Testing for positive quadrant dependence

Chuan-Fa Tang[1], **Dewei Wang**[2], **Hammou El Barmi**[3], **Joshua M. Tebbs**[2,*]

[1]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

[2]Department of Statistics, University of South Carolina, Columbia, SC 29208, USA

[3]Department of Information Systems and Statistics, City University of New York, New York, NY 10010, USA

SUMMARY:

We develop an empirical likelihood approach to test independence of two univariate random variables X and Y versus the alternative that X and Y are strictly positive quadrant dependent (PQD). Establishing this type of ordering between X and Y is of interest in many applications, including finance, insurance, engineering, and other areas. Adopting the framework in Einmahl and McKeague (2003, *Bernoulli*), we create a distribution-free test statistic that integrates a localized empirical likelihood ratio test statistic with respect to the empirical joint distribution of X and Y. When compared to well known existing tests and distance-based tests we develop by using copula functions, simulation results show the EL testing procedure performs well in a variety of scenarios when X and Y are strictly PQD. We use three data sets for illustration and provide an online R resource practitioners can use to implement the methods in this article.

Keywords

Bivariate data; Copula function; Empirical likelihood; Independence; Kendall's rank test; Spearman's rank test

1. Introduction

Positive quadrant dependence refers to the joint behavior of two random variables when they are likely to assume small (or large) values simultaneously. Specifically, let (X, Y) denote a continuous random vector with cumulative distribution function (cdf) H, and let F and G denote the marginal cdfs of X and Y, respectively. From Lehmann (1966), univariate random variables X and Y are positive quadrant dependent (PQD) if and only if

$$pr(X \leq x, Y \leq y) \geq pr(X \leq x)pr(Y \leq y), \text{ for all } (x, y) \in \mathbb{R}^2, \tag{1}$$

or, equivalently, if and only if $pr(X > x, Y > y) \geq pr(X > x) pr(Y > y)$ for all $(x, y) \in \mathbb{R}^2$. Note that positive quadrant dependence implies the correlation of X and Y is positive but that this relationship does not hold in reverse. In addition, if the inequality in (1) is replaced with an equality; i.e., if $pr(X \leq x, Y \leq y) = pr(X \leq x) pr(Y \leq y)$, then X and Y are

*Corresponding author Joshua M. Tebbs tebbs@stat.sc.edu.

independent. We say X and Y are "strictly PQD" when the inequality in (1) is strict for at least one $(x, y) \in \mathbb{R}^2$.

Applications of positive quadrant dependence are replete in finance, insurance, engineering, and other areas. For example, if X and Y are the returns of two stocks in a portfolio, cautious investors might prefer X and Y to be independent rather than strictly PQD, as the probability of simultaneously large losses for two strictly PQD assets is larger than it would be under independence (Malevergne and Sornette, 2006). In engineering, if X and Y denote the lifetimes of two components, then incorrectly assuming X and Y are independent (when they are strictly PQD) would lead one to underestimate the system reliability in a series system and overestimate it in a parallel system (Lai and Xie, 2006). Therefore, having knowledge of this type of dependence structure can help to improve calculations needed for establishing maintenance schedules. Finally, in diagnostic screening, incorrectly assuming two biomarkers X and Y are independent (when they are strictly PQD) could underestimate the sensitivity of an assay when testing diseased individuals. This, in turn, would overestimate the number of false negative diagnoses and potentially compromise inference when estimating the probability of disease in a population (Hanson, Johnson, and Gardner, 2003).

In this article, motivated by the applications in the previous paragraph and elsewhere, we are interested in testing independence of X and Y versus the alternative that X and Y are strictly PQD; i.e., testing $\mathscr{H}_0$ versus $\mathscr{H}_1 - \mathscr{H}_0$, where

$$\mathscr{H}_0 : H(x, y) = F(x)G(y), \text{ for all } (x, y) \in \mathbb{R}^2$$

$$\mathscr{H}_1 : H(x, y) \geq F(x)G(y), \text{ for all } (x, y) \in \mathbb{R}^2.$$

This testing problem has been considered before in the literature and there are well known procedures available for it. Kochar and Gupta (1987) proposed a class of tests for $\mathscr{H}_0$ versus $\mathscr{H}_1 - \mathscr{H}_0$ based on U-statistics, for which Kendall's one-sided rank test is a special case. Janic-Wróblewska, Kallenberg, and Ledwina (2004) extended the independence test in Kallenberg and Ledwina (1999) to the restricted PQD setting, embedding Spearman's one-sided rank test within a larger family of testing procedures. Güven and Kotz (2008) assumed a parametric model for H; specifically, that (X, Y) follows a generalized Farlie-Gumbel-Morgenstern (FGM) distribution (Schucany, Parr, and Boyer, 1978). In this model, testing $\mathscr{H}_0$ versus $\mathscr{H}_1 - \mathscr{H}_0$ reduces to testing the value of a "dependence parameter," which is zero under independence. There is a larger literature on formulating goodness-of-fit tests for PQD through the use of copula functions; see, e.g., Scaillet (2005), Gijbels, Omelka, and Sznajder (2010), Gijbels and Sznajder (2013), and Ledwina and Wyłupek (2014).

Our primary goal in this article is to evaluate an empirical likelihood (EL) procedure for testing $\mathscr{H}_0$ versus $\mathscr{H}_1 - \mathscr{H}_0$. EL methods were popularized by Owen (1988) and Owen (1990) to construct confidence regions for parameters in estimating equations. Einmahl and McKeague (2003) later developed a general framework for hypothesis testing using EL and

illustrated this technique with numerous one- and two-sample problems, including testing for symmetry, exponentiality, and the problem of testing whether X and Y are independent (versus the omnibus alternative that X and Y are not independent). More recently, EL testing has been applied to nonparametric problems in order-restricted inference, including testing for stochastic ordering (El Barmi and McKeague, 2013; Chang and McKeague, 2016) and uniform stochastic ordering (El Barmi and McKeague, 2016; El Barmi, 2017).

The advantages of EL–both for estimation and for hypothesis testing–are well documented in the statistics literature. In general, EL-based testing combines the flexibility of nonparametric methods with the efficiency of likelihood ratio-based inference. When applied to the PQD testing problem we consider, the EL approach is easy to implement and performs well under a variety of dependence structures. In Section 2, we develop the EL test, propose natural distance-based competitor tests formed from estimating copula functions, and provide implementation details. In Section 3, we summarize a simulation study that compares the EL test to the distance tests and existing approaches. In Section 4, we use three data sets to illustrate our methods. In Section 5, we offer a summary discussion and describe future research. Note that our R code available on GitHub (https://github.com/cftang9/pqd) will reproduce all simulations and calculations in this article.

2. Testing procedures

2.1. Empirical likelihood formulation

Suppose $\{(X_i, Y_i)\}_{i=1}^{n}$ is an independent and identically distributed (iid) sample from H and we wish to test $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$ as defined in Section 1. Following Einmahl and McKeague (2003), the approach taken in this subsection translates the problem into testing a family of "local" hypotheses of the form $\mathcal{H}_0^{x,y}$ versus $\mathcal{H}_1^{x,y} - \mathcal{H}_0^{x,y}$, where $\mathcal{H}_0^{x,y}: H(x,y) = F(x)G(y)$, $\mathcal{H}_1^{x,y}: H(x,y) \geq F(x)G(y)$, and (x, y) is fixed. The EL test then combines all local tests, forming one overall statistic to test $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$.

We now describe how the localized test is performed for fixed (x, y). Our description is casual, and we relegate the more technical aspects to the supplementary materials. Let $\Theta_0^{x,y}$ denote the collection of all bivariate cdfs $\widetilde{H}$ where $\widetilde{H}(x,y) = \widetilde{F}(x)\widetilde{G}(y)$, and let $\Theta_1^{x,y}$ denote the analogous collection where $\widetilde{H}(x,y) \geq \widetilde{F}(x)\widetilde{G}(y)$. The localized empirical likelihood ratio at (x, y) is

$$\mathcal{R}_n(x,y) = \frac{\sup\left\{\mathcal{L}(\widetilde{H}): \widetilde{H} \in \Theta_0^{x,y}\right\}}{\sup\left\{\mathcal{L}(H): H \in \Theta_1^{x,y}\right\}},$$

where $\mathcal{L}(\widetilde{H}) = \prod_{i=1}^{n} \tilde{p}_i$ is the EL function (Owen, 2001) and where $\tilde{p}_i$ is the probability $\widetilde{H}$ assigns to (x_i, y_i), for $i = 1, 2, \ldots, n$. Because $\mathcal{L}(\widetilde{H}) = 0$ when $\widetilde{H}$ is continuous, it suffices to consider only discrete distributions $\widetilde{H}$ with $\tilde{p}_i > 0$ when maximizing $\mathcal{L}(\widetilde{H})$. For example, the

empirical cdf $H_n(a, b) = n^{-1} \sum_{i=1}^{n} I(X_i \leq a, Y_i \leq b)$ assigns probability $1 / n$ to each (x_i, y_i) and maximizes $\mathscr{L}(\widetilde{H})$ over the collection of all bivariate cdfs.

Analogous to forming likelihood ratio statistics in a parametric framework, calculating $\mathscr{R}_n(x, y)$ requires that we maximize the EL function $\mathscr{L}(\widetilde{H})$ over both $\Theta_0^{x, y}$ and $\Theta_1^{x, y}$. To provide a casual description of how this is done, refer to Figure 1 which displays the four sets

$$A_{11} = (-\infty, x] \times (-\infty, y] \quad A_{12} = (-\infty, x] \times (y, \infty)$$

$$A_{21} = (x, \infty) \times (-\infty, y] \quad A_{22} = (x, \infty) \times (y, \infty)$$

along with an iid sample of $n = 50$ observations from H with marginals F and G. Define the true probabilities associated with these four sets to be, respectively,

$$\phi_{11} = H(x, y) \quad \phi_{12} = F(x) - H(x, y)$$

$$\phi_{21} = G(y) - H(x, y) \quad \phi_{22} = 1 - F(x) - G(y) + H(x, y).$$

As shown in the supplementary materials, calculating $\mathscr{R}_n(x, y)$, the ratio of the maximized EL functions, reduces to finding the maximum likelihood estimator (MLE) of the multinomial parameter $\boldsymbol{\phi} = (\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22})$ in two ways: one subject to the restriction that $H(x, y) = F(x)G(y)$ and one subject to the restriction that $H(x, y) \quad F(x)G(y)$.

Closed-form expressions for both sets of estimators are available. For example, under $\mathscr{H}_0^{x, y}$, the MILF of $\phi_{11} = H(x, y)$ is $\hat{\phi}_{11}^{(0)} = F_n(x)G_n(y)$, where $F_n(a) = n^{-1} \sum_{i=1}^{n} I(X_i \leq a)$ and $G_n(b) = n^{-1} \sum_{i=1}^{n} I(Y_i \leq b)$ are the marginal empirical cdfs. The remaining multinomial probabilities under $\mathscr{H}_0^{x, y}$ are estimated with $\hat{\phi}_{12}^{(0)} = F_n(x)\{1 - G_n(y)\}$, $\hat{\phi}_{21}^{(0)} = \{1 - F_n(x)\}G_n(y)$, and $\hat{\phi}_{22}^{(0)} = \{1 - F_n(x)\}\{1 + G_n(y)\}$. Under $\mathscr{H}_1^{x, y}$, the estimators are $\hat{\phi}_{11}^{(1)} = \max\{\hat{\phi}_{11}, \hat{\phi}_{11}^{(0)}\}$, $\hat{\phi}_{12}^{(1)} = \min\{\hat{\phi}_{12}, \hat{\phi}_{12}^{(0)}\}$, $\hat{\phi}_{21}^{(1)} = \min\{\hat{\phi}_{21}, \hat{\phi}_{21}^{(0)}\}$, and $\hat{\phi}_{22}^{(1)} = \max\{\hat{\phi}_{22}, \hat{\phi}_{22}^{(0)}\}$, where $\hat{\phi}_{rs}$ is unrestricted MLE of ϕ_{rs}, for $r, s \in \{1, 2\}$. For the simulated data shown in Figure 1, both sets of multinomial estimates are given in the figure caption. Step-by-step instructions for calculation are given in the supplementary materials.

With both sets of multinomial estimates, the localized empirical likelihood ratio statistic $\mathscr{R}_n(x, y)$ can be written as

$$\mathcal{R}_n(x, y) = \prod_{r=1}^{2} \prod_{s=1}^{2} \left(\frac{\hat{\phi}_{rs}^{(0)}}{\hat{\phi}_{rs}^{(1)}} \right)^{n\hat{\phi}_{rs}} = I\left(\hat{\phi}_{11} \le \hat{\phi}_{11}^{(0)}\right) + I\left(\hat{\phi}_{11} > \hat{\phi}_{11}^{(0)}\right) \prod_{r=1}^{2} \prod_{s=1}^{2} \left(\frac{\hat{\phi}_{rs}^{(0)}}{\hat{\phi}_{rs}} \right)^{n\hat{\phi}_{rs}},$$

with small values of $\mathcal{R}_n(x, y)$ leading to the rejection of $\mathcal{H}_0^{x,y}$. The expression for $\mathcal{R}_n(x, y)$ makes sense intuitively. If the unrestricted MLE $\hat{\phi}_{11} = H_n(x, y)$ is too small; i.e., less than $\hat{\phi}_{11}^{(0)} = F_n(x)G_n(y)$, then this would provide virtually no evidence against (local) independence; hence, $\mathcal{R}_n(x, y) = 1$ and $\mathcal{H}_0^{x,y}$ would not be rejected. It is only when $\hat{\phi}_{11} > \hat{\phi}_{11}^{(0)}$ does one potentially have any evidence against $\mathcal{H}_0^{x,y}$. Of course, rejecting $\mathcal{H}_0^{x,y}$ when $\mathcal{R}_n(x, y)$ is small is equivalent to rejecting $\mathcal{H}_0^{x,y}$ when $-2\ln\mathcal{R}_n(x, y)$ is large.

The EL test of $\mathcal{H}_0$ versus $\mathcal{H}_1 - \mathcal{H}_0$ is subsequently formed by aggregating all of the local tests, which leads to the test statistic

$$\mathrm{EL}_n = \int_{\mathbb{R}^2} -2\ln\mathcal{R}_n(x, y)\, dF_n(x)\, dG_n(y) = -\frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \ln\mathcal{R}_n\left(X_i, Y_j\right),$$

with large values of EL_n providing evidence against $\mathcal{H}_0$. The asymptotic distribution of EL_n under $\mathcal{H}_0$ closely resembles the asymptotic distribution of the test statistic provided in Einmahl and McKeague (2003) for the "two-sided," omnibus test for independence of X and Y. Specifically, under $\mathcal{H}_0$,

$$\mathrm{EL}_n \xrightarrow{d} \int_{[0,1]^2} \frac{\left\{ [\xi\mathcal{B}(u, v) - u\mathcal{B}(1, v) - v\mathcal{B}(u, 1)]_+ \right\}^2}{u(1-u)v(1-v)}\, du\, dv,$$

as $n \to \infty$, where $\mathcal{B}(\cdot, \cdot)$ is a standard bivariate Brownian bridge; i.e., a mean-zero Gaussian process on $[0,1]^2$ with covariance $cov\left(\mathcal{B}(u_1, v_1), \mathcal{B}(u_2, v_2)\right) = \min\{u_1, u_2\}\min\{v_1, v_2\} - u_1 v_1 u_2 v_2$, and where $a_+ = \max\{a, 0\}$ for $a \in \mathbb{R}$. We provide a rigorous derivation of this asymptotic result in the supplementary materials.

2.2.   Distance-based copula tests

There is a substantive literature on goodness-of-fit tests for PQD, that is, testing that X and Y are PQD versus the complementary alternative that X and Y are not PQD. Essentially, a goodness-of-fit test is a test of our $\mathcal{H}_1$ versus $\mathcal{H}_2 - \mathcal{H}_1$, where $\mathcal{H}_2$ places no dependence restriction on the joint cdf H and marginal cdfs F and G. Evidence for PQD is obtained when one does not reject $\mathcal{H}_1$. Of course, not rejecting $\mathcal{H}_1$ does not allow one to disentangle independence from X and Y being strictly PQD as is our goal herein.

Most nonparametric approaches testing $\mathscr{H}_1$ versus $\mathscr{H}_2 - \mathscr{H}_1$ frame the problem in terms of copula functions. From Sklar's Theorem (Sklar, 1959), a copula C is a function such that $C(F(x), G(y)) = H(x, y)$, for all $(x, y) \in \mathbb{R}^2$, when F and G are continuous; i.e., C is a bona fide joint distribution on $[0, 1]^2$ with uniform marginals. Recognizing $\Pi(u, v) = u v$ as the independence copula, testing $\mathscr{H}_1$ versus $\mathscr{H}_2 - \mathscr{H}_1$ is equivalent to testing $C(u, v) \quad u v$ for all $(u, v) \in [0,1]^2$ versus $C(u, v) < u v$ or some $(u, v) \in [0,1]^2$. Using various copula estimators, Gijbels et al. (2010) formulated three distance-based statistics to test $\mathscr{H}_1$ versus $\mathscr{H}_2 - \mathscr{H}_1$, generalizing the earlier work of Scaillet (2005). These statistics, which are based on Kolmogorov-Smirnov (KS), Cramér-von Mises (CvM), and Anderson-Darling (AD) distances, can be modified to accommodate testing $\mathscr{H}_0$ versus $\mathscr{H}_1 - \mathscr{H}_0$ in our situation.

As in Gijbels et al. (2010), define the pseudo-observations $U_i = nF_n(X_i)/(n+1)$ and, $V_i = n G_n(Y_i)/(n+1)$, for $i = 1, 2, \ldots, n$. Scaling the empirical cdfs by $n/(n+1)$ alleviates potential problems in estimation near the boundaries in $[0, 1]^2$. Plotting the psuedo-observations can be helpful. When $\mathscr{H}_0$ is true, the (U_i, V_i) values should be approximately uniformly distributed over the unit square. A natural estimator of C is $C_n(u, v) = n^{-1} \sum_{i=1}^{n} I(U_i \leq u, V_i \leq v)$, and distance-based tests can be formed by comparing the empirical estimator $C_n(u, v)$ to the independence copula $\Pi(u, v) = uv$ under a given distance. The KS, CvM, and AD distance statistics adapted to our testing problem are

$$KS_n = \sqrt{n} \sup_{(u, v) \in [0, 1]^2} \{C_n(u, v) - uv\}_+$$

$$CvM_n = n \int_{[0, 1]^2} [\{C_n(u, v) - uv\}_+]^2 dC_n(u, v)$$

$$AD_n = n \int_{[0, 1]^2} \frac{[\{C_n(u, v) - uv\}_+]^2}{u(1-u)v(1-v)} dC_n(u, v).$$

Clearly, large values of any distance statistic are evidence for X and Y being strictly PQD. Note that both CvM_n and AD_n are based on L_2 distances, but AD_n gives more weight to those psuedo-observations residing near the boundaries of $[0, 1]^2$. Asymptotic distributions of KS_n, CvM_n, and AD_n and under $\mathscr{H}_0$ are given in the supplementary materials.

2.3. Implementation

Because the finite-sample distributions of EL_n and the distance-based statistics do not depend on the marginal distributions F and G, it is ultimately unnecessary to rely on the asymptotic results in Sections 2.1 and 2.2 for implementation. In fact, we have found that using critical values from the asymptotic distributions leads to tests that are unduly anti-conservative, especially with small to moderately sized samples. Therefore, we advocate using critical values from the corresponding finite-sample distributions which we determine as follows. For a given sample size n, we generate 10,000 observations from a bivariate

uniform distribution over the unit square, denoted by $H =^{\mathscr{U}} [0, 1]^2$. We then calculate the value of each statistic (E L $_n$, K S $_n$, C vM $_n$ and A D n) for each of the 10,000 data sets and record the upper ath quantile from the empirical distributions. Each test rejects $\mathscr{H}_0$ when its test statistic exceeds its corresponding quantile. For a user-specified sample size, our R code on GitHub will automate critical value calculations for the practitioner. This code can also be used to reproduce simulation results reported in the next section and quickly implement all tests with real data (see Section 4).

3. Simulation evidence

We compare the EL test and the distance-based copula tests in Section 2.2 to the tests in Kochar and Gupta (1987) and Janic-Wróblewska et al. (2004), which were referenced in Section 1 and are also nonparametric in nature. Kochar and Gupta (1987), hereafter abbreviated KG, proposed a class of U-statistics, U_{k+1}, which estimate $\pi_{1k} = \int_{\mathbb{R}^2} H^k(x, y) dH(x, y)$, for any positive integer k. KG show the test of $\mathscr{H}_0$ versus $\mathscr{H}_1 - \mathscr{H}_0$ reduces to testing $\pi_{1k} = (1 + k)^{-2}$ under independence versus $\pi_{1k} > (1 + k)^{-2}$ under strict PQD, so that a large value of U_{k+1} is evidence against $\mathscr{H}_0$. Kendall's (one-sided) rank statistic arises as a special member of this class when $k = 1$; for larger k, computation can become overwhelming when n is large (e.g., $n = 100$, etc.) as KG's general U-statistic involves calculating $\binom{n}{k+1}$ kernel functions. We therefore restrict attention to $k = 1$ in our comparison for computational reasons.

The approach taken in Janic-Wróblewska et al. (2004), hereafter abbreviated JKL, relies on the fact that X and Y are PQD if and only if co v$\{ s(F(X)), t(G(Y)) \}$ 0 for all nondecreasing functions s and t. JKL describe parametrically this collection of functions using a sequence of Legendre polynomials $\{ b_j^* \}_{j-1}^k$, suitably projected and normalized, which then model the joint probability density function of $(F(X), G(Y))$ under PQD. After utilizing this reparameterization and estimating F and G with their respective empirical cdfs, JKL's approach leads to a family of rank statistics. For example, when $k = 1$, the test statistic for $\mathscr{H}_0$ versus $\mathscr{H}_1 - \mathscr{H}_0$ is $V(1, 1) = n^{-1/2} \sum_{i=1}^{n} b_1^*\{(R_i - 1/2)/n\} b_1^*\{(S_i - 1/2)/n\}$, where $b_1^*(u) = \sqrt{3}(2u - 1)$ and R_i and S_i are the ranks of $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$, respectively. Interestingly, rejecting $\mathscr{H}_0$ when $V(1, 1)$ is large is identical to performing Spearman's one-sided rank test. One can formulate test statistics using a larger number of polynomials (i.e., for larger k); however, the difficulty with computation and implementation also increases substantially. We restrict attention to the $k = 1$ version of JKL herein.

All of the test statistics described so far, including the Kendall and Spearman statistics, do not depend on the marginal distributions F and G. Therefore, to investigate small-sample properties, we use copulas to simulate data. We start by considering common copula families (i.e., Clayton, Frank, Gaussian, and Gumbel), which were also used by Gijbels et al. (2010) in their evaluations. The Clayton, Frank, and Gumbel copulas belong to the well known Archimedean class and are widely used to model data with varying degrees of right-tail dependence. Each copula in this class is indexed by Kendall's tau parameter T, which

equals zero under independence and is larger than zero when X and Y are strictly PQD. The same feature holds for the Gaussian copula, which is indexed by the correlation parameter ρ. In what follows, we used finite-sample critical values for all tests calculated using the approach described in Section 2.3. All simulations are based on $B = 10000$ Monte Carlo data sets.

Table 1 displays the results when $n = 100$ and $\alpha = 0.05$. The same tables for sample sizes $n = 50$ and $n = 200$ are shown in the supplementary materials. All six tests generally perform at the nominal level when $\mathcal{H}_0$ is true. In terms of power, we make the following general observations. First, although there is no test that is uniformly more powerful than its competitors under all scenarios, it is probably safe to say that KS is the least powerful overall. The remaining five tests are more comparable. For the heavier-tailed copulas, EL clearly performs best under the Clayton, whereas AD is generally preferred for the Gumbel, most notably when the PQD signal is weaker. On the other hand, for the more "well-behaved" Frank and Gaussian copulas, the EL test closely rivals the Spearman and Kendall tests and is generally as powerful as or more powerful than the distance-based tests.

The four copula models used in Table 1 are commonly used in applications but do represent a similar dependence structure over the unit square. An anonymous referee has suggested we broaden the scope of our comparison to examine different types of dependence and distributions with heavier tails. In the supplementary materials, we describe and completely summarize the same comparison among the six tests using three additional models:

- the FGM and Cuadras-Augé (CA) copulas, motivated by theoretical characterizations of heavy-tailed distribution families in Weng and Zhang (2012), and

- a restricted bivariate t distribution family, motivated by the simulation designs in Vexler, Tsai, and Hutson (2014) and Vexler, Chen, and Hutson (2017).

The bivariate t distribution used in Vexler et al. (2014, 2017) does not satisfy PQD, but we prove in the supplementary materials a restricted version of this distribution (arising by considering only the first quadrant in $\mathbb{R}^2$) does satisfy PQD.

Finally, our R programs on GitHub will reproduce the results in Table 1 and those in the supplementary materials. For example, running the $n = 100$ simulation in Table 1 took approximately 73 minutes on a computer with a 3.1GHz processor and 16GB of memory. This time increases substantially for larger n because sample-size specific critical values must be obtained first; for example, with $n = 200$, the same simulation takes slightly over 8 hours.

4. Applications

We apply all tests in this article to three data applications. Table 2 summarizes the relevant results for each application; i.e., test statistics, critical values, and p-values. Figures 1–3 are used to show scatterplots of the data and psuedo-observations.

4.1. Twins data

A natural application of the PQD testing problem is in twins studies. Ashenfelter and Krueger (1994) summarize an observational study examining hourly wages for $n = 149$ identical twins in the United States, where each individual (within each twin pair) had different levels of education. Each twin pair in the study included individuals aged 18 years and older. The goal of this study was to evaluate the impact of education level on wages; restricting the study to identical twins allowed the investigators to mitigate the impact of lurking variables such as genetic differences and family background. We assess the bivariate relationship of wages in each twin pair. Let X and Y denote the hourly wage (measured in log US dollars) for each twin sibling, respectively. A scatterplot of the data (x_i, y_i), $i = 1, 2, \ldots, 149$, is given in Figure 1, along with the corresponding plot of psuedo-observations (u_i, v_i) as defined in Section 2.2 The psuedo-observations certainly appear to be incongruous with independence, as the (u_i, v_i) values tend to cluster about the main diagonal of the unit square. Not surprisingly, each of the six tests in Table 2 provides strong evidence in favor of strict PQD.

4.2. Education data

In 2018, high school teachers from eight states in the US participated in formal "walk out" protests, motivated largely by state governments' continued cuts in spending for educational support. Perhaps interestingly, all states involved in the protests were "conservative leaning" or "battleground states" in the contentious 2016 and 2018 elections. Are the striking teachers' demands justified? The authors of this article are reluctant to weigh in on this question; however, we can apply the tests in this article to a data set which partially addresses this issue. Using resources from the United States Census Bureau and the Department of Education, we constructed a data set on the high school graduation rate in 2016 (X) and the average amount of state government spending per student during the 2015–2016 academic year (Y, measured in log US dollars). Both variables were recorded on each state separately plus the District of Columbia ($n = 51$) and included only data from statewide public schools. The data and psuedo-observations appear in Figure 2. None of the tests (see Table 2) reject independence at the $a = 0.05$ level, although the EL, CvM, Kendall, and Spearman tests are borderline significant. Of course, this analysis does not answer the question posed initially, but it also does not significantly strengthen the argument for increased funding, at least if that argument is based on improving statewide graduation rates.

4.3. Stock data

We collected the closing prices of three stocks in the United States during 125 consecutive trading days between January 4, 2016 and June 30, 2016: Apple (APPL), Google (GOOGL), and Walmart (WMT). For each stock, the 125 closing prices are obviously not independent, but the first difference of the log prices; i.e., the $n = 124$ log returns, should be approximately independent. We checked this by using the Ljung-Box test (Ljung and Box, 1978) for independence on the returns, which provided p-values of 0.577, 0.635, and 0.082, respectively. Because these companies are common investments in 401K retirement plans, we find it of interest to assess the pairwise dependence structure among these three

company's returns. A brief discussion on assessing trivariate dependence as an extension is given in Section 5.

Table 2 (bottom) includes the test statistics and p-values for assessing independence of the returns versus strict PQD for each pair of stocks. Upon first glance at Figure 3, one might suspect that independence is reasonable between the Apple and Walmart returns; however, 5 of the 6 tests reject it at the $\alpha = 0.05$ level, including the EL test albeit barely. There is overwhelming evidence of strict PQD in the other pairs (Apple and Google; Google and Walmart). For investors concerned about diversification and reducing volatility, including all three stocks in the same portfolio might be an unwise decision.

5. Discussion

This article presents a new EL test for independence versus strict PQD by aggregating localized test statistics as espoused by Einmahl and McKeague (2003). Although the EL test is not uniformly more powerful than existing tests, our simulations show it emerges as being consistently competitive under a variety of scenarios. Our online resources are designed with reproducibility in mind and make it easy for practitioners to implement all tests in this article. The data sets in Section 4 are available with R code on GitHub along with instructions on how to analyze them.

We envision two useful extensions. First, as noted in Section 2.2, there is a healthy literature on testing X and Y are PQD versus the alternative that X and Y are not PQD; i.e., testing $\mathscr{H}_1$ versus $\mathscr{H}_2 - \mathscr{H}_1$. Is it possible to generalize our EL approach for this problem? The answer is not immediately obvious. Following the same approach in Section 2.1, it is easy to create the EL statistic for $\mathscr{H}_1$ versus $\mathscr{H}_2 - \mathscr{H}_1$; it is $\tilde{T}_n = n^{-1} \sum_{i=1}^{n} -2\ln \mathscr{R}_n^*(X_i, Y_i)$, where the localized statistic for rejecting $\mathscr{H}_1^{x,y} : H(x, y) \geq F(x)G(y)$ is

$$\mathscr{R}_n^*(x, y) = I\left(\hat{\phi}_{11} \geq \hat{\phi}_{11}^{(0)}\right) + I\left(\hat{\phi}_{11} < \hat{\phi}_{11}^{(0)}\right) \prod_{r=1}^{2} \prod_{s=1}^{2} \left(\frac{\hat{\phi}_{rs}^{(0)}}{\hat{\phi}_{rs}}\right)^{n\hat{\phi}_{rs}}.$$

However, as $\mathscr{H}_1$ can be true in many ways, one needs to first find the least favorable configuration; i.e., the configuration of F and G in $\mathscr{H}_1$ that maximizes the probability of Type I error. Our intuition strongly suggests independence of X and Y is, in fact, the least favorable configuration in $\mathscr{H}_1$; however, we have not been able to prove this rigorously from within an EL framework. Most existing approaches to test $\mathscr{H}_1$ versus $\mathscr{H}_2 - \mathscr{H}_1$ using copulas (see, e.g., Gijbels and Sznajder, 2013) often utilize the independence copula $\Pi(u, v) = uv$ for critical value calculation but do not provide proof as to why this is justified.

Second, it should be possible to generalize the EL test in Section 2.1 to test for positive orthant dependence (POD) in higher dimensions, as described in Gijbels and Sznajder (2013) and elsewhere. For example, with three random variables X_1, X_2, and X_3, positive orthant dependence would require both

$$pr(X_1 \leq x_1, X_2 \leq x_2, X_3 \leq x_3) \geq pr(X_1 \leq x_1)pr(X_2 \leq x_2)pr(X_3 \leq x_3)$$

$$pr(X_1 > x_1, X_2 > x_2, X_3 > x_3) \geq pr(X_1 > x_1)pr(X_2 > x_2)pr(X_3 > x_3)$$

to hold for all $(x_1, x_2, x_3) \in \mathbb{R}^3$. One approach to test for strict POD might involve calculating two aggregated EL statistics–one for testing independence versus each ordering above separately–and then combining these tests (e.g., by implementing a Bonferroni correction). Finding maximizers under independence is straightforward. Constrained optimization methods could be used to find restricted estimates under each ordering separately, but we do not believe these estimates will exist in closed form even for three random variables.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Ashenfelter O and Krueger A (1994). Estimates of the economic return to schooling from a new sample of twins. American Economic Review 84, 1157–1173.

Chang H and McKeague I (2016). Empirical likelihood based tests for stochastic ordering under right censorship. Electronic Journal of Statistics 10, 2511–2536. [PubMed: 31178947]

Einmahl J and McKeague I (2003). Empirical likelihood based hypothesis testing. Bernoulli 9, 267–290.

El Barmi H (2017). Testing for uniform stochastic ordering via empirical likelihood under right censoring. Statistica Sinica 27, 645–664.

El Barmi H and McKeague I (2013). Empirical likelihood-based tests for stochastic ordering. Bernoulli 19, 295–307. [PubMed: 23874142]

El Barmi H and McKeague I (2016). Testing for uniform stochastic ordering via empirical likelihood. Annals of the Institute of Statistical Mathematics 68, 955–976.

Gijbels I, Omelka M, and Sznajder D (2010). Positive quadrant dependence tests for copulas. Canadian Journal of Statistics 38, 555–581.

Gijbels I and Sznajder D (2013). Positive quadrant dependence testing and constrained copula estimation. Canadian Journal of Statistics 41, 36–64.

Güven B and Kotz S (2008). Test of independence for generalized Farlie-Gumbel-Morgenstern distributions. Journal of Computational and Applied Mathematics 212, 102–111.

Hanson T, Johnson W, and Gardner I (2003). Hierarchical models for estimating herd prevalence and test accuracy in the absence of a gold standard. Journal of Agricultural, Biological, and Environmental Statistics 8, 223–239.

Janic-Wróblewska A, Kallenberg W, and Ledwina T (2004). Detecting positive quadrant dependence and positive function dependence. Insurance: Mathematics and Economics 34, 467–487.

Kallenberg W and Ledwina T (1999). Data-driven rank tests for independence. Journal of the American Statistical Association 94, 285–301.

Kochar S and Gupta R (1987). Competitors of the Kendall-tau test for testing independence against positive quadrant dependence. Biometrika 74, 664–666.

Lai C and Xie M (2006). Stochastic Ageing and Dependence for Reliability. Springer, New York.

Ledwina T and Wyłupek G (2014). Validation of positive quadrant dependence. Insurance: Mathematics and Economics 56, 38–47.

Lehmann E (1966). Some concepts of dependence. Annals of Mathematical Statistics 37, 1137–1153.

Ljung G and Box G (1978). On a measure of lack of fit in time series models. Biometrika 65, 297–303.

Malevergne Y and Sornette D (2006). Extreme Financial Risks: From Dependence to Risk Management. Springer, Berlin.

Owen A (2001). Empirical Likelihood Chapman and Hall/CRC, Boca Raton.

Owen A (1988). Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.

Owen A (1990). Empirical likelihood ratio confidence regions. Annals of Statistics 18, 90–120.

Scaillet O (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. Canadian Journal of Statistics 33, 415–427.

Schucany W, Parr W, and Boyer J (1978). Correlation structure in Farlie-Gumbel-Morgenstern distributions. Biometrika 65, 650–653.

Sklar A (1959). Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de L'Université de Paris 8, 229–231.

Vexler A, Chen X, and Hutson A (2017). Dependence and independence: Structure and inference. Statistical Methods in Medical Research 26, 2114–2132. [PubMed: 26229085]

Vexler A, Tsai W, and Hutson A (2014). A simple density-based empirical likelihood ratio test for independence. American Statistician 68, 158–169.

Weng C and Zhang Y (2012). Characterization of multivariate heavy-tailed distribution families via copula. Journal of Multivariate Analysis 106, 178–186.
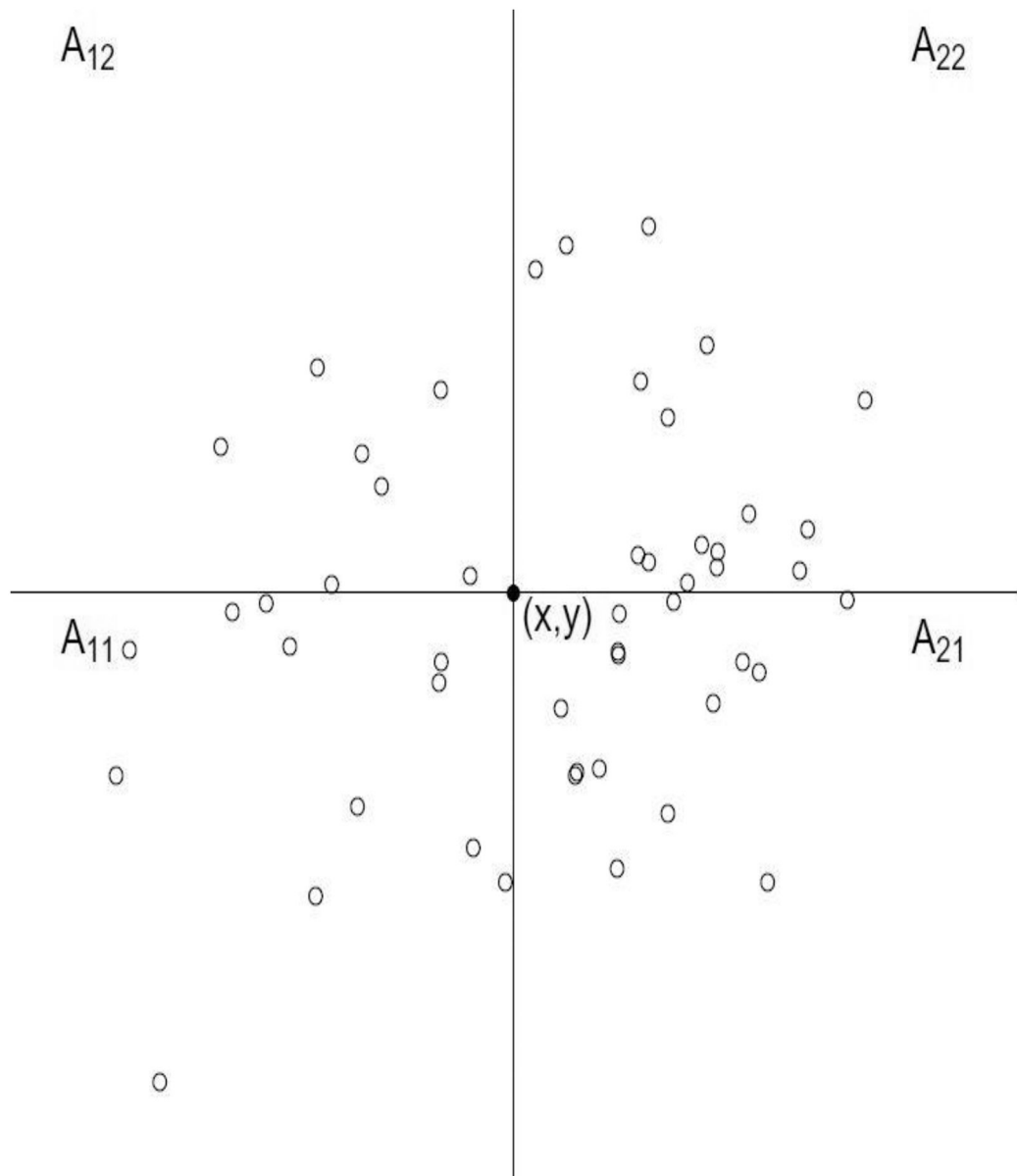
Fig. 1.
Localized test at (x, y). An iid sample of $n = 50$ is observed from the joint distribution H with marginals F and G. For these data, the MLE of ϕ subject to $H(x, y) = F(x)G(y)$ is $\left(\hat{\phi}_{11}^{(0)}, \hat{\phi}_{12}^{(0)}, \hat{\phi}_{21}^{(0)}, \hat{\phi}_{22}^{(0)}\right) = (0.2052, 0.1748, 0.3348, 0.2852)$. The MLE of ϕ subject to $H(x, y) \neq F(x)G(y)$ is $\left(\hat{\phi}_{11}^{(1)}, \hat{\phi}_{12}^{(1)}, \hat{\phi}_{21}^{(1)}, \hat{\phi}_{22}^{(1)}\right) = (0.24, 0.14, 0.30, 0.32)$. These calculations are shown in the supplementary materials.
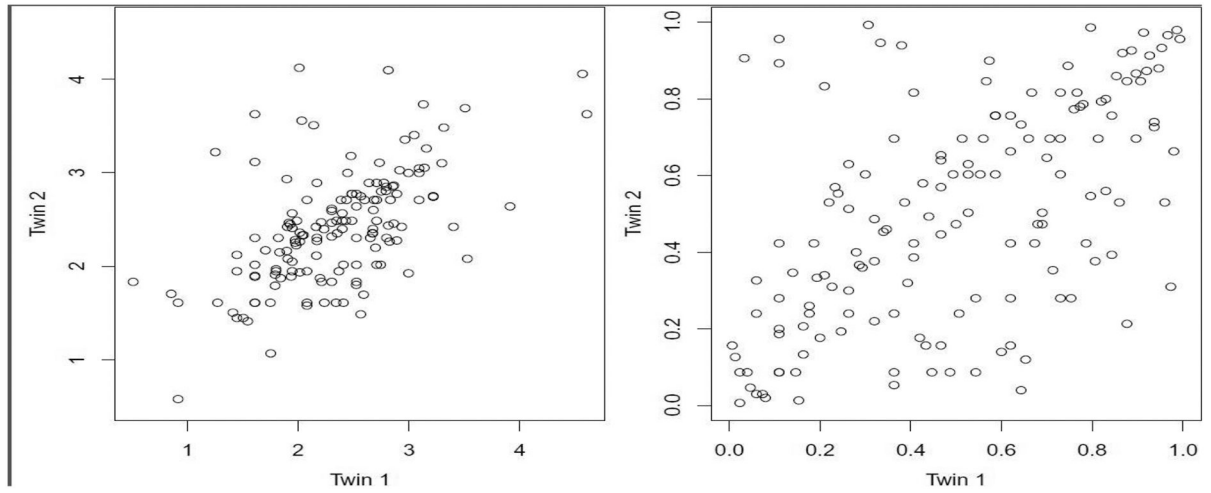
Fig. 2.
Twins data. Left: Scatterplot of hourly wages (in log US dollars) for $n = 149$ pairs of identical twins. Right: Scatterplot of psuedo-observations.
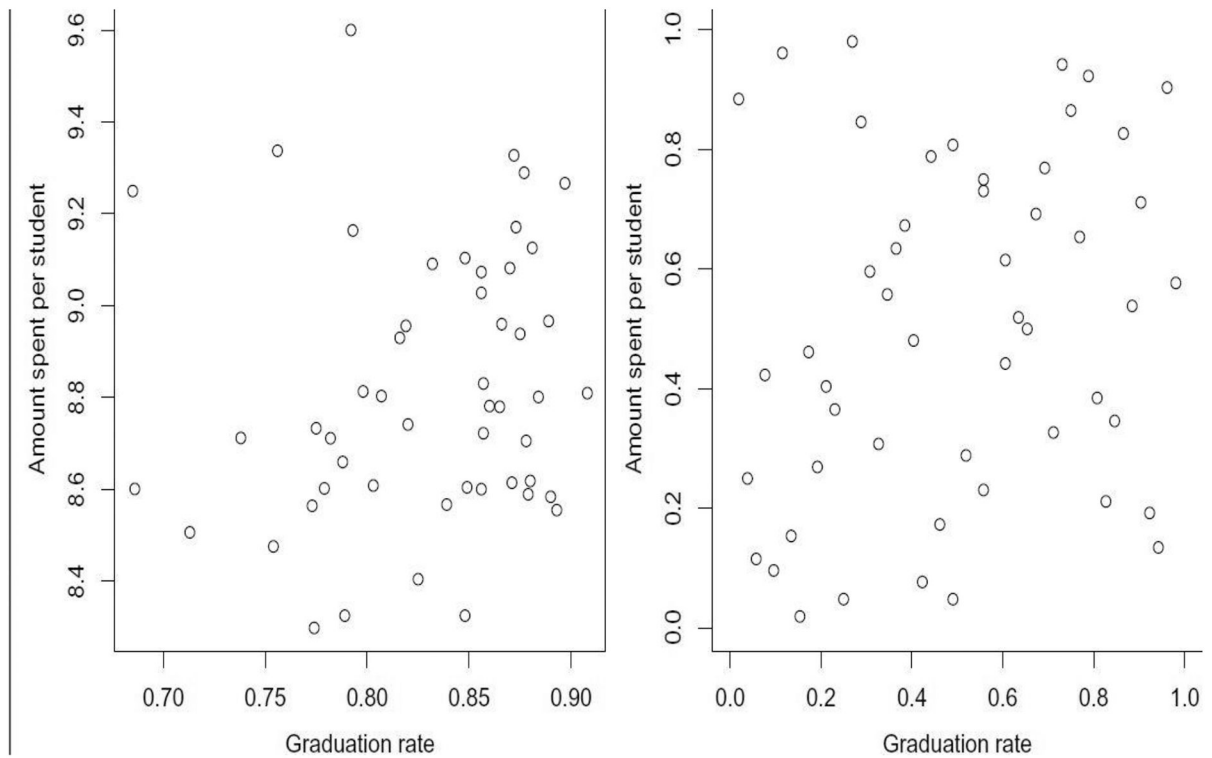
Fig. 3.
Education data. Left: Scatterplot of amount spent per student (in log US dollars) in 2015–2016 versus high school graduation rate in 2016. Public school data are included from each state and the District of Columbia ($n = 51$). Right: Scatterplot of psuedo-observations.

Fig. 4.
Stock data. Top: Pairwise scatterplots of $n = 124$ log returns for Apple, Google, and Walmart during January 4, 2016 through June 30, 2016. Bottom: Scatterplot of psuedo-observations.

Table 1

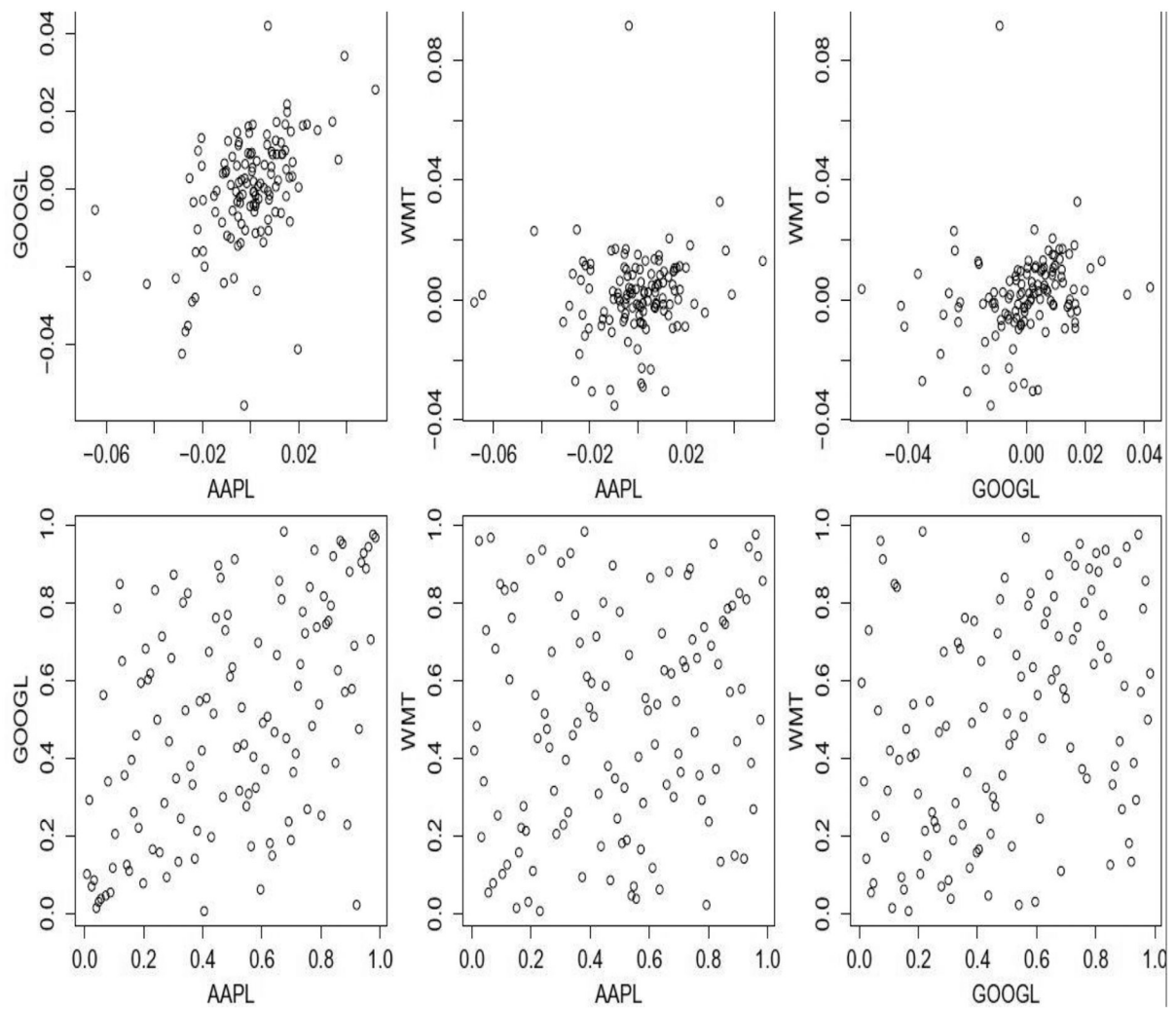Simulation study. Estimated size and power of six tests when $n = 100$ and $\alpha = 0.05$. All results are based on $B = 10000$ Monte Carlo data sets. For the Clayton, Gumbel, and Frank copulas, $\mathscr{H}_0$ is true when $T = 0$. For the Gaussian copula, $\mathscr{H}_0$ is true when $\rho = 0$. The margin of error in the size estimates, assuming a 99% confidence level, is approximately 0.006. The same tables for sample sizes $n = 50$ and $n = 200$ are shown in the supplementary materials.

	Test	$T = 0$	0.10	0.20	0.30	0.40
Clayton	EL	0.052	0.457	0.923	0.999	1.000
	KS	0.053	0.311	0.780	0.983	0.999
	CvM	0.053	0.407	0.886	0.996	1.000
	AD	0.050	0.366	0.877	0.996	1.000
	Spearman	0.051	0.440	0.902	0.997	1.000
	Kendall	0.049	0.432	0.898	0.997	1.000
Frank	EL	0.049	0.426	0.902	0.998	1.000
	KS	0.050	0.342	0.815	0.990	1.000
	CvM	0.050	0.429	0.905	0.998	1.000
	AD	0.049	0.314	0.817	0.992	1.000
	Spearman	0.049	0.441	0.914	0.998	1.000
	Kendall	0.048	0.435	0.913	0.998	1.000
Gumbel	EL	0.049	0.446	0.902	0.997	1.000
	KS	0.051	0.326	0.780	0.978	0.999
	CvM	0.050	0.428	0.892	0.996	1.000
	AD	0.050	0.506	0.919	0.997	1.000
	Spearman	0.049	0.434	0.898	0.997	1.000
	Kendall	0.049	0.432	0.896	0.997	1.000
	Test	$\rho = 0$	0.10	0.20	0.30	0.40
Gaussian	EL	0.048	0.237	0.599	0.897	0.991
	KS	0.054	0.189	0.445	0.740	0.937
	CvM	0.051	0.234	0.571	0.875	0.986
	AD	0.048	0.198	0.520	0.850	0.983
	Spearman	0.049	0.245	0.605	0.897	0.991
	Kendall	0.048	0.239	0.599	0.893	0.991

Table 2

Data analysis. Test statistics, critical values, and p-values for the twins data (Section 4.1), education data (Section 4.2), and stock data (Section 4.3).

		Test	**Test statistic**	**Critical value**	**p-value**
Twins data		EL	13.176	1.374	<0.001
		KS	1.465	0.683	<0.001
		CvM	0.756	0.067	<0.001
		AD	23.476	3.163	<0.001
		Spearman	0.559	0.135	<0.001
		Kendall	0.421	0.091	<0.001
Education data		EL	1.144	1.433	0.096
		KS	0.470	0.672	0.330
		CvM	0.083	0.091	0.068
		AD	2.773	4.716	0.224
		Spearman	0.215	0.235	0.068
		Kendall	0.150	0.159	0.062
		Test	Test statistic	Critical value	p-value
Stock data	APPL v GOOGL	EL	8.253	1.403	<0.001
		KS	1.018	0.679	<0.001
		CvM	0.381	0.069	<0.001
		AD	17.420	3.367	<0.001
		Spearman	0.476	0.148	<0.001
		Kendall	0.341	0.100	<0.001
	APPL v WMT	EL	1.410	1.403	0.049
		KS	0.662	0.679	0.059
		CvM	0.089	0.069	0.021
		AD	3.801	3.367	0.032
		Spearman	0.155	0.148	0.042
		Kendall	0.112	0.100	0.031
	GOOGL v WMT	EL	4.266	1.403	<0.001
		KS	1.028	0.679	<0.001
		CvM	0.247	0.069	<0.001
		AD	6.863	3.367	0.002
		Spearman	0.336	0.148	<0.001
		Kendall	0.236	0.100	<0.001