Chiu, 2010). Examples are the Yerkes-Dodson inverted-U model (Yerkes & Dodson, 1908) or the more recently proposed "too much of a good thing" effect (Pierce & Aguinis, 2013). Both suggest an inverted U-shape relationship, where monotonic positive relations between two variables reach a maximum after which the relation either turns negative or levels off at a plateau. In longitudinal data analysis, logistic and exponential growth models have also been proposed to study non-linear within-person trajectories over time (Grimm, Zhang, Hamagami, & Mazzocco, 2013). Yet others have used oscillations to describe neuropsychological processes, such as memory and perception (Gray, König, Engel, & Singer, 1989; Rutishauser, Ross, Mamelak, & Schuman, 2010). However, such nonlinear or non-monotone functional relationships might often remain undetected by the traditional correlation measures used in psychological research. As a consequence, the true, more complex dependencies that can occur between variables may be underestimated.

Numerous statistical methods have been developed to study nonlinear and non-monotone psychological phenomena. Early suggestions included simple extensions of the linear regression model, such as polynomial regression and step functions (Fox, 2008). Later, more complex approaches like splines, local regression and generalized additive models were proposed (James, Witten, Hastie, & Tibshirani, 2000). Researchers have modelled nonlinearities in time-series data with nonlinear multilevel models, and in psychometrics, they have started to explore nonlinear effects in single- and multilevel structural equation models (MacCallum, Kim, Malarkey, & KiecoltGlaser, 1997; Schermelleh-Engel, Kerwer, &

Klein, 2014). Moreover, the emergence of nonlinear dynamics and chaos theory in the 1980s and 1990s led to the application of nonlinear differential equations in psychology (Guastello, 2001; Heath, 2000).

However, many of the methods mentioned above can model only specific types of nonlinear and non-monotone relationships and assume that one knows the particular form one wishes to model. Non-monotone dependence measures, on the other hand, can detect any non-monotone relation. Moreover, if one is less interested in knowing the exact relationships that exist between variables but rather in demonstrating the existence of dependence between two variables, dependence measures may be more suitable than intricate statistical models.

A well-established non-monotone dependence measure in probability theory and information theory is mutual information (MI), which reflects the mutual dependence between two random variables (MacKay, 2003). The crucial property of MI for this study is that in contrast to the conventional correlation measures, MI can capture any type of relation - linear, monotone and periodic dependence - and when the data are indeed linearly related, MI can be reduced to the Pearson correlation coefficient (Kinney & Atwal, 2014).

Another, relatively new yet promising and well cited, approach to detecting nonlinear or non-monotone dependence structures is the distance correlation (Dcor). It generalizes the idea of correlation in two ways (Székely, Rizzo, & Bakirov, 2007). Unlike the Pearson correlation coefficient, which measures linear dependence, it captures the degree of dependence of two random variables and it can be computed for two variables in arbitrary dimensions.

Both MI and Dcor have been investigated empirically by several authors and shown to be powerful methods compared to traditional measures of dependence. For example, a simulation study by Simon and Tibshirani (2011) compared the power of the Maximal Information Coefficient (MIC), as a variant of the mutual information proposed by Reshef et al. (2011), to that of the distance correlation and Pearson correlation. Distance correlation performed better than MIC in almost every situation. Only in the case of the high-frequency sine wave relationship, the MIC had higher power, and in the linear case, it performed worse than the Pearson correlation. The authors concluded that the distance correlation measure was the most powerful technique overall and should be considered for general use. In the context of gene expression signal analysis, a simulation study comparing Dcor and MI to several other methods, including the Heller-Heller-Gorfin measure (HHG), found that the HHG measure and Hoeffding's D had the highest power on non-linear, and non-monotone relationships (de Siqueira Santos, Takahashi, Nakata, & Fujita, 2014). The authors

recommended that the HHG measure and Hoeffding's D are suitable for general use and sample sizes with fewer than 30 observations, Dcor, MI, and MIC could be powerful as well. Another simulation study by Clark (2013) compared the three correlation measures to the distance correlation and MIC as well. It showed that the Pearson correlation was good at detecting linear relationships, while distance correlation was best at finding quadratic relationships, also when the data was noisy. The MIC found strong sine wave patterns for a wide range of noise levels. More recently, the MIC, MI, distance correlation, and linear correlation were compared to the copula correlation, as one of the newer methods for measuring non-monotone dependence (Ding & Li, 2013). No single test dominated in power in all cases. The linear correlation was best at detecting linear relationships. Concerning the sine wave pattern, MIC was very good at detecting higher frequencies, whereas the copula correlation was best at detecting lower frequencies. In a later study, Kinney and Atwal (2014) compared the MI, MIC, distance correlation and Hoeffding's D – a more traditional nonlinear dependence measure – to the Pearson correlation. Similar to the previous findings, MIC had low statistical power in almost all cases except on the sine wave relationship, and Pearson correlation, distance correlation and Hoeffding's D performed considerably better than the MI and MIC on the linear relationship.

Most of these studies were conducted in the field of molecular biology and used larger sample sizes of 300, 500, 1000 or 10000 that are common for gene-expression data, for example. However, none of these previous studies has investigated the performance of MI and Dcor on data that is typical of psychological research where usually, the number of observations is small, and relationships are noisy. Also, not all of the studies accounted for the noise level when examining the performance of the relevant measures. Thus, this study will shed light on two questions that remain unanswered:

1. How well can MI and Dcor capture dependencies typically found in psychological research? and

2. How does each of them compare to traditional correlation measures?

The results from this study could provide valuable insight into alternative dependence measures that could be used in psychology to capture relationships that might otherwise be missed by the commonly used Pearson or rank correlation coefficients.

To compare the dependence measures directly, a standard metric is necessary since each measure is measured on a different scale. Therefore, we will examine the statistical power of each method for different functional relationships and under different scenarios

typically encountered in psychological settings via a simulation study. We will also use standard correlation measures and Hoeffding's independence test, a well-established nonparametric dependence test, as a benchmark. Based on the previous findings, we expect MI to be less powerful than Dcor and the traditional correlation measures at detecting linear relationships, and more powerful at detecting the sine wave relationship.

The remainder of this thesis is organized as follows: in the next section, we will describe the dependence measures used in this study, as well as the simulation study in detail. In section 3, the results of the simulation study are presented, followed by a discussion and conclusion in sections 4 and 5.

Method

Non-monotone Dependence Measures

In what follows, we describe the nonlinear dependence measures and their corresponding estimation procedures. For simplicity, only continuous variables and bivariate relationships are considered, but all measures presented here can be computed for discrete variables and the multivariate case as well.

Mutual information, distance correlation, and Hoeffding's D are based on different but equivalent mathematical characterizations of the statistical independence between two random variables X and Y with a similar form:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \tag{1}$$

for all x, y, where $f_{X,Y}$ can be either the joint probability density function $p_{X,Y}$, the joint characteristic function $\varphi_{X,Y}$, or the joint cumulative distribution function (CDF) $F_{X,Y}$. Similarly, f_X and f_Y can be either the corresponding marginal probability density functions (PDF) p_X and p_Y, the characteristic functions φ_X and φ_Y, or the CDFs F_X and F_Y. Thus, to establish the dependence between X and Y, all measures quantify the discrepancy or distance between some version of each one of the terms on the left- and right-hand side of Equation (1).

Rényi's axioms. Before turning to the description of each of the three nonparametric dependence measures, here we outline several conditions that any suitable dependence measure should satisfy. One of the most widely used criteria to evaluate the properties of dependence measures are the seven axioms proposed by Rényi (1959):

A. $\lambda(X,Y)$ is defined for any pair of random variables X and Y, where neither one of the random variables is constant for all values;

B. $\lambda(X,Y) = \lambda(Y,X)$;

C. $0 \leq \lambda(X,Y) \leq 1$;

D. $\lambda(X,Y) = 0$ if and only if X and Y are independent;

E. $\lambda(X,Y) = 1$ if there is a strict dependency between X and Y, i.e. either $Y = f(X)$ or $X = g(Y)$ for some functions f, g;

F. It is invariant under marginal one-to-one transformations of the random variables;

G. If the joint distribution of X and Y is normal, then $\lambda(X,Y) = \rho(X,Y)$, where $\rho(X,Y)$ is the Pearson correlation coefficient of X and Y.

In the next two sections, we evaluate to what extent MI and Dcor fulfill these axioms.

Mutual information. The first measure of interest to consider is the mutual information (MI), which reflects how similar the joint *probability density* function of two random variables X and Y is to the product of their marginal PDFs. To illustrate this, consider two single events x and y of the random variables X and Y. In order to determine whether x and y are independent, one can compute the so-called *pointwise mutual information (PMI)*, which is defined as:

$$PMI(x;y) = log\left(\frac{p(x,y)}{p(x)p(y)}\right), \tag{2}$$

where $p(x,y)$ denotes the joint PDF, and $p(x)$ and $p(y)$ are the marginal PDFs. Thus, if x and y are independent, then $p(x,y) = p(x)p(y)$ and $PMI(x;y) = \log(1) = 0$. Mutual information is then defined as the expected value of *PMI* over all possible events $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ of the random variables X and Y (Kraskov, Stögbauer, & Grassberger, 2004):

$$I(X;Y) = E\left(log\left(\frac{p(x,y)}{p(x)p(y)}\right)\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\, log\left(\frac{p(x,y)}{p(x)p(y)}\right) dx\, dy. \tag{3}$$

The base of the logarithm determines the units in which information is measured. Taking base 2, in particular, leads to information measured in bits. MI is nonnegative and symmetric, i.e. $I(X;Y) = I(Y;X)$. Moreover, when the random variables X and Y are independent, their joint PDF will be the product of their marginal PDFs and $log \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}$ in Equation (3)

becomes $\log(1) = 0$. Thus, in the case of independence, $I(X;Y) = 0$. Conversely, if X can be completely determined by Y, then $\log\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}$ will be greater than zero and may become very large. Consequently, $I(X;Y)$ lies in the interval $[0; +\infty)$. Because this unbounded range of values makes MI hard to interpret and difficult to compare to other dependence measures, many normalized variants have been suggested (Cahill, 2010). One of the normalized variants of MI for continuous variables is the *informational coefficient of correlation*, which is defined as

$$\lambda\widehat{(X,Y)} = \sqrt{1 - \exp\left(-2I\widehat{(X;Y)}\right)}. \tag{4}$$

Note that the unstandardized mutual information $I(X;Y)$ from Equation (3) satisfies all of Rényi's axioms except C, E and G. That is, it is not bounded from above, it is not equal to 1 when there is a perfect relationship between random variables X and Y, and it cannot be reduced to the Pearson correlation coefficient. In contrast to $I(X;Y)$, the informational coefficient of correlation $\lambda\widehat{(X,Y)}$ satisfies all seven postulates for a suitable measure of dependence and will be used as a normalized MI measure in this study.

Kernel density estimation. As mentioned before, to compute the MI, one needs to estimate the probability density functions in Equation (3). Because these are usually not known, they have to be estimated from the data. Several MI estimation approaches have been proposed over the years, such as histogram-based techniques (Moon, Rajagopalan, & Lall, 1995; Steuer, Kurths, Daub, Weise, & Selbig, 2002), *k*-nearest neighbor estimation (KNN) (Kraskov et al., 2004), Edgeworth approximation of differential entropy (Hulle, 2005), and Bayesian estimation approaches (Gencaga, Malakar, & Lary, 2014).

In the present study, kernel density estimation (KDE) will be used as it has been shown to outperform other mutual information estimation approaches for small sample sizes (Khan et al., 2007). The idea of KDE is that a non-negative function, also known as a *kernel*, is placed on each observation x_i of the random variable X and these kernels are then summed to obtain the kernel density estimator. The kernel function can take on many forms, e.g. normal or uniform. For example, KDE with a Gaussian kernel may be pictured as putting Gaussian "bumps" on each observation x_i, and the kernel density estimator is obtained by taking the sum of these "bumps". In that sense, KDE is similar to histograms. However, KDE estimates have been shown to be superior to the histogram-based techniques in terms of (a) a better mean square error rate of convergence of the estimate to the underlying density, (b) an

insensitivity to the choice of origin, and (c) the ability to specify more sophisticated window shapes than the rectangular window shapes for frequency counting (Moon et al., 1995; Steuer et al., 2002). Besides, KDE has been found to perform better than KNN when the sample size is small, and noise levels are high (Khan et al., 2007). Since we typically deal with small sample sizes and high noise levels in psychological research, KDE is an adequate choice for this study.

Assume a sample $x_1, \dots, x_n$ of independently identically distributed observations from a random variable X with density f. The kernel density estimator for the marginal probability density of X with a generalized weight or kernel function $K(x)$ is given by

$$\hat{f}(x) = \frac{1}{nh}, \tag{5}$$

where h is a so-called bandwidth parameter that determines the "bumpiness" or smoothness of the probability density function. Hence, two decisions need to be made when using KDE. First, choosing a kernel function K and, second, the bandwidth h. Both decisions are commonly made based on minimizing the discrepancy of the density estimator $\hat{f}$ from the true density f (Silverman, 1986). The mean integrated squared error (MISE) is one of the most widely used measures for this discrepancy. It has been shown that, if the bandwidth parameter h is chosen optimally, the most efficient kernel in the MISE sense will be the Epanechnikov kernel, which can be written as:

$$K_e(x) = \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}x^2\right). \tag{6}$$

Note, however, that most other commonly used kernels (e.g. gaussian, triangular, rectangular) are nearly as efficient, which is why the particular choice of a kernel has been deemed rather unimportant in KDE (Harpole, Woods, Rodebaugh, Levinson, & Lenze, 2014; Silverman, 1986). Nonetheless, we will use the Epanechnikov kernel, also because it is already efficiently implemented as the standard kernel in the R software package 'mpmi' that computes MI with KDE.

To choose the optimal bandwidth, we employed the following rule of thumb by Silverman (1986):

$$h = 0.9An^{-1/5} \tag{7}$$

where $A = min\ (standard\ deviation, interquartile\ range/1.34)$, representing an adaptive estimate for the spread of the data. The value given by Equation (7) has been shown to recover a wide range of densities reasonably well with respect to MISE, also for the Epanechnikov kernel and especially for smaller sample sizes in the range of 50 to 100, and it is easy to evaluate (Harpole et al., 2014; Silverman, 1986).

Distance Correlation. As mentioned earlier, all non-monotone dependence measures in this study quantify the dissimilarity of the joint distribution and the product of the marginal distributions of two random variables X and Y. While MI captures the discrepancy directly in terms of the probability density functions, distance correlation transforms the joint and marginal PDFs first via so-called *characteristic functions* and then computes the distance between these transformed quantities.

The characteristic function, (denoted as φ in the following) of a random variable is a function, which takes on values in the complex numbers (DasGupta, 2011; Karr, 1993). Complex numbers can be written in the form $z = x + iy$, where x and y are real numbers, and i is the imaginary number $\sqrt{-1}$. More precisely, a characteristic function of a random variable X is defined as the expected value of e^{itX} with i as the imaginary part and t as the real part of the complex-valued function:

$$\varphi_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} p_X(x)dx, \tag{8}$$

where p_X denotes the probability density function of X. Characteristic functions have several advantages. First, they are defined for all real-valued random variables and can always be computed. Second, they uniquely identify the underlying probability distribution function, which means that the distribution function of X can be determined when one knows its characteristic function. Furthermore, in many cases, it is easier to make computations with characteristic functions of random variables than estimating their probability density functions. It can be shown that, if X and Y are independent, then the joint characteristic function will be equal to the product of the marginal characteristic functions: $\varphi_{X,Y} = \varphi_X \varphi_Y$ (Karr, 1993).

Thus, distance correlation, as our second dependence measure of interest, is the standardized *distance covariance*. The distance covariance computes the distance between the joint characteristic function $\varphi_{X,Y}(s,t)$ and the product of the marginal characteristic functions $\varphi_X(s)\varphi_Y(t)$ of two random variables X and Y:

$$dCov(X,Y) = \frac{1}{c^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{|\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2}{|s|^2|t|^2} dt ds \tag{9}$$

where s and t are the real parameters of the characteristic functions and c is a constant (Székely et al., 2007). The distance correlation is then defined as the standardized value of $dCov(X,Y)$:

$$dCor(X,Y) = \begin{cases} \dfrac{dCov(X,Y)}{dVar(X)dVar(Y)}, & dVar(X)dVar(Y) > 0 \\ 0, & dVar(X)dVar(Y) = 0. \end{cases} \tag{10}$$

where $dVar(X) = dCov(X,X)$ and $dVar(Y) = dCov(Y,Y)$ are the distance variances of random variables X and Y, respectively.

With respect to Rényi's postulates: $dCor(X,Y) = dCor(Y,X)$, i.e. the distance correlation is symmetric, $0 \leq dCor(X,Y) \leq 1$, $dCor(X,Y) = 0$ if and only if X and Y are independent, and $dCor(X,Y) = 1$ when there is a vector a, a nonzero real number b, and an orthogonal matrix C such that $Y = a + bXC$.

The estimation of the theoretical distance covariance via the computation of distance matrices is relatively straightforward. For details, the interested reader is referred to Székely et al. (2007). We used the 'energy' package in R to compute the distance correlation.

Hoeffding's D. The mutual information and distance correlation will be compared against Hoeffding's D, which is one of the earliest methods to detect non-monotone dependences, as a benchmark. Hoeffding's D measures the distance between the joint CDF $F_{X,Y}$ and the product of the marginal distributions $F_X F_Y$:

$$H(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y)(F_{X,Y}(x,y) - F_X(x)F_Y(y))^2 dx dy \tag{11}$$

for all x, y (Hoeffding, 1948). Similar to Spearman's ρ and Kendall's τ, the empirical estimate $\widehat{H(x,y)}$ is a rank-based measure. It considers the difference between the joint ranks of the observed data for the random variables X and Y and the product of their marginal ranks. However, unlike the monotone dependence measures, Hoeffding's D can also identify non-monotonic relationships. It can be shown that the random variables X and Y are independent if and only if $D(x,y) = 0$ (Hoeffding, 1948). We will use the test statistic $\widehat{D}$

from the 'Hmisc' R package developed by Harrell and Dupont (2006), which returns a value for $\widehat{D}$ that is 30 times larger than the original $\widehat{D}$ described in Hoeffding (1948). This adapted version of $\widehat{D}$ ranges between -0.5 and 1, where larger values indicate a stronger relationship between the variables. For more details on the computation of $\widehat{D}$, the interested reader is referred to Hoeffding (1948) as well as Harrell and Dupont (2006) and the references therein.

Simulation Study

In this section, a series of Monte Carlo simulations is presented in which the performance of MI and distance correlation in capturing dependencies typically found in psychological research was evaluated. We compared the performance of these two methods with that of another, more traditional non-monotone dependence measure and the three conventional correlation measures: (i) Hoeffding's D, (ii) the Pearson correlation coefficient, (iii) Spearman's rank correlation coefficient, and (iv) Kendall's τ (Kendall, 1938; Pearson, 1895; Spearman, 1987). We varied the factors, sample size, type of functional relationship, and noise level. To be able to compare all measures directly, a standard metric was necessary since each measure has a different scale. Therefore, the statistical power of each method was assessed for different functional relationships under different scenarios that are typically encountered in psychological research.

Sample size. The sample size was varied according to those typically found in four leading APA journals from 1995-2006 (Marszalek, Barber, Kohlhart, & Cooper, 2011): 10, 20, 30, 40, 50, 60, 100 and 150. This range of values contains more than 75% of sample sizes reported in psychological journals within that period, and the sample sizes below 60 were covered more densely because they occurred more often and included more than 50% of the sample sizes reported in psychological journals. Initially, we also wanted to investigate a sample size of approximately 10,000 observations, which occurred quite often in the Journal of Applied Psychology. However, this would have taken much computational time, so we decided to exclude this extreme case.

Type of relationship. We simulated pairs of variables that were related linearly, quadratically, exponentially or periodically via a sine wave. For the quadratic case, we chose an inverted U-shape relationship, as this particular quadratic shape has been suggested in several psychological theories (Pierce & Aguinis, 2013; Yerkes & Dodson, 1908). Similarly, we modelled negative exponential growth based on theoretical and empirical examinations of within-person trajectories over time as in longitudinal psychological research. Thus, for each

cell of the design, a random variable $X \sim \mathcal{N}(0,1)$ was generated, based on which several Y variables were created with the following functional forms:

Linear: $\qquad\qquad Y_1 = 4 + 0.7 \times X + \varepsilon$ (12)

Quadratic: $\qquad\qquad Y_2 = -0.5 \times X^2 + \varepsilon$ (13)

Exponential: $\qquad\qquad Y_3 = -e^{-X} + \varepsilon$ (14)

Periodic: $\qquad\qquad Y_4 = \sin(4 \times X) + \varepsilon$ (15)

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ denotes random normally distributed error with error variance σ^2 dependent on the noise level, which will be described in the next subsection. Figure 1 depicts the different functional forms graphically for $\varepsilon \sim \mathcal{N}(0,0)$.
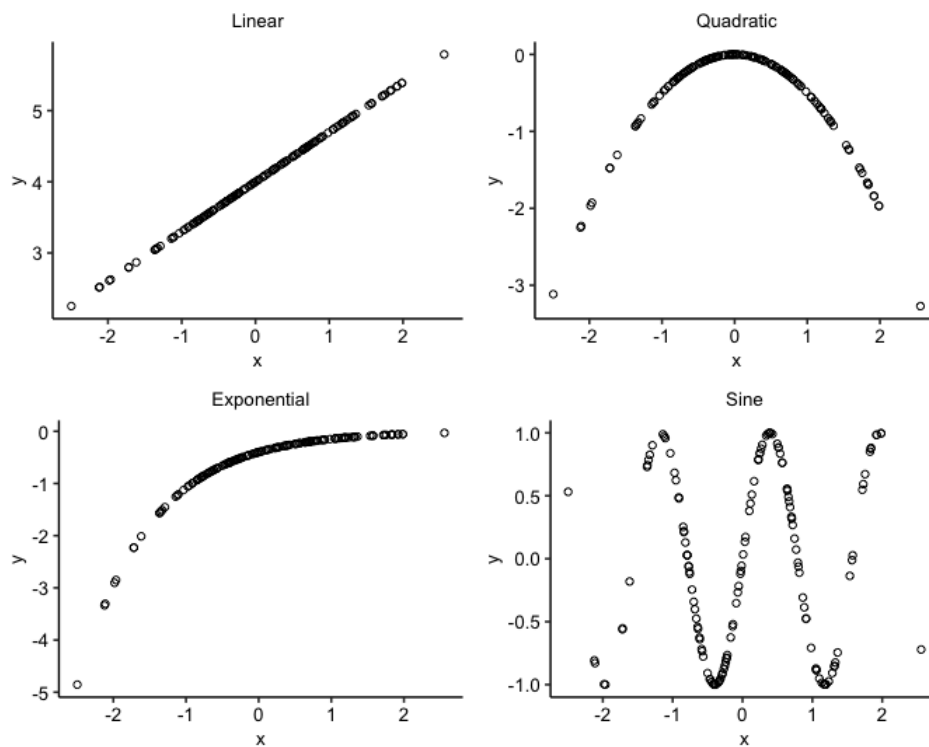
Figure 1. The four different functional relationships specified in Equations (12) – (15).

Noise level. As described in the previous section, varying levels of noise were added to each equation that specified the type of functional relationship in Equations (12) – (15). This was done to explore the influence of the strength of the relationships on the power of each measure. The aim was to cover the range of possible values in a relatively dense fashion, similar to Simon and Tibshirani (2011) who chose 30 different noise levels between

0 and 1. However, due to a trade-off between a dense coverage and computational burden, we chose the 10-fold range of values, 0, 0.1, …, 0.9.

Design and replications. Thus, the following data generating parameters were manipulated in a $4 \times 10 \times 8$ full factorial design with 320 design cells in total:

1. Functional form: linear, quadratic, exponential, and periodic
2. Noise level: 0, 0.1, 0.2, 0.3, 0.4, 0.5 0.6, 0.7, 0.8, 0.9
3. Sample size N: 10, 20, 30, 40, 50, 60, 100, and 150.

For each design cell, the data generating process and estimations for the dependence measures were repeated 1000 times.

Evaluating Performance of Methods

Permutation test. We used Monte-Carlo permutation tests as a nonparametric approach to hypothesis testing. In a regular permutation test, the distribution of the test statistic of interest under the null hypothesis that the random variables X and Y are independent is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points. To illustrate this, suppose we have paired data $(X_i, Y_i) \sim F$ for $i = 1, …, n$, where F is some unknown bivariate distribution. Because we have labels $i = 1, …, n$, there are $n!$ different possible orderings of y_i. The basic idea of the permutation test is that, if the null hypothesis of independence is correct, then reordering of y_i will not affect the value of our dependence measure. The p-value is then computed as the relative number of these hypothetical test statistics under the null hypothesis that exceed the observed value of the test statistic.

A common problem is that the number of possible rearrangements $n!$ is too large and computing the corresponding test statistics becomes infeasible within reasonable computational constraints. This problem can be solved by using a Monte Carlo approach, where the permutation-based p-value is derived based on a random sample B of all possible permutations of y_i. Here we chose $B=1000$ as a commonly used number of permutations that delivers precise enough estimates. Due to computational limitations, the permutation test was used to test the power of MI and Dcor only. For Hoeffding's independence test, and the three tests based on Pearson's r, Spearman's ρ, and Kendall's τ, we used the hypothesis tests that are already implemented efficiently in the 'Hmisc' and standard 'base' R packages.

Empirical Power. In order to compare the different dependence measures of interest, we conducted hypothesis tests with each measure as the corresponding test statistic and then computed the statistical power across all simulations as a standard metric. The statistical

power of a hypothesis test is defined as the probability that the test rejects the null hypothesis (H_0: X and Y are independent) when the alternative hypothesis (H_1: X and Y are not independent) is correct. We estimated power as the average number of p-values that exceeded the significance level $\alpha = .05$ across replications n_{sim}:

$$\widehat{power} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} 1(p_i \geq \alpha), \tag{16}$$

where 1 is the indicator function that takes on the value of 1 if the p-value p_i is larger than or equal to the significance level α and the value of 0 if p_i is smaller than α.

Exact test. Apart from having high power, a good test of a given size α should also maintain a low type I error rate that is equal or smaller to the nominal alpha level ($\leq \alpha$). Such a test is called an exact test. Thus, to assess whether the tests included in the simulation were exact, we also computed the power of each method for the case in which there was no relationship between X and Y. The calculated powers, in this case, are the empirical type I error rates that reflect how often the null hypothesis was rejected although it was true. Due to computational limitations, the type I error rates were computed for only the two smallest reported sample sizes in psychological research of $n = 10$ and $n = 20$.

Unbiased test. We also assessed whether the dependence measure tests were unbiased. An unbiased test is a statistical test which is characterized by the fact that the probability of a type I error does not exceed a given significance level α but the power, i.e. the probability that the test rejects under the alternative hypothesis, is always higher than the type I error. Unbiasedness was assessed for each dependence measure-based test included in the simulation.

It should be noted here that there are also more strict evaluation criteria for hypothesis tests, with the strictest being whether a test is uniformly most powerful (UMP). A UMP test is a statistical test at significance level α whose power is not less than that of any other statistical test of the same significance level (Lehmann, 2005; Upton & Cook, 2014). However, UMP tests exist only under very restricted conditions. Therefore, we do not expect any of the dependence measure-based tests included in this study to be UMP. Furthermore, while it can be checked mathematically whether a given test is UMP, it is empirically and practically impossible as one would have to compare the type I error rates and power level of a given test against all other statistical tests of the same significance level that exist in theory.

Next to UMP tests, there are also uniformly most powerful (UMP) unbiased tests. A UMP unbiased test is uniformly most powerful amongst all unbiased tests of the same significance level. For the same reasons outlined above, we do not expect any of the hypothesis tests included in our study to be UMP unbiased.

Results

In this section, the results of the simulation study are presented for each data generating mechanism, the underlying type of relationship between variables X and Y. To illustrate the general trend of each measure's performance, we provide the results for only a few sample sizes. For the complete results, we refer the interested reader to Figures A1-A4 and Tables A1-A4 in the Appendix.

Concerning the question of how well MI and Dcor can capture the different types of dependencies that are typically found in psychological research, we evaluated whether they can detect them *in principle*. That is, we assessed whether the methods reached a power of 1 in the limit, at the largest sample size. Moreover, to see how each of them compared to the traditional correlation measures, we checked whether there was a stable ranking of methods in terms of power across the different noise levels.

Exactness

The simulated powers for the case, in which X and Y are independent, are the type-I error rates presented in Table 1. All empirical type I error rates were significantly close to the nominal type I error rate of $\alpha = .05$. That is, all binomial tests testing the null hypothesis that the type I error rate was greater or equal to the nominal level α were significant at the 1%-level. The rates were estimated using 1000 simulations for the two smallest sample sizes of n = 10 and $n = 20$. Due to computational limitations, we ran the simulations for these two sample sizes only.

Table 1

Empirical type I error rate for each dependence measure-based hypothesis test

	MI	Dcor	Hoeffding	Pearson	Spearman	Kendall
$N = 10$	.049***	.040***	.052***	.047***	.044***	.043***
$N = 20$	.057***	.048***	.050***	.051***	.050***	.051***

Notes. MI denotes the mutual information, Dcor the distance correlation, Hoeffding the Hoeffding's D, Pearson the Pearson's r, Spearman the Spearman's ρ, and Kendall the Kendall's τ based hypothesis test, respectively.

Type I error was estimated via 1000 simulations with $\alpha = .05$.

*** $p < .001$

Empirical Power by Type of Relationship

Linear. Figure 2 shows the power of each of the six dependence measures as a function of the noise level and two selected sample sizes for the scenario, in which x and y were related linearly. As can be seen in the right plot, all methods except for MI seemed to be able to detect the linear relationship in principle. That is, Dcor, Hoeffding's D, Pearson's r, Spearman's ρ, and Kendall's τ all reached a power of 1 at the largest sample size of $n = 150$ for all noise levels. Conversely, the power of MI deteriorated as noise increased. This trend was almost the same for all other sample sizes of $n > 10$ (see Figure A1, Appendix).

Figure 2. Linear Relationship: Power of distance correlation (Dcor), Mutual Information (MI), Hoeffding's *D* (Hoeffd), Pearson's r (Pearson), Spearman's ρ (Spearman), and Kendall's τ (Kendall) as a function of the level of noise added, for two selected sample sizes. Power is estimated via 1000 simulations with $\alpha = .05$.

For the smallest sample size of $n = 10$, there was a stable ranking of methods across the different noise levels, as can be seen in the plot on the left of Figure 2: Pearson > Dcor > Spearman > Kendall > Hoeffding > MI. Pearson correlation and distance correlation had the highest power with values ranging between .70 (high noise) and 1 (no noise), followed by the

rank correlations and MI. The MI had the lowest power for all noise levels greater than 0, with values ranging between .45 (high noise) and .90 (low noise) but it performed equally well compared to all other measures when there was no noise at all (=1). Thus, the biggest difference in power was observed between the MI and Pearson correlation, which was approximately =.30 at moderate noise levels.

   Quadratic. The cases in which x and y were related via an inverted U-shape are depicted in Figure 3. In principle, both MI and Dcor, as well as Hoeffding's D, were able to detect the quadratic relationship. As expected, the three conventional correlation measures (Pearson, Spearman, and Kendall) were not able to detect the underlying non-monotone dependence in any case. That is, even for the largest sample size of $n = 150$ and a noise level of zero the power of any of those methods was not higher than .28.

Figure 3. Quadratic Relationship: Power of distance correlation (Dcor), Mutual Information (MI), Hoeffding's D (Hoeffd), Pearson's r (Pearson), Spearman's ρ (Spearman), and Kendall's τ (Kendall) as a function of the level of noise added, for two selected sample sizes. Power is estimated via 1000 simulations with $\alpha = .05$.

Among the three non-monotone dependence measures, there was a stable ranking across noise levels and sample sizes. Distance correlation performed best at all sample sizes, but the superiority was most evident at $n = 20, 30, 40$ and 50 where the difference in power compared to Hoeffding's D, as the second-best performing measure, was close to .40. MI performed considerably worse with power values ranging only between .24 and .35.

Although the power of Dcor was higher compared to all other measures at the smallest sample size of $n = 10$, it was only moderate in absolute terms, and it even fell below .50 as noise increased. For sample sizes larger than 20, the power of Dcor ranged between .63 (high noise) and 1 (no noise) and at $n = 50$, it started to stabilize at values between .90 (high noise) and 1 (no noise). For the largest sample size of $n = 150$, Hoeffding's D and MI performed almost as good as Dcor with power values ranging between .95 (high noise) and 1 (no noise).

Exponential. As shown in the lower right plot of Figure 4, all six methods included in the simulation seemed to able to detect the exponential relationship in principle. That is, all methods had power of 1 in the limit. Moreover, similar to the linear case, there was a consistent ranking of the methods across the different noise levels: Pearson > Dcor > Spearman > Kendall > Hoeffding > MI, with only slight differences between Spearman and Kendall.

At the smallest sample size of $n = 10$ and when there was no noise, the exponential relationship was detected by all measures except for MI. That is, all measures had power of 1, except for the MI which had a power of only .54. As the noise level increased, Pearson's r and Dcor had higher power than the other measures, albeit to a moderate degree.

Except for the MI, all measures were able to detect the exponential relationship at medium to large sample sizes with power values equal or close to 1. MI performed worse at $n = 30$ and $n = 50$ with increasing noise levels and approached the other measures' power of 1 only at the largest sample size of $n = 150$.

As expected, the two rank correlation coefficients were able to capture the non-monotonicity of the exponential relationship. However, more surprising is the fact that the Pearson correlation, which is only supposed to detect linearity in theory, was as powerful, if not more, than Spearman and Kendall. A possible explanation is that the Pearson correlation captured the linear part of the exponential graph, as depicted in Figure 5. The dotted blue line represents the line of best linear fit, where the slope of the line can be interpreted as the Pearson correlation coefficient. As one can see, the fitted line is a reasonably good

approximation of the underlying relationship in this case, which could explain the relatively high power of the linear correlation in the case of the exponential relationship.

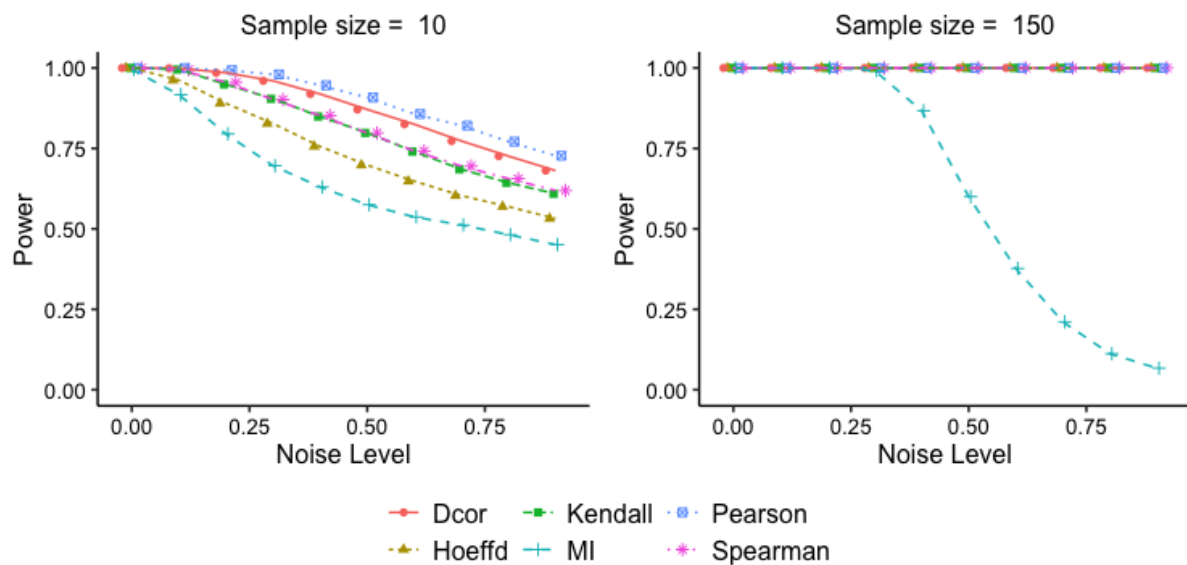*Figure 4.* Exponential Relationship: Power of distance correlation (Dcor), Mutual Information (MI), Hoeffding's D (Hoeffd), Pearson's r (Pearson), Spearman's ρ (Spearman), and Kendall's τ (Kendall) as a function of the level of noise added, for two selected sample sizes. Power is estimated via 1000 simulations with $\alpha = .05$.
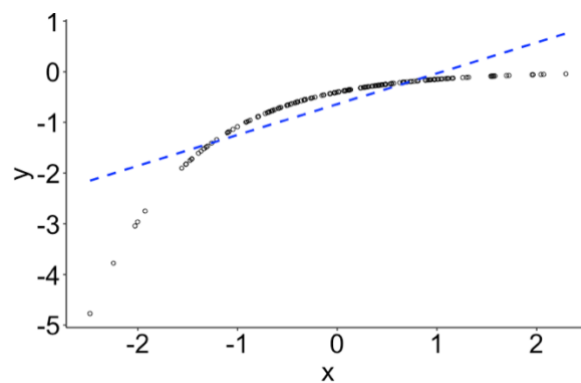
Figure 5. The exponential relationship between *x* and *y.* Line of best linear fit represented as dotted blue line.

Sine wave. Figure 6 presents the results for the sine wave relationship. As expected, only the three non-monotone dependence measures were able to detect the periodic relation in principle. That is, MI, Dcor, and Hoeffding's D all reached a power of 1 at the largest sample size (see lower right plot).

Similar to the quadratic case, there was a stable ranking across noise levels and sample sizes among these three nonmonotone methods of the following: MI > Hoeffding > Dcor. MI was most powerful with values ranging between .70 (high noise) and 1 (no noise) at moderate and higher sample sizes. Hoeffding's D and Dcor, as the second and third best performing methods, reached the power level of MI only at the higher sample sizes n = 100, and 150. Note that at the smallest sample size, none of the three measures were very powerful. That is, none of the methods reached a power higher than .20 at any noise level.

As expected, none of the conventional correlation measures (Pearson, Spearman, Kendall) had power higher than .08. Thus, none of these methods were able to identify the nonmonotonicity of the sine wave irrespective of sample size or noise level.
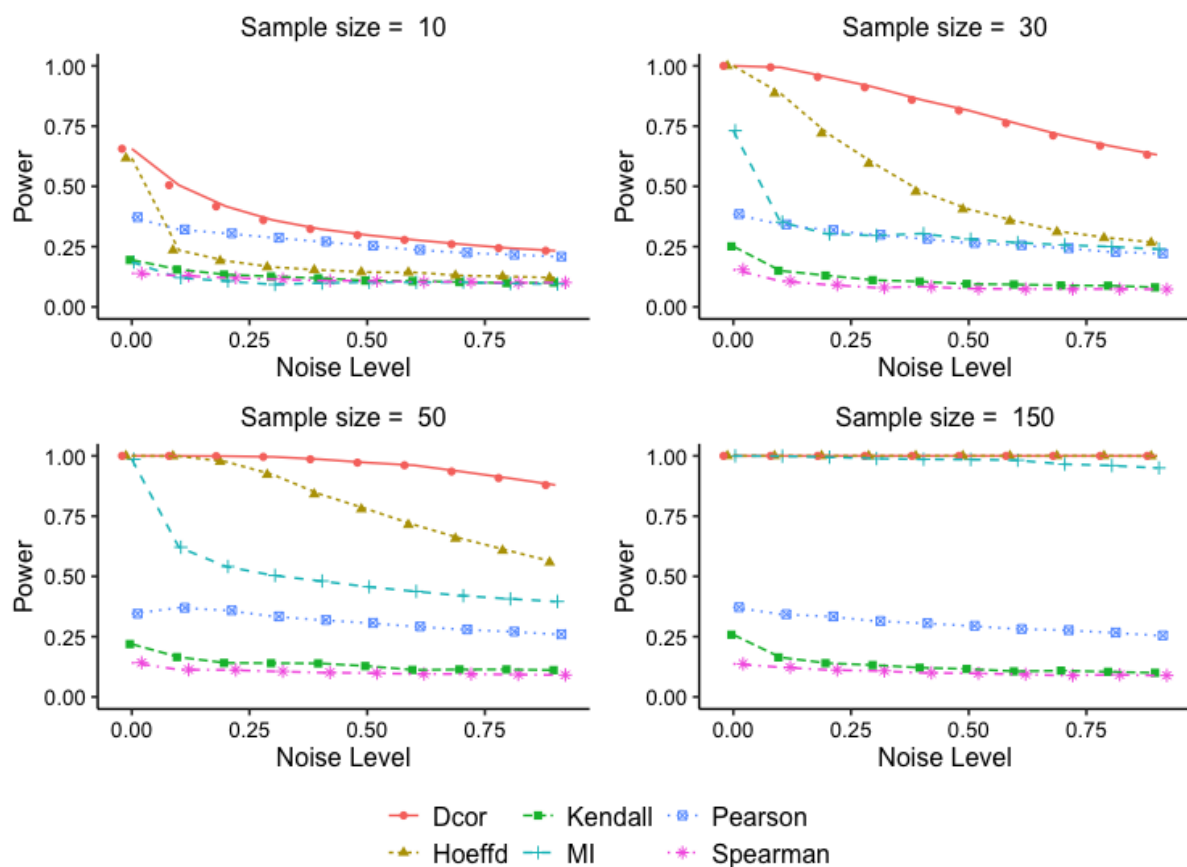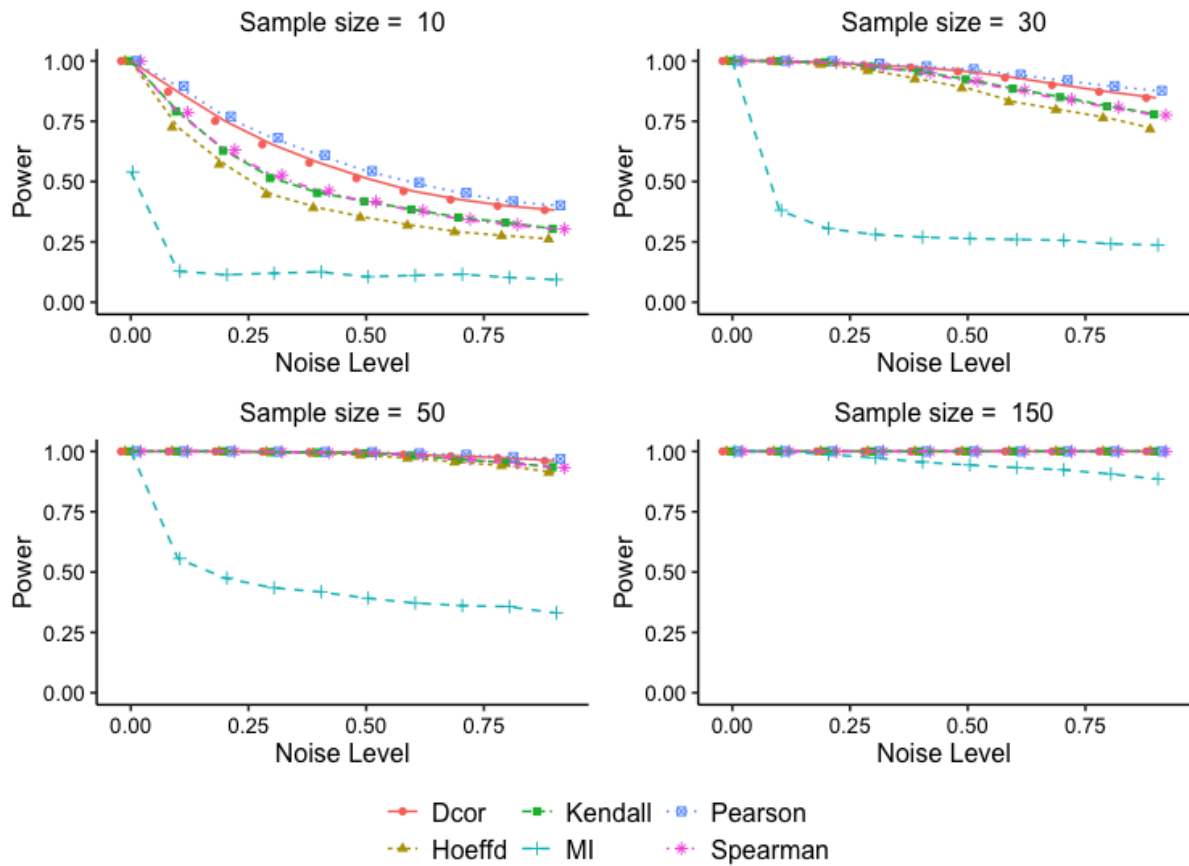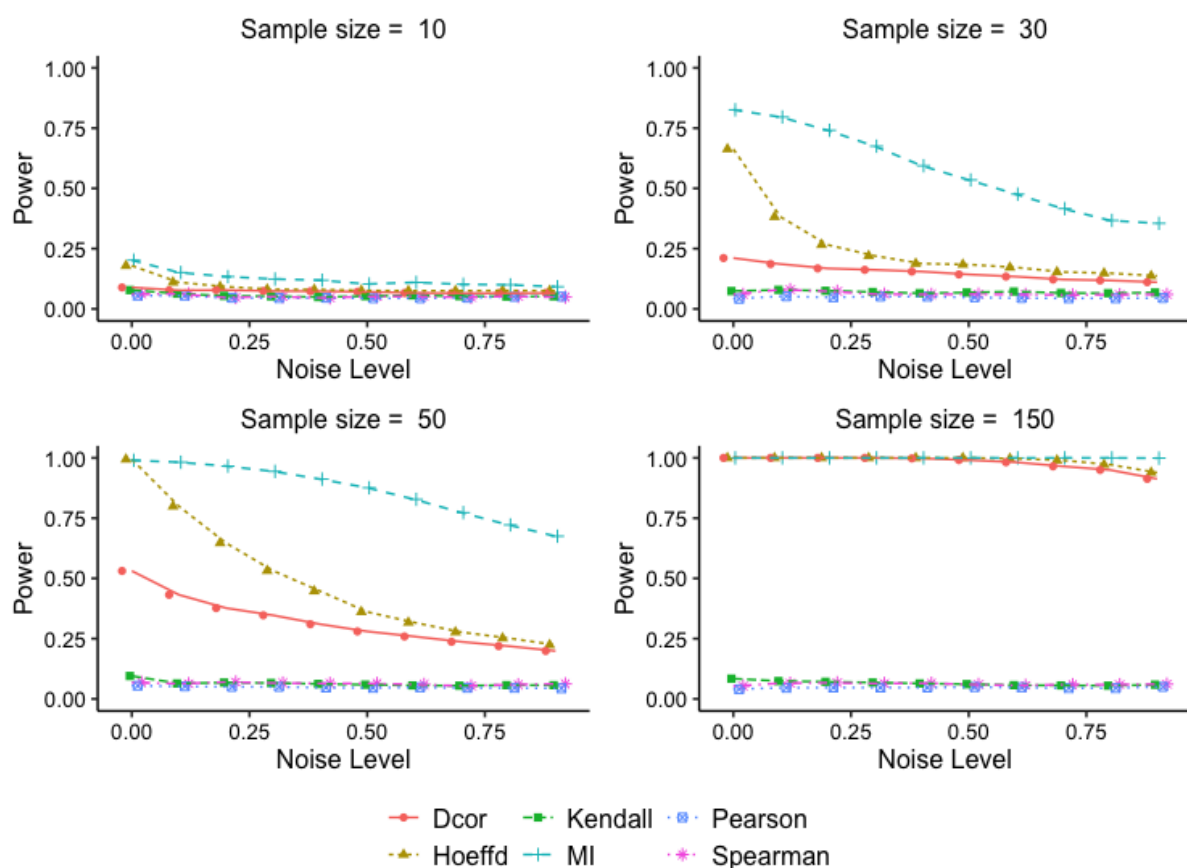

Figure 6. Sine Wave Relationship: Power of distance correlation (Dcor), Mutual Information (MI), Hoeffding's *D* (Hoeffd), Pearson's r (Pearson), Spearman's ρ (Spearman), and Kendall's τ (Kendall) as a function of the level of noise added, for two selected sample sizes. Power is estimated via 1000 simulations with α = .05.

Overall Performance

Table 2 presents the mean power of each method over all simulation designs. Regarding monotone dependence, Pearson correlation had the highest power in detecting the linear (=.99) and exponential (=.93) relationships on average. Note, however, that Dcor, Hoeffding's *D* and the two other conventional rank correlation measures performed almost equally well in both cases, with only slightly smaller power values in the range of .01 to .05. Dcor was most powerful at capturing the quadratic relationship (= .84), and MI captured the sine wave relationship the best (=.70) on average.

One drawback of the summary presented in Table 2 is that it aggregates information over all design cells such that differences in sample size and noise level are ignored. Thus, to summarize the findings from the previous sections and with data that is typical for psychological research in mind, it should be noted that for small sample sizes and high noise, Dcor was better than MI or Hoeffding's *D* at detecting linear, quadratic, and exponential relationships. MI detected the sine wave relationship better than Dcor or Hoeffding's *D*. For large sample sizes, the distance correlation was able to detect all relationships. At small and moderate sample sizes, MI could detect the sine wave relationship better than all other measures. Noisy linear relationships, however, were not captured well by MI regardless of the sample size and are better captured by the Pearson or Dcor.

Table 2

Average power over all 320 simulation design cells for each method

Relationship	MI	Dcor	Hoeffd.' D	Pearson	Spearman	Kendall
Linear	.69	.98	.96	**.99**	.98	.98
Quadratic	.51	**.84**	.69	.29	.11	.14
Exponential	.50	.92	.88	**.93**	.90	.90
Sine	**.70**	.38	.48	.05	.06	.07

Notes. Power was estimated via 1000 simulations with $\alpha = .05$ Highest power values in bold face.

Unbiasedness. Before, we learned that in order for a test to satisfy the conditions of an unbiased test, the probability of a type I error does not exceed a given significance level α but the power, i.e. the probability that the test rejects under the alternative hypothesis, is always higher than the type I error. The tests based on MI, Dcor, and Hoeffding's *D* were all unbiased, as power was always higher than the type I error rate for these tests across all factor

combinations of the simulation design (see Tables A1-A4, Appendix). Conversely, the tests based on the standard correlation measures, Pearson's r, Spearman's ρ, and Kendall's τ, had power values smaller than $\alpha = .05$ for some design cells and, therefore, did not meet the unbiased test criterion.

Discussion

This study aimed to investigate how well the mutual information and distance correlation can capture dependencies typically found in psychological research and how they compare to each other. It was also explored how each of the two methods compares to the Hoeffding's D, one of the earliest non-monotone dependence measures, and the three conventional correlation measures, Pearson's r, Spearman's ρ and Kendall's τ.

Summary

The distance correlation could detect all functional relationships in principle (in the limit, i.e. at the largest sample size). However, in the case of the periodic sine wave relationship, Dcor did not perform well at smaller sample sizes ($n < 100$). On average (across all factor combinations of the simulation design), Dcor performed best at detecting the quadratic relationship.

MI could detect all relationships in principle, except for the linear one. Regarding the periodic relationship for smaller sample sizes, MI outperformed all other methods.

In the linear and exponential case, the comparison of MI and Dcor with the traditional correlation measures and Hoeffding's D showed a stable ranking in performance: Pearson > Dcor > Spearman > Kendall > Hoeffding > MI. For the quadratic case, there was a stable ranking among the three non-monotone dependence measures of the following: Dcor > Hoeffding > MI. There was a different yet also stable ordering of these three methods for the periodic relationship: MI > Hoeffding > Dcor.

Finally, all dependence measure-based tests were exact, maintaining the nominal α level. However, only the hypothesis tests based on the three non-monotone dependence measures MI, Dcor, and Hoeffding's D were unbiased.

Comparison to previous research

The present findings are in line with previous simulation studies that found the distance correlation to have higher power than mutual information on the linear and quadratic relationship, as well as the mutual information to be more powerful on the periodic

relationships (Clark, 2013; Ding & Li, 2013; Kinney & Atwal, 2014; Simon & Tibshirani, 2014). As previous authors noted as well, the mutual information fails to detect the linear relationship and the Pearson correlation outperforms other measures. The findings presented in this paper also extend previous research that had found similar results for only larger sample sizes with 500 observations and more. In addition, we checked more strict performance criteria than authors of some previous studies by evaluating whether the measures met exact and unbiased test conditions.

Recommendations

Our recommendations contradict those from de Siqueira Santos et al. (2013), who suggested that for smaller sample sizes (< 30), only Dcor and Hoeffding's D should be used on nonmonotonic relationships. We recommend using MI, particularly on the sine wave relationship. Also, as opposed to the authors, we do not recommend using MI on the linear relationship for larger sample sizes (> 30). Because de Siqueira Santos et al. (2013) also used KDE to estimate MI, this contradiction might be caused by the fact that, unlike in this study, the authors did not account for different noise levels in the data.

Limitations and Future Research

It should be noted that in real-life scenarios, one typically does not know the level of noise present in the data. Therefore, the usefulness of our recommendations based on the results for different noise levels is limited. Nonetheless, it can be valuable to know whether a measure has high power for any level of noise or vice versa.

Also, the result that most methods included in the simulation were able to detect the underlying relationships in principle irrespective of the noise level should be evaluated with caution. It might be that the maximum noise level we chose in this study was too small, as all methods performed well at the largest sample size of only 150 observations. Therefore, future work needs to examine whether the results can be reproduced with higher noise levels.

Another limitation is the fact that mutual information was estimated by kernel density estimation. Any other estimation method may have produced different results, and because many different parameters have to be chosen in KDE, the estimation may be far from the true MI value.

Due to computational limits, we could only run 1000 simulations, which may have been too few to obtain precise enough estimations, and the extreme case of $n = 10,000$ reflecting a sample size typically reported in Applied psychology research had to be excluded

for the same reason. Simulation studies with a higher number of replications and extreme cases should be investigated in the future.

A disadvantage of simulation studies, in general, is that it remains unclear how well results extend to other situations beyond the ones chosen by a particular simulation design. Nevertheless, it should be noted that the simulation design in this study was carefully selected based on past research on sample sizes and typical relationships in psychological research. As such, this study is also the first to apply this selection of measures on data that is typical for psychological research.

Finally, the ultimate interest lies in finding out how well the measures can quantify an underlying relationship of any kind, and not merely in how well they can detect it. To examine this question and to evaluate how close estimations are to the true values, future work should, therefore, assess other performance measures such as bias and MSE. It should be noted, however, that using power as a performance measure allowed us to evaluate all methods with a common metric and to compare them with each other directly.

Conclusion

Overall, this study suggests that no method is uniformly more powerful across different sample sizes and noise levels. We recommend using the distance correlation as the standard measure for detecting any relationship in psychological research when sample sizes are large enough (≥ 150). For smaller sample sizes, and when there is a strong belief that the underlying relationship is periodic, MI should be used instead. Potential reasons for why MI performed best on the sine wave relationship should be studied in the future. Future research should also shed light on how well these measures can quantify, and not just detect, the underlying relationship between two variables.

References

Berentsen, G., & Tjøstheim, D. (2014). Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation. *Statistics and Computing*, *24*(5), 785–801.

Bohon, C., Hembacher, E., Moller, H., Moody, T. D., & Feusner, J. D. (2012). Nonlinear relationships between anxiety and visual processing of own and others' faces in body dysmorphic disorder. *Psychiatry Research*, *204*(2–3), 132–139. Retrieved from http://search.proquest.com/docview/1237502959/

Cahill, N. D. (2010). Normalized Measures of Mutual Information with General Definitions of Entropy for Multimodal Image Registration. In B. Fischer, B. M. Dawant, & C. Lorenz (Eds.), *Biomedical Image Registration* (pp. 258–268). Berlin, Heidelberg: Springer Berlin Heidelberg.

Clark, M. (2013). A Comparison Of Correlation Measures. *Center for Social Research, University of Notre Dame*, *4*. Retrieved from http://www3.nd.edu/~mclark19/learn/CorrelationComparison.pdf

Colman, A. M. (2015). *Oxford dictionary of psychology* (Fourth edi). Oxford: Oxford University Press.

DasGupta, A. (2011). *Probability for statistics and machine learning fundamentals and advanced topics*. New York [etc.]: Springer.

de Siqueira Santos, S., Takahashi, D. Y., Nakata, A., & Fujita, A. (2014). A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings in Bioinformatics*, *15*(6), 906–918.

Ding, A. A., & Li, Y. (2013). Copula Correlation: An Equitable Dependence Measure and Extension of Pearson's Correlation.

Fox, J. (2008). *Applied regression analysis and generalized linear models*. *Applied regression analysis and generalized linear models*.

Gencaga, D., Malakar, N. K., & Lary, D. J. (2014). Survey On The Estimation Of Mutual Information Methods as a Measure of Dependency Versus Correlation Analysis, *1636*(1).

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, *338*(6213).

Grijalva, E., Harms, P. D., Newman, D. A., Gaddis, B. H., & Fraley, R. C. (2015).

Narcissism and Leadership: A Meta-Analytic Review of Linear and Nonlinear Relationships. *Personnel Psychology*, *68*(1), 1–47.

Grimm, K., Zhang, Z., Hamagami, F., & Mazzocco, M. (2013). Modeling Nonlinear Change via Latent Change and Latent Acceleration Frameworks: Examining Velocity and Acceleration of Growth Trajectories. *Multivariate Behavioral Research*, *48*(1), 117–143.

Guastello, S. J. (2001). Nonlinear Dynamics in Psychology. *Discrete Dynamics in Nature and Society*, *6*(1), 11–29.

Harpole, J. K., Woods, C. M., Rodebaugh, T. L., Levinson, C. A., & Lenze, E. J. (2014). How Bandwidth Selection Algorithms Impact Exploratory Data Analysis Using Kernel Density Estimation. *Psychological Methods*, *19*(3), 428–443.

Heath, R. A. (2000). *Nonlinear dynamics: techniques and applications in psychology*. Mahwah, N.J. [etc.]: Lawrence Erlbaum Associates.

Heller, R., Heller, Y., & Gorfine, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika*, *100*(2), 503–510.

Hoeffding, W. (1948). A Non-Parametric Test of Independence. *The Annals of Mathematical Statistics*, *19*(4), 546–557.

Hulle, M. M. Van. (2005). Edgeworth Approximation of Multivariate Differential Entropy. *Neural Computation*, *17*(9), 1903–1910.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2000). *An introduction to Statistical Learning. Current medicinal chemistry*. https://doi.org/10.1007/978-1-4614-7138-7

Karr, A. F. (1993). *Probability*. New York, NY: Springer New York.

Kendall, M. G. (1938). A New Measure of Rank Correlation. *Biometrika*, *30*(1/2), 81–93.

Khan, S., Bandyopadhyay, S., Ganguly, A. R., Saigal, S., Erickson, D. J., Protopopescu, V., & Ostrouchov, G. (2007). Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *76*(2 Pt 2), 26209. Retrieved from http://search.proquest.com/docview/68369014/

Kinney, J. B., & Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(9), 3354–3359. Retrieved from http://search.proquest.com/docview/1504739897/

Klassen, R. M., & Chiu, M. M. (2010). Effects on Teachers' Self-Efficacy and Job Satisfaction: Teacher Gender, Years of Experience, and Job Stress. *Journal of*

Educational Psychology, *102*(3), 741–756.

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *69*(6 Pt 2), 66138.

Lehmann, E. L. (2005). *Testing statistical hypotheses* (3rd ed..). New York, NY: Springer.

MacCallum, R. C., Kim, C., Malarkey, W. B., & KiecoltGlaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioural research*, *32*(3), 215–253. https://doi.org/10.1207/s15327906mbr3203_1

MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms.* Cambridge university press.

Marszalek, J. M., Barber, C., Kohlhart, J., & Cooper, B. H. (2011). Sample Size in Psychological Research over the Past 30 Years. *Perceptual and Motor Skills*, *112*(2), 331–348.

Moon, Rajagopalan, & Lall. (1995). Estimation of mutual information using kernel density estimators. *Physical Review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *52*(3), 2318–2321. Retrieved from http://search.proquest.com/docview/1859237228/

Pearson, K. (1895). Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London (1854-1905)*, *58*(347), 240–242.

Pierce, J. R., & Aguinis, H. (2013). The Too-Much-of-a-Good-Thing Effect in Management. *Journal of Management*, *39*(2), 313–338.

Rényi, A. (1959). On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungarica*, *10*(3), 441–451.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., Mcvean, G., Turnbaugh, P. J., … Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science (New York, N.Y.)*, *334*(6062), 1518–1524. Retrieved from http://search.proquest.com/docview/911949650/

Rutishauser, U., Ross, I. B., Mamelak, A. N., & Schuman, E. M. (2010). Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature*, *464*(7290).

Schermelleh-Engel, K., Kerwer, M., & Klein, A. G. (2014). Evaluation of model fit in nonlinear multilevel structural equation modeling. *Frontiers in Psychology*, *5*, 181. Retrieved from http://search.proquest.com/docview/1507792051/

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London [etc.]: Chapman and Hall.

Simon, N., & Tibshirani, R. (2014). Comment on "Detecting Novel Associations In Large Data Sets" by Reshef Et Al, Science Dec 16, 2011.

Spearman, C. (1987). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *100*(3/4), 441. https://doi.org/10.2307/1422689

Steuer, R., Kurths, J., Daub, C. O., Weise, J., & Selbig, J. (2002). The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, *18*, S231--S240.

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, *35*(6), 2769–2794. https://doi.org/10.1214/009053607000000505

Upton, G., & Cook, I. (2014). uniformly most powerful test. *A Dictionary of Statistics*. Oxford University Press. Retrieved from http://www.oxfordreference.com/view/10.1093/acref/9780199679188.001.0001/acref-9780199679188-e-1690

Wang, Y., Li, Y., Liu, X., Pu, W., Wang, X., Wang, J., … Jin, L. (2017). Bagging Nearest-Neighbor Prediction independence Test: an efficient method for nonlinear dependence of two continuous variables. *Sci Rep*, *7*(1), 12736.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482.

Appendix. Power Figures Tables for Complete Simulation Design
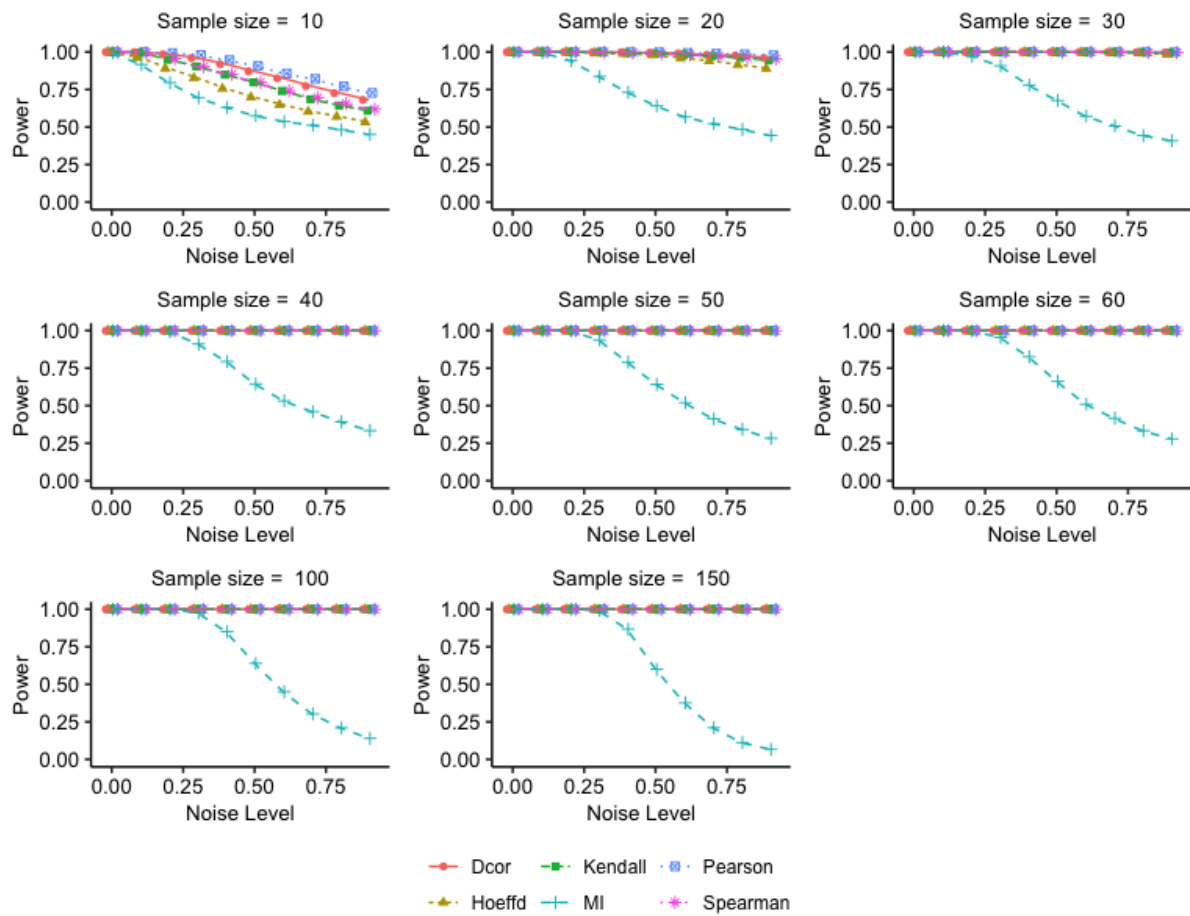
Figure A1. <u>Linear Relationship:</u> Power of distance correlation (Dcor), Mutual Information (MI), Hoeffding's *D* (Hoeffd), Pearson correlation (Pearson), Spearman's rank correlation (Spearman), and Kendall's tau (Kendall) as a function of the level of noise added, for *all* sample sizes. Power is estimated via 1000 simulations with α = .05.
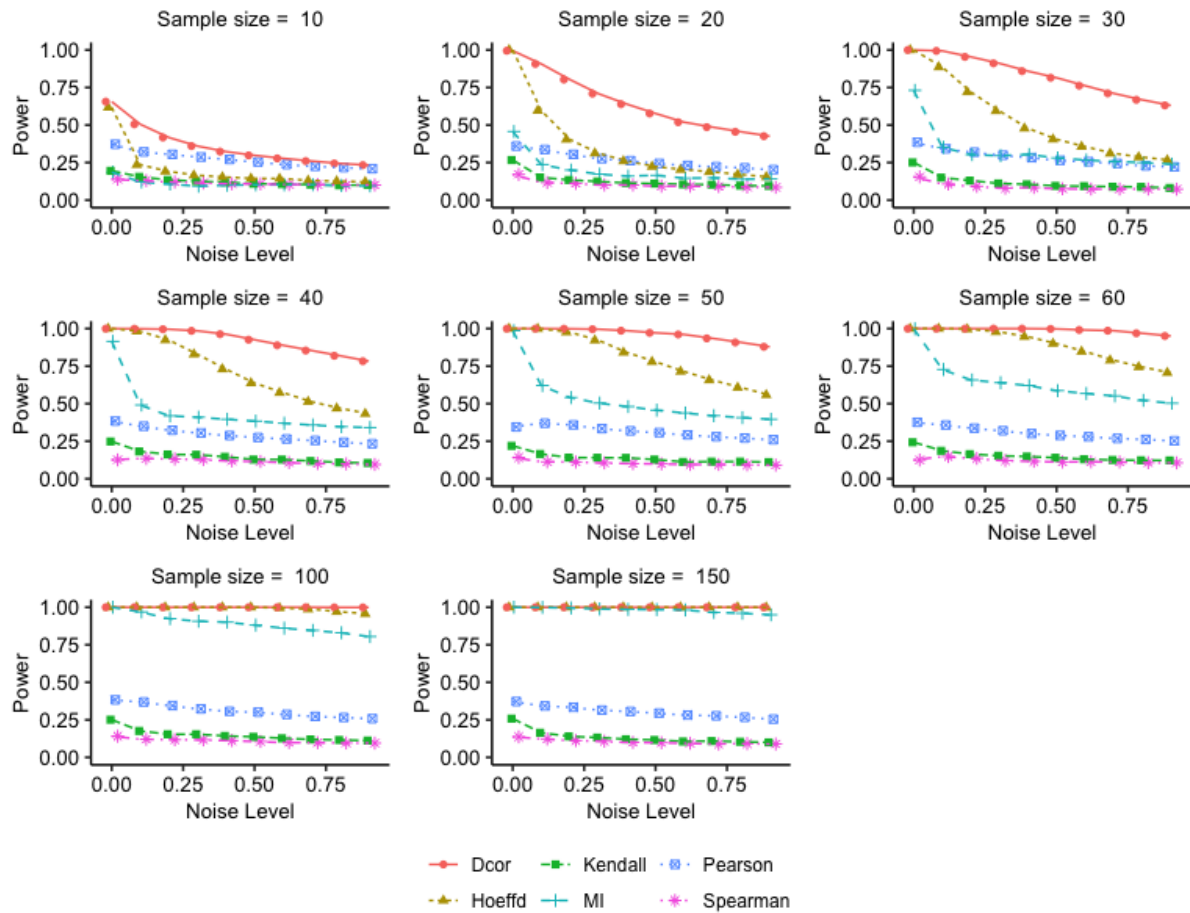
Figure A2. Quadratic Relationship: Power of distance correlation (Dcor), Mutual Information (MI), Hoeffding's *D* (Hoeffd), Pearson correlation (Pearson), Spearman's rank correlation (Spearman), and Kendall's tau (Kendall) as a function of the level of noise added, for *all* sample sizes. Power is estimated via 1000 simulations with α = .05.
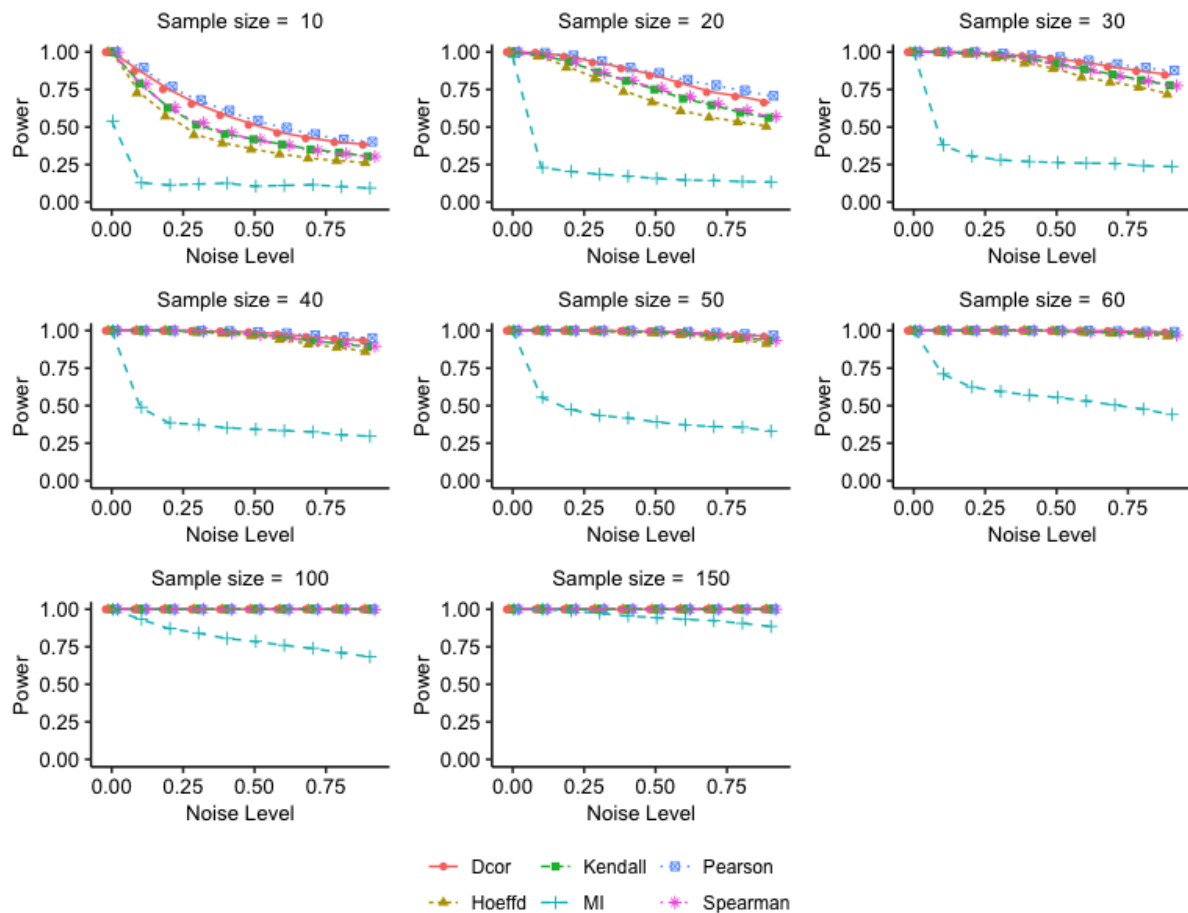
Figure A3. Exponential Relationship: Power of distance correlation (Dcor), Mutual Information (MI), Hoeffding's *D* (Hoeffd), Pearson correlation (Pearson), Spearman's rank correlation (Spearman), and Kendall's tau (Kendall) as a function of the level of noise added, for *all* sample sizes. Power is estimated via 1000 simulations.
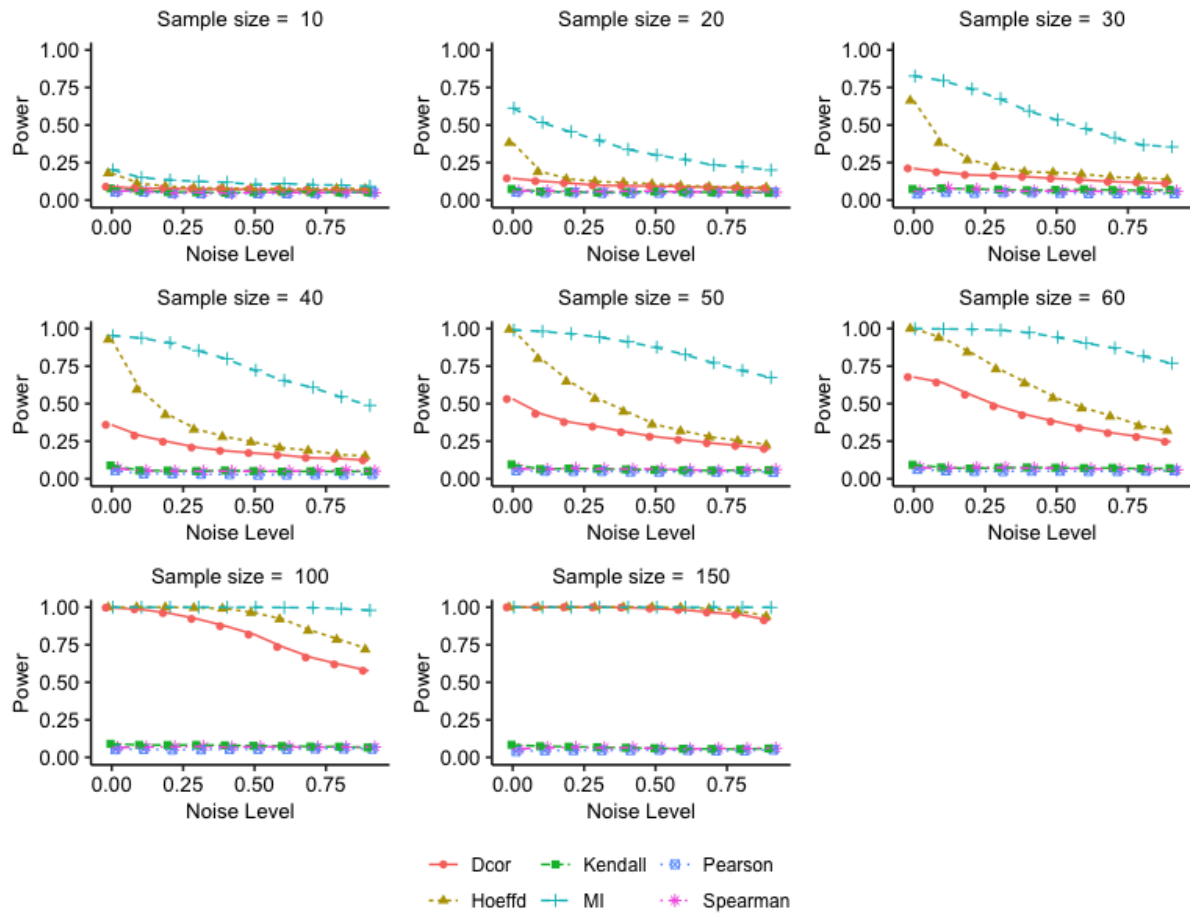
Figure A4. <u>Sine Wave Relationship:</u> Power of distance correlation (Dcor), Mutual Information (MI), Hoeffding's *D* (Hoeffd), Pearson correlation (Pearson), Spearman's rank correlation (Spearman), and Kendall's tau (Kendall) as a function of the level of noise added, for all sample sizes. Power is estimated via 1000 simulations.

Table A1.

Power simulation results for six different dependency measures where underlying

relationship is linear

N	Noise level	Mutual Information	Distance Correlation	Hoeffding's D	Pearson's r	Spearman's ρ	Kendall's τ
10	0	0.995	1	1	1	1	1
	0.1	0.917	0.999	0.962	1	0.995	0.995
	0.2	0.795	0.985	0.890	0.993	0.955	0.949
	0.3	0.696	0.960	0.827	0.980	0.902	0.904
	0.4	0.630	0.919	0.756	0.946	0.852	0.849
	0.5	0.575	0.871	0.698	0.908	0.798	0.798
	0.6	0.537	0.825	0.648	0.857	0.742	0.740
	0.7	0.512	0.773	0.603	0.821	0.695	0.685
	0.8	0.482	0.726	0.569	0.771	0.656	0.643
	0.9	0.451	0.681	0.532	0.727	0.619	0.609
20	0	1	1	1	1	1	1
	0.1	0.996	1	1	1	1	1
	0.2	0.943	1	0.998	1	1	1
	0.3	0.837	0.999	0.991	1	0.998	0.998
	0.4	0.731	0.998	0.988	0.999	0.994	0.993
	0.5	0.641	0.995	0.979	0.998	0.988	0.99
	0.6	0.567	0.991	0.958	0.994	0.982	0.983
	0.7	0.522	0.981	0.94	0.989	0.972	0.971
	0.8	0.484	0.977	0.914	0.985	0.964	0.962
	0.9	0.444	0.96	0.889	0.978	0.953	0.944
30	0	1	1	1	1	1	1
	0.1	1	1	1	1	1	1
	0.2	0.971	1	1	1	1	1
	0.3	0.908	1	1	1	1	1
	0.4	0.78	1	0.999	1	1	1
	0.5	0.677	1	0.999	1	0.999	0.999
	0.6	0.572	1	0.998	1	0.999	0.999
	0.7	0.508	1	0.996	1	0.999	0.998
	0.8	0.443	0.998	0.992	1	0.997	0.997
	0.9	0.409	0.996	0.983	0.999	0.996	0.995
40	0	1	1	1	1	1	1

	0.1	1	1	1	1	1	1
	0.2	0.993	1	1	1	1	1
	0.3	0.911	1	1	1	1	1
	0.4	0.794	1	1	1	1	1
	0.5	0.645	1	1	1	1	1
	0.6	0.532	1	1	1	1	1
	0.7	0.46	1	1	1	1	1
	0.8	0.392	1	1	1	1	1
	0.9	0.334	1	0.999	1	1	1
50	0	1	1	1	1	1	1
	0.1	1	1	1	1	1	1
	0.2	0.998	1	1	1	1	1
	0.3	0.938	1	1	1	1	1
	0.4	0.79	1	1	1	1	1
	0.5	0.642	1	1	1	1	1
	0.6	0.52	1	1	1	1	1
	0.7	0.412	1	1	1	1	1
	0.8	0.342	1	1	1	1	1
	0.9	0.283	1	0.999	1	1	1
60	0	1	1	1	1	1	1
	0.1	1	1	1	1	1	1
	0.2	0.999	1	1	1	1	1
	0.3	0.955	1	1	1	1	1
	0.4	0.826	1	1	1	1	1
	0.5	0.662	1	1	1	1	1
	0.6	0.51	1	1	1	1	1
	0.7	0.416	1	1	1	1	1
	0.8	0.333	1	1	1	1	1
	0.9	0.279	1	1	1	1	1
100	0	1	1	1	1	1	1
	0.1	1	1	1	1	1	1
	0.2	1	1	1	1	1	1
	0.3	0.976	1	1	1	1	1
	0.4	0.85	1	1	1	1	1
	0.5	0.639	1	1	1	1	1
	0.6	0.45	1	1	1	1	1
	0.7	0.303	1	1	1	1	1
	0.8	0.21	1	1	1	1	1

	0.9	0.14	1	1	1	1	1
150	0	1	1	1	1	1	1
	0.1	1	1	1	1	1	1
	0.2	1	1	1	1	1	1
	0.3	0.992	1	1	1	1	1
	0.4	0.867	1	1	1	1	1
	0.5	0.6	1	1	1	1	1
	0.6	0.377	1	1	1	1	1
	0.7	0.211	1	1	1	1	1
	0.8	0.112	1	1	1	1	1
	0.9	0.067	1	1	1	1	1

Notes: Power was calculated via 1000 simulations for each design cell with $\alpha = .05$

Table A2.

Power simulation results for six different dependency measures where underlying relationship is quadratic

N	Noise level	Mutual Information	Distance Correlation	Hoeffding's D	Pearson's r	Spearman's ρ	Kendall's τ
10	0	0.186	0.656	0.617	0.372	0.139	0.195
	0.1	0.121	0.505	0.235	0.321	0.13	0.155
	0.2	0.107	0.417	0.191	0.305	0.121	0.134
	0.3	0.093	0.360	0.165	0.287	0.115	0.126
	0.4	0.099	0.323	0.153	0.271	0.109	0.119
	0.5	0.102	0.298	0.144	0.254	0.108	0.109
	0.6	0.102	0.278	0.143	0.237	0.104	0.107
	0.7	0.103	0.260	0.13	0.225	0.104	0.103
	0.8	0.097	0.244	0.126	0.217	0.1	0.099
	0.9	0.095	0.233	0.121	0.209	0.101	0.101
20	0	0.457	0.995	1	0.36	0.171	0.267
	0.1	0.237	0.907	0.596	0.337	0.117	0.151
	0.2	0.2	0.803	0.404	0.305	0.112	0.134
	0.3	0.173	0.709	0.312	0.276	0.101	0.124
	0.4	0.16	0.639	0.260	0.263	0.101	0.114
	0.5	0.165	0.577	0.223	0.244	0.093	0.11

	0.6	0.146	0.519	0.203	0.231	0.093	0.105
	0.7	0.148	0.486	0.188	0.223	0.091	0.103
	0.8	0.139	0.455	0.17	0.215	0.087	0.097
	0.9	0.141	0.426	0.154	0.203	0.085	0.095
30	0	0.731	1	1	0.385	0.155	0.251
	0.1	0.352	0.994	0.886	0.342	0.107	0.15
	0.2	0.303	0.954	0.72	0.319	0.092	0.13
	0.3	0.296	0.911	0.594	0.299	0.08	0.111
	0.4	0.304	0.859	0.478	0.283	0.084	0.105
	0.5	0.281	0.815	0.405	0.265	0.076	0.095
	0.6	0.267	0.762	0.356	0.256	0.074	0.093
	0.7	0.257	0.711	0.311	0.244	0.074	0.089
	0.8	0.25	0.668	0.286	0.228	0.074	0.088
	0.9	0.24	0.631	0.266	0.222	0.073	0.081
40	0	0.915	1	1	0.384	0.126	0.248
	0.1	0.491	0.999	0.982	0.35	0.135	0.182
	0.2	0.419	0.996	0.924	0.323	0.132	0.162
	0.3	0.41	0.985	0.829	0.305	0.128	0.16
	0.4	0.396	0.963	0.73	0.288	0.119	0.146
	0.5	0.384	0.926	0.637	0.275	0.115	0.128
	0.6	0.37	0.889	0.573	0.265	0.108	0.128
	0.7	0.358	0.854	0.513	0.255	0.102	0.118
	0.8	0.347	0.819	0.469	0.243	0.099	0.109
	0.9	0.341	0.783	0.433	0.233	0.095	0.105
50	0	0.986	1	1	0.345	0.142	0.219
	0.1	0.621	1	0.999	0.37	0.113	0.165
	0.2	0.541	0.999	0.976	0.359	0.112	0.141
	0.3	0.504	0.996	0.922	0.334	0.106	0.14
	0.4	0.481	0.986	0.841	0.319	0.101	0.139
	0.5	0.457	0.972	0.779	0.307	0.099	0.128
	0.6	0.438	0.961	0.714	0.292	0.095	0.112
	0.7	0.42	0.935	0.657	0.28	0.095	0.114
	0.8	0.407	0.908	0.607	0.271	0.093	0.114
	0.9	0.396	0.879	0.558	0.26	0.091	0.111
60	0	0.999	1	1	0.376	0.125	0.243
	0.1	0.728	1	1	0.358	0.148	0.186
	0.2	0.658	1	0.996	0.337	0.136	0.164
	0.3	0.64	1	0.978	0.32	0.122	0.153

N	Noise level	Mutual Information	Distance Correlation	Hoeffding's D	Pearson's r	Spearman's r	Kendall's τ
	0.4	0.621	1	0.946	0.301	0.116	0.148
	0.5	0.587	0.998	0.9	0.289	0.112	0.14
	0.6	0.568	0.991	0.847	0.28	0.113	0.13
	0.7	0.552	0.986	0.789	0.269	0.112	0.125
	0.8	0.522	0.969	0.744	0.262	0.107	0.123
	0.9	0.503	0.951	0.706	0.253	0.106	0.122
100	0	1	1	1	0.383	0.141	0.251
	0.1	0.969	1	1	0.369	0.121	0.176
	0.2	0.924	1	1	0.345	0.118	0.154
	0.3	0.907	1	1	0.323	0.117	0.154
	0.4	0.901	1	1	0.307	0.112	0.143
	0.5	0.881	1	1	0.301	0.106	0.137
	0.6	0.861	1	0.995	0.286	0.099	0.127
	0.7	0.846	1	0.987	0.273	0.098	0.12
	0.8	0.830	0.999	0.97	0.266	0.096	0.116
	0.9	0.805	0.999	0.956	0.258	0.095	0.112
150	0	1	1	1	0.372	0.137	0.258
	0.1	0.999	1	1	0.343	0.123	0.164
	0.2	0.995	1	1	0.334	0.112	0.14
	0.3	0.989	1	1	0.315	0.108	0.132
	0.4	0.986	1	1	0.306	0.1	0.121
	0.5	0.985	1	1	0.295	0.098	0.116
	0.6	0.982	1	1	0.282	0.094	0.106
	0.7	0.966	1	1	0.277	0.089	0.109
	0.8	0.960	1	1	0.267	0.092	0.104
	0.9	0.950	1	0.999	0.255	0.089	0.1

Notes: Power was calculated via 1000 simulations for each design cell with $\alpha = .05$

Table A3.

Power simulation results for six different dependency measures where underlying relationship is exponential

		Mutual Information	Distance Correlation	Hoeffding's D	Pearson's r	Spearman's r	Kendall's τ
N	Noise level						
10	0	0.54	1	1	1	1	1

	0.1	0.129	0.872	0.725	0.896	0.786	0.791
	0.2	0.114	0.751	0.571	0.770	0.631	0.628
	0.3	0.120	0.654	0.446	0.681	0.527	0.515
	0.4	0.126	0.578	0.391	0.610	0.462	0.453
	0.5	0.106	0.514	0.351	0.544	0.417	0.418
	0.6	0.111	0.460	0.318	0.496	0.380	0.384
	0.7	0.116	0.425	0.291	0.453	0.345	0.351
	0.8	0.103	0.399	0.275	0.418	0.322	0.330
	0.9	0.094	0.382	0.261	0.402	0.303	0.304
20	0	0.966	1	1	1	1	1
	0.1	0.230	0.995	0.970	0.994	0.980	0.979
	0.2	0.205	0.969	0.896	0.977	0.938	0.934
	0.3	0.186	0.929	0.821	0.940	0.862	0.865
	0.4	0.173	0.890	0.735	0.896	0.811	0.807
	0.5	0.158	0.843	0.665	0.860	0.760	0.750
	0.6	0.147	0.786	0.606	0.816	0.703	0.689
	0.7	0.144	0.732	0.563	0.779	0.655	0.646
	0.8	0.137	0.703	0.532	0.745	0.611	0.596
	0.9	0.134	0.663	0.503	0.708	0.570	0.562
30	0	1	1	1	1	1	1
	0.1	0.381	1	0.998	1	0.999	0.999
	0.2	0.306	0.994	0.985	0.999	0.994	0.994
	0.3	0.281	0.983	0.959	0.989	0.977	0.976
	0.4	0.27	0.973	0.924	0.978	0.951	0.956
	0.5	0.264	0.956	0.887	0.966	0.917	0.923
	0.6	0.26	0.931	0.83	0.944	0.88	0.885
	0.7	0.257	0.9	0.797	0.92	0.841	0.85
	0.8	0.242	0.872	0.763	0.896	0.808	0.812
	0.9	0.237	0.847	0.717	0.877	0.775	0.779
40	0	1	1	1	1	1	1
	0.1	0.489	1	1	1	1	1
	0.2	0.386	1	1	1	1	1
	0.3	0.374	0.998	0.99	0.999	0.995	0.993
	0.4	0.353	0.995	0.98	0.997	0.985	0.987
	0.5	0.343	0.99	0.965	0.991	0.972	0.975
	0.6	0.334	0.975	0.939	0.984	0.958	0.958
	0.7	0.326	0.958	0.908	0.97	0.93	0.933
	0.8	0.306	0.944	0.886	0.958	0.912	0.913

	0.9	0.297	0.934	0.859	0.95	0.896	0.892
50	0	1	1	1	1	1	1
	0.1	0.556	1	1	1	1	1
	0.2	0.475	1	0.999	1	1	1
	0.3	0.435	0.999	0.997	0.999	0.998	0.997
	0.4	0.418	0.997	0.992	0.998	0.995	0.995
	0.5	0.392	0.996	0.984	0.997	0.992	0.992
	0.6	0.372	0.987	0.971	0.992	0.984	0.983
	0.7	0.361	0.98	0.955	0.986	0.969	0.966
	0.8	0.357	0.973	0.94	0.977	0.952	0.954
	0.9	0.331	0.961	0.912	0.969	0.933	0.935
60	0	1	1	1	1	1	1
	0.1	0.712	1	1	1	1	1
	0.2	0.625	1	1	1	1	1
	0.3	0.595	1	1	1	1	1
	0.4	0.571	1	1	1	1	1
	0.5	0.556	1	0.995	1	0.997	0.999
	0.6	0.532	1	0.99	1	0.991	0.993
	0.7	0.505	0.998	0.983	0.996	0.989	0.986
	0.8	0.478	0.994	0.975	0.995	0.982	0.983
	0.9	0.442	0.988	0.962	0.993	0.97	0.976
100	0	1	1	1	1	1	1
	0.1	0.934	1	1	1	1	1
	0.2	0.872	1	1	1	1	1
	0.3	0.842	1	1	1	1	1
	0.4	0.806	1	1	1	1	1
	0.5	0.786	1	1	1	1	1
	0.6	0.76	1	1	1	1	1
	0.7	0.74	1	1	1	1	1
	0.8	0.711	1	1	1	1	1
	0.9	0.684	1	1	1	0.998	0.999
150	0	1	1	1	1	1	1
	0.1	1	1	1	1	1	1
	0.2	0.988	1	1	1	1	1
	0.3	0.974	1	1	1	1	1
	0.4	0.956	1	1	1	1	1
	0.5	0.944	1	1	1	1	1
	0.6	0.933	1	1	1	1	1

	0.7	0.924	1	1	1	1	1
	0.8	0.907	1	1	1	1	1
	0.9	0.886	1	1	1	1	1

Notes: Power was calculated via 1000 simulations for each design cell with $\alpha = .05$

Table A4.

Power simulation results for six different dependency measures where underlying relationship is the sine wave

N	Noise level	Mutual Information	Distance Correlation	Hoeffding's D	Pearson's r	Spearman's r	Kendall's τ
10	0	0.202	0.089	0.179	0.055	0.061	0.078
	0.1	0.151	0.077	0.110	0.055	0.062	0.063
	0.2	0.134	0.078	0.092	0.046	0.047	0.054
	0.3	0.124	0.074	0.081	0.045	0.051	0.052
	0.4	0.119	0.073	0.080	0.046	0.046	0.049
	0.5	0.103	0.071	0.076	0.044	0.051	0.053
	0.6	0.110	0.066	0.075	0.044	0.052	0.056
	0.7	0.102	0.065	0.075	0.046	0.048	0.053
	0.8	0.100	0.065	0.076	0.050	0.052	0.051
	0.9	0.092	0.066	0.075	0.051	0.049	0.052
20	0	0.613	0.145	0.382	0.054	0.061	0.074
	0.1	0.518	0.126	0.188	0.050	0.062	0.057
	0.2	0.456	0.113	0.140	0.051	0.059	0.054
	0.3	0.398	0.098	0.123	0.048	0.057	0.050
	0.4	0.338	0.096	0.117	0.046	0.059	0.053
	0.5	0.301	0.093	0.107	0.048	0.060	0.058
	0.6	0.271	0.089	0.096	0.050	0.054	0.054
	0.7	0.234	0.085	0.089	0.053	0.055	0.052
	0.8	0.222	0.081	0.088	0.052	0.053	0.053
	0.9	0.200	0.077	0.084	0.053	0.053	0.048
30	0	0.826	0.211	0.663	0.042	0.06	0.074
	0.1	0.796	0.186	0.382	0.052	0.08	0.079
	0.2	0.741	0.168	0.266	0.048	0.073	0.075
	0.3	0.675	0.162	0.22	0.051	0.061	0.07

	0.4	0.594	0.155	0.188	0.051	0.06	0.065
	0.5	0.535	0.143	0.184	0.048	0.061	0.068
	0.6	0.476	0.134	0.172	0.045	0.06	0.072
	0.7	0.416	0.122	0.153	0.044	0.058	0.066
	0.8	0.367	0.118	0.148	0.044	0.059	0.065
	0.9	0.355	0.111	0.137	0.045	0.06	0.068
40	0	0.953	0.359	0.927	0.056	0.071	0.089
	0.1	0.938	0.287	0.593	0.034	0.054	0.055
	0.2	0.906	0.246	0.425	0.035	0.051	0.054
	0.3	0.854	0.208	0.328	0.031	0.05	0.051
	0.4	0.799	0.186	0.279	0.029	0.051	0.055
	0.5	0.723	0.171	0.243	0.026	0.05	0.051
	0.6	0.655	0.157	0.206	0.027	0.05	0.047
	0.7	0.609	0.139	0.187	0.028	0.053	0.05
	0.8	0.546	0.134	0.16	0.028	0.056	0.047
	0.9	0.489	0.123	0.151	0.029	0.052	0.05
50	0	0.990	0.531	0.994	0.055	0.068	0.095
	0.1	0.982	0.432	0.799	0.052	0.063	0.064
	0.2	0.966	0.377	0.647	0.050	0.070	0.068
	0.3	0.944	0.347	0.532	0.050	0.066	0.066
	0.4	0.912	0.310	0.447	0.047	0.064	0.062
	0.5	0.876	0.280	0.361	0.046	0.064	0.059
	0.6	0.828	0.259	0.317	0.048	0.060	0.056
	0.7	0.773	0.237	0.277	0.046	0.054	0.055
	0.8	0.722	0.219	0.252	0.046	0.059	0.057
	0.9	0.675	0.199	0.224	0.044	0.062	0.057
60	0	0.999	0.677	1	0.064	0.076	0.093
	0.1	0.998	0.641	0.938	0.054	0.069	0.075
	0.2	0.996	0.559	0.842	0.050	0.074	0.068
	0.3	0.990	0.482	0.730	0.049	0.075	0.073
	0.4	0.974	0.424	0.635	0.052	0.071	0.075
	0.5	0.943	0.380	0.537	0.052	0.071	0.073
	0.6	0.904	0.337	0.469	0.050	0.067	0.071
	0.7	0.871	0.304	0.415	0.050	0.063	0.072
	0.8	0.817	0.278	0.350	0.052	0.063	0.067
	0.9	0.770	0.246	0.320	0.054	0.060	0.070
100	0	1	0.997	1	0.053	0.065	0.090
	0.1	1	0.986	1	0.055	0.069	0.083

	0.2	1	0.961	1	0.051	0.070	0.082
	0.3	1	0.922	0.996	0.052	0.073	0.083
	0.4	1	0.874	0.991	0.055	0.068	0.080
	0.5	1	0.818	0.963	0.055	0.071	0.077
	0.6	0.998	0.735	0.92	0.055	0.07	0.075
	0.7	0.997	0.665	0.844	0.059	0.068	0.072
	0.8	0.99	0.619	0.784	0.059	0.069	0.071
	0.9	0.98	0.577	0.719	0.057	0.069	0.066
150	0	1	1	1	0.040	0.053	0.084
	0.1	1	1	1	0.046	0.065	0.074
	0.2	1	1	1	0.047	0.067	0.071
	0.3	1	1	1	0.047	0.064	0.068
	0.4	1	0.998	1	0.047	0.065	0.064
	0.5	1	0.991	1	0.049	0.060	0.061
	0.6	1	0.982	0.997	0.048	0.056	0.057
	0.7	1	0.965	0.99	0.045	0.059	0.056
	0.8	1	0.95	0.972	0.045	0.059	0.055
	0.9	0.999	0.913	0.939	0.050	0.061	0.059

Notes: Power was calculated via 1000 simulations for each design cell with $\alpha = .05$