Dependence and Independence: Structure and Inference

Journal:	*Statistical Methods in Medical Research*
Manuscript ID:	SMM-14-0311.R2
Manuscript Type:	Review Article
Keywords:	empirical likelihood, random effect, test of independence, Pearson's correlation coefficient, rank test
Abstract:	Evaluations of relationships between pairs of variables, including testing for independence, are increasingly important. Erich Leo Lehmann noted that ``the study of the power and efficiency of tests of independence is complicated by the difficulty of defining natural classes of alternatives to the hypothesis of independence''. This paper presents a general review, discussion and comparison of classical and novel tests of independence. We investigate a broad spectrum of dependence structures with/without random effects, including those that are well-addressed in both the applied and the theoretical scientific literatures as well as scenarios when the classical tests of independence may break down completely. Motivated by practical considerations, the impact of random effects in dependence structures are studied in the additive and multiplicative forms. A novel index of dependence is proposed based on the area under the Kendall plot. In conjunction with the scatterplot and the Kendall plot, the proposed method provides a comprehensive presentation of the data in terms of graphing and conceptualizing the dependence. We also present a graphical methodology based on heat maps to effectively compare the powers of various tests. Practical examples illustrate the use of various tests of independence and the graphical representations of dependence structures.

Dependence and Independence: Structure and Inference

Albert Vexler,[1] **Xiwei Chen**[1] **and Alan Hutson**[1]

Abstract

Evaluations of relationships between pairs of variables, including testing for independence, are increasingly important. Erich Leo Lehmann noted that "the study of the power and efficiency of tests of independence is complicated by the difficulty of defining natural classes of alternatives to the hypothesis of independence". This paper presents a general review, discussion and comparison of classical and novel tests of independence. We investigate a broad spectrum of dependence structures with/without random effects, including those that are well-addressed in both the applied and the theoretical scientific literatures as well as scenarios when the classical tests of independence may break down completely. Motivated by practical considerations, the impact of random effects in dependence structures are studied in the additive and multiplicative forms. A novel index of dependence is proposed based on the area under the Kendall plot. In conjunction with the scatterplot and the Kendall plot, the proposed method provides a comprehensive presentation of the data in terms of graphing and conceptualizing the dependence. We also present a graphical methodology based on heat maps to effectively compare the powers of various tests. Practical examples illustrate the use of various tests of independence and the graphical representations of dependence structures.

Keywords

data-driven test, dependence measure, empirical likelihood, Pearson's correlation coefficient, random effect, rank test, test of independence

1 Introduction

In statistical analyses of medical data, exploring dependence between variables plays a fundamental role. For example, it is important to detect and quantify the dependence between the disease status and a variety of potential predictors to determine significant risk factors. In genetic epidemiology, genome-wide association studies are commonly conducted in order to examine the dependence between common genetic variants, e.g. single-nucleotide polymorphisms, and traits like major diseases. The study of dependence was developed extensively at the beginning of the 19^{th} century by Karl Pearson and George Yule[1] under multivariate normal assumptions and most often referred

[1]Department of Biostatistics, University at Buffalo, Buffalo, NY 14226, USA.
Corresponding author:
Albert Vexler, Department of Biostatistics, University at Buffalo, Buffalo, NY 14226, USA
Email: avexler@buffalo.edu

to as the theory of correlation.[2] Since that time various nonparametric tests have been proposed with the exchangeable, although not always technically correct, usage of the terms correlation and dependence. For the purpose of this article, we define dependence to be the case where the bivariate distribution function $F(x, y) \neq F_X(x)F_Y(y)$ for some $(x, y) \in \mathbb{R}^2$.

Classical measures, such as the Pearson, Spearman and Kendall rank correlation coefficients, and accompanying tests of independence technically target at certain dependence structures. For example, Pearson's correlation coefficient[3] measures the degree of the linear correlation while Spearman correlation coefficient[4] focuses on a monotone relationship between two random variables, both of which may fail in the broader sense when the underlying dependence structure is non-monotone.[5] Furthermore, Pearson and Spearman correlation coefficients may not be well-suited for measuring dependence in several cases, e.g. in the models $Y = 1/X$ or $Y = \varepsilon/X^k$, $k = 1, 2$, where ε is a random variable and X, Y are dependent in an inverse manner. In the second case, $E(XY)$ may not even exist. The Kendall rank correlation coefficient quantifies the concordance in ranks between pairs of random variables[6,7] but shows relatively low power in many cases as compared to Pearson and Spearman correlation coefficients.[8] In this lies the difficulty of interpreting the classical correlation coefficients as general dependence measures. Kendall and Stuart claimed that "In general, the problem of joint variation is too complex to be comprehended in a single coefficient".[2] In practice, the dependence structure may be more complex than those detectable by classical correlation coefficients. For example, in dose-response studies, the dependence between the change in effect on an organism's metabolism and levels of exposure (or doses) of a stressor (e.g. amount of a drug) may be described by a polynomial[9] or U-shaped curve[10]. More recently an emphasis has been placed on developing methods that go beyond the limitations of classical approaches towards capturing a broad spectrum of dependence structures. For example, rank-based tests of independence are proposed against a wide class of alternatives assuming an exponential family of distributions, which have great power stability and are sensitive to both grade linear correlation and grade correlations of higher-order polynomial.[11] Einmahl et al. proposed a test of independence based on the empirical likelihood (EL) methodology,[12] which has greater power than the corresponding Cramer-von Mises type statistics.[13] Vexler et al. developed a highly efficient nonparametric likelihood ratio test that has very favorable and robust power properties against linear and non-monotone forms of dependence.[14]

Once the existence of dependence is established, it is important to quantify dependence in an informative and interpretable fashion. A wide range of dependence measures have been developed in addition to those classical correlation coefficients described above. For example, Hoeffding's D

measure describes nonlinear dependence.[15] The maximal information coefficient (MIC) provides similar scores to different dependence structures with equal noise.[16] In practice, it is also desirable to visualize the dependence structure. In contrast with formal tests, which provide at best a single piece of information about a single form of association, graphical tools provide a rich source of information about dependence.[17] A traditional scatterplot of the raw data is widely used as a graphical tool for examining dependence between two variables. However, when we visualize dependence structures via a scatterplot, the null model of independence is a random scatter of points and is difficult to characterize. In order to manifest independence in a characteristic manner, Genest and Boies proposed a Kendall plot (or K-plot for short) by adapting the concept of a probability plot[18] to aid in the detection of dependence.[19] The K-plot has many advantages, including the property of invariance with respect to monotone transformations of the marginal distributions, easy interpretation, and extendibility to the multivariate context. In a similar manner to the receiver operating characteristic (ROC) curve methodology,[20,21] we propose a dependence measure by calculating the area under the K-plot (AUK). The proposed AUK measure, in conjunction with the scatterplot and the K-plot, provides a consistent unification of quantification and a graphical presentation of dependence.

It should be emphasized that when random effects exist in the model structure they should be accounted for when examining dependence. Random effects, which may arise from a variety of sources such as instrumentation specificity or biological variations, are common in practice and should be considered to avoid inconsistent and/or invalid statistical inference. For instance, Gu et al. investigated the inconsistency in results of different studies in the association between lung cancer and two polymorphisms rs1051730, rs8034191 and suggested considering random effects to address heterogeneity and publication bias.[22] Furthermore, Rakovski et al. showed that the power of testing dependence of DNA methylation between patient groups can be improved when the random effects due to sampling variation is taken into account.[23]

With a variety of tests and measures of dependence available, an inevitable question may arise regarding how to choose an appropriate one. In the statistical literature, comparisons of powers of tests of independence are rarely performed.[11] Toward this end, we compare powers of several classical methods and some nonparametric approaches under a variety of complex dependence structures, focusing on their usefulness in practice. The following aspects are investigated at the interface between statistical methodology and areas of application: (1) suggest the choice of nonparametric test of independence to use under different dependence structures as illustrated via scatterplots; (2) point out situations where classical methods may break down; (3) demonstrate the

influence of various random effects may have on the test of independence; (4) show an efficient way to visualize comparisons of powers of different tests; and (5) add additional situations to classically described scenarios such as some particular forms of heteroscedasticity. Using the paper results, practitioners can easily choose an efficient test of independence.

The paper is organized as follows. Section 2 presents various tests of independence. Section 4 introduces a broad spectrum of structures of dependence. Dependence measures that can be applied in general cases are presented in Section 3. In Section 5, we conduct simulation studies to compare powers of the tests of independence. Application of tests and measures of dependence to real data examples will be discussed in Section 6. Section 7 presents a broader discussion on choosing appropriate tests of independence. Additional discussions and one more data example regarding vascular endothelial growth factor expression are presented in the supplemental material (thereafter SM for abbreviation).

2 Tests of Independence

Assume we obtain a random sample of n independent and identically distributed (i.i.d.) pairs of observations (X_1, Y_1), ..., (X_n, Y_n) from a continuous bivariate population. Let $F_X(x)$ and $F_Y(y)$ denote the marginal distribution functions of X and Y, respectively, and let $F_{XY}(x, y) = \Pr(X \leq x, Y \leq y)$ denote the joint distribution function of the (X, Y) pairs. The hypothesis testing for bivariate independence between X and Y can be formally stated as

$$H_0 : F_{XY}(x, y) = F_X(x)F_Y(y), \text{ for all } (x, y) \in \mathbb{R}^2,$$

against $H_a : F_{XY}(x, y) \neq F_X(x)F_Y(y)$, for some $(x, y) \in \mathbb{R}^2$. Throughout this paper, we use the following notations. Let R_i and S_i denote the rank of X_i and Y_i, $i = 1, \ldots, n$, respectively, and F_{Xn}, F_{Yn} and F_n be the empirical distribution functions of F_X, F_Y and F_{XY}, respectively.

2.1 Classical Methods

In this section, we briefly review three well-known correlation coefficients. Pearson's correlation coefficient was developed to measure the linear dependence between a pair of variables (X, Y) under bivariate normality assumptions. It is defined in terms of moments as $\rho = \rho(X, Y) = \sigma_{XY}/(\sigma_X\sigma_Y)$, where σ_{XY} represents the covariance of X and Y, and σ_X, $\sigma_Y > 0$ denote the standard deviation of X and Y, respectively. By applying the Cauchy-Schwarz inequality to the σ_{XY}, it can be obtained that $-1 \leq \rho \leq 1$, where $\rho = 1$ indicates a perfect increasing linear relationship and $\rho = -1$ shows a perfect decreasing linear relationship. Spearman's rank correlation coefficient ρ_S is Pearson's correlation between ranks of X and Y, i.e. $\rho_S = \rho_S(X, Y) = \rho(F_X(X), F_Y(Y))$, $-1 \leq \rho_S \leq 1$. It accesses the monotonic relationship between two variables. Kendall's rank correlation coefficient has

the form of $\tau = \Pr((X_1 - X_2)(Y_1 - Y_2) > 0) - \Pr((X_1 - X_2)(Y_1 - Y_2) < 0)$, $-1 \leq \tau \leq 1$. It measures the strength of monotonic dependence. Under those specific dependence structures, the null hypothesis of independence can be rejected for large values of the absolute values of corresponding correlation coefficients. For details regarding estimators of correlation coefficients described above as well as the asymptotic distribution, we refer the reader to Section 1 of SM as well as Balakrishnan and Lai[24].

2.2 Data-Driven Rank Tests

In order to take into account not only linear correlation but also correlations of higher-order polynomials of random variables, Kellenberg and Ledwina[11] used approximations to the joint distribution of $(F_X(X), F_Y(Y))$ via Fourier coefficients to propose data-driven rank tests of independence between X and Y. In this case the joint distribution of $(F_X(X), F_Y(Y))$, called the copula function[25] or the grade representation[26], is assumed to be in an exponential family, which is defined with growing dimension and with sufficient statistics in terms of Legendre polynomials. Then the test of independence converges to the powerful maximum likelihood ratio hypothesis testing regarding parameters in a Neyman-Pearson[27] manner. Note that two random variables X and Y are independent if and only if $cov(g(X), h(Y)) = 0$ for all functions g and h ranging over a separating class of functions.[28]. By means of Legendre polynomial approximations to functions g and h, the data-driven approach, which essentially origins from Neyman and is the basis for the smooth tests, provides tests with very high power levels where the degree of the polynomials involved is determined by the data.

Define $X^* = F_X(X)$ and $Y^* = F_Y(Y)$. First consider for the joint distribution $Pr\{X^* < x^*, Y^* < y^*\}$ of (X^*, Y^*) in an exponential family restricted to the "diagonal", which is symmetric in x^* and y^* and contains only elements $\theta_j b_j(x^*) b_j(y^*)$, i.e. products of the parameter θ_j and Legendre polynomials with the same order in both variables, where b_j, $j = 1, \ldots, k$ denotes the j^{th} orthonormal Legendre polynomial with order k (see Section 2 in SM for details). It is of particular interest if the correlation of $\{X^*\}^r$ and $\{Y^*\}^s$ is the same as that of $\{X^*\}^s$ and $\{Y^*\}^r$. For most symmetric distributions used in practice, restriction to the diagonal models is sufficient.[11] In this case, the null hypothesis corresponds to $\boldsymbol{\theta} = 0$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T$. Replacing the unknown distribution functions F_X and F_Y by their corresponding empirical distribution functions in the score test statistic (see Section 2 in SM for details) and then applying a correction for continuity, smooth test statistic statistics can be obtained in the form

$$T_k = \sum_{j=1}^{k} V(j,j), \text{ with } V(r,s) = \left\{ n^{-\frac{1}{2}} \sum_{i=1}^{n} b_r\left(\frac{R_i - 1/2}{n}\right) b_s\left(\frac{S_i - 1/2}{n}\right) \right\}^2 .$$

According to the modified Schwarz's rule (similar to the one presented in Inglot et al.),[29] a "diagonal" test statistic $TS2 = T_{S2}$ can be obtained where the order is chosen as

$$S2 = \arg\min_k \{T_k - k\log(n) \geq T_j - j\log(n),\ 1 \leq j,k \leq d(n)\},$$

and $d(n)$ is a sequence of numbers, $d(n) \to \infty$ as $n \to \infty$. To simplify notation, Kellenberg and Ledwina[11] gave the test statistic only in the case $d(n) = 2$. We reject H_0 for large values of $TS2$. On the other hand, in the more general "mixed" product case without assuming the aforementioned symmetry, we consider k-dimensional exponential families containing always $b_1(x^*)b_l(y^*)$ with the index $[1,1]$, and $k-1$ other products $b_i(x^*)b_j(y^*)$ with the index $[i,j]$, $i,j = 1,\ldots d(n)$, $(i,j) \neq (1,1)$ (see Section 2 in SM for details). Let Λ denote a set of indices, $T_\Lambda = \sum_{(r,s)\in\Lambda} V(r,s)$ and $|\Lambda|$ denote the cardinality of Λ. We search for a model

$$\Lambda^* = \arg\max_\Lambda \{T_\Lambda - |\Lambda|\log(n)\}.$$

Then we reject H_0 for large values of the "mixed" statistic $V = T_{\Lambda^*}$. If Λ^* is not unique, the first among those Λ^*s that have the smallest cardinality is chosen. For details and discussions regarding data-driven rank tests, we refer the reader to Section 2 of SM.

2.3 Empirical Likelihood-Based Method

Statistical literature shows that if certain key assumptions are met, parametric likelihood ratio tests provide the most powerful statistical decision rules.[27] When these key assumptions are not met, parametric approaches may be suboptimal as compared to their robust counterparts across the many features of statistical inferences. Thus, it is important to preserve efficiency through the use of robust likelihood type methods, while concurrently minimizing assumptions about underlying distributions. To this end, various "approximate" likelihood-based techniques are developed by employing empirical likelihood (EL) $L(\tilde{F}) = \prod_{i=1}^{n} \tilde{P}\{(X_i, Y_i)\}$, where $\tilde{P}\{.\}$ is the probability measure corresponding to the distribution function $\tilde{F}$. As a pioneer of nonparametric likelihood-based tests of independence, Einmahl et al.[13] constructed a very powerful test statistic by integrating the localized EL over variables. Einmahl et al.[13] considered the maximum EL ratio

$$R(x,y) = \sup\left\{L(\tilde{F}) : \tilde{F}(x,y) = \tilde{F}_1(x)\tilde{F}_2(y)\right\} / \sup\left\{L(\tilde{F})\right\}$$

to show that the local log-EL ratio test statistic is

$$\log R(x,y) = nP_n(A_{11})\log\frac{F_{Xn}(x)F_{Yn}(y)}{P_n(A_{11})} + nP_n(A_{12})\log\frac{F_{Xn}(x)\left(1-F_{Yn}(y)\right)}{P_n(A_{12})}$$
$$+ nP_n(A_{21})\log\frac{\left(1-F_{Xn}(x)\right)F_{Yn}(y)}{P_n(A_{21})} + nP_n(A_{22})\log\frac{\left(1-F_{Xn}(x)\right)\left(1-F_{Yn}(y)\right)}{P_n(A_{22})},$$

for $(x, y) \in \mathbf{R}^2$ (see Section 3 of SM for more details). Here P_n is the empirical probability measure, and $A_{11} = (-\infty, x] \times (-\infty, y]$, $A_{12} = (-\infty, x] \times (y, \infty)$, $A_{21} = (x, \infty) \times (-\infty, y]$, $A_{22} = (x, \infty) \times (y, \infty)$ and $0 \log(\cdot/0) = 0$. The distribution-free test statistic T_n is then constructed by integrating the local log-EL ratio over variables $(x, y) \in \mathbf{R}^2$, where

$$T_n = -2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \log R(x, y) dF_{Xn}(x) dF_{Yn}(y).$$

We reject H_0 for large values of T_n. For details regarding the computation of the test statistic and the critical value, see Section 3 of SM.

2.4 Density-Based Empirical Likelihood Ratio Test

In Section 2.3, the classical EL approach was introduced. This EL test of independence is distribution-based.[30] According to the Neyman-Pearson lemma, parametric density-based likelihood ratio tests provide uniformly most powerful decision-making rules. In this section, we present a distribution-free test statistic via the density-based EL method proposed by Vexler et al., which approximates the parametric Neyman-Pearson statistic, to test the null hypothesis of bivariate independence against a wide class of alternatives.[14] The density-based EL test statistic is

$$VT_n = \prod_{i=1}^{n} n^{1-\beta_2} \tilde{\Delta}_i \left([0.5n^{\beta_2}, 0.5n^{\beta_2}] \right),$$

where the function $[x]$ denotes the nearest integer to x,

$$\tilde{\Delta}_i(m, r) = \left(F_{Xn}(X_{(R_i+r)}) - F_{Xn}(X_{(R_i-r)}) \right)^{-1} \left(F_n(X_{(R_i+r)}, Y_{(i+m)}) \right.$$
$$\left. - F_n(X_{(R_i-r)}, Y_{(i+m)}) - F_n(X_{(R_i+r)}, Y_{(i-m)}) + F_n(X_{(R_i-r)}, Y_{(i-m)}) + n^{-\beta_1} \right),$$

and $\beta_1 \in (0, 0.5)$ and $\beta_2 \in (0.75, 0.9)$. We reject H_0 for large values of $\log(VT_n)$. The power of the density-based EL test does not depend significantly on the choice of β_1 and β_2. For more details, see Section 4 of SM.

Remark 1. The data-driven techniques, the EL based method and the density-based EL ratio test described above are all exact. Their critical values can be calculated numerically using Monte Carlo techniques. For example, the critical values of the density-based EL ratio test can be determined by drawing 50,000 samples of $X_1, \ldots, X_n \sim Uniform[0, 1]$ *and* $Y_1, \ldots, Y_n \sim Uniform[0, 1]$ *and then calculating the* $1 - \alpha$ *quantile of generated values of the test statistic* $\log(VT_n)$ *at each sample size* n*, where* α *is the significance level.*

3 Indices of Dependence

Indices of dependence indicate how closely the variables X and Y are related in some particular manner. Among these indices, classical correlation coefficient-based measures of dependence, especially Pearson's correlation coefficient, are by far the most prominent. These classical measures

of dependence are geared towards specific dependence structures.[24] However, more complex forms of dependence lie outside the scope of these measures. In order to capture more general dependence structures, we introduce the maximal information coefficient[16] and propose a novel measure of dependence based on the Kendell plot[19].

3.1 Classical Measures of Dependence

The classical correlation coefficient-based measures of dependence, as was discussed in detail in Section 2.1, are widely used but can only work well for linear/monotone dependence structures. We refer the reader to Section 2.1 above and Section 1 in SM for detailed discussions. Hoeffding's measure of dependence, D, is a rank based dependence measure that detects more general departures from independence.[15] It approximates a weighted sum over observations of chi-square statistics for 2-by-2 classification tables by setting each set of (X_i, Y_i), $i = 1, \ldots, n$ values as cut points. However, when ties occur, the Hoeffding's D measure may result in a smaller value.[31] For more details, see Section 5 of SM. These traditional dependence measure can be difficult (i) to decide when a particular value indicates association strong enough for a given purpose, and (ii) in a given situation, to weigh the losses involved in obtaining more strongly associated variables against the gains.[32] Therefore, it is suggested using functions of these traditional dependence measure that can be interpreted as the probability of making a wrong decision in certain situations, e.g. the cube of correlation coefficient, correlation ratio, maximal correlation (sup correlation), monotone correlations, as well as concordance measures such as Gini index and Blomqvist's β; we refer the reader to Balakrishnan and Lai[24] for more details.

3.2 The Maximal Information Coefficient

The maximal information coefficient (MIC)[16] is a measure of pair-wise dependence defined by constructing a grid to create bins on the scatterplot of (X, Y) variables and encapsulating dependence within the grid. It can detect general dependence structures.

For a grid G among all grids up to a maximal grid resolution, dependent on the sample size n, let $I_G = \sum_{(x,y)\in G} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$ denote the mutual information of the probability distribution induced on the boxes of G, where the probability of a box $p(x, y)$ is the proportion of data points falling inside the bin box. Define the characteristic matrix $M = (m_{x,y})$. For every ordered pair of integers (x, y), the maximum normalized mutual information achieved by any x-by-y grids G is $m_{x,y} = \max_G \{I_G / \log \min \{x, y\}\}$. Normalizing by $\log \min \{x, y\}$ ensures modified values between 0 and 1 in addition to a fair comparison between grids of different dimensions and therefore across different distributions. Then MIC is defined to be $\max_{xy<B(n)} \{m_{x,y}\}$, where $B(n)$ is a function of sample size and usually set to be $n^{0.6}$.

The value of MIC is between zero for independence and one for a noiseless functional relationship. The MIC measure focuses on equitability, i.e. relationships with similar noise levels may result in similar scores regardless of the type of relationship. However, MIC works only for bivariate data and may fail to detect some important scenarios, e.g. linear dependence with a certain amount of noisy data due to discrepancy in the concept of equitability.[33] The result represented in this paper related to the calculation of MIC are obtained via the function *mine* in the R package *minerva*.

3.3 Area Under the Kendall Plot

Genest and Boies proposed a Kendall plot (or K-plot for short) as a rank-based graphical tool to detect dependence, which adopts the concept of the probability plot (the P-P plot).[19] Similar to the standard P-P plot in which a lack of linearity indicates nonnormality of the distribution of a random variable, the K-plot is close to a straight line in the independent bivariate case. The amount of curvature in the K-plot is characteristic of the strength of dependence in the data. Similar to data-driven rank tests which express the dependence via Fourier coefficients of the copula function, the K-plot essentially characterize the underlying dependence structure based on the copula function. It plots the Kendall function $K(w) = \Pr(C(U,V) \leq w)$ versus $\Pr(UV \leq w)$, where U and V are independent uniform random variables on the interval $[0,1]$ and C is the underlying copula corresponding to the joint cumulative distribution function of (X,Y), i.e. $C(U,V) = F_{XY}(F_X^{-1}(U), F_Y^{-1}(V))$. The K-plot plot has the property of invariance with respect to monotone transformations of marginal distributions, and can be extended to the multivariate context.

We proposed a new index of dependence that borrows strength from the K-plot in $[0,1] \times [0,1]$ space.[19] To this end, we use principles related to the receiver operating characteristic (ROC) curve analysis that are extensively used in biostatistical literature. The ROC curves graphically represent differences between two distributions, e.g. the cumulative distribution functions F_X and F_Y of two populations. The ROC curve $R(t)$ can be defined as $R(t) = 1 - F_X(F_Y^{-1}(1-t))$, where $t \in [0,1]$. Accordingly, the area under ROC curve (AUC), which can be calculated by $\int_0^1 R(t)dt$, is defined as an indicator of a distance between two distributions.[20,34] It is a convenient way to compare two distributions based on the ROC curve since it places tests on the same scale where they can be compared for accuracy. For several details regarding the ROC methodology, see Section 6 of SM. In analogy to the AUC, we propose the area under the K-plot (AUK) as an index of dependence. Consider a bivariate sample (X_i, Y_i), $i = 1, \ldots, n$, we define a random variable H_i as

$$H_i = \hat{H}_n(X_i, Y_i) = \frac{1}{n-1} \sum_{j \neq i} I\{X_j \leq X_i, Y_j \leq Y_i\},$$

where $I\{\cdot\}$ is the indicator function. Note that the empirical distribution function $\hat{H}_n$ based on the (X_j, Y_j), $j \neq i$, converges to H as $n \to \infty$, where $H = F_{XY}$ is the bivariate cumulative distribution function of (X, Y). Order the H_i to get $H_{(1)} \leq \cdots \leq H_{(n)}$ and define $K(w) = \Pr(H(X, Y) \leq w)$. Under $H_0 : F_{XY} = F_X F_Y$, we can obtain $K_0(w) = \Pr(UV \leq w) = w - w\log(w), w \in [0, 1]$, where U and V are independent uniform random variables on the interval $[0, 1]$. Then by definition of the density of an order statistic, the expectation of the i^{th}th order statistic in a random sample of size n from the distribution K_0 of the H_i under the null hypothesis of independence can be expressed as

$$W_{i:n} = n\binom{n-1}{i-1}\int_0^1 w\left\{K_0(w)\right\}^{i-1}\left\{1 - K_0(w)\right\}^{n-i} dK_0(w).$$

A K-plot can be obtained via plotting the pairs $(W_{i:n}, H_{(i)})$, $1 \leq i \leq n$. For a large enough sample, the K-plot will be a plot of $K^{-1}\{K_0(w)\}$ versus w, $w \in [0, 1]$. The area under the K-plot (AUK) is $1 - \int_0^1 K(w)dK_0(w)$, i.e. $1 + \int_0^1 K(w)\log(w)dw$. In the context of the Frchet-Hoeffding copula bounds[24], it is clear that we have AUK=0 when $\tau(X, Y) = -1$, and AUK= 3/4 when $\tau(X, Y) = 1$. We refer the reader to Section 7 in SM for proofs.

In the case of an unknown joint distribution of (X, Y), the empirical cumulative distribution function can be used and the corresponding empirical area under the K-plot can be estimated. Several theoretical properties of the AUK measure as well as R code for obtaining the empirical AUK are presented in Section 7 of SM. Section 7 of SM shows that the AUK measure presents "more associated-ordering" for bivariate distributions. We refer the reader to Section 7 in SM for more details. We also provide a data example regarding vascular endothelial growth factor expression to show the case where AUK measures are very robust and efficient in Section 9 of SM.

4 Structures of Dependence

Researchers within the earlier scientific literature, e.g. Reshef et al.[16] and Johnson[35], thoroughly addressed the examination of a variety of dependence structures. Included in the coverage were linear, quadratic, cubic, reciprocal, logarithmic, trigonometric, and bivariate distributions such as bivariate normal distributions, bivariate Pearson distribution families, bivariate normal offset distributions, bivariate Morgenstern distributions, bivariate Placket distributions, bivariate Cauchy distributions, bivariate exponential power distribution families, copula families, and the reciprocal-normal distribution, among others. In this section, details about these dependence structures as well as random-effect-type dependencies will be presented.

4.1 Basic Structures

Regression analysis is one of the commonly used techniques that provide an estimate of the formulaic dependence between variables. For example, we can consider $Y = f(X) + \varepsilon$ in general, where the function $f(X) = E(Y|X)$ describes the formulaic dependence between X and Y and ε represents an unobserved random effect with mean zero conditioned on the variable X.

Linear dependence: Linear dependence is very common in our everyday life. For example, the average weights for humans in the population of American women of age 30 to 39 can be expected to be linearly dependent on their height. The simple linear model $Y = \alpha + \beta X + \varepsilon$ can be applied, where the linear dependence $f(X) = \alpha + \beta X$ between X and Y are considered.[36]

Quadratic dependence: In many settings, linear dependence between X and Y may not hold. For example, in the field of chemistry, it is often found that the yield of a chemical synthesis improves by increasing amounts for each unit increase in temperature at which the synthesis takes place. In this case, a quadratic dependence between X and Y can be assumed and a quadratic model can be proposed in the form of $Y = f(X) + \varepsilon$, where the conditional expectation of Y is a quadratic function of X and $f(X) = \beta_0 + \beta_1 X + \beta_k X^2$. The test of independence is conducted via the null hypothesis $H_0 : \beta_1 = 0, \beta_2 = 0$. In general, we can consider a polynomial regression model $Y = \beta_0 + \beta_1 X + \cdots + \beta_k X^k + \varepsilon$, where the dependence between X and Y is modeled as a k^{th} (k is a positive integer) order polynomial and ε is a random effect with mean zero conditioned on X.

Reciprocal dependence: In the case when the variable Y descends down to a floor, or ascends up to a ceiling as the variable X increases (e.g. approaches an asymptote), a reciprocal dependence between X and Y can be expected. For example, in economics, the marginal utility of each homogeneous unit decreases as the supply of units increases. In this case, a reciprocal model can be considered, where $f(X) = 1/X^k$ and the order k is a positive integer.

Logarithmic dependence: Let us next consider logarithmic growth, e.g. a phenomenon whose size or cost can be described as a logarithm function of some input. For example in microbiology, bacterial growth can be modeled as a logarithm function of the number of colony forming units. In this case, a linear-log model can be specified, where $f(X) = \alpha + \beta \log(X)$. This model is also commonly used in econometrics. e.g. the dependence of the consumption spending on income. Note that any logarithm base can be used since they are convertible by multiplying the function by a fixed constant.

Trigonometric dependence: Trigonometric models can be applied in modeling periodic phenomena. For example, tidal experts and meterologists use sine waves to help predict tides. Electromagnetic radiation also can be modeled with sine waves. There are many familiar trigonometric functions and different trigonometric models can be considered, e.g. the sinusoidal with non-Fourier frequency

$Y = \sin(9\pi x)$. In particular, we have two-dimensional spirals for curves winding around a fixed center point at a continuously increasing or decreasing distance, which are common in plants and in some animals. For example, chambers of the shell in the nautilus (a cephalopod mollusc) are arranged in a logarithmic spiral.[37] Given a modern understanding of fractals, a growth spiral can be viewed as a special case of self-similarity.[38] Two-dimensional spirals can be described via polar coordinates. For example, the above-mentioned logarithmic spiral can be described in parametric form as $X(\theta) = aexp(b\theta)\cos(\theta)$, and $Y(\theta) = aexp(b\theta)\sin(\theta)$ with the angle θ and real numbers a, b.

Bivariate distributions: Knowledge of the marginal distributions is known to be inadequate in terms of determining the joint distribution function. Instead of modeling $E(Y|X)$ as a function of X to consider the dependence structure between X and Y, we can consider the joint distribution of (X, Y) directly. A variety of joint distribution functions have been developed.[35,39,40]

The bivariate (or multivariate, for more than two random variables) normal distribution is the most useful joint distributions in probability. Assume that $(X, Y)^T$ follow the bivariate normal distribution with the mean vector $(\mu_X, \mu_Y)^T$ and the covariance matrix $\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$, where ρ is the correlation between X and Y, and σ_X, $\sigma_Y > 0$ denote the standard deviation of X and Y, respectively. the conditional expectation of Y given X is $\mu_Y + \rho(X - \mu_X)\sigma_Y/\sigma_X$. In this case, Pearson's correlation coefficient ρ is fully informative about their joint dependence and zero correlation is equivalent to independence. Note that the bivariate normal distribution is a special case of elliptical distributions. For other elliptical distributions, such as bivariate Pearson II distributions and Pearson VII distributions that share some properties of the bivariate normal distributions, the dependence structure is also fully determined by the correlation matrix[41]. When the data exhibit skewness, the multivariate skew-normal distribution may be considered, where the marginal densities are scalar skew-normal.[42] We refer the reader to Section 8 of SM for several important bivariate distributions in practice.

4.2 Random-effect-type Dependence

Random effects are common in practice and should be considered to avoid inconsistent and/or invalid statistical inference. For example, the sample Pearson's correlation coefficient calculated from a small sample may be totally misleading if not viewed in the context of its likely sampling error[24]. Random effects can represent different types of measurement error (ME), that may arise from a variety of sources, e.g. instrumentation specificity, biological variations and/or questionnaire based self-report data. For example, the observed response may be formulated as $Y = log(|X + \varepsilon|)$, where measurements of X contain additive random effect ε, or formulated as $Y = log(|X + \varepsilon_X|) + \varepsilon_Y$, where both X and Y contain additive random effect ε_X and ε_Y, respectively.

Standard regression procedures assume that the predictor variables are measured without error, and random effects inherent in the model are associated with the response variable only. For example, in the regression models considered in Section 4.1, additive random effects on Y are considered, i.e. $Y = Y_0 + \varepsilon$, where Y is the observed variable, $Y_0 = f(X)$ is the true value, and ε is a random variable with mean zero and independent of X.

In the case when predictor variables are measured with errors, a common situation in practice, estimation based on the standard assumption of no error leads to bias or inconsistency of statistical inferences. In this case, models that account for random effects in both predictor variables and response variables, e.g. errors-in-variables models, should be applied. In the two-variable case involving X and Y, random effects in either variable will result in conservative inferences in view of attenuated correlations in magnitude[43] and will reduce the power of the test of independence in linear models[44]. Various models and designs are developed to tackle the errors-in-variables issue. For example, the simple linear errors-in-variables model considers additive random effects in both the predictor variable and the response variable, i.e. $Y = \alpha + \beta X_0 + \varepsilon_Y$ and $X = X_0 + \varepsilon_X$, where X_0 and Y_0 are the unobserved true values for X and Y, respectively.

To illustrate, we show the scatterplots ($n = 1000$) of observed (X, Y) in Figure 1 for some dependence structures with no random effect and with additive normally distributed random effects with mean 0 and variance 0.25 or 2. In Figure 1, each row corresponds to the case where the true values $(X_0, Y_0)^T$ follow: (M3) the polynomial-type dependence with the order $k = 2$ (abbreviated as LinearQuadratic); (M4) Cubic; (M5) Reciprocal with the order $k = 1$; (M6) Reciprocal with the order $k = 2$; (M8) Logarithm $Y_0 = log(1 + |X_0|)$; and (M9) sinusoidal $Y_0 = \sin(\pi X_0)$; see in Table 2 in Section 4.1 for more details. Each column in Figure 1 displays various random effects, where (a) represents the case of no random effect in either X or Y, i.e. $X = X_0$, $Y = Y_0$; (b) and (b') show the case where X has no random effect and Y has additive effects, i.e. $X = X_0$, $Y = Y_0 + \varepsilon_Y$ with the variance of noise $var(\varepsilon_Y)$ assumed to be 0.25 (shown in (b)) or 2 (shown in (b')), respectively; (c) and (c') show the case where Y has no random effect and X has additive random effects, i.e. $Y = Y_0$, $X = X_0 + \varepsilon_X$ with the variance of noise $var(\varepsilon_X)$ assumed to be 0.25 or 2, respectively; (d) and (d') show the case where both X and Y have additive random effect, i.e. $X = X_0 + \varepsilon_X$, $Y = Y_0 + \varepsilon_Y$, with the variance of noise $var(\varepsilon_X) = var(\varepsilon_Y)$ assumed to be 0.25 or 2, respectively. Note that in Figure 1, scatterplots with different random effects are shown in the same scale for each type of dependence structures.

It can be observed that not only the variance of random effects, but also the type of random effects, affect the dependence structure. In addition, Table 1 shows the corresponding MIC and AUK

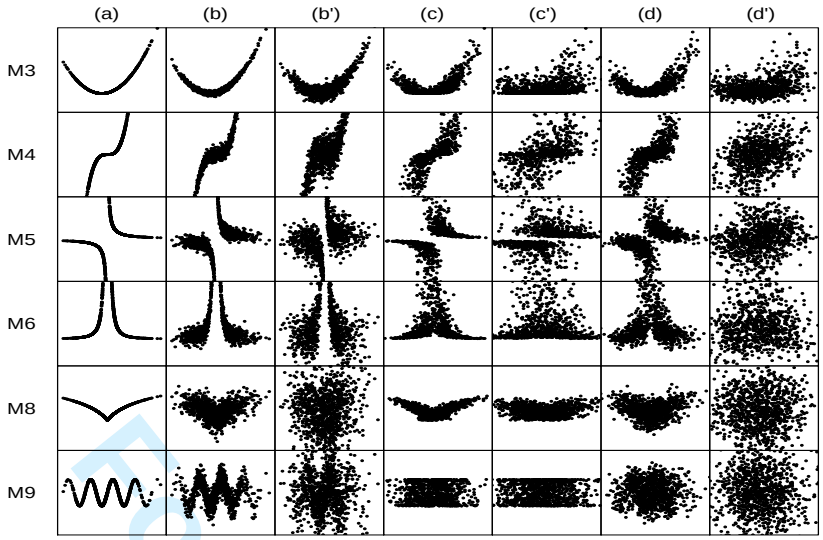


Figure 1: Scatterplots ($n = 1000$) of measurements of (X, Y) under different dependence structures with normally distributed random effects with mean 0, where the unobserved true values of X_0 and Y_0 can be formulated as $Y_0 = f(X_0)$. Each row corresponds to each of the following structures: (M3) the polynomial dependence with the order $k = 2$ (abbreviated as LinearQuadratic); (M4) Cubic; (M5) Reciprocal with the order $k = 1$; (M6) Reciprocal with the order $k = 2$; (M8) Logarithm $Y_0 = log(1 + |X_0|)$; and (M9) sinusoidal $Y_0 = \sin(\pi X_0)$. Each column displays various random-effect schemes, where (a) shows the case $X = X_0$, $Y = Y_0$; (b) and (b') show the case $X = X_0$, $Y = Y_0 + \varepsilon_Y$ with the variance of noise $var(\varepsilon_Y)$ assumed to be 0.25 or 2, respectively; (c) and (c') show the case $Y = Y_0$, $X = X_0 + \varepsilon_X$ with the variance of noise $var(\varepsilon_X)$ assumed to be 0.25 or 2, respectively; (d) and (d') show the case $X = X_0 + \varepsilon_X$, $Y = Y_0 + \varepsilon_Y$, with the variance of noise $var(\varepsilon_X) = var(\varepsilon_Y)$ assumed to be 0.25 or 2, respectively.

measures of dependence with $n = 1000$. We further show measures of MIC and AUK for sample size $n = 100$ under various dependence structures and random effects in Section 5. Note that the MIC can range from 0 for the independence case to 1 for the perfect dependence case, while the AUK can range from 0 for the perfect negatively dependent case to 0.75 for the perfect positively dependent case, where the value 0.5 indicates independence. When the variance of random effects increases, the MIC decreases showing comparatively weaker dependence while the AUK measures only vary by a small amount, showing a lightly weaker dependence. In addition, we present the scatterplots ($n = 1000$) of several bivariate distributions described in Section 4.1 in Figure 4 of SM. In the context of considered bivariate distributions, the presence of additive random effects in one variable of bivariate distributions does not change the dependence structure as much as the case where one variable is simple a function of the other variable as shown in Figure 1.

Instead of the additive random effect, the true value could also be modified proportionally, for

Structure	Measures	(a)	(b)	(b')	(c)	(c')	(d)	(d')
M3	MIC[0,1]	1	0.814	0.497	0.591	0.234	0.541	0.202
(LinearQuadratic)	AUK[0,0.75]	0.664	0.652	0.63	0.633	0.587	0.628	0.578
M4	MIC[0,1]	1	0.807	0.623	0.753	0.427	0.608	0.307
(Cubic)	AUK[0,0.75]	0.75	0.717	0.69	0.708	0.649	0.693	0.616
M5	MIC[0,1]	1	0.904	0.501	0.745	0.375	0.583	0.162
(Inverse)	AUK[0,0.75]	0.404	0.488	0.523	0.492	0.525	0.523	0.539
M6	MIC[0,1]	1	0.842	0.699	0.567	0.221	0.497	0.169
(Inverse2)	AUK[0,0.75]	0.377	0.38	0.395	0.431	0.472	0.422	0.476
M8	MIC[0,1]	1	0.32	0.185	0.589	0.207	0.28	0.124
(log)	AUK[0,0.75]	0.582	0.537	0.512	0.557	0.53	0.542	0.507
M9	MIC[0,1]	1	0.572	0.25	0.21	0.125	0.166	0.129
(sin)	AUK[0,0.75]	0.492	0.484	0.492	0.498	0.511	0.495	0.493
M10	MIC[0,1]	0.373	0.15	0.136	0.193	0.125	0.135	0.136
(Spiral)	AUK[0,0.75]	0.491	0.498	0.505	0.495	0.51	0.5	0.501

Table 1: Dependence measures of MIC and AUK under different dependence structures and normally distributed random effects with mean 0, where the unobserved true values of the predictor variable X_0 and the response variable Y_0 can be formulated as $Y_0 = f(X_0)$. Each row corresponds to each of the following structures: (M3) the polynomial-type dependence with the order $k = 2$ (abbreviated as LinearQuadratic); (M4) Cubic; (M5) Reciprocal with the order $k = 1$; (M6) Reciprocal with the order $k = 2$; (M8) Logarithm $Y_0 = log(1 + |X_0|)$; and (M9) sinusoidal $Y_0 = \sin(\pi X_0)$. Each column displays various random-effect schemes, where (a) shows the case $X = X_0$, $Y = Y_0$; (b) and (b') show the case $X = X_0$, $Y = Y_0 + \varepsilon_Y$ with the variance of noise $var(\varepsilon_Y)$ assumed to be 0.25 or 2, respectively; (c) and (c') show the case $Y = Y_0$, $X = X_0 + \varepsilon_X$ with the variance of noise $var(\varepsilon_X)$ assumed to be 0.25 or 2, respectively; (d) and (d') show the case $X = X_0 + \varepsilon_X$, $Y = Y_0 + \varepsilon_Y$, with the variance of noise $var(\varepsilon_X) = var(\varepsilon_Y)$ assumed to be 0.25 or 2, respectively. Note that the MIC can range from 0 for the independence case to 1 for the perfect dependence case, while the AUK can range from 0 for the perfect negatively dependent case to 0.75 for the perfect positively dependent case, where the value 0.5 indicates independence.

example, exposure by chemicals or radiation. In such cases, the multiplicative random effect can be assumed, i.e. $Y = \varepsilon Y_0$, where Y and Y_0 is the observed and unobserved truth variable, respectively, and the multiplicative noise ε is a random variable independent of X. Similarly, random effect of X can also present. As a concrete example, we consider the reciprocal dependence $f(X) = 1/X$ and the multiplicative random effects in the response variable. The dependence between X and Y can be formulated as $Y = \varepsilon/X$, where $var(Y|X) = var(\varepsilon)/X^2$. Also in the model $Y = \alpha + \beta X + \varepsilon/X$, the random effect affect the signal in a proportinally inverse manner. In such cases, random effects cause heteroscedasticity, a common situation in areas such as epidemiological studies[45], and extended least squares and/or weighted least squares techniques are often used for estimation[46]. Another example would be generalized linear mixed models[47]. In a special case of generalized linear mixed models, $Y = \alpha + \gamma X + \varepsilon$, where γ is a random variable independent of X and the random error ε, both additive and multiplicative random effects are considered. To illustrate the impact of

various types of random effects, we demonstrate the scatterplots ($n = 1000$) of several bivariate distributions including additive and/or multiplicative random effects in Figure 5 of SM. It can be seen that the shapes of the scatterplots are heavily affected by the types of random effects, even when the variance of random effect is relatively small. Note that parameter of the bivariate distributions can be affected by random effects. For example, when measurements of X and Y consist of multiplicative random effects γ, which follow the normal distribution with mean a and variance 1, and the underlying true values $(X_0, Y_0)^T \sim \text{MVN}_2(\mu, \Sigma)$, then the mean parameter of the observed measurements $(X, Y)^T$ is $a\mu$, affected by the random effects; so does the covariance matrix of the observed bivariate measurements.

5 Monte Carlo Comparisons of Tests of Independence

We carried out an extensive Monte Carlo study to compare powers of tests of independence under different dependence structures and various random effects at the 0.05 significance level. Classical methods such as Pearson's product moment correlation coefficient (P), Kendall's tau (K) and Spearman's rho (S), as well as nonparametric tests including empirical likelihood-based hypothesis testing (EMK), data-driven rank tests of test statistics TS and V in the cases $d(n) = 1, 2, 3$, and density-based empirical likelihood ratio test (dbEL) described in Section 2 were evaluated. Without loss of generality, in the density-based empirical likelihood ratio test, we set $\beta_1 = 0.45$ and $\beta_2 = 0.8$, respectively.

We attended to general forms of dependence,[16,35] which have been commonly pointed out with respect to bivariate data, including linear dependence, nonlinear dependence and bivariate distributions as shown in Table 2. The estimates of indices of dependence based on samples of sizes 1 000 are given in Table 3 of SM.

Motivated by the necessity to consider random-effect-type dependencies between two sets of observations, we investigate comparisons of above-mentioned tests of independence of X and Y via the Monte Carlo simulations, under the representative cases where X has no random effect and Y has the following situations: (1) no random effect in Y; (2) additive random effect in Y; (3) multiplicative random effect in Y; and (4) both additive and multiplicative random effects in Y. All random effects are assumed to follow the normal distribution with mean 0 and variance σ_ε^2; see Table 3 for details. For instance, when $f(X) = 1/X$ as considered in M5 in Table 2, the case that X has no random effect and Y has both additive and multiplicative random effects (as described as scenario (4) in Table 3) lead to the model $Y = \varepsilon_M/X + \varepsilon_A$.

Designs	Models/Description		Location of Results
	X	$f(X)$	
Linear			
M1(Linear)	$N(0,2)$	$0.5X$	SM
Nonlinear			
M2(Quadratic)	$N(0,2)$	$2+X^2$	SM
M3(LinearQuadratic)	$N(0,2)$	$2+X+X^2$	Main
M4(Cubic)	$N(0,2)$	X^3	SM
M5(Reciprocal $k=1$)	$N(0,2)$	$1/X$	Main
M6(Reciprocal $k=2$)	$N(0,2)$	$1/X^2$	Main
M7(Reciprocal $k=3$)	$N(0,2)$	$1/X^3$	SM
M8(Logarithm)	$N(0,2)$	$\log(1+\lvert X\rvert)$	Main
M9($\sin(\pi x)$)	$N(0,2)$	$\sin(\pi X)$	SM
M10(Spiral)	$exp(0.2\pi\theta)\cos(\pi\theta)$	$exp(0.2\pi\theta)\sin(\pi\theta)$, where $\theta\sim N(0,2)$.	SM
Bivariate Distributions			
M11(Ellipse)	an ellipse of a 95% probability region for $\text{MVN}_2(\mathbf{0},\left(\begin{smallmatrix}1 & 0.8\\ 0.8 & 1\end{smallmatrix}\right))$		SM
M12(Normal)	bivariate normal distribution $\text{MVN}_2(\mathbf{0},\left(\begin{smallmatrix}1 & 0.5\\ 0.5 & 1\end{smallmatrix}\right))$		SM
M13(NormOffset)	bivariate normOffset distribution		SM
M14(Morg)	bivariate Morgenstern distribution with $\alpha=1$		SM
M15(Placket)	bivariate Placket distribution with $\phi=3.5$		SM
M16(PearsonII)	bivariate Pearson Type II distribution with $\mu=\mathbf{0}$, $\Sigma=\mathbf{I}$ and $m=1.1$		SM
M17(PearsonVII)	bivariate Pearson Type VII distribution with $\mu=\mathbf{0}$, $\Sigma=\mathbf{I}$ and $m=1.1$		SM
M18(Cauchy)	bivariate Cauchy distribution		SM
M19(EP)	bivariate exponential power distribution		SM
M20(Gumbel)	two marginally normal distributions coupled with the Gumbel copula		SM
M21(Clayton)	two marginally normal distributions coupled with the Clayton copula		SM
M22(RecipN)	reciprocal-normal type distribution		Main
M23(CondN)	Conditional distributions are each normal and $f_{X,Y}(x,y)=Cexp\{-(x^2+y^2+2xy(x+y+xy))\}$		SM

Table 2: Dependence structures that considered in the main text or in SM.

Figures 2a through 4 show powers of different methods via graphs under different alternative relationships assuming $\sigma_\varepsilon^2=1$, over 10,000 simulations at the 5% nominal significance level for various sample sizes $n=10,25,50,100$, where the MIC and AUK indices of dependence are also presented in the caption corresponding to different dependence structures and random effects. Each figure can be divided into six sub-panels corresponding to scenarios (1) to (4) described in Table 3. In each sub-panel, the top level is the heat map and the bottom level shows the scatterplot (in the left) and K-plot (in the right). A heat map manifests itself with comparisons of powers with column

Scenarios	X	Y
(1)	$X=X_0$	$Y=Y_0$
(2)	$X=X_0$	$Y=Y_0+\varepsilon_A$
(3)	$X=X_0$	$Y=\varepsilon_M Y_0$
(4)	$X=X_0$	$Y=\varepsilon_M Y_0+\varepsilon_A$

Table 3: Random effect schemes considered for the Monte Carlo comparison of tests of independence between X and Y, where X_0 and Y_0 are unobserved true values. The additive random effects ε_A and multiplicative random effects ε_M follow the $N(0,1)$ distribution.

names presenting tests for dependence considered and row names presenting sample sizes. A dendrogram at the top side of the heat map clusters tests for dependence considered. A grey-scaled color bar shows the magnitude of powers on the right of the heat map. It can easily tell what methods perform the best and how similar methods are. As an example, in Table 4, we summarize considered tests based on powers in descending order at the 0.05 significance level for different sample sizes n for the M5 (Reciprocal with order $k = 1$) and the M9 ($\sin(\pi X)$) dependence structure, considered the above-mentioned random effects (1) to (4) described in Table 3. For each dependence structure and sample size, considered tests of independence are clustered in 3 clusters. Within each cluster, results are ordered in descending order in terms of powers as shown in parenthesis. Between clusters, results are ordered in descending order in terms of powers, separated by $>$. We further show the scenario of M3 (LinearQuadratic) in Section 11 of SM. For example, in the case of $Y = 1/X$ with additive random effects, when the sample size is 50, dbEL, EMK and data driven tests of statistics $TS2$ and V with $d(n) = 2, 3$ have the high power to detect dependence, while Pearson and Kendall tests break down, as also can be seen in Figure 2b. We considered other dependence structures in Section 10 of SM, such as the case $X \sim Uniform[-2, 2]$, $Y = \sin(\pi X)$, where it differs from M9 defined in Table 2 in the distribution of X. In this case, only the dbEL and EMK remains similar power when the underlying distribution of X changes.

6 Data Examples

6.1 TBARS

In this section, we illustrate the practical application and comparisons of different tests of independence with data from a study related to thiobarbituric acid-reactive substances (TBARS). Thiobarbituric acid-reactive substances (TBARS) is commonly used to summarize antioxidant status process of an individual in laboratory research[48], but its use as a discriminant factor between individuals with and without myocardial infarction (MI) disease is still controversial[49]. In the study investigating the discriminant ability of TBARS with regard to the MI disease, dependencies between TBARS and other antioxidant status related to the MI disease are evaluated. The sampling frame for adults between the ages of 35 and 65 was the New York State Department of Motor Vehicles drivers' license rolls. Also a randomly selected elderly sample (age $65 - 79$) from the Health Care Financing Administration database was taken.[49] Participants provided a 12-hour fasting blood specimen for biochemical analysis. A number of antioxidants were examined from fresh blood samples at baseline, including TBARS, high-density lipoprotein (HDL)-cholesterol and vitamin E (a fat-soluble antioxidant vitamin). We implemented tests of independence between TBARS and

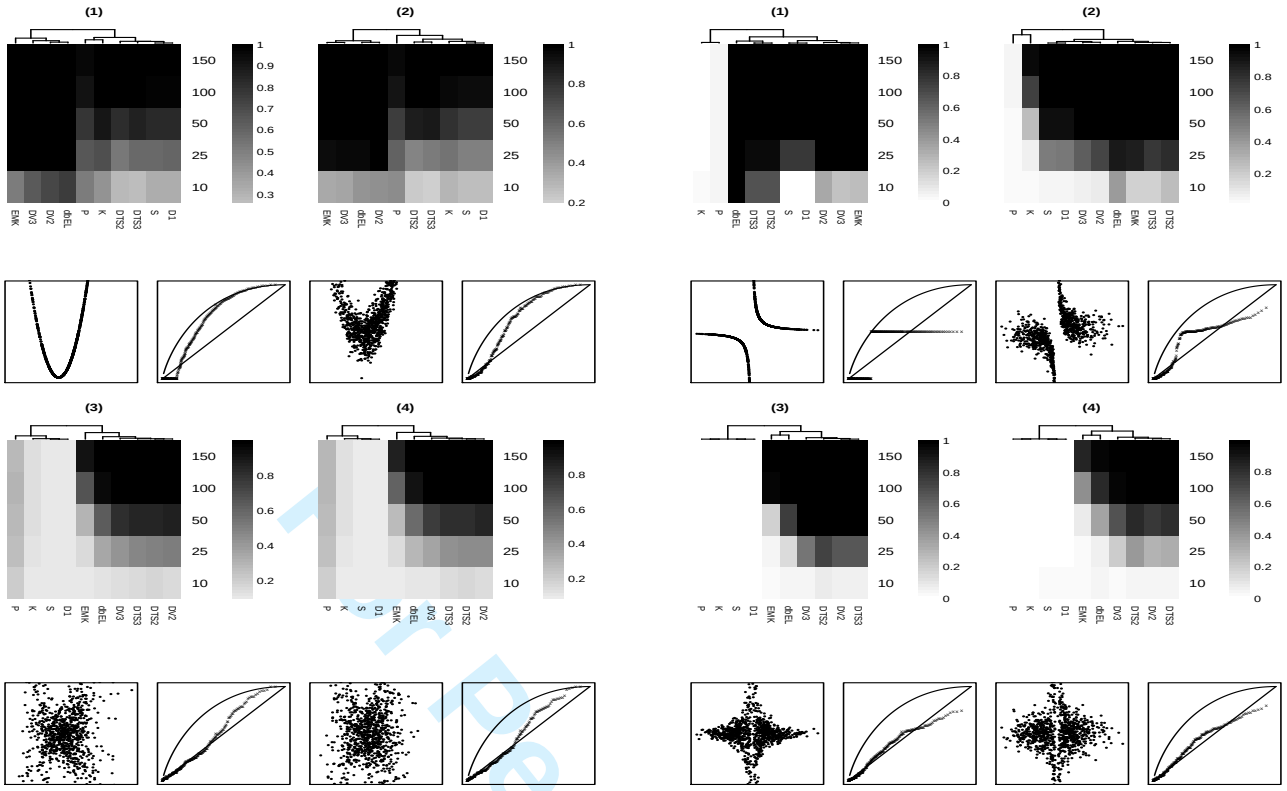
Designs (RE)	n	Clustering of the test powers (order within/between clusters) in descending order for different sample sizes (n)
M5 (Figure 2b): $f(X) = 1/X$		
(1)	25	(EMK, D_{V2}, D_{V3}, dbEL, D_{TS2}, D_{TS3}) > (S, D_1) > (P, K)
	50	(EMK, D_{V2}, D_{V3}, dbEL, D_{TS2}, D_{TS3}, S, D_1) > (P) > (K)
	100	(S, EMK, D_1, D_{TS2}, D_{V2}, D_{TS3}, D_{V3}, dbEL) > (P) > (K)
(2)	25	(dbEL, EMK, D_{TS2}, D_{TS3}, D_{V2}, D_{V3}, D_1, S) > (K) > (P)
	50	(dbEL, EMK, D_{V2}, D_{V3}, D_{TS2}, D_{TS3}) > (S, D_1) > (K, P)
	100	(EMK, D_{TS2}, D_{V2}, D_{TS3}, D_{V3}, dbEL, S, D_1) > (K) > (P)
(3)	25	(D_{TS2}, D_{TS3}, D_{V2}, D_{V3}) > (dbEL) > (EMK, D_1, S, K, P)
	50	(D_{TS2}, D_{TS3}, D_{V2}, D_{V3}, dbEL) > (EMK) > (S, D_1, K, P)
	100	(D_{TS2}, D_{V2}, D_{TS3}, D_{V3}, dbEL) > (EMK) > (D_1, S, K, P)
(4)	25	(D_{TS2}, D_{TS3}, D_{V2}) > (D_{V3}) > (dbEL, EMK, D_1, S, K, P)
	50	(D_{TS2}, D_{TS3}, D_{V2}, D_{V3}) > (dbEL) > (EMK, S, D_1, K, P)
	100	(D_{TS2}, D_{TS3}, D_{V2}, D_{V3}, dbEL) > (EMK) > (D_1, S, K, P)
M9 (Figure 11 in SM): $f(X) = \sin(\pi X)$		
(1)	25	(D_{V3}, EMK) > (dbEL, D_{V2}) > (K, D_{TS2}, D_1, S, D_{TS3}, P)
	50	(EMK, dbEL) > (D_{V3}) > (D_{V2}, K, D_{TS3}, D_{TS2}, S, D_1, P)
	100	(EMK, dbEL) > (D_{V3}) > (D_{V2}, K, D_{TS3}, D_{TS2}, D_1, S, P)
(2)	25	(D_{V3}, EMK, dbEL) > (D_{V2}, D_1, D_{TS2}, P, S) > (D_{TS3}, K)
	50	(dbEL) > (EMK, D_{V3}) > (D_{V2}, D_{TS3}, D_{TS2}, K, S, D_1, P)
	100	(dbEL) > (EMK) > (D_{V3}, D_{V2}, D_{TS3}, D_{TS2}, D_1, S, K, P)
(3)	25	(D_{TS3}, D_{V3}, D_{TS2}) > (D_{V2}, D_1, dbEL) > (S, P, EMK, K)
	50	(D_{TS2}, D_{TS3}, EMK, D_{V3}) > (P, S, D_1, D_{V2}, K) > (dbEL)
	100	(D_{TS2}, EMK, D_{TS3}) > (D_{V2}, D_{V3}, D_1, P, S, K) > (dbEL)
(4)	25	(EMK, D_{TS2}) > (D_1, D_{TS3}, P, S) > (dbEL, D_{V2}, K, D_{V3})
	50	(D_{V3}) > (EMK, D_{V2}, K, dbEL, D_{TS3}) > (D_{TS2}, S, D_1, P)
	100	(K, D_1, D_{V2}, D_{TS3}) > (S, EMK, D_{TS2}, D_{V3}) > (dbEL, P)

Table 4: Summarization of different tests based on powers in descending order at the 0.05 significance level for different sample sizes n, under the dependence structures M5 ($f(X) = 1/X$) and M9 ($f(X) = \sin(\pi X)$) as well as different random effects, where (1) represents no random effect in Y; (2) represents additive random effects in Y; (3) represents multiplicative random effects in Y; (4) represents both additive and multiplicative random effects in Y. For each dependence structure and sample size, considered tests for independence are clustered in 3 clusters. Within each cluster, results are ordered in descending order in terms of powers as shown in parenthesis. Between clusters, results are ordered in descending order in terms of powers, separated by the symbol >.

HDL-cholesterol, with a random sample of 100 individuals with the MI disease.

Figure 5a presents the scatterplot of the "TBARS" data in the left and the K-plot in the right. The estimates of MIC and AUK measures are 0.210 and 0.484, respectively.

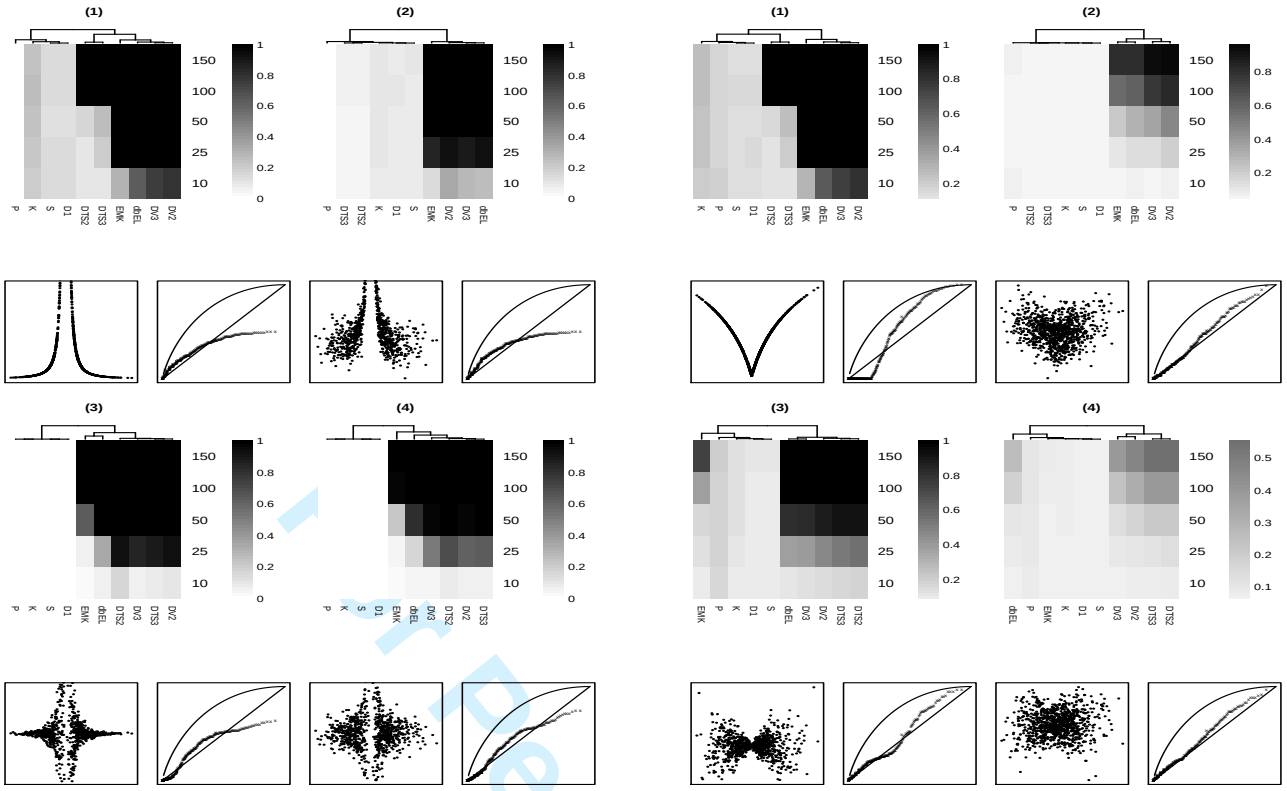
In accordance with ideas introduced by Stigler[50], we organized a jackknife type procedure to examine different Tests of Independence, including Pearson's product moment correlation coefficient (P), Kendall's tau (K), Spearman's rho (S), the empirical likelihood-based hypothesis testing (EMcK) method[13], data-driven rank tests[11], and density-based empirical likelihood ratio test (dbEL)[14]. The strategy was that a sample with size $n_{\text{sub}} = 20$, 30, 50, 70 and 100 was randomly selected from the data to be tested for independence at 5% level of significance. For each test of



(a) M3 (LinearQuadratic) (b) M5 (Reciprocal with $k = 1$)

Figure 2: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure M3 (LinearQuadratic) in the left panel (a) and M5 (Reciprocal with $k = 1$) in the right panel (b) as well as various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and K-plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. For M3 (LinearQuadratic), the measures of MIC and (AUK) for each type of random effects are (1) 1(0.664); (2) 0.664(0.639); (3) 0.276(0.534); and (4) 0.271(0.533). For M5 (Reciprocal with $k = 1$), the measures of MIC and (AUK) for each type of random effects are (1) 1(0.438); (2) 0.69(0.522); (3) 0.32(0.471); and (4) 0.278(0.483).

independence, we repeated this strategy $1,000$ times calculating the mean of p-values. As shown in Table 5, this jackknife type procedure showed that the empirical likelihood-based hypothesis testing method and density-based empirical likelihood ratio test have stable power property and diagnose dependency faster than other methods with respect to the sample size. The empirical likelihood-based hypothesis testing method and density-based empirical likelihood ratio test start to detect the dependence based on small samples, e.g. $n_{\text{sub}} = 70$. Among those classical correlation-based methods, Kendall's rank correlation test can be recommended. The Pearson's product moment correlation test completely fails.



(a) M6 (Reciprocal with $k = 2$) (b) M8 (Logarithm)

Figure 3: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure M3 (LinearQuadratic) in the left panel (a) and M5 (Reciprocal with $k = 1$) in the right panel (b) as well as various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and K-plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. For M6 (Reciprocal with $k = 2$), the measures of MIC and (AUK) for each type of random effects are (1) 1(0.407); (2) 0.756(0.416); (3) 0.429(0.458); and (4) 0.379(0.473). For M8 (Logarithm), the measures of MIC and (AUK) for each type of random effects are (1) 1(0.576); (2) 0.281(0.526); (3) 0.277(0.519); and (4) 0.237(0.511).

6.2 Periodontal Disease

The cross-sectional study[51] evaluated the association between radiographic evidence of alveolar bone loss and the concentration of host-derived bone resorptive factor interleukin-1 beta (IL-1β) and markers of bone turnover (osteonectin) in stimulated human whole saliva collected from 100 untreated dental patients. In order to investigate candidate salivary biomarkers associated with alveolar bone loss, researchers are often interested in testing the independence between IL-1β and osteonectin to establish a biological linkage. Figure 5b presents the scatterplot of the data in the left and the K-plot in the right. The estimates of MIC and AUK measures are 0.287 and 0.429,

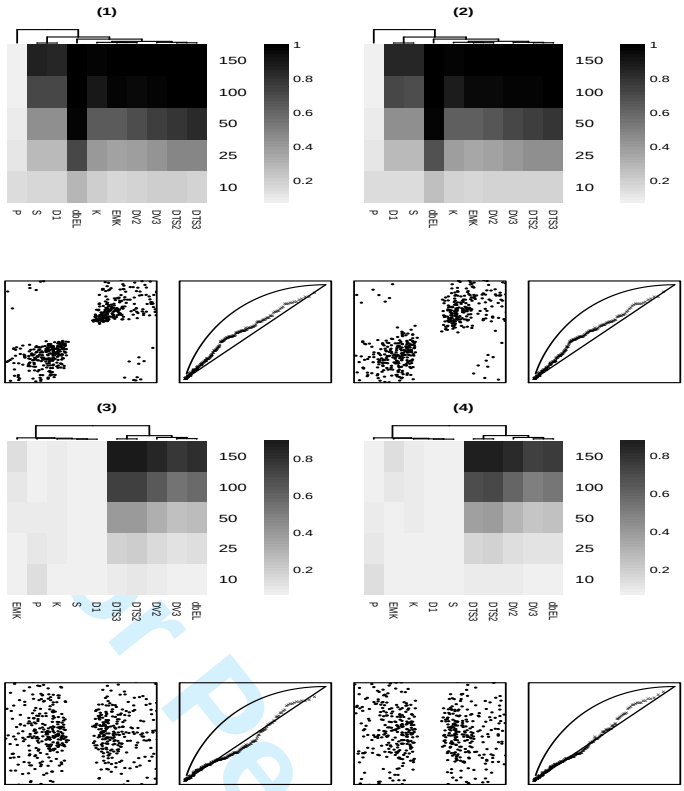


Figure 4: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure M24 (Reciprocal-Normal) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and K-plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.522(0.571); (2) 0.506(0.569); (3) 0.241(0.512); and (4) 0.241(0.512).

respectively. In a similar manner described in the "TBARS" data, a jackknife type procedure is conducted to examine different Tests of Independence with randomly selected subsamples of the sample size $n_{\text{sub}} = 20$, 30, 50, 70, 85 and 100, repeating $1,000$ times. As shown in Table 5, among those classical correlation-based methods, Kendall's rank correlation test and Spearman correlation test can be recommended. All methods but the data-driven test using V statistic with $d(n) = 2, 3$ can detect the dependence starting from $n_{\text{sub}} = 85$.

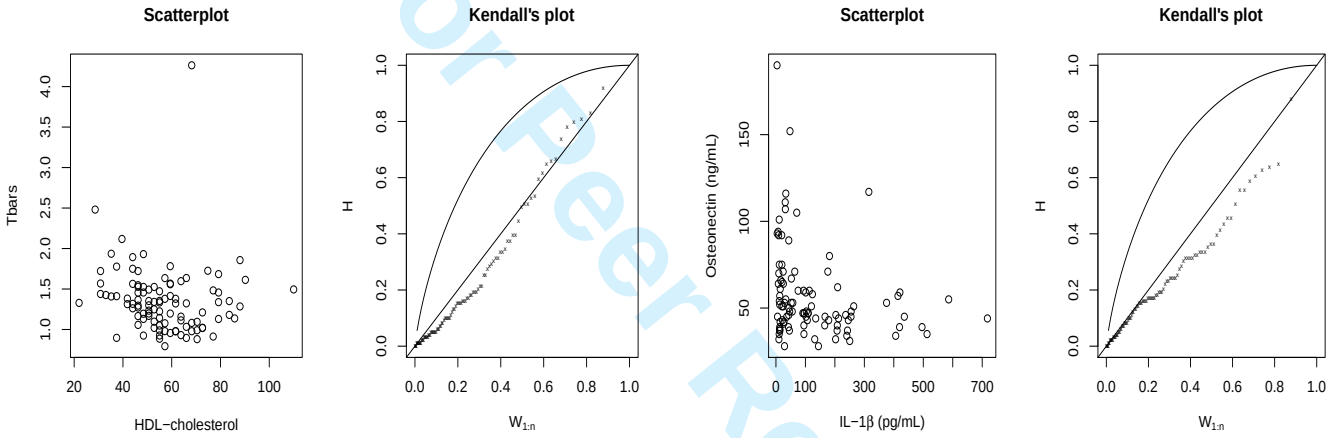
Section 9 of SM considers an additional data example regarding vascular endothelial growth factor expression.

7 Discussion

In the set of the classical tests of dependence, including Pearson's correlation, Spearman's rank correlation and Kendall's rank correlation, it is clear that Kendall's test demonstrates powerful

Method	The "TBARS" data						The "periodontal" data					
	20	30	50	60	70	100	20	30	50	70	85	100
P	0.426	0.438	0.432	0.422	0.392	0.245	0.318	0.244	0.124	0.056	0.031	0.016
K	0.343	0.266	0.134	0.090	0.053	0.007	0.294	0.209	0.093	0.028	0.009	0.002
S	0.356	0.287	0.155	0.109	0.067	0.010	0.305	0.219	0.098	0.031	0.010	0.002
EMK	0.301	0.222	0.100	0.059	0.032	0.003	0.306	0.223	0.103	0.033	0.011	0.002
D_{TS1}	0.355	0.288	0.156	0.109	0.065	0.010	0.304	0.220	0.099	0.030	0.011	0.003
D_{V1}	0.355	0.288	0.156	0.109	0.065	0.010	0.304	0.220	0.099	0.030	0.011	0.003
D_{TS2}	0.380	0.322	0.189	0.139	0.095	0.028	0.316	0.236	0.117	0.048	0.023	0.008
D_{V2}	0.489	0.431	0.238	0.167	0.080	0.006	0.539	0.488	0.358	0.227	0.077	0.010
D_{TS3}	0.384	0.330	0.196	0.145	0.100	0.034	0.319	0.244	0.124	0.052	0.027	0.014
D_{V3}	0.432	0.387	0.236	0.165	0.097	0.014	0.472	0.467	0.387	0.268	0.114	0.022
dbEL	0.346	0.264	0.133	0.080	0.044	0.001	0.344	0.257	0.171	0.093	0.045	0.013

Table 5: P-values of the tests of independence in a jackknife type procedure based on the "TBARS" data and the "periodontal" data.



(a) (b)

Figure 5: The panels (a) and (b) display the scatterplot (in the left side of each panel) and the K-plot (in the right of each panel) based on the "TBARS" and the "periodontal" data, respectively.

properties across most of the evaluated scenarios and practical examples; Pearson's test is very good for linear dependencies. The classical methods, especially Pearson's correlation, work in the structure of dependence in the linear domain. However, in the non-monotonic dependence structure, classical correlation-based tests tend to break down. Moreover, the correlation-based dependence measures calculated from a small sample may be totally misleading if not viewed in the context of its likely sampling error.

The empirical likelihood-based method[13] and density-based empirical likelihood ratio test[14] possess the good property of stable power and is suggested for use in the considered cases and our real data examples. The approaches based on the empirical likelihood technique, i.e. the empirical likelihood-based method and the density-based empirical likelihood ratio test, avoid specifying a

specific dependence structure. These empirical likelihood-based approaches and data-driven tests work very efficient and are fast in the detection of linear, non-linear, and/or random-effect forms of dependence structures. In the case of the coexistence of additive and multiplicative random effects, the dependence is mostly difficult to be recognized as compared to other random-effect formulation. We introduce a new and efficient dependence measure, AUK, based on the ROC curve concept. We present theoretical properties and applications of AUK. Furthermore, the graphical method of heat maps can be efficiently used in order to visually compare powers and various tests, providing clustering of methods.

8 Supplemental Material

The supplemental material consists of the following aspects: additional details describing the classical and modern tests of independence with supporting R code for implementation; more details in classical measures of dependence; brief description of the receiver operating characteristic curve; theoretical properties of the novel AUK measure with supporting R code for implementation; more details in dependence structure for several bivariate distributions; an additional data example; and more details in comparison of the tests reviewed in this paper under extended dependence structures and random effects (SupplementaryMaterials.pdf).

Acknowledgments

We gratefully acknowledge the editor and the two anonymous referees for their very helpful and constructive comments and suggestions, which led to significant improvements to the paper.

References

1. Beeton M, Yule G and Pearson K. Data for the problem of evolution in man. v. on the correlation between duration of life and the number of offspring. *Proc Roy Soc London* 1900; **67**(435-441): 159–179.

2. Kendall M and Stuart A. Inference and Relationship. In *The Advanced Theory of Statistics*. London: Griffin, 1961.

3. Pearson K. Notes on the history of correlation. *Biometrika* 1920; **13**(1): 25–45.

4. Spearman C. The proof and measurement of association between two things. *Am J Psychol* 1904; **15**(1): 72–101.

5. Embrechts P, McNeil A and Straumann D. Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond* 2002; pages 176–223.

6. Kendall MG. A new measure of rank correlation. *Biometrika* 1938; .

7. Kendall MG. *Rank Correlation Methods*. Griffin, 1948.

8. Mudholkar GS and Wilding GE. On the conventional wisdom regarding two consistent tests of bivariate independence. *Journal of the Royal Statistical Society: Series D (The Statistician)* 2003; **52**(1): 41–57.

9. Harrell F, Lee KL and Mark DB. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 1996; **15**: 361–387.

10. May S and Bigelow C. Modeling nonlinear dose-response relationships in epidemiologic studies: statistical approaches and practical challenges. *Dose-Response* 2005; **3**(4): 474–490.

11. Kallenberg WC and Ledwina T. Data-driven rank tests for independence. *J Amer Statist Assoc* 1999; **94**(445): 285–301.

12. Owen A. Empirical likelihood ratio confidence regions. *Ann Statist* 1990; **18**(1): 90–120.

13. Einmahl JH, McKeague IW et al. Empirical likelihood based hypothesis testing. *Bernoulli* 2003; **9**(2): 267–290.

14. Vexler A, Tsai WM and Hutson AD. A simple density-based empirical likelihood ratio test for independence. *Amer Statist* 2014; **68**(3): 158–169.

15. Hoeffding W. A non-parametric test of independence. *The Annals of Mathematical Statistics* 1948; pages 546–557.

16. Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M and Sabeti PC. Detecting novel associations in large data sets. *Science* 2011; **334**(6062): 1518–1524.

17. Fisher N and Switzer P. Graphical assessment of dependence: Is a picture worth 100 tests? *Amer Statist* 2001; **55**(3): 233–239.

18. Wilk MB and Gnanadesikan R. Probability plotting methods for the analysis for the analysis of data. *Biometrika* 1968; **55**(1): 1–17.

19. Genest C and Boies JC. Detecting dependence with kendall plots. *Amer Statist* 2003; **57**(4): 275–284.

20. Vexler A, Schisterman EF and Liu A. Estimation of roc curves based on stably distributed biomarkers subject to measurement error and pooling mixtures. *Stat Med* 2008; **27**(2): 280–296.

21. Pepe MS. A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika* 1997; **84**(3): 595–608.

22. Gu M, Dong X, Zhang X, Wang X, Qi Y, Yu J and Niu W. Strong association between two polymorphisms on 15q25. 1 and lung cancer risk: a meta-analysis. *PloS One* 2012; **7**(6): e37970.

23. Rakovski C, Weisenberger DJ, Marjoram P, Laird PW and Siegmund KD. Modeling measurement error in tumor characterization studies. *BMC Bioinformatics* 2011; **12**(1): 284.

24. Balakrishnan N and Lai CD. *Continuous Bivariate Distributions*. Springer, 2009.

25. Schweizer B and Wolff EF. On nonparametric measures of dependence for random variables. *The Annals of Statistics* 1981; pages 879–885.

26. Kendall MG, Buckland WR et al. A dictionary of statistical terms. *A Dictionary of Statistical Terms* 1957; .

27. Neyman J and Pearson ES. *On the Problem of the Most Efficient Tests of Statistical Hypotheses*. Springer, 1992.

28. Breiman L. *Probability*. Addison-Wesley, Reading, MA, 1968.

29. Inglot T, Kallenberg WC and Ledwina T. Data driven smooth tests for composite hypotheses. *Ann Statist* 1997; pages 1222–1250.

30. Owen AB. *Empirical Likelihood*. CRC press, 2001.

31. Hollander M, Wolfe DA and Chicken E. *Nonparametric Statistical Methods*, **751**. John Wiley & Sons, 2013.

32. Elffers H. On interpreting the product moment correlation coefficient. *Statist Neerlandica* 1980; **34**(1): 3–11.

33. Simon N and Tibshirani R. Comment on "Detecting Novel Associations In Large Data Sets" by Reshef Et Al, Science Dec 16, 2011. *ArXiv e-prints* January 2014; .

34. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psych* 1975; **12**(4): 387–415.

35. Johnson ME. *Multivariate Statistical Simulation: A Guide to Selecting and Generating Continuous Multivariate Distributions*. John Wiley & Sons, 2013.

36. Edwards AL. *An Introduction to Linear Regression and Correlation*. WH Freeman, 1976.

37. Maor E and King JP. *The Story of a Number*, **38**. Oxford Univ Press, 1994.

38. Ball P. *Shapes: Nature's Patterns: A Tapestry in Three Parts*, **1**. Oxford University Press, 2011.

39. Mardia KV. *Families of Bivariate Distributions*, **27**. Griffin London, 1970.

40. Kotz S, Balakrishnan N and Johnson NL. *Continuous Multivariate Distributions, Models and Applications*, **1**. John Wiley & Sons, 2004.

41. Fang K and Zhang Y. *Generalized Multivariate Analysis*. Springer, Berlin, 1990.

42. Azzalini A and Dalla Valle A. The multivariate skew-normal distribution. *Biometrika* 1996; **83**(4): 715–726.

43. Johnston J and DiNardo J. *Econometric Methods*. New York: McGraw-Hill Book Co., 1963.

44. Cochran WG. Errors of measurement in statistics. *Technometrics* 1968; **10**(4): 637–666.

45. Kulathinal S, Kuulasmaa K and Gasbarra D. Estimation of an errors-in-variables regression model when the variances of the measurement errors vary between the observations. *Stat Med* 2002; **21**(8): 1089–1101.

46. Peck CC, Beal SL, Sheiner LB and Nichols AI. Extended least squares nonlinear regression: a possible solution to the choice of weights problem in analysis of individual pharmacokinetic data. *J Pharmacokinet Biopharm* 1984; **12**(5): 545–558.

47. Wood SN. A simple test for random effects in regression models. *Biometrika* 2013; **100**(4): 1005–1010.

48. Armstrong D. *Free Radicals in Diagnostic Medicine: A Systems Approach to Laboratory Technology, Clinical Correlations, and Antioxidant Therapy*. Plenum Press, 1994.

49. Schisterman EF, Faraggi D, Browne R, Freudenheim J, Dorn J, Muti P, Armstrong D, Reiser B and Trevisan M. Tbars and cardiovascular disease in a population-based sample. *J Cardiovasc Risk* 2001; **8**(4): 219–225.

50. Stigler SM. Do robust estimators work with real data? *Ann Statist* 1977; **5**(6): 1055–1098.

51. Ng PYB, Donley M, Hausmann E, Hutson AD, Rossomando EF and Scannapieco FA. Candidate salivary biomarkers associated with alveolar bone loss: cross-sectional and in vitro studies. *FEMS Immunology & Medical Microbiology* 2007; **49**(2): 252–260.

Supplemental Materials for Dependence and Independence: Structure and Inference

Albert Vexler, Xiwei Chen and Alan Hutson
Department of Biostatistics
University at Buffalo, 3435 Main St, Buffalo, NY 14214, USA
email: avexler@buffalo.edu

1 Testing Approaches Based on Classical Correlation Coefficients

In this section, we present in detail testing procedures for independence based on three correlation coefficients introduced in Section 2.1 of the main article. Test of independence under assumed specific model assumptions, e.g. bivariate normality, are conducted via H_0 : no correlation against H_a : non-zero correlation coefficient. To test a corresponding positive or negative dependence between X and Y, a one-sided test alternative can be constructed in a standard manner similarly.

Pearson Correlation Coefficient

The estimator of the Pearson correlation coefficient ρ is obtained by replacing the population moments with their sample counterparts and given as

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}, -1 \leq r \leq 1.$$

Under bivariate normality assumptions and in the large sample setting, the test statistic $t = ((n-2)/(1-r^2))^{1/2} r$ has an asymptotic t distribution with $n-2$ degrees of freedom under the null hypothesis[1]. Accordingly, we reject H_0 if $|t| \geq t_{\alpha/2,n-2}$, where $t_{\alpha/2,n-2}$ is the $(1-\alpha/2)^{\text{th}}$ quantile of t distribution with $n-2$ degrees of freedom.

In practice, it is not uncommon to deal with heavy-tailed distributions whose second moment does not exist (i.e., is infinite), e.g., Student's t distribution with degrees of freedom equal to 2 or 1. For example, many financial time series data sets have heavy tails and tend to correspond to distributions that do not possess finite second moments or higher[2]. In this case, the Pearson correlation is not defined. Another limitation of the Pearson correlation is that it is invariant only with respect to linear transformations of the variables but not with respect to strictly increasing nonlinear transformations. Furthermore, zero correlation only requires $cov(X, Y) = 0$, whereas independence requires $cov(f_1(X), f_2(Y)) = 0$ for any functions f_1 and f_2. Therefore, independence implies zero correlation, but not vice versa, manifesting a drawback of using the correlation as a measure of dependence. For example, if the random variable X is symmetrically distributed about 0, and $Y = X^2$, then $\rho = 0$ but X and Y are perfectly dependent. Only in the specific case where X and Y are from a bivariate normal distribution does zero correlation imply independence.

These limitations and robustness considerations have motivated the development of rank-based correlation measures, alternative measures of dependence, which are considered in the next subsection.

Spearman's Rank Correlation Coefficient

Testing of independence based on Spearman's rank correlation coefficient ρ_S is equivalent to testing $H_0 : \rho_S = 0$ versus $H_a : \rho_S \neq 0$. The estimator of Spearman's rank correlation coefficient ρ_S is given as

$$r_s = 1 - 6 \sum_{i=1}^{n} (R_i - S_i)^2 \Big/ (n(n^2 - 1)).$$

Note that if there are tied X values and/or tied Y values, each observation in the tied group is assigned with the average of the ranks associated with the tied group.

At the significance level α, we reject H_0 if $|r_s| \geq r_{s,\alpha/2}$, where $r_{s,\alpha/2}$ can be found by *qSpearman*[3] in R[4]. For large sample sizes, one can also conduct the test based on the asymptotic t distribution of the Pearson correlation coefficient between the ranked variables.

Kendall's Rank Correlation Coefficient

Testing of independence based on Kendall's rank correlation coefficient τ is equivalent to testing $H_0 : \tau = 0$ versus $H_a : \tau \neq 0$. The Kendall statistic is defined as

$$K = \sum_{i=1}^{n} \sum_{j=i+1}^{n} sgn\{(Y_j - Y_i)(X_j - X_i)\},$$

where $sgn\{x\} = 1$, if $x > 0$; 0, if $x = 0$; and -1, if $x < 0$. Accordingly, at the significance level α, an exact test can be conducted, and we reject H_0 if $\overline{K} \geq k_{\alpha/2}$, where $\overline{K} = K/(n(n-1)/2)$ and $k_{\alpha/2}$ can be found by *qKendall*[3] in R[4]. Alternatively, the test can be conducted based on the asymptotic standard normal distribution of the standardized $K^* = (n(n-1)(2n+5)/18)^{-1/2} K$ under H_0. The null hypothesis is rejected if $|K^*| \geq z_{\alpha/2}$, where $z_{\alpha/2}$ is the $100(1-\alpha/2)^{\text{th}}$ quantile of the standard normal distribution.

Note that although the rank correlation measures ρ_S and τ are invariant under monotonic transformations and can capture perfect monotone dependence, they are not simple functions of moments and therefore are more computationally involved.

2 Data-driven Rank Tests

Data-driven rank tests can capture a variety of dependence structures, including correlations of linear and higher-order polynomials. Within the exponential families that is restricted to the

"diagonal", the joint distribution of (X^*, Y^*) has the form

$$h(x^*, y^*) = c(\boldsymbol{\theta}) exp\left\{\sum_{j=1}^{k} \theta_j b_j(x^*) b_j(y^*)\right\},$$

where b_j denotes the j^{th} orthonormal Legendre polynomial, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T$, and $c(\boldsymbol{\theta})$ is a normalizing constant. The null hypothesis of independence corresponds to $\boldsymbol{\theta} = 0$. If F_X and F_Y are known, the score test for testing $\boldsymbol{\theta} = 0$ against $\boldsymbol{\theta} \neq 0$ is given by rejecting for large values of $\left\{n^{-\frac{1}{2}} \sum_{i=1}^{n} b_r\left(F_X(X_i)\right) b_s\left(F_Y(Y_i)\right)\right\}^2$. When F_X and F_Y are unknown, the null hypothesis is rejected for large values of the "diagonal" test statistic $TS2 = T_{S2}$ defined in Section 2.2 in the main body. In the more general "mixed" model, we consider k-dimensional exponential families containing always $b_1(x*)b_l(y*)$ and $k-1$ other products $b_i(x*)b_j(y*)$. For example, $[(1,1),(2,1),(2,3)]$ refers to the exponential family $c(\boldsymbol{\theta})exp\{\theta_1 b_1(x^*)b_1(y^*) + \theta_2 b_2(x^*)b_1(y^*) + \theta_3 b_2(x^*)b_3(y^*)\}$, with $c(\boldsymbol{\theta})$ a normalizing constant. For the "mixed" test statistic, we consider the one- dimensional model $[(1,1)]$, the two-dimensional models $[(1,1),(i,j)]$ with $i,j = 1,\ldots d(n)$, $(i,j) \neq (1,1)$, the three-dimensional models $[(1,1),(i,j),(k,l)]$ with $i,j,k,l = 1,\ldots,d(n)$, all pairs being different, and so on.

Remark 1. *Under expondential families of data distributions, it is most probable that model-free tests of independence have less power than the data-driven rank tests.*[5] *However, many familiar classes of distributions are non-exponential families.*[6] *A concrete example is the reciprocal-normal type distribution, which can be generated using the R-command*[4]*: 1/mvrnorm(n, mu, Sigma), where mu is the mean vector and Sigma is a positive-definite symmetric matrix specifying the covariance matrix of the variables. Furthermore, by virtue of the tests structures, to be consistent, the data-driven rank test requires that* $E_{H_1} b_j(F_X(X)) b_s(F_Y(Y)) \neq 0$ *for some* j *and* s*. For example, when* $X_i \sim Uniform[0,1]$ *and* $Y_i = \arg\min_z \left\{\left(\sum_{j=1}^{2} b_j(X) b_j(z)\right)^2\right\}$, $i = 1, \ldots n$*, the data-driven rank tests of independence may render low power performance. We will show the comparison of powers of various tests of independence via Monte Carlo simulations in Section 5 of the main article.*

R codes for obtaining the data-driven rank test statistic and the critical values are presented below.

```
# some basic functions #
funs<-list(b1<-function(x) sqrt(3)*(2*x-1),
b2<-function(x) sqrt(5)*(6*x^2-6*x+1),
b3<-function(x) sqrt(7)*(20*x^3-30*x^2+12*x-1),
b4<-function(x) 3*(70*x^4-140*x^3+90*x^2-20*x+1) )

DataDriven.ts<-function(x,y,d.n){
n<-length(x)
R<-rank(x)
```

```
S<-rank(y)
## eq(7)
V_rs<-function(r,s)  sum(funs[[r]]((R-1/2)/n)*funs[[s]]((S-1/2)/n))^2/n

######### test TS2 ###############
## eq(4): order k (first smooth test statistics)
T_k<-function(k) sum(sapply(1:k,function(j) V_rs(j,j)))
## choice of the order k is done by the modified Schwarz's rule
S2_d.n <- sapply(1:d.n,function(k) T_k(k)-k*log(n))
S2<- which.max(S2_d.n)[1]
TS2<-T_k(S2)

######### test V ###############
## the SECOND data-driven smooth test stat V
## consider one and two dimension models for now
if (d.n==1) V<-V_rs(1,1) else {
rs_all<-expand.grid(1:d.n,1:d.n)
V_all=apply(rs_all,1,function(x) c(x[1],x[2],V_rs(x[1],x[2])))
other.max=max(V_all[-1])
V<-ifelse(other.max<log(n),V_all[1],V_all[1]+other.max)
}
return(cbind(TS2,V))
}

############ get critival value ###############
# mc: The number of the Monte Carlo simulations
DataDriven.crit<-function(n,alpha,d.n,mc){
x <- matrix(rnorm(n*mc), nrow =n)
y <- matrix(rnorm(n*mc), nrow =n)
# Obtain the test statistic value following Eq. (5) (after taking log)
vt <- sapply(seq_len(mc), function(j) {
DataDriven.ts(x[,j],y[,j],d.n=d.n)
})
crit<-apply(vt,1,function(ts) as.numeric(quantile(ts, 1-alpha)))
names(crit)<-c("TS2","V")
return(crit )
}
##  DataDriven.crit(n,alpha,d.n) # an exaple
```

3 Empirical Likelihood-based Method

Let $L(\tilde{F}_{XY}) = \prod_{i=1}^{n} \tilde{P}\left(\{(X_i, Y_i)\}\right)$, where $\tilde{P}$ is the probability measure corresponding to F_{XY}. For $(x, y) \in \mathbf{R}^2$, the local likelihood ratio test statistic is

$$R(x,y) = \frac{sup\left\{L(\tilde{F}_{XY}) : \tilde{F}_{XY}(x,y) = \tilde{F}_X(x)\tilde{F}_Y(y)\right\}}{sup\left\{L(\tilde{F}_{XY})\right\}}.$$

Note that the unrestricted likelihood in the denominator is maximized by setting $\tilde{F}_{XY}$ as the corresponding empirical joint distribution function. Based on the constraint of independence, the supremum in the numerator is attained by product of corresponding empirical marginal distributions. Then the local log-likelihood ratio test statistic and the distribution-free test statistic T_n of testing of independence defined in Section 2.3 can be obtained accordingly.

Clearly T_n is distribution-free and its small-sample null distribution can be approximated easily by simulation. If F_X and F_Y are continuous, then under H_0, the limit distribution of T_n is given by

$$T_n \xrightarrow{D} \int_0^1 \int_0^1 \frac{W_0^2(u,v)}{uv(1-u)(1-v)} dudv,$$

where W_0 is a four-sided tied-down Wiener process on $[0,1]^2$.

R codes for obtaining the empirical likelihood-based test statistic and the critical values are presented below.

```
############# test statistics value for EMK ######################
EMcK.ts<-function(x,y){
Fn.x<-ecdf(x)
Fn.y<-ecdf(y)
n<-length(x)
zz<-expand.grid(x,y)

logR<-apply(zz,1,function(xy){
u<-xy[1]
v<-xy[2]
f1n<-Fn.x(u)   #f1n<-sum(  (x<=u) )/n
f2n<-Fn.y(v)   #f2n<-sum(  (y<=v) )/n

pa11<-sum((x<=u)*(y<=v))/n
pa12<-sum((x<=u)*(y>v))/n
pa21<-sum((x>u)*(y<=v))/n
pa22<-sum((x>u)*(y>v))/n

if (pa11!=0) {a<-pa11*log(  (f1n*f2n)/pa11 )} else {a<-0}
if (pa12!=0) {b<-pa12*log(  (f1n*(1-f2n))/pa12 )} else {b<-0}
if (pa21!=0) {c<-pa21*log(  ((1-f1n)*f2n)/pa21 )} else {c<-0}
if (pa22!=0) {d<-pa22*log(  ((1-f1n)*(1-f2n))/pa22 )} else {d<-0}

return(n*(a+b+c+d))
})
Tn<-(-(2*sum(logR))/(n^2))
return(Tn)
}

############ get critival value ###############
```

```
# mc: The number of the Monte Carlo simulations
EMcK.crit<-function(n,alpha,mc){
x <- matrix(rnorm(n*mc), nrow =n)
y <- matrix(rnorm(n*mc), nrow =n)
# Obtain the test statistic value following Eq. (5) (after taking log)
vt <- sapply(seq_len(mc), function(j) {
EMcK.ts(x[,j],y[,j])
})
return( as.numeric(quantile(vt, 1-alpha)))
}
```

4 Density-based Empirical Likelihood Ratio Test

The power of the density-based empirical likelihood approach does not depend significantly on values of $\beta_1 \in (0, 0.5)$ and $\beta_2 \in (0.75, 0.9)$ under various alternative distributions applied to the hypothesis of bivariate independence; see Vexler et al.[7] for more details. The null hypothesis is rejected if $\log(VT_n) > C_\alpha$, where C_α is an α-level test threshold. It follows that

$$P_{H_0}\left(\log(VT_n) > C_\alpha\right) = P_{\{X_i\}_{i=1}^n,\{Y_i\}_{i=1}^n \sim \text{Uniform}_{[0,1]}}\left(\log(VT_n) > C_\alpha\right).$$

Note that the term $n^{-\beta_1}$, $0 < \beta_1 < 0.5$ ensures the consistency of the density-based empirical likelihood ratio test. This test is exact, and the critical values for the proposed test can be accurately approximated using Monte Carlo techniques.

R codes for obtaining the density-based empirical likelihood ratio test statistic and the critical values were presented in Vexler et. al[7] and are omitted here.

5 Hoeffding's Measure of Dependence

Hoeffiding's D is defined as

$$D = 30\left((n-2)(n-3)\sum_{i=1}^{n}\prod_{j=1}^{2}(Q_i - j) + \sum_{i=1}^{n}\prod_{j=1}^{2}(R_i - j)(S_i - j) - 2(n-2)D_*\right)\left(\prod_{i=0}^{4}(n-i)\right)^{-1},$$

where R_i is the rank of X_i, S_i is the rank of Y_i, the bivariate rank Q_i is $1 + \sum_{j\neq i} I\{X_j < X_i, Y_j < Y_i\}$, $i = 1, \ldots, n$, and $D_* = \sum_{i=1}^{n}(R_i - 2)(S_i - 2)(Q_i - 1)$. A point that is tied on both X and Y contributes $1/4$ to Q_i. A point that is tied on only the X value or Y value contributes $1/2$ to Q_i if the other value is less than the corresponding value for the i^{th} point. Without ties in the data set, the Hoeffding's D measure of dependence is on the interval $[-0.5, 1]$, with one indicating perfect dependence.

6 Brief Comments Regarding Receiver Operating Characteristic Curve

Receiver operating characteristic curve (ROC) is a well-accepted statistical method for evaluating the discriminatory ability of biomarkers (e.g., Vexler et al.[8]). An ROC curve plots the true positive

rates of a biomarker versus its false-positive rates for various threshold of the test result. It is a convenient way to compare diagnostic biomarkers because the ROC curve places tests on the same scale where they can be compares for accuracy. The area under the ROC curve (AUC) is a common and well developed index that summarizes the information contained in the ROC curve. Bamber[9] showed that

$$AUC = \int (1 - F_X(u))dF_Y(u) = \Pr(X > Y).$$

Obviously, the closer the AUC is to one, the better the diagnostic accuracy of the biomarker in terms of the distances between the distribution functions F_X and F_Y . The case $F_X = F_Y$ is reflected by AUC$= 0.5$ that corresponds to the area under the diagonal line. The AUC is also closely related to the Gini coefficient,[10] which is twice the area between the diagonal and the ROC curve.

7 Area Under the Kendall Plot (AUK)

7.1 Theoretical Properties of AUK

The following propositions provide a way to consider the AUK in the light of the well-addressed in the statistical literature AUC analysis. Define U and V to be independent uniformly distributed random variables between zero and one.

Proposition 1. *Suppose continuous random variable X and Y are distributed according to a bivariate distribution function H. Then*

$$AUK = \Pr\{UV < H(X,Y)\} = E\{H(X,Y) - H(X,Y)\log(H(X,Y))\}.$$

Proof. Based on the definition, we have

$$AUK = 1 - \int_0^1 \Pr\{H(X,Y) < t\}d\Pr\{UV < t)\} = \Pr\{UV < H(X,Y)\}$$
$$= E\Pr\{UV < H(X,Y)|H(X,Y)\} = E(H(X,Y) - H(X,Y)\log H(X,Y)).$$

□

Proposition 1 shows a simple method for expressing the AUK. It is very useful both for direct evaluations and for simulations. This proposition presents a result that is similar to that obtained regarding the AUC.[9] Suppose, for example, that a random sample $(X_i, Y_i), i = 1, ..., n$, has been drawn from a bivariate distribution H. By virtue of Proposition 1, we can estimate the AUK in a nonparametric manner via the statistic

$$\overline{AUK} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{H}(X_i,Y_i) - \hat{H}(X_i,Y_i)\log(\hat{H}(X_i,Y_i))\right\},$$

where we replace the distribution function H by its empirical estimator $\hat{H}$. The form of $\overline{AUK}$, the estimator of AUK, is much simpler than that can be obtained directly by using the definition of the AUK.

Schriever defined the "more associated-ordering" for bivariate distributions.[11] By virtue of Proposition 1, we have $AUK = \int\int J(H(x,y))dH(x,y)$, where the function $J(u) = u - u\log(u)$ increases and its convex for $u \in [0,1]$. Then Example 3.2 in Schriever[11] can be directly adapted to show that the proposed AUK based measure preserve more concordant-ordering for dependence. We refer the reader to Schriever[11] for details regarding the ordering for dependence.

The well-known Fechet-Hoeffding results regarding copula bounds implies the following proposition.

Proposition 2. *For any continuous random variables X and Y distributed according to a bivariate distribution function $H(x,y)$, the measurements AUK satisfy $0 \leq AUK \leq 3/4$, where the case with $(X, Y = X)$ provides AUK to reach the upper bound $3/4$ and $(X, Y = -X)$ leads to the lower bound zero.*

Proof. By virtue of the Fechet-Hoeffding upper-bound, we have

$$\Pr\{H(X,Y) < t\} \geq \Pr\{\min(F_X(X), F_Y(Y)) < t\} \geq \Pr\{F_X(X) < t\} = t.$$

It is clear that $\Pr\{H(X,Y) < t\} \leq 1$. Then

$$0 \leq AUK = \int_0^1 \Pr\{H(X,Y) > t\}\, d\,(t - t\log(t)) \leq \int_0^1 (1-t)(-\log(t))dt = 3/4.$$

The range of AUK is [0, 3/4]. In the case $(X, Y = X)$, we have $H(x,y) = \Pr(X < x, X < y)$ and $H(X,Y) = H(X,X)$. Thus $\Pr\{H(X,Y) > t\} = \Pr\{F_X(X) > t\}$ and

$$AUK = \int_0^1 \Pr\{F_X(X) > t\}\, d\,(t - t\log(t)) = \int_0^1 (1-t)d(-\log(t)) = 3/4$$

In the case $(X, Y = -X)$, we have $H(x,y) = \Pr(X < x, -X < y)$ and $H(X,Y) = H(X,-X)$, which leads to $\Pr\{H(X,Y) > t\} = 0$ and $AUK = 0$. □

7.2 Algorithms in R for Obtaining the Empirical AUK

The R codes to implement the Kendall plot and the computation of the AUK measure of dependence are presented below.

```
####### to obtain AUK based on the definition #######
# x and y two samples to be tested for independence
# plot: Logical; whether the Kendall plot is presented.
# installing package 'CDVine' for the first time:
if (!require('CDVine')) install.packages('CDVine')
require(CDVine)
```

```
get.AUK<-function(x,y,plot = TRUE){
tmp <- BiCopKPlot(x,y,PLOT=FALSE)
if (plot == TRUE) BiCopKPlot(x,y,PLOT=TRUE)
W <- c(0,tmp$W.in,1)
H <- c(0,tmp$Hi.sort,1)
idx <- 2:length(W)
area <- as.double((W[idx]-W[idx-1])%*%(H[idx]+H[idx-1]))/2
return(min(area,0.75))
}

###### to obtain AUK.bar based on Proposition 1 in Section 7 of SM ######
get.AUK.bar<-function(x,y,plot = TRUE){
get.H<-function(u1,u2) mean(1*(x<u1)&(y<u2))
get.H.V<-Vectorize(get.H)
G<-get.H.V(x,y)
GlogG<-array(0,length(x)*length(y))
GlogG[G!=0]<-G[G!=0]*log(G[G!=0])
area <- mean(G-GlogG)
return(min(area,0.75))
}

# an example
x=rnorm(300)
y=x+3
get.AUK(x,y,plot=TRUE)
get.AUK.bar(x,y,plot=TRUE)
```

8 Several Bivariate Distributions

Generalizing the normal distribution by adding a "non-normality" parameter $\kappa > 0$, multivariate exponential power distributions can be presented.[12]The parameter κ represents the kurtosis departure from the multivariate normal distribution. As a special case, the multivariate normal distribution is derived when $\kappa = 2$.

Plackett's distribution[12] can include a full range of dependence from completely negative though independence to completely positive. With normal marginals, it shares considerable similarity to the bivariate normal distribution, especially when the correlation parameter in the bivariate normal distribution $\rho = -cos(\pi\phi^{1/2}/(1+\phi^{1/2}))$, where the parameter ϕ governs the dependence between Plackett distributed $(X, Y)^T$.

Weak dependence properties are presented in the Morgenstern distribution[12]. The bivariate distribution function of (X, Y) is $F_{XY}(x, y) = F_X(x)F_Y(y)(1 + \alpha((1 - F_X(x))(1 - F_Y(y))))$, $0 \leq x, y \leq 1$ and the parameter $-1 \leq \alpha \leq 1$ controls the dependence structure between X and Y. In

particular, for uniform marginals, the Pearson's correlation coefficient is $\rho = \alpha/3$ implying $|\rho| \leq 1/3$. The maximum is $1/\pi$ for normal marginals and $1/4$ for exponential marginals.

Note that the joint distributions can also be determined by the incorporation of marginal and conditional specifications. It is possible to construct distributions where the two conditional densities $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ are each normal and yet the joint distribution is not bivariate normal.[13,14] As a simple example, we consider the joint density function $f_{X,Y}(x,y) = Cexp\{-(x^2+y^2+2xy(x+y+xy))\}$, where $C > 0$ is the normalizing constant. It may be verified that $E(Y|X=x) = -x^2/(1+2x+2x^2)$ and $var(Y|X=x) = 1/(2+4x+4x^2)$.

9 Data Example of Vascular Endothelial Growth Factor Expression

The gynecologic oncology group study evaluated the association between the relative expression of the N-terminally truncated isoform and debulking status, the relative expression of vascular endothelial growth factor (VEGF) and VEGF receptor-1 (VEGFR-1) based on 60 cases.[15] It is of great interest to test of independence between VEGF expression and VEGF receptor 1. Figure 1 presents the scatterplot of the data containing VEGF and VEGF receptor 1 expression measurements in the left panel and the K-plot in the right panel. The estimates of MIC and AUK measures are 0.706 and 0.687, respectively. In this dataset, measurements of VEGF expression and VEGF receptor 1 are highly dependent. In a similar manner described in the Tbars data, a jackknife type procedure is conducted to examine different tests of independence with randomly selected subsamples of the sample size ranging from 10 to 19, repeating $5,000$ times. As shown in Table 1, among those classical correlation-based methods, the Kendall's rank correlation performs the best. Except for the data-driven test using V statistic with $d(n) = 2, 3$, all methods can detect the dependence starting from $n_{\text{sub}} = 15$. Figure 2 displays measures of dependence (in the left panel) including AUK (black solid line) and MIC (grey dashed line) as well as p-values obtained via a jackknife-type procedure (in the right panel) based on the data containing VEGF expression and VEGF receptor 1 measurements, , where P (—○—), K (− − ○ − −), S (· · · ○ · · ·), EMK (—△—), D_{TS2} (− − △ − −), D_{V2} (· · · △ · · ·), D_{TS3} (—+—), D_{V3} (− − + − −), dbEL (· · · + · · ·). It can be noted that the MIC measure of dependence slightly decreases as the sample size increases, demonstrating an inferior performance compared to the AUK measure of dependence.

10 Extended Dependence Structures

Except for the dependence structures defined in Table 2 of the main article, we further considered the following structures as shown in Table 2 below.

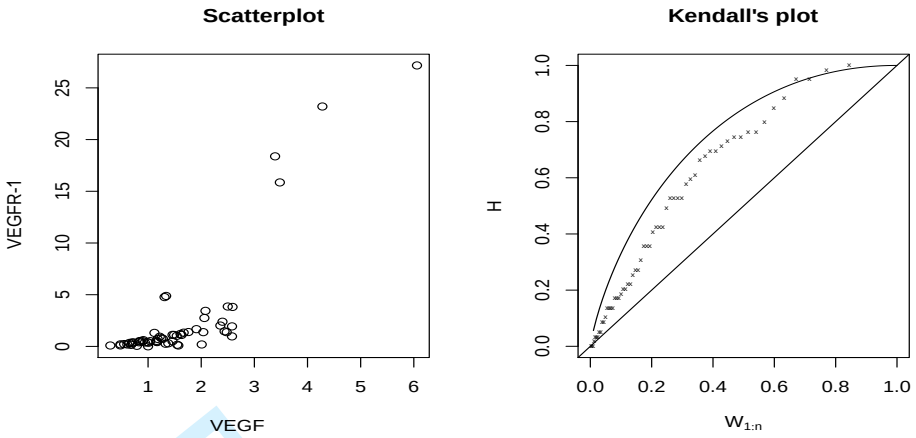


Figure 1: The scatterplot (in the left) and the K-plot (in the right) based on the data containing VEGF expression and VEGF receptor 1 measurements.

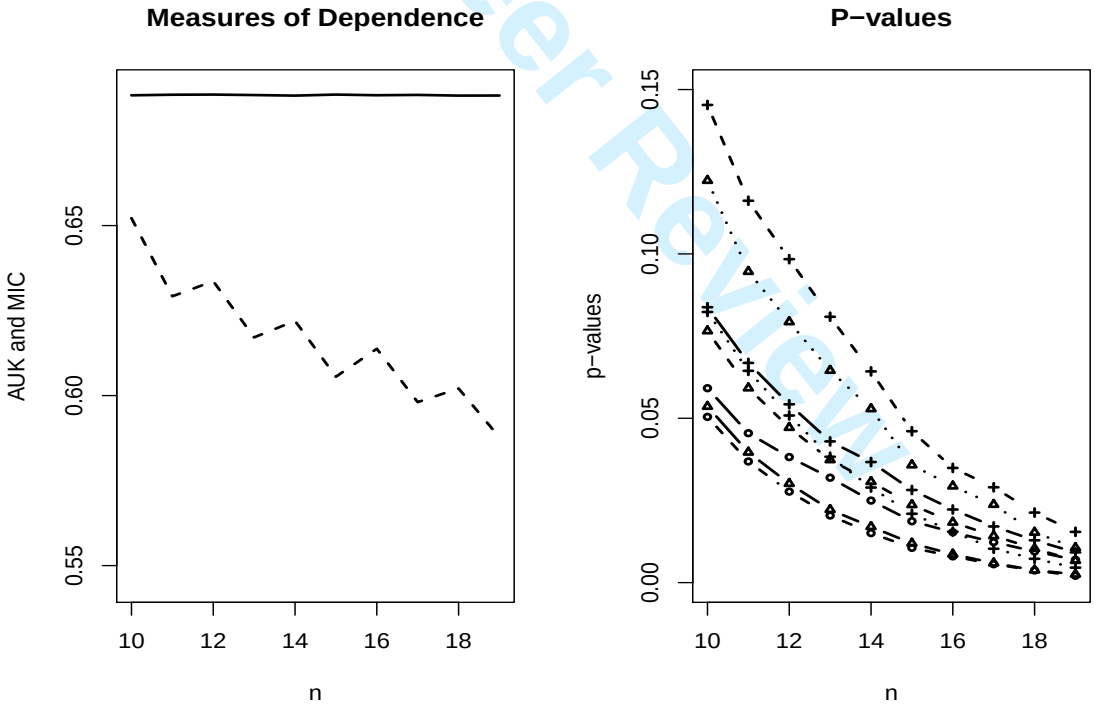


Figure 2: Measures of dependence (in the left panel) including AUK (black solid line) and MIC (grey dashed line) as well as p-values obtained via a jackknife-type procedure (in the right panel) based on the data containing VEGF expression and VEGF receptor 1 measurements, where P (—○—), K ($--\circ--$), S ($\cdots\circ\cdots$), EMK (—△—), D_{TS2} ($--\triangle--$), D_{V2} ($\cdots\triangle\cdots$), D_{TS3} (—+—), D_{V3} ($--+--$), dbEL ($\cdots+\cdots$).

Method	Sample Size									
	10	11	12	13	14	15	16	17	18	19
P	0.059	0.045	0.038	0.032	0.025	0.019	0.015	0.012	0.009	0.007
S	0.057	0.044	0.034	0.025	0.020	0.014	0.011	0.008	0.006	0.003
K	0.050	0.037	0.028	0.020	0.015	0.011	0.008	0.006	0.004	0.002
EMK	0.054	0.040	0.030	0.022	0.017	0.012	0.009	0.006	0.004	0.003
D_{TS1}	0.057	0.043	0.034	0.025	0.020	0.015	0.011	0.009	0.006	0.004
D_{V1}	0.057	0.043	0.034	0.025	0.020	0.015	0.011	0.009	0.006	0.004
D_{TS2}	0.077	0.059	0.047	0.037	0.031	0.024	0.018	0.014	0.010	0.007
D_{V2}	0.122	0.095	0.079	0.065	0.053	0.036	0.029	0.024	0.015	0.011
D_{TS3}	0.084	0.067	0.054	0.043	0.037	0.028	0.022	0.017	0.013	0.009
D_{V3}	0.145	0.116	0.098	0.081	0.064	0.046	0.035	0.029	0.021	0.015
dbEL	0.082	0.064	0.051	0.038	0.029	0.021	0.016	0.010	0.007	0.005

Table 1: P-values of the tests of independence obtained via a jackknife-type procedure based on the data containing VEGF expression and VEGF receptor 1 measurements.

Figure 3 shows the scatterplots of observed (X, Y) $(n = 1000)$ for dependence structures considered but not presented in Figure 1 of the main article, with no random effect and various additive normally distributed random effects with mean 0 and variance 0.25 or 2. Each row corresponds to the case where the true values (X_0, Y_0) follow: (M1) Linear; (M2) Quadratic; (M7) Reciprocal with order $k = 3$; (M10) Spirals; (M24) DDF1; and (M25) DDF2; see in Table 2 in Section 4.1 of the main article and Table 2 for more details. Each column in Figure 3 displays various random effects, where (a) represents the case of no random effect in either X or Y, i.e. $X = X_0$, $Y = Y_0$; (b) and (b') show the case where X has no random effect and Y has additive effects, i.e. $X = X_0$, $Y = Y_0 + \varepsilon_Y$ with the variance of noise $var(\varepsilon_Y)$ assumed to be 0.25 (shown in (b)) or 2 (shown in (b')), respectively; (c) and (c') show the case where Y has no random effect and X has additive random effects, i.e. $Y = Y_0$, $X = X_0 + \varepsilon_X$ with the variance of noise $var(\varepsilon_X)$ assumed to be 0.25 or 2, respectively; (d) and (d') show the case where both X and Y have additive random effect, i.e. $X = X_0 + \varepsilon_X$, $Y = Y_0 + \varepsilon_Y$, with the variance of noise $var(\varepsilon_X) = var(\varepsilon_Y)$ assumed to be 0.25 or 2, respectively. Note that in Figure 3, scatterplots with different random effects are shown in the same scale for each type of dependence structures. It can be observed that not only the variance of random effects but also the type of random effects affect the dependence structure.

In Figure 4, we show the scatterplots $(n = 1000)$ of several bivariate distributions described above and in Section 4.1 of the main article. Each row represent the bivariate distributions where the true values of (X_0, Y_0) follow: (M11) ellipse; (M12) bivariate Normal distribution with correlation 0.5; (M15) bivariate Placket distribution with $\phi = 3.5$; (M21) two marginally normal distributions coupled with the Clayton copula; (M22) reciprocal-normal type distribution; and (M23) normal conditional distributions; see in Table 2 in Section 4.1 of the main article for more details. Here we

considered additive random effects only in Y since X and Y are exchangeable for most considered bivariate distributions. Each column of Figure 4 represents the case $X = X_0$, $Y = Y_0 + \varepsilon_Y$, where the random effects $\varepsilon_Y \sim N(0, \sigma^2)$ and $\sigma^2 = 0.25$, 0.5, 1, and 2, from left to right. In such case, the presence of additive random effects in one variable of bivariate distributions does not change the dependence structure as much as the case where one variable is simple a function of the other variable as shown in Figure 1 of the main article.

Designs	Models/Description	
	X	$f(X)$
Nonlinear		
M9'($\sin(\pi x)$)	$Unif[-2, 2]$	$\sin(\pi X)$
M24(DDF1)	$Unif[0, 1]$	$\arg\min_z(\sum_{j=1}^{2} b_j(X)b_j(z))^2$
M25(DDF2)	$Unif[0, 1]$	$\arg\min_z (b_1(X)b_1(z) + 0.3b_1(X)b_2(z) + 0.5b_2(X)b_1(z) + b_2(X)b_2(z))^2$

Table 2: Dependence structures.

Figure 5 considers a variety of random effects, including additive and multiplicative random effects, illustrating the scatterplots ($n = 1000$) of several bivariate distributions demonstrated in Figure 4. Each row represent the bivariate distributions where the true values of $(X_0, Y_0)^T$ follow: (M11) ellipse; (M12) bivariate Normal distribution with correlation 0.5; (M15) bivariate Placket distribution with $\phi = 3.5$; (M21) two marginally normal distributions coupled with the Clayton copula; (M22) reciprocal-normal type distribution; and (M23) normal conditional distributions. Each column in Figure 5 displays various random-effect schemes in a same scale for each type of bivariate distribution, where (a) shows the case $X = X_0$, $Y = Y_0$; (b) displays the case $X = X_0$, $Y = \varepsilon_{Y_M} Y_0$; (c) considers the case $Y = Y_0$, $X = \varepsilon_{X_M} X_0$; (d) demonstrates the case $X = \varepsilon_{X_M} X_0$, $Y = \varepsilon_{Y_M} Y_0$; (e) studies the case $X = X_0 + \varepsilon_{X_A}$, $Y = \varepsilon_{Y_M} Y_0$; and (e) depicts the case $X = \varepsilon_{X_M} X_0 + \varepsilon_{X_A}$, $Y = \varepsilon_{Y_M} Y_0 + \varepsilon_{Y_A}$. In this case, all random effects ε_{j_k}, $j = X$ or Y, $k = A$ or M are assumed to be normally distributed with mean 0 and variance 0.25. We further illustrate the scatterplots under various dependence structures with additive and/or multiplicative random effects in Supplementary Materials. It can be seen that the shapes of the scatterplots are heavily affected by the types of randome effects, even when the variance of random effect is relatively small.

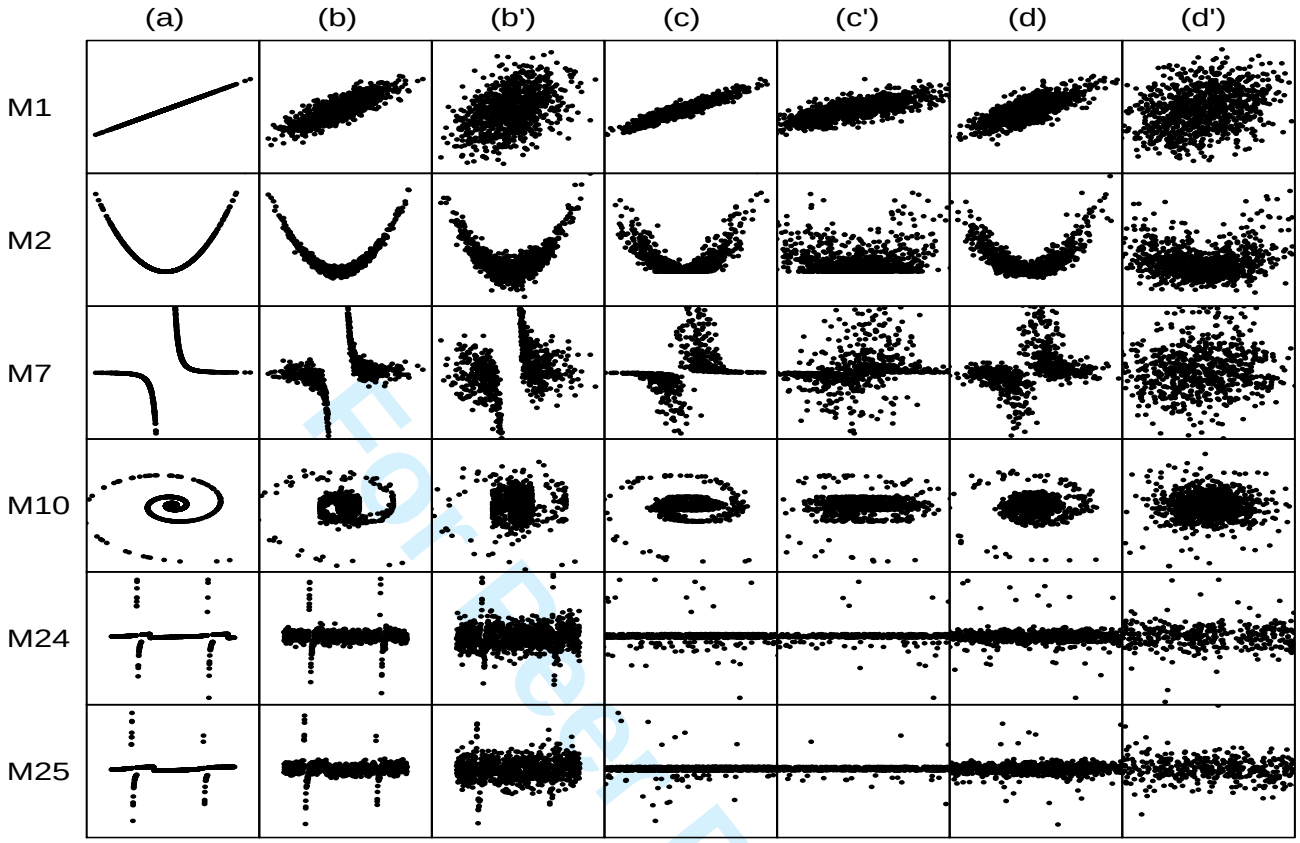


Figure 3: Scatterplots ($n = 1000$) of measurements of (X, Y) under different dependence structures and normally distributed random effects with mean 0, where the unobserved true values of the predictor variable X_0 and the response variable Y_0 can be formulated as $Y_0 = f(X_0)$. Each row corresponds to each of the following structures: (M1) Linear; (M2) Quadratic; (M7) Reciprocal with order $k = 3$; (M10) Spirals; (M24) DDF1; and (M25) DDF2. Each column displays various random-effect schemes, where (a) shows the case $X = X_0$, $Y = Y_0$; (b) and (b') show the case $X = X_0$, $Y = Y_0 + \varepsilon_Y$ with the variance of noise $var(\varepsilon_Y)$ assumed to be 0.25 or 2, respectively; (c) and (c') show the case $Y = Y_0$, $X = X_0 + \varepsilon_X$ with the variance of noise $var(\varepsilon_X)$ assumed to be 0.25 or 2, respectively; (d) and (d') show the case $X = X_0 + \varepsilon_X$, $Y = Y_0 + \varepsilon_Y$, with the variance of noise $var(\varepsilon_X) = var(\varepsilon_Y)$ assumed to be 0.25 or 2, respectively.

11 Monte Carlo Comparisons of Tests of Independence

In this section, Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with dependence structures that are not represents in Section 5 of the main article as well as comparisons of the considered tests at the 0.05 significance level are shown here, with various types of random effects. Table 3 shows the estimates of indices of

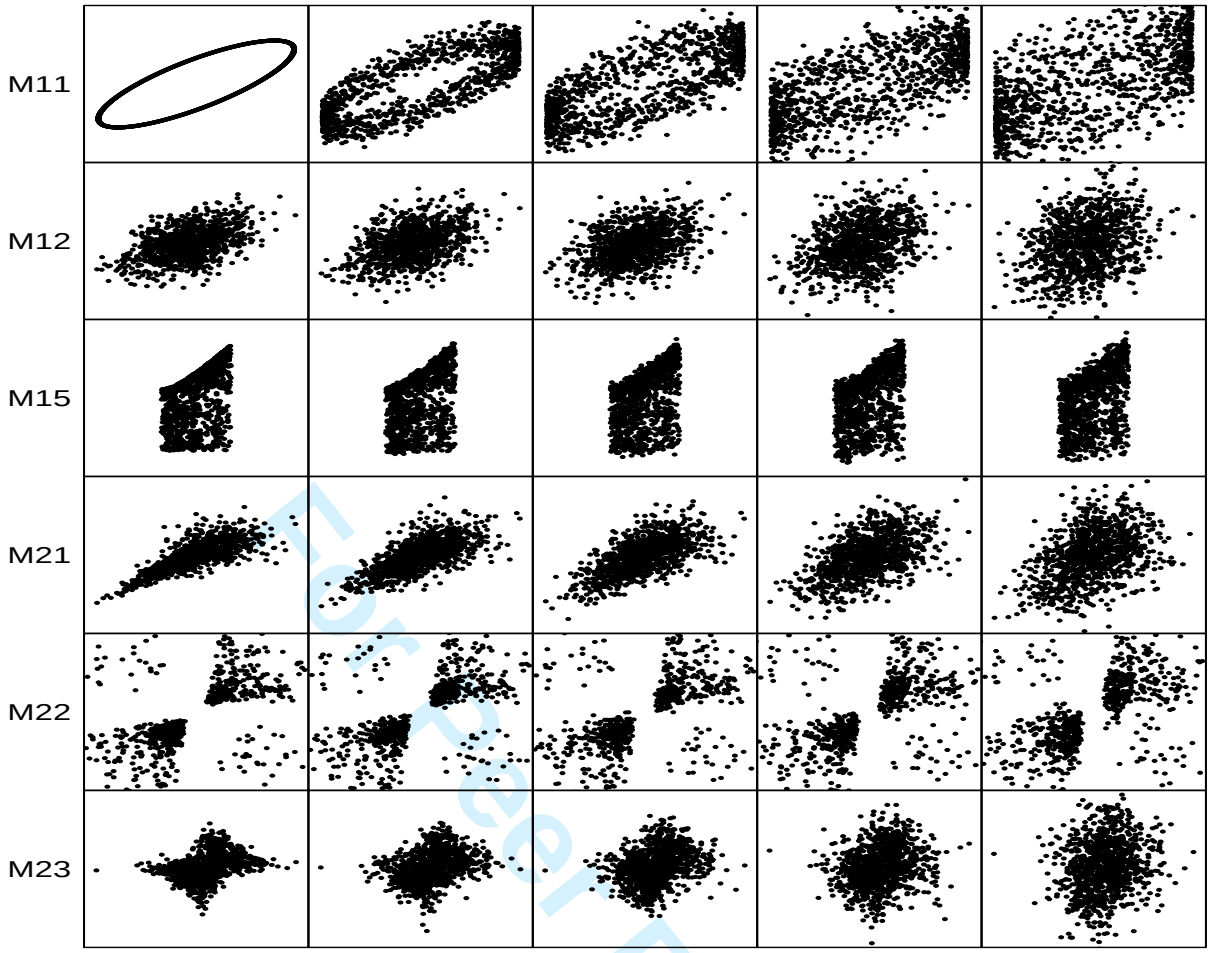


Figure 4: *Scatterplots ($n = 1000$) of dependence structures of (X, Y) that can be represented via bivariate distributions with additive random effects of mean 0. Each row represent the bivariate distributions where the true values of $(X_0, Y_0)^T$ follow: (M11) ellipse; (M12) bivariate Normal distribution with correlation 0.5; (M15) bivariate Placket distribution with $\phi = 3.5$; (M21) two marginally normal distributions coupled with the Clayton copula; (M22) reciprocal-normal type distribution; and (M23) normal conditional distributions. And each column represents the case $X = X_0$, $Y = Y_0 + \varepsilon_Y$, where the random effects $\varepsilon_Y \sim N(0, \sigma^2)$ and $\sigma^2 = 0.25$, 0.5, 1, and 2, from left to right.*

dependence based on samples of sizes 1000.

As an example, in Table 4, we summarize the considered tests based on powers in descending order at the 0.05 significance level for different sample sizes n for the dependence structures M3 where $f(X) = 2 + X + X^2$. Random effects are also considered, where (1) represents no ME in Y; (2) represents additive random effects in Y; (3) represents multiplicative random effects in Y; (4) represents both additive and multiplicative random effects in Y. For each dependence structure and sample size, considered Tests of Independence are clustered in 3 clusters. Within each cluster, results are ordered in descending order in terms of powers as shown in parenthesis. Between clusters, results are ordered in descending order in terms of powers, separated by $>$.

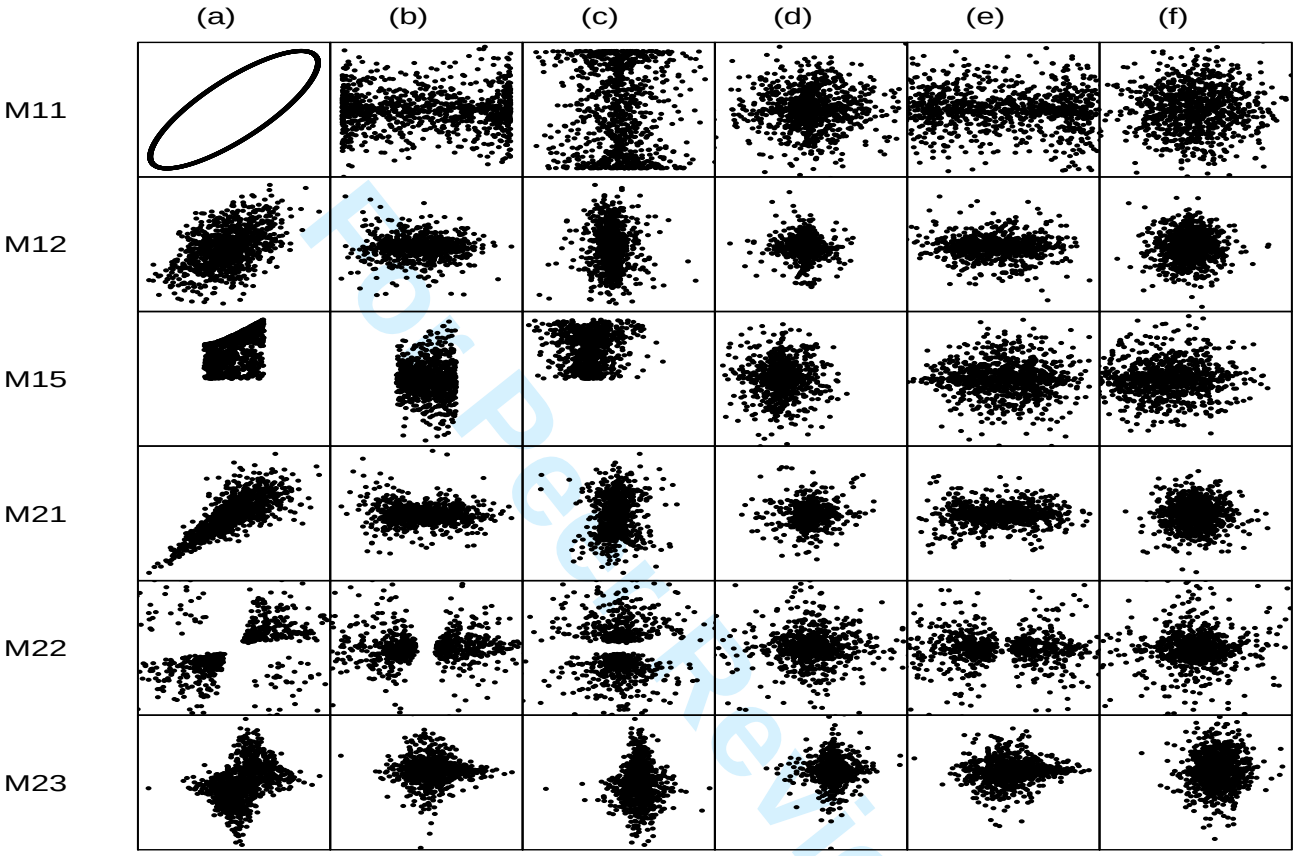


Figure 5: *Scatterplots ($n = 1000$) of bivariate dependence structures with additive and/or multiplicative random effects. Each row represent the bivariate distributions where the true values of $(X_0, Y_0)^T$ follow: (M11) ellipse; (M12) bivariate Normal distribution with correlation 0.5; (M15) bivariate Placket distribution with $\phi = 3.5$; (M21) two marginally normal distributions coupled with the Clayton copula; (M22) reciprocal-normal type distribution; and (M23) normal conditional distributions. Each column displays the following random-effect schemes in a same scale for each type of bivariate distribution: (a) $X = X_0$, $Y = Y_0$; (b) $X = X_0$, $Y = \varepsilon_{Y_M} Y_0$; (c) $Y = Y_0$, $X = \varepsilon_{X_M} X_0$; (d) $X = \varepsilon_{X_M} X_0$, $Y = \varepsilon_{Y_M} Y_0$; (e) $X = X_0 + \varepsilon_{X_A}$, $Y = \varepsilon_{Y_M} Y_0$; and (e) $X = \varepsilon_{X_M} X_0 + \varepsilon_{X_A}$, $Y = \varepsilon_{Y_M} Y_0 + \varepsilon_{Y_A}$, where all random effects $\varepsilon_{j_k} \sim N(0, 0.25)$, $j = X$ or Y, $k = A$ or M.*

Structures [Range];(Measure$_{H_0}$)	Pearson $[-1,1];(0)$	Spearman $[-1,1];(0)$	Kendall $[-1,1];(0)$	MIC $[0,1];(0)$	AUK $[0,0.75];(0.5)$
M1	1.000	1.000	1.000	1.000	0.750
M2	-0.000	0.001	0.001	1.000	0.578
M3	0.447	0.383	0.443	1.000	0.664
M4	0.781	1.000	1.000	1.000	0.750
M5	0.028	-0.000	0.498	1.000	0.404
M6	-0.000	-0.001	-0.001	1.000	0.377
M7	0.000	-0.000	0.498	1.000	0.419
M8	0.000	0.001	0.001	1.000	0.582
M9	0.001	0.004	0.005	1.000	0.590
M10	-0.099	-0.025	-0.015	0.373	0.491
M11	0.800	0.592	0.785	0.623	0.625
M12	0.500	0.333	0.482	0.267	0.603
M13	-0.001	-0.001	-0.001	1.000	0.386
M14	0.319	0.223	0.333	0.205	0.566
M15	0.452	0.356	0.487	0.392	0.622
M16	0.000	0.000	0.000	0.142	0.496
M17	-0.018	0.000	0.001	0.447	0.533
M18	0.005	-0.000	-0.000	0.159	0.523
M19	0.500	0.333	0.491	0.279	0.599
M20	0.858	0.667	0.848	0.610	0.691
M21	0.775	0.600	0.785	0.549	0.644
M22	0.000	0.079	0.102	0.248	0.525
M23	0.361	0.290	0.476	0.333	0.560

Table 3: *Indices of dependence, including Pearson correlation coefficient, Spearman's rank correlation coefficient, Kendall's rank correlation coefficient, the MIC measure and the AUK measure. Note that Pearson correlation coefficient, Spearman's rank correlation coefficient and Kendall's rank correlation coefficient can range from -1 for the perfect negative dependence case to 1 for the perfect dependence case, where the value 0 stands for the independence case. The MIC measure can range from 0 for the independence case to 1 for the perfect dependence case, while the AUK can range from 0 for the perfect negatively dependent case to 0.75 for the perfect positive dependence case, where the value 0.5 indicates independence.*

Designs (RE)	n	Clustering of the test powers (order within/between clusters) in descending order for different sample sizes (n)
M3 (Figure 2a of the main article): $f(X) = 2 + X + X^2$		
(1)	25	(EMK, D_{V2}, D_{V3}, dbEL) > (K, P) > (D_1, S, D_{TS3}, D_{TS2})
	50	(EMK, D_{V2}, D_{V3}, dbEL) > (K, D_{TS3}) > (S, D_1, D_{TS2}, P)
	100	(EMK, D_{V2}, D_{TS3}, D_{V3}, dbEL, D_{TS2}) > (K, D_1, S) > (P)
	150	(EMK, D_{TS2}, D_{V2}, D_{TS3}, D_{V3}, dbEL, K) > (S, D_1) > (P)
(2)	25	(D_{V2}, D_{V3}, dbEL, EMK) > (P) > (K, D_{TS3}, D_1, S, D_{TS2})
	50	(EMK, D_{V2}, D_{V3}, dbEL) > (D_{TS3}, D_{TS2}) > (K, P, S, D_1)
	100	(EMK, D_{V2}, D_{V3}, dbEL) > (D_{TS3}, D_{TS2}) > (K, D_1, S, P)
	150	(EMK, D_{TS2}, D_{V2}, D_{TS3}, D_{V3}, dbEL) > (K, S, D_1) > (P)
(3)	25	(D_{V2}, D_{TS2}, D_{TS3}, D_{V3}) > (dbEL, P) > (EMK, K, D_1, S)
	50	(D_{V2}, D_{TS2}, D_{TS3}, D_{V3}, dbEL) > (EMK, P) > (K, S, D_1)
	100	(D_{V2}, D_{V3}, D_{TS2}, D_{TS3}, dbEL) > (EMK) > (P, K, D_1, S)
	150	(D_{V3}, D_{V2}, D_{TS3}, D_{TS2}, dbEL) > (EMK) > (P, K, S, D_1)
(4)	25	(D_{V2}, D_{TS2}, D_{TS3}, D_{V3}) > (dbEL, P) > (EMK, K, D_1, S)
	50	(D_{V2}, D_{TS2}, D_{TS3}, D_{V3}) > (dbEL) > (EMK, P, K, S, D_1)
	100	(D_{V2}, D_{TS2}, D_{TS3}, D_{V3}, dbEL) > (EMK) > (P, K, D_1, S)
	150	(D_{V2}, D_{V3}, D_{TS2}, D_{TS3}, dbEL) > (EMK) > (P, K, S, D_1)

Table 4: Summarization of different tests based on powers in descending order at the 0.05 significance level for different sample sizes n, under the formulaic dependence structure $f(X) = 2 + X + X^2$ and different random effects, where (1) represents no ME in Y; (2) represents additive random effects in Y; (3) represents multiplicative random effects in Y; (4) represents both additive and multiplicative random effects in Y. For each dependence structure and sample size, considered tests for independence are clustered in 3 clusters. Within each cluster, results are ordered in descending order in terms of powers as shown in parenthesis. Between clusters, results are ordered in descending order in terms of powers, separated by >.

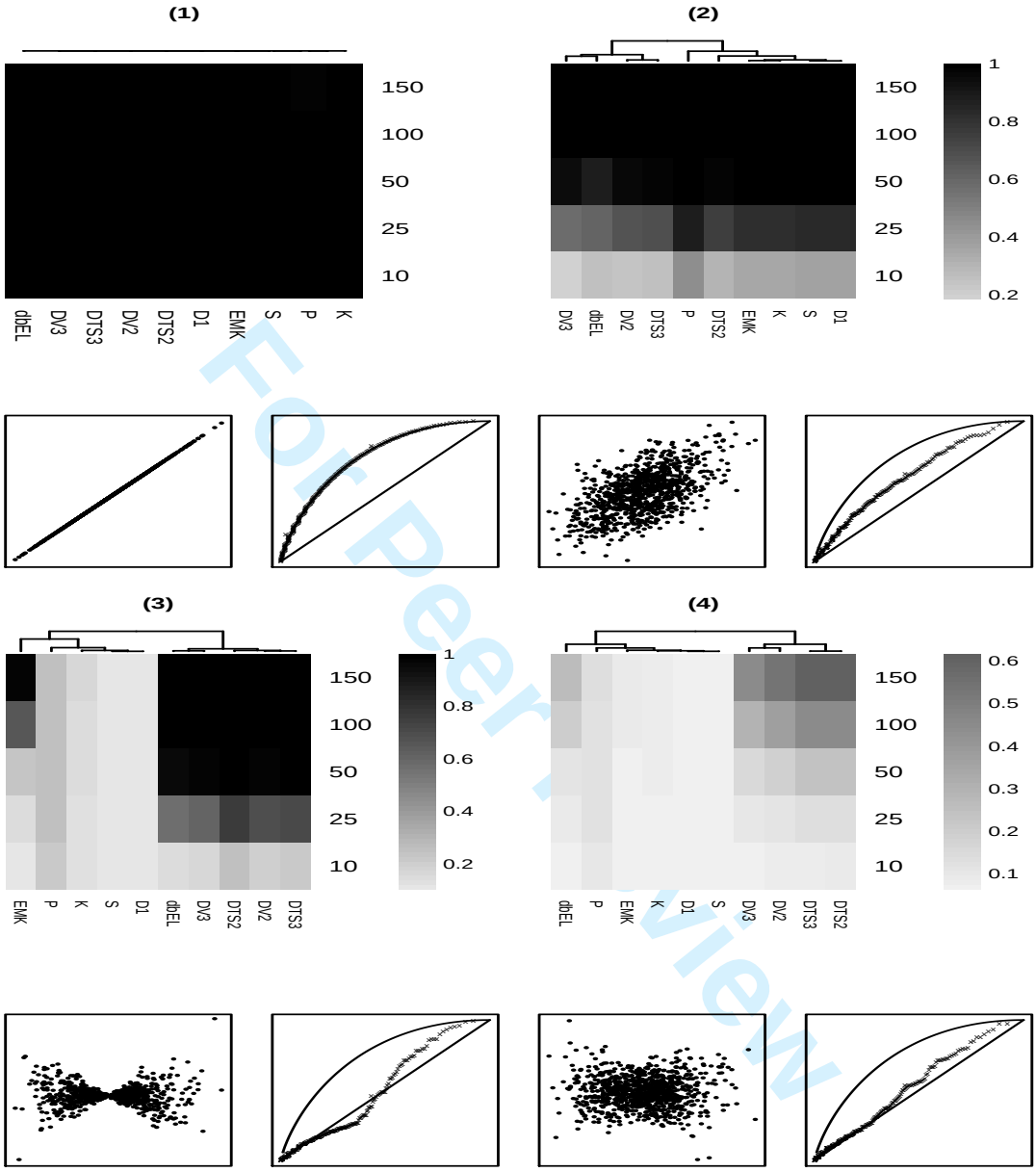


Figure 6: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M1 (Linear) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 1(0.75); (2) 0.405(0.619); (3) 0.305(0.522); and (4) 0.237(0.511).

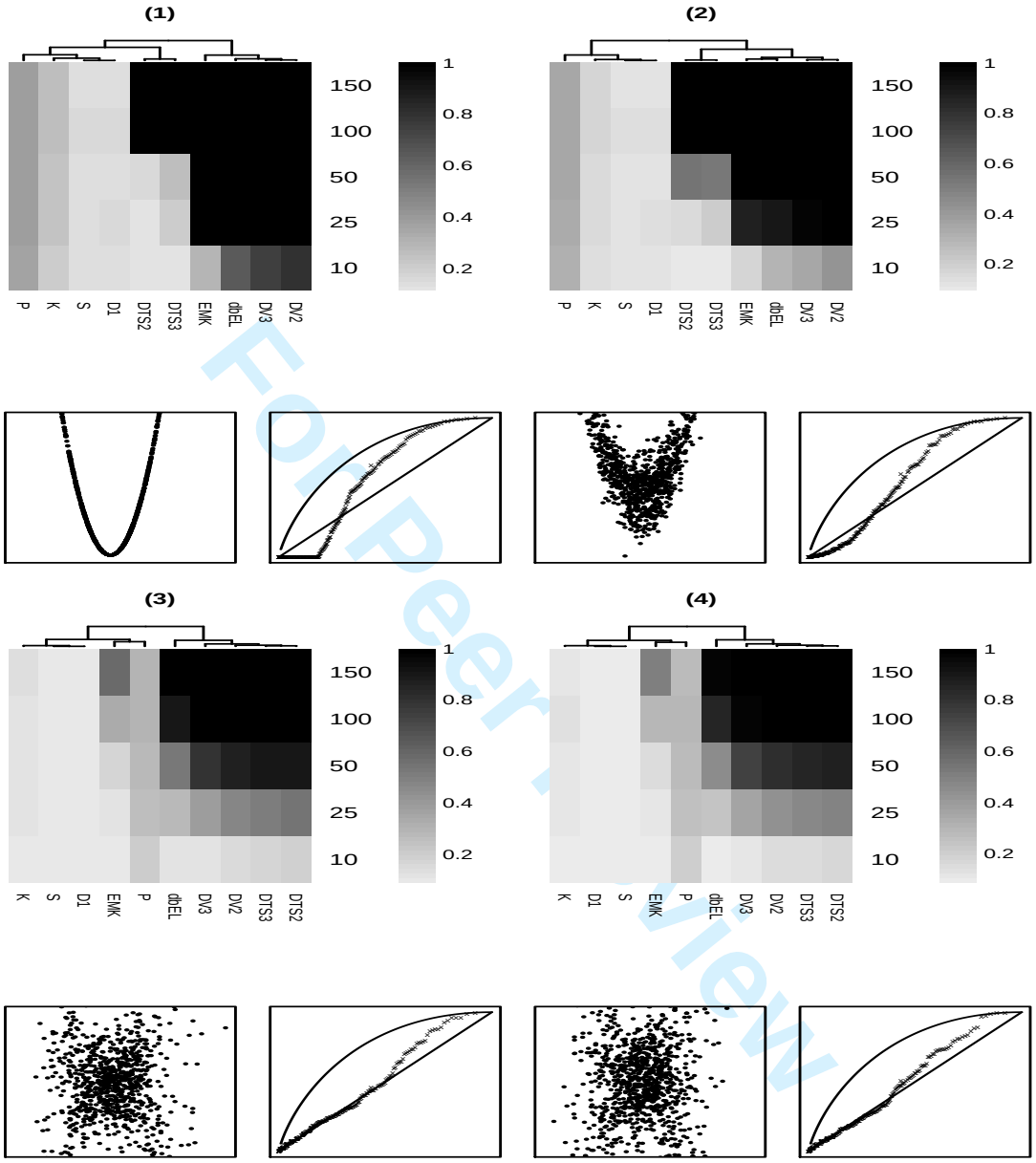


Figure 7: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M2 (Quadratic) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 1(0.576); (2) 0.636(0.563); (3) 0.244(0.52); and(4) 0.242(0.519).

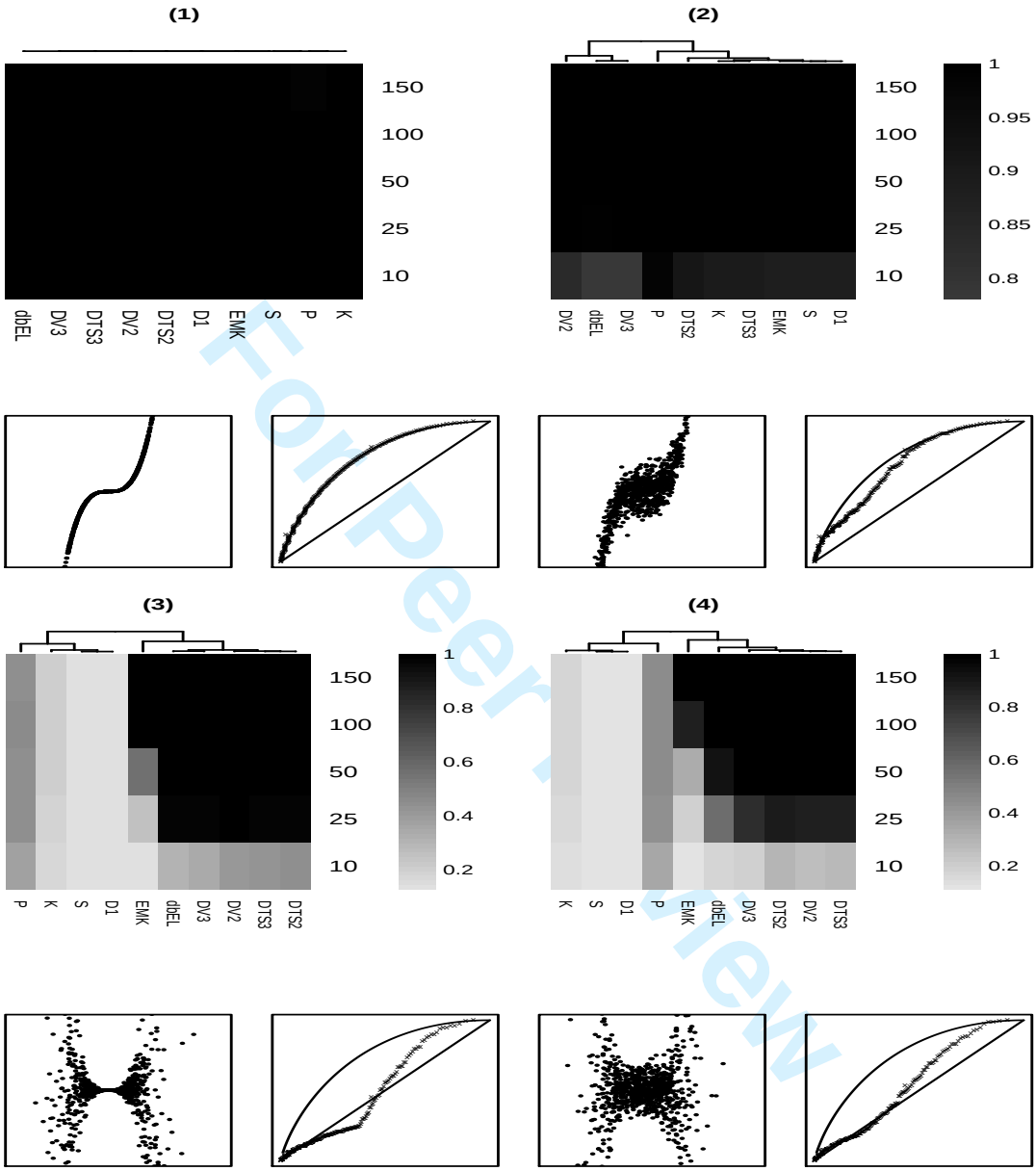


Figure 8: Heat maps of different tests at the 0.05 significance level under the dependence structure of M4 (Cubic) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and Kendall plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 1(0.75); (2) 0.708(0.703); (3) 0.484(0.529); and (4) 0.305(0.526).

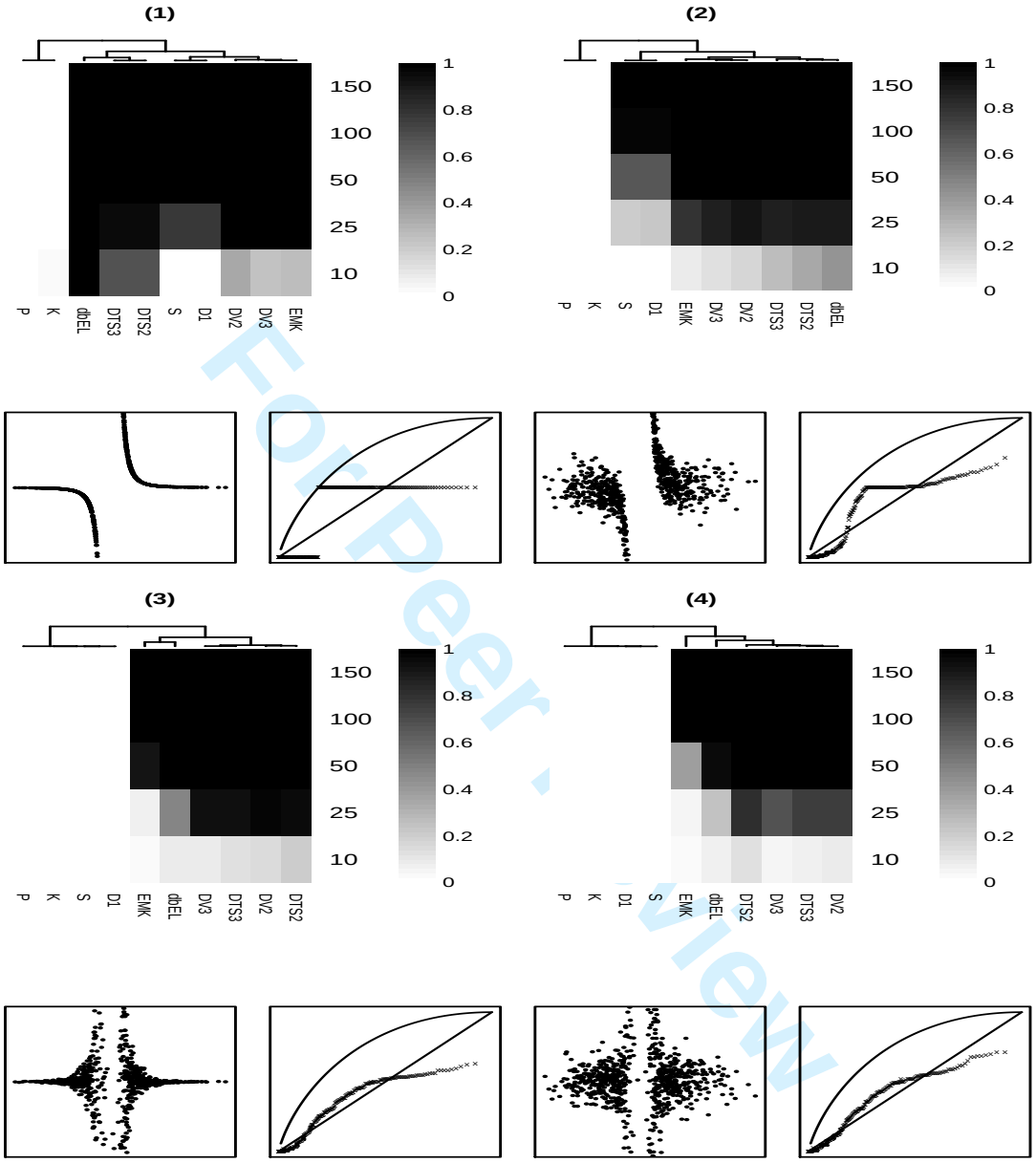


Figure 9: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M7 (Reciprocal with $k = 3$) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 1(0.438); (2) 0.708(0.48); (3) 0.488(0.452); and (4) 0.445(0.469).

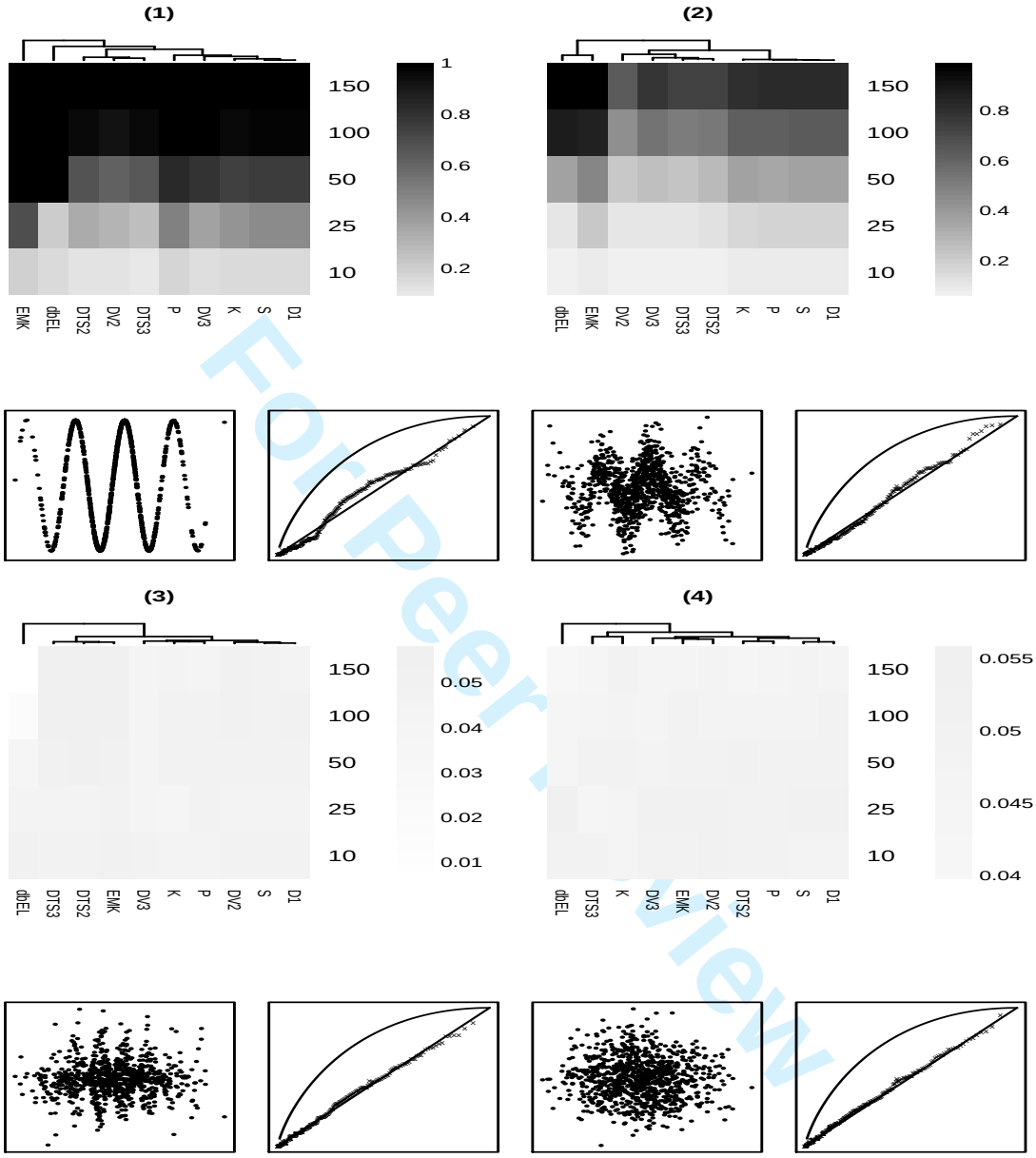


Figure 10: Heat maps of different tests at the 0.05 significance level under the dependence structure of M9' ($\sin(\pi X)$ where $X \sim Uniform[-2, 2]$) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 1.000(0.397); (2) 0.391(0.443); (3) 0.242(0.501); and (4) 0.237(0.503).

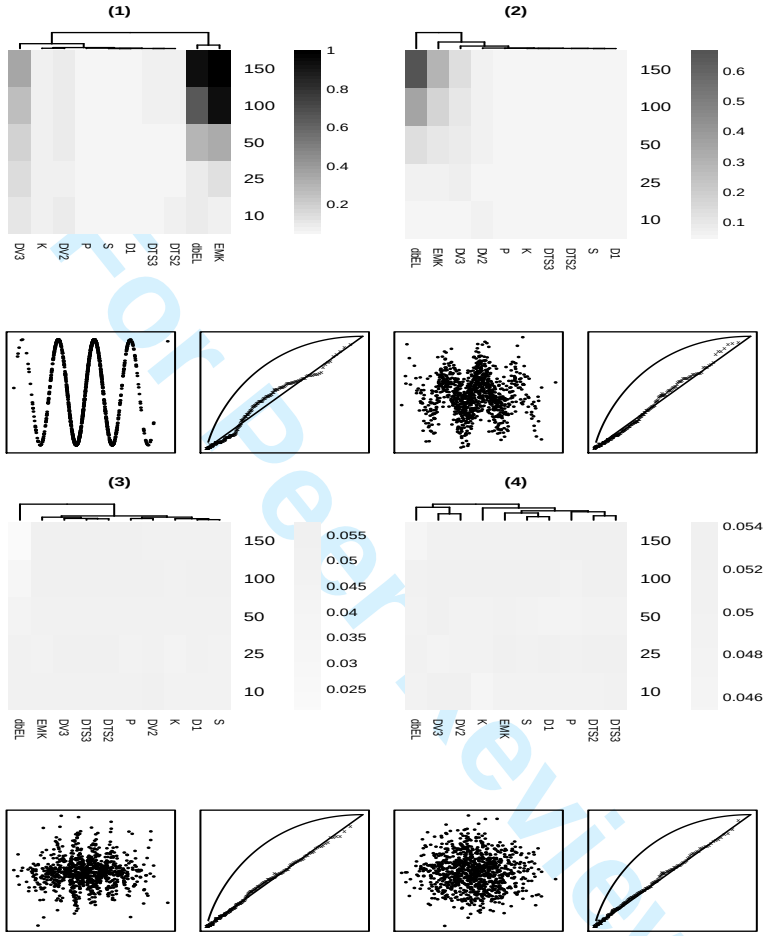


Figure 11: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure M9 ($\sin(\pi x)$) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and K-plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.966(0.503); (2) 0.368(0.502); (3) 0.24(0.503); and (4) 0.237(0.503).

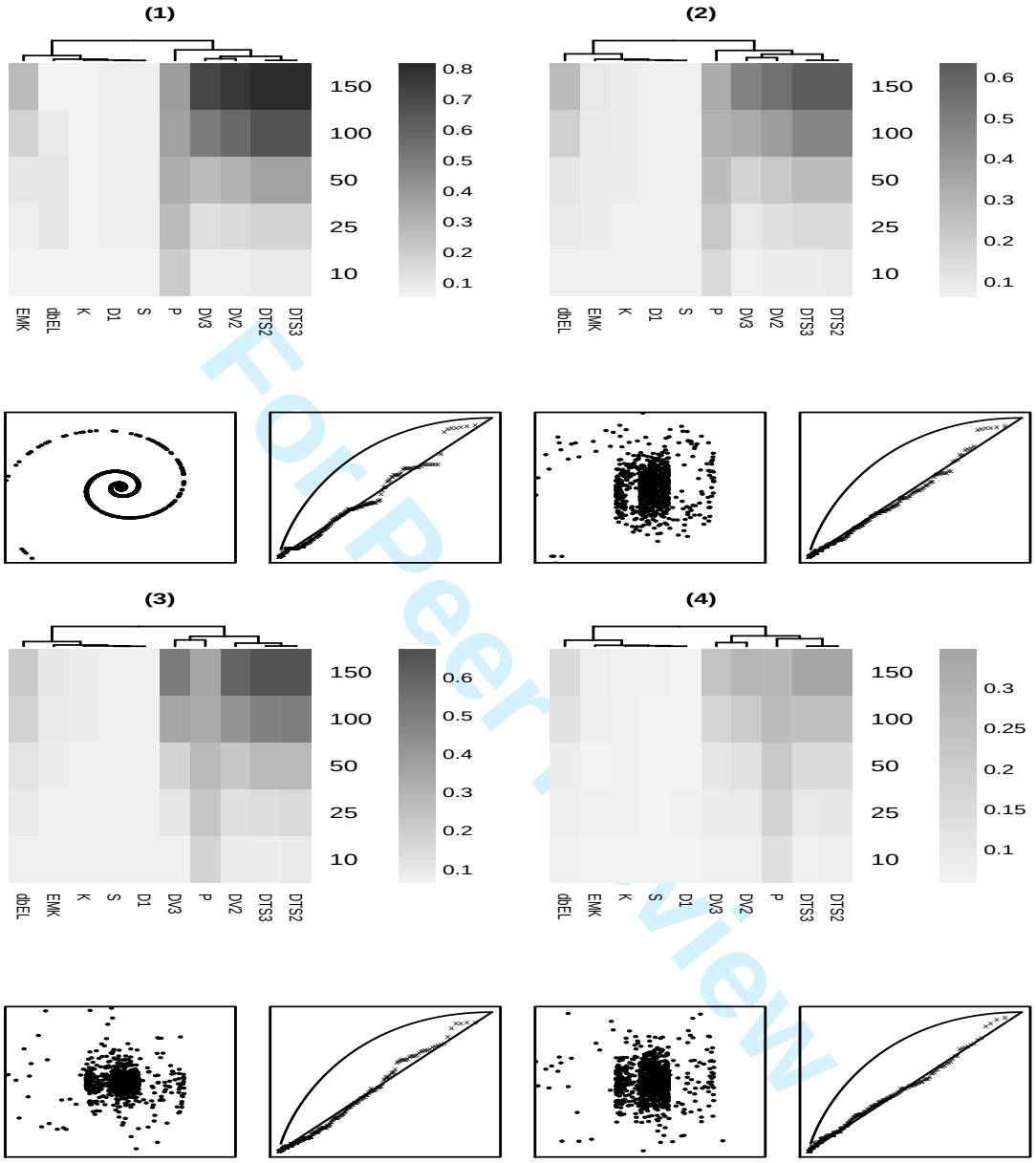


Figure 12: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M11 (Spiral) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.287(0.508); (2) 0.241(0.512); (3) 0.239(0.512); and (4) 0.237(0.509).

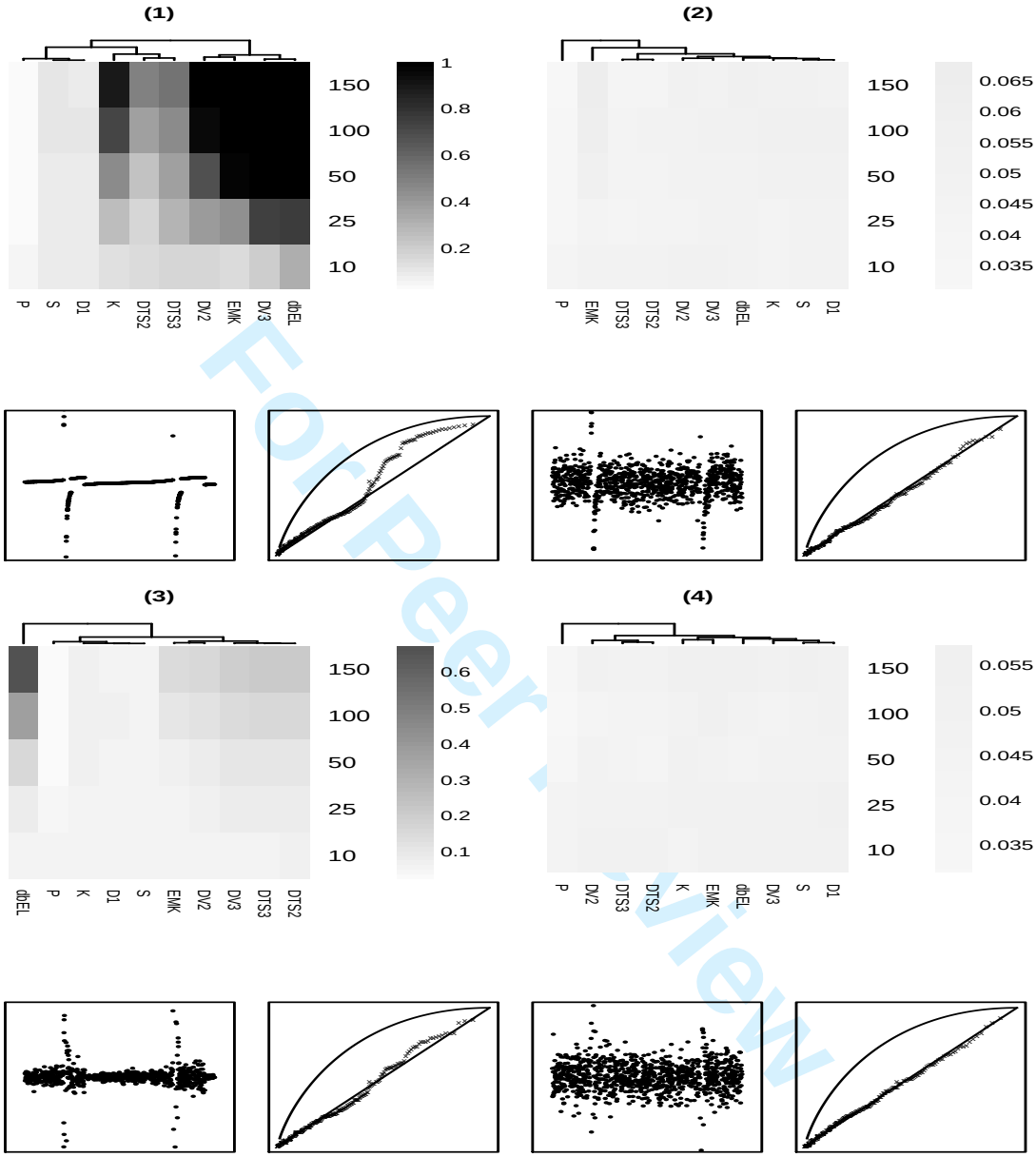


Figure 13: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M24 (DDF1) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.909(0.576); (2) 0.248(0.509); (3) 0.245(0.508); and (4) 0.237(0.505).

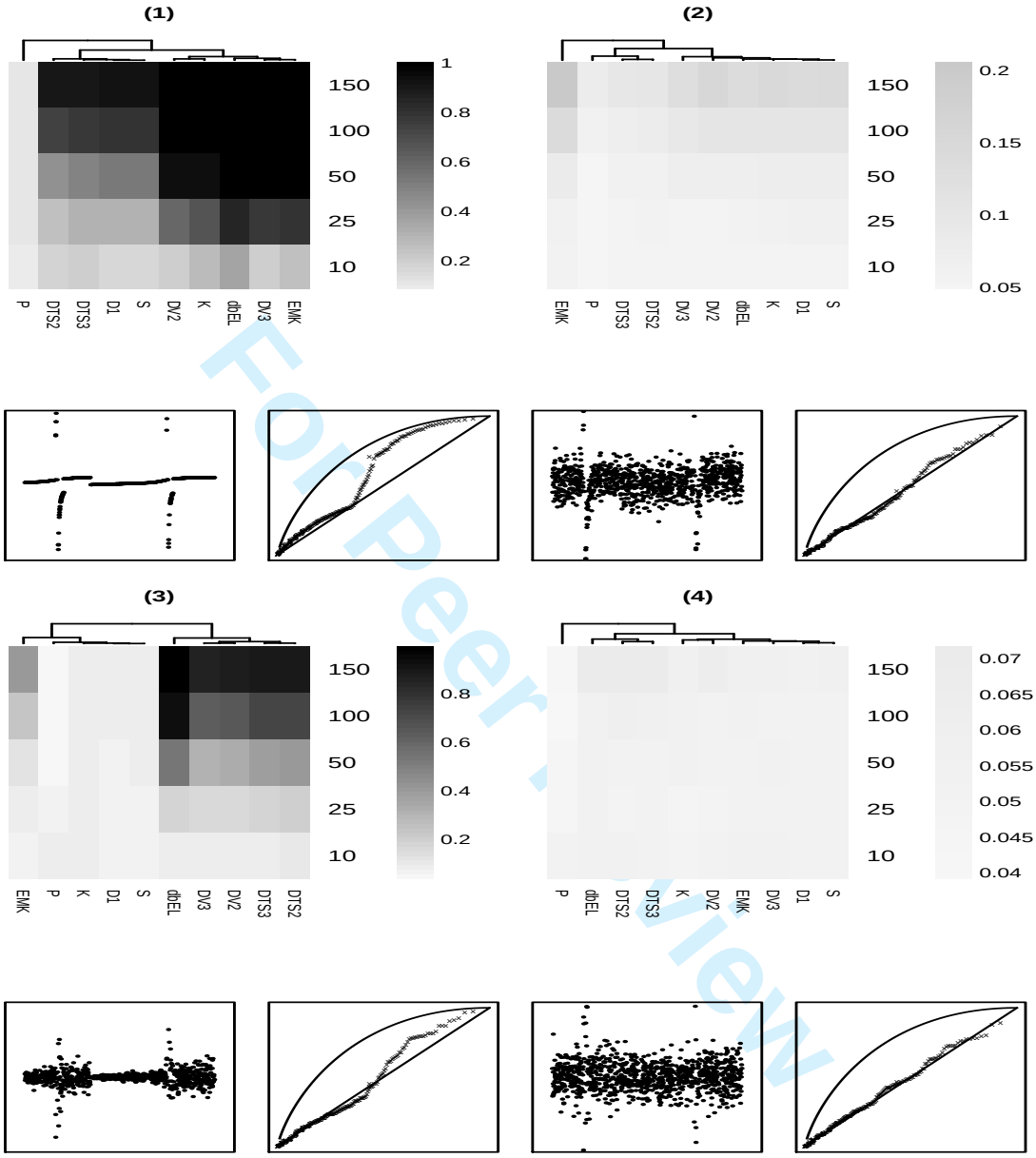


Figure 14: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M25 (DDF2) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.941(0.628); (2) 0.255(0.524); (3) 0.271(0.519); and (4) 0.237(0.506).

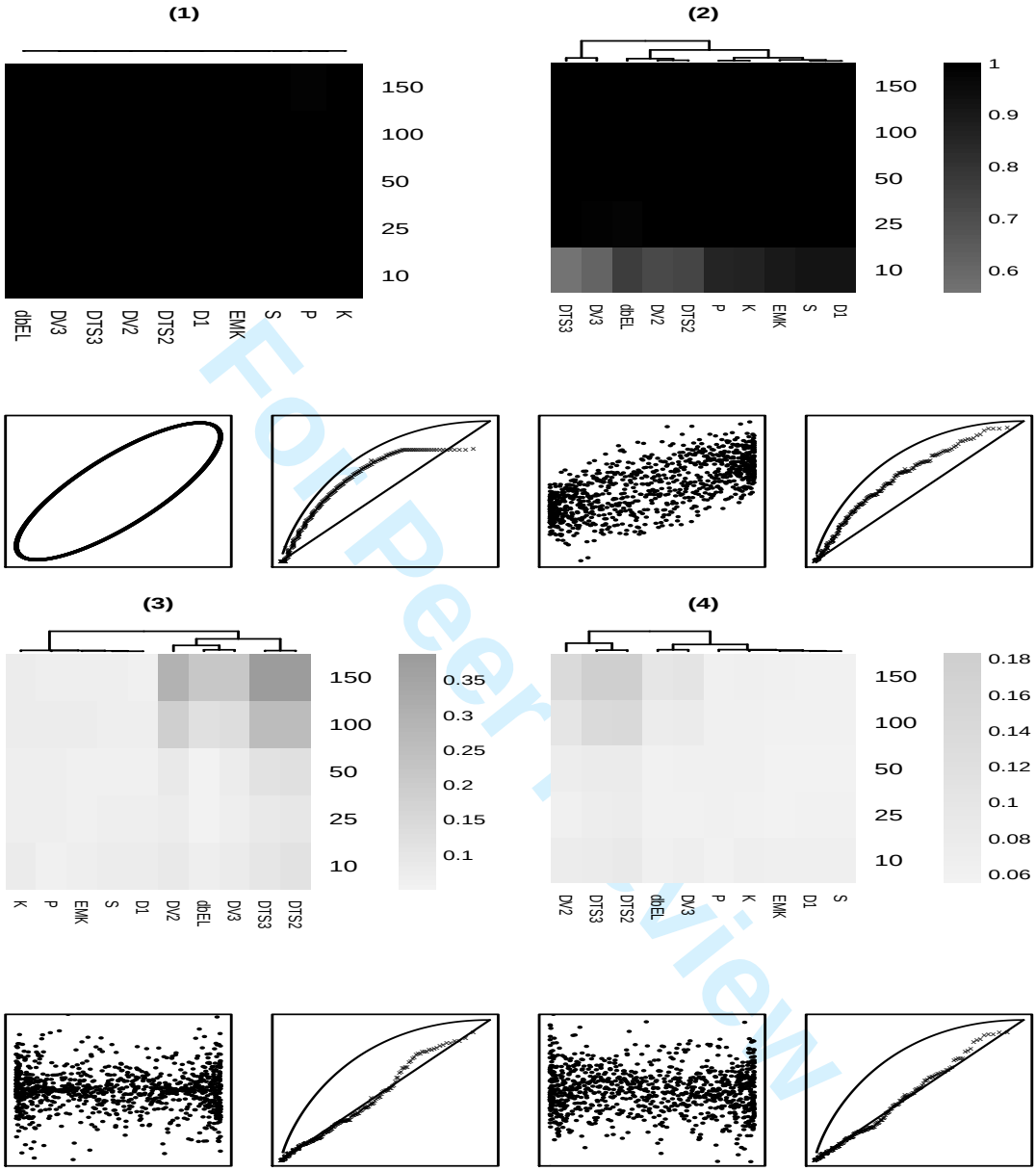


Figure 15: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M11 (ellipse) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.64(0.628); (2) 0.558(0.63); (3) 0.235(0.51); and (4) 0.237(0.508).

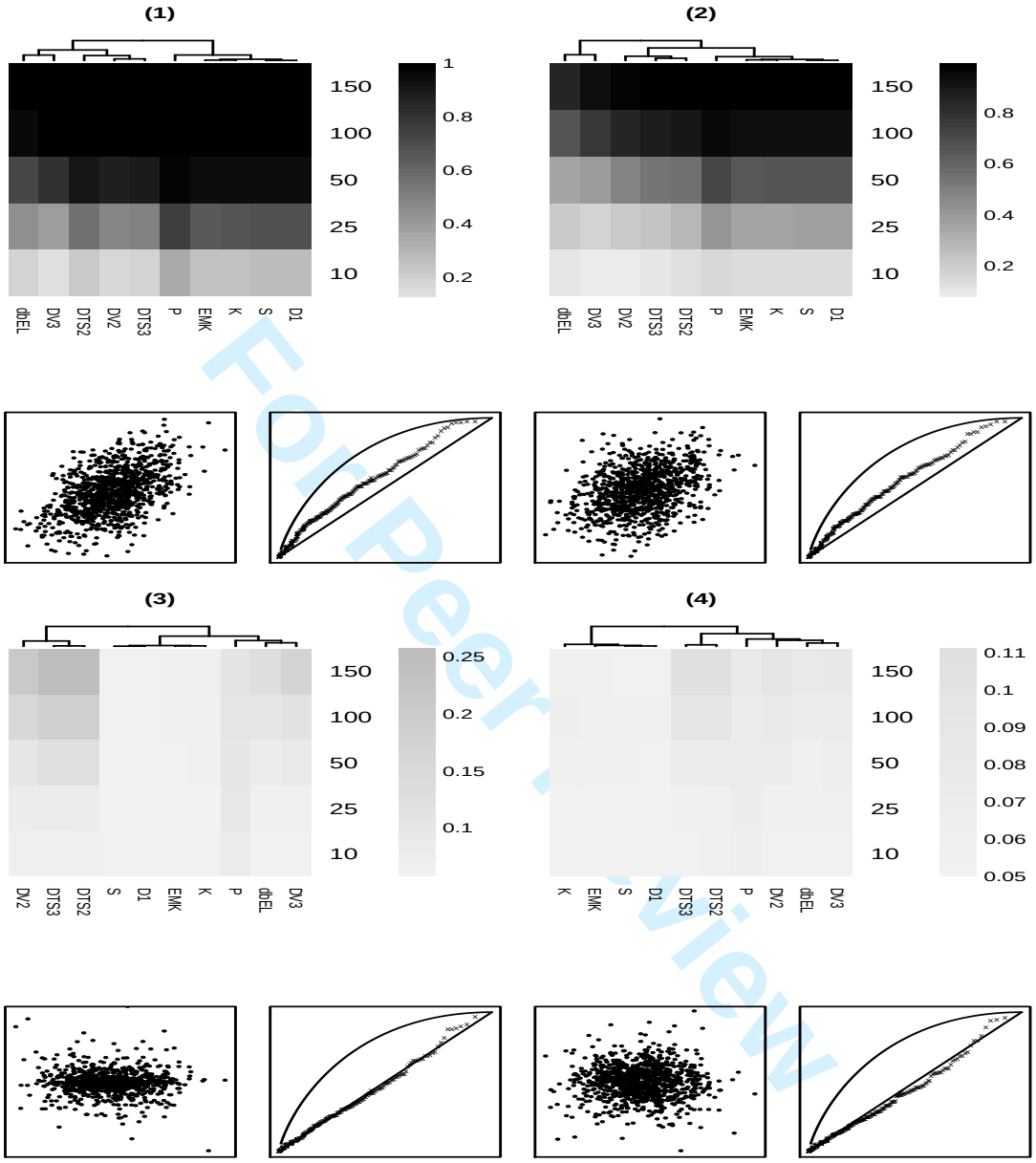


Figure 16: Heat maps of different tests at the 0.05 significance level under the dependence structure of M12 (Normal with $\rho = 0.5$) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and Kendall plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.359(0.603); (2) 0.295(0.574); (3) 0.238(0.508); and (4) 0.237(0.506).

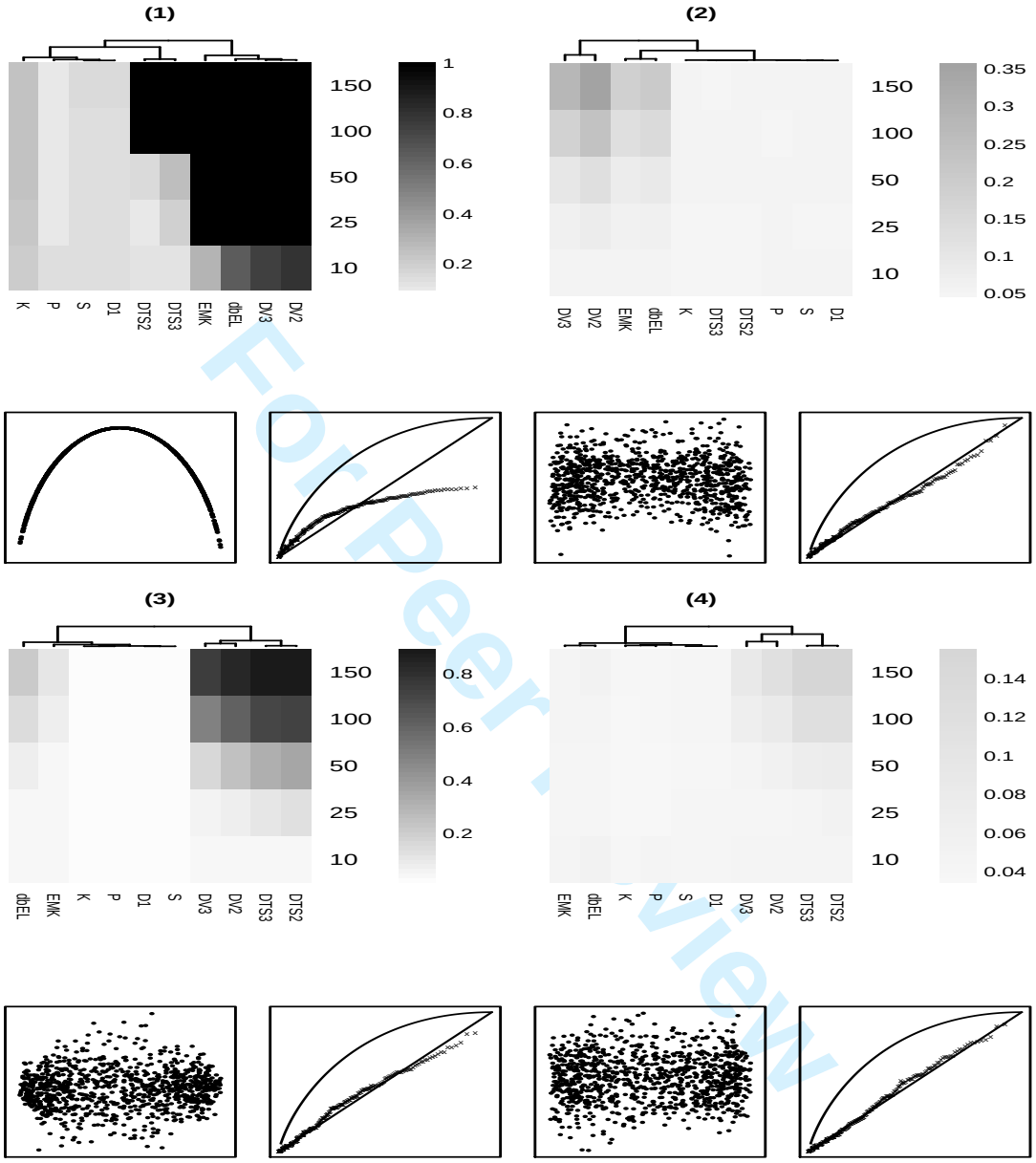


Figure 17: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M13 (NormOffset) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 1(0.406); (2) 0.246(0.491); (3) 0.238(0.492); and (4) 0.237(0.5).

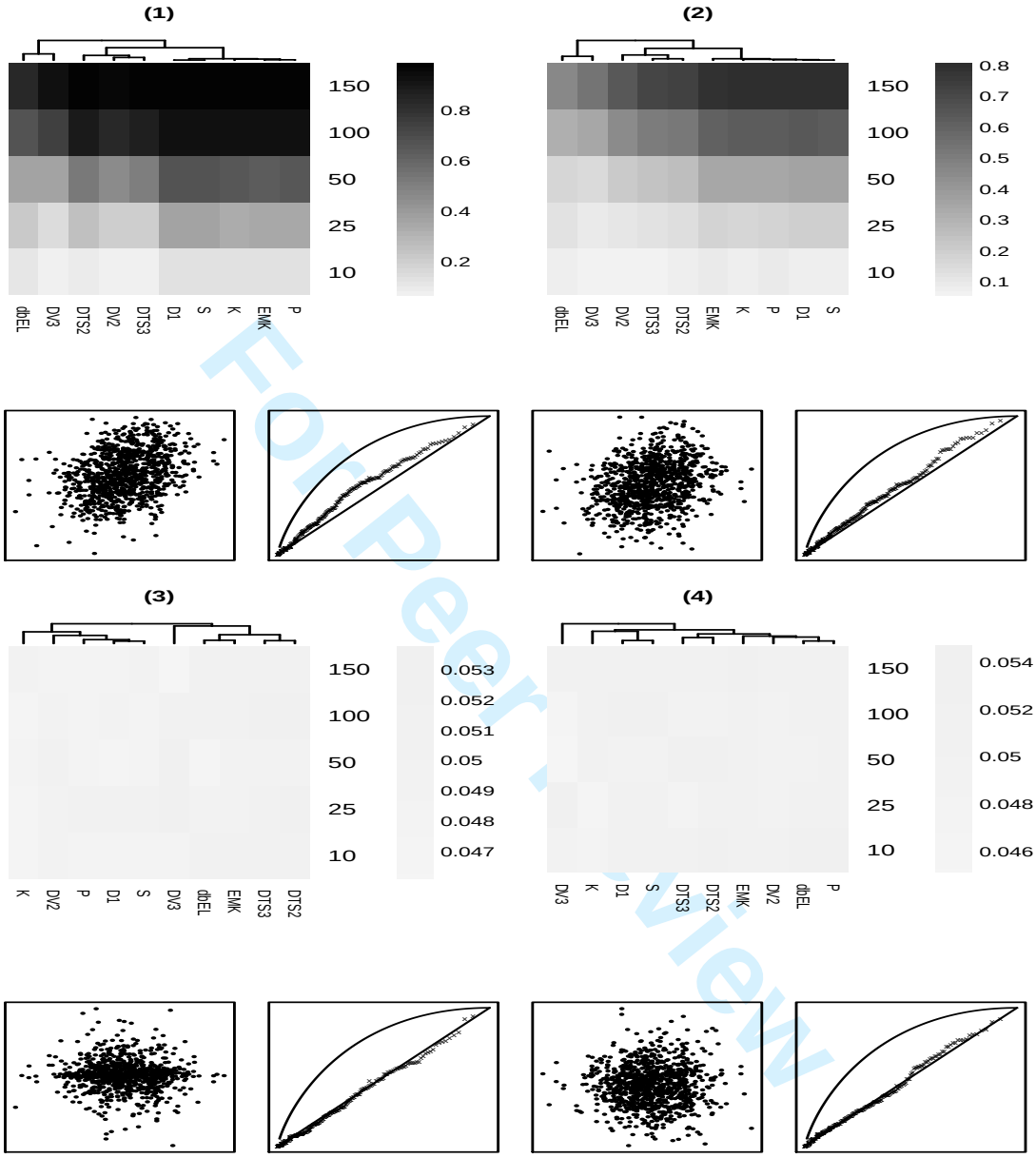


Figure 18: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M14 (Morgenstern) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.298(0.568); (2) 0.265(0.55); (3) 0.238(0.504); and (4) 0.238(0.504).

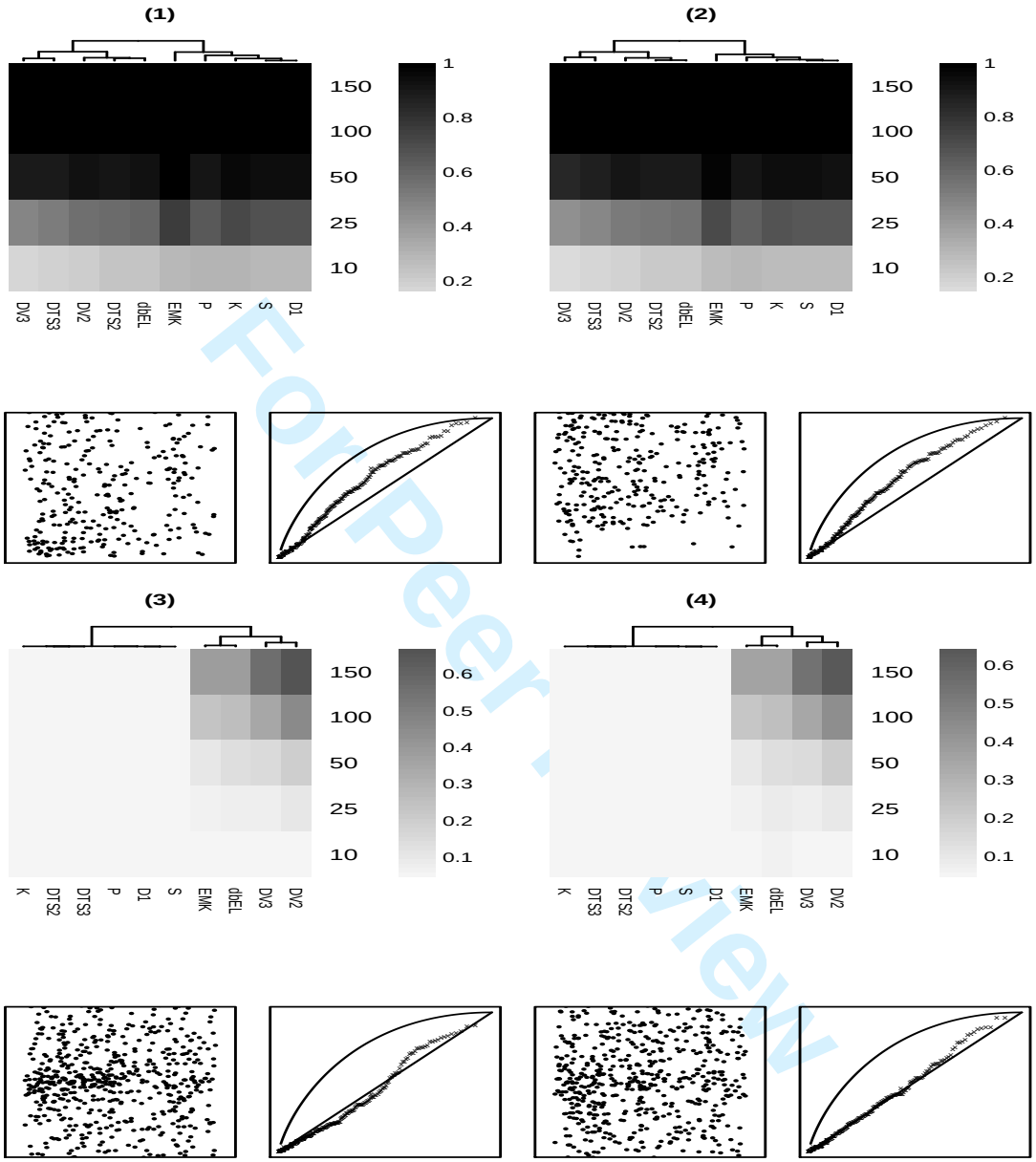


Figure 19: Heat maps of different tests at the 0.05 significance level under the dependence structure of M15 (Placket) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and Kendall plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.455(0.622); (2) 0.42(0.616); (3) 0.257(0.519); and (4) 0.256(0.519).

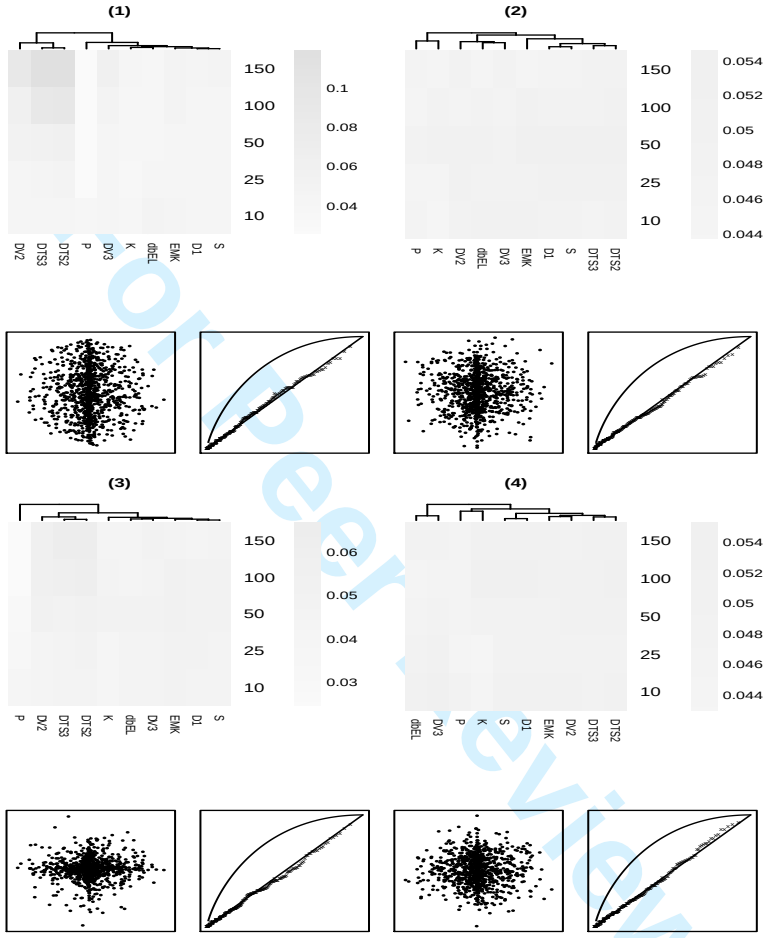


Figure 20: Heat maps of different tests at the 0.05 significance level under the dependence structure of M16 (PearsonII) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and Kendall plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.237(0.5); (2) 0.237(0.503); (3) 0.238(0.502); and (4) 0.237(0.503).

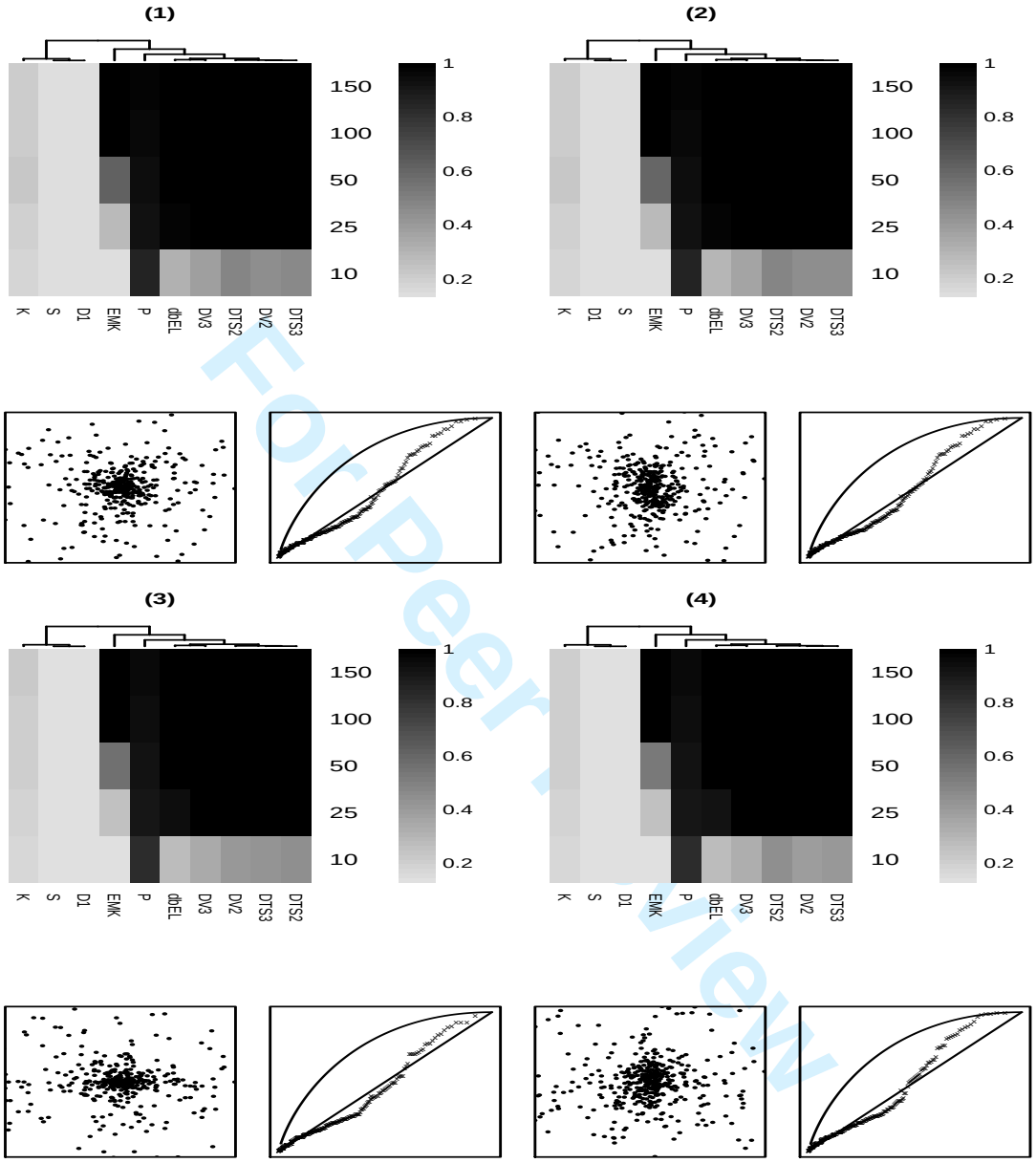


Figure 21: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M17 (PearsonVII) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.463(0.532); (2) 0.459(0.532); (3) 0.425(0.531); and (4) 0.42(0.53).

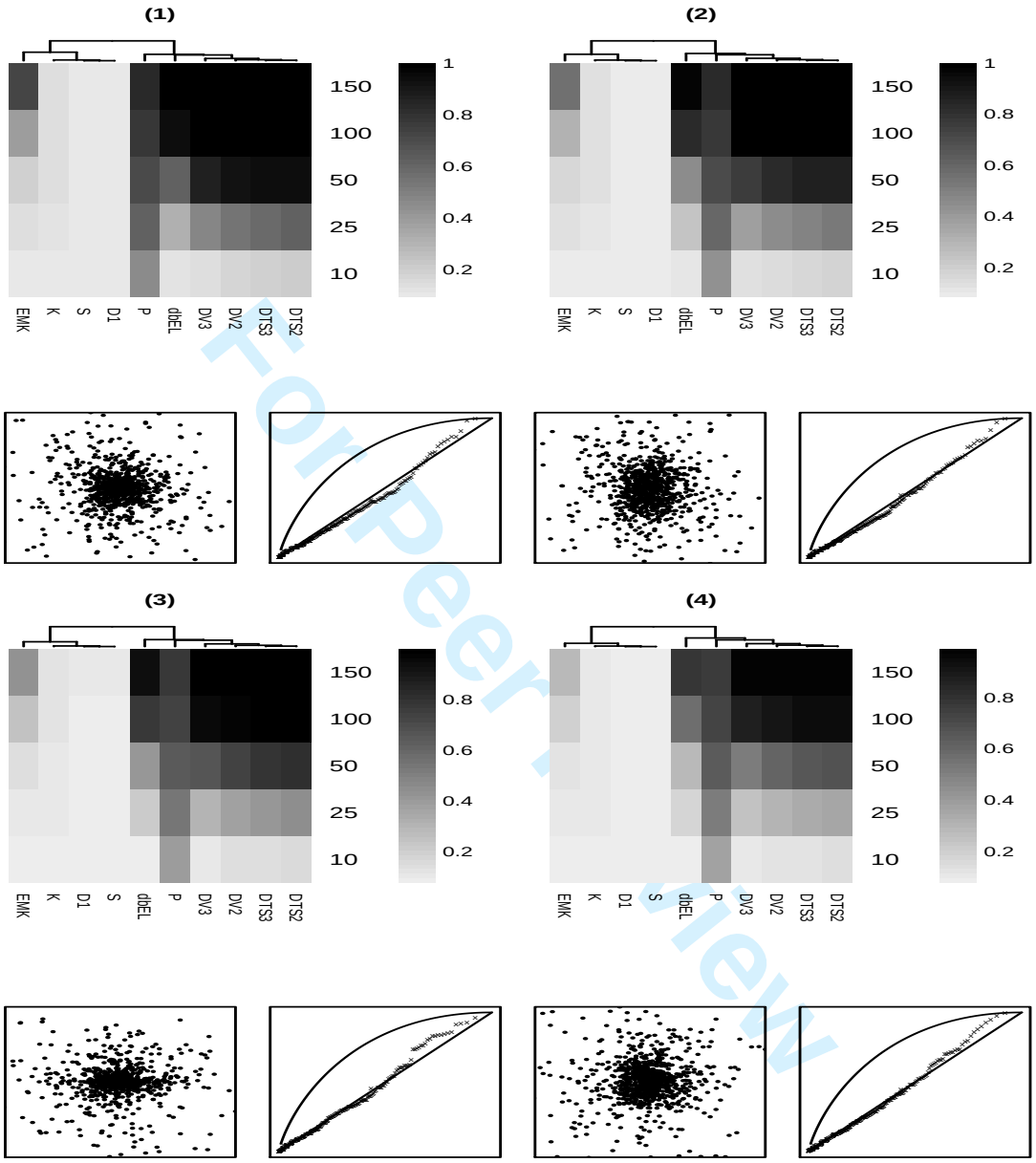


Figure 22: Heat maps of different tests at the 0.05 significance level under the dependence structure of M18 (Cauchy) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and Kendall plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.245(0.522); (2) 0.241(0.52); (3) 0.24(0.519); and (4) 0.238(0.518).

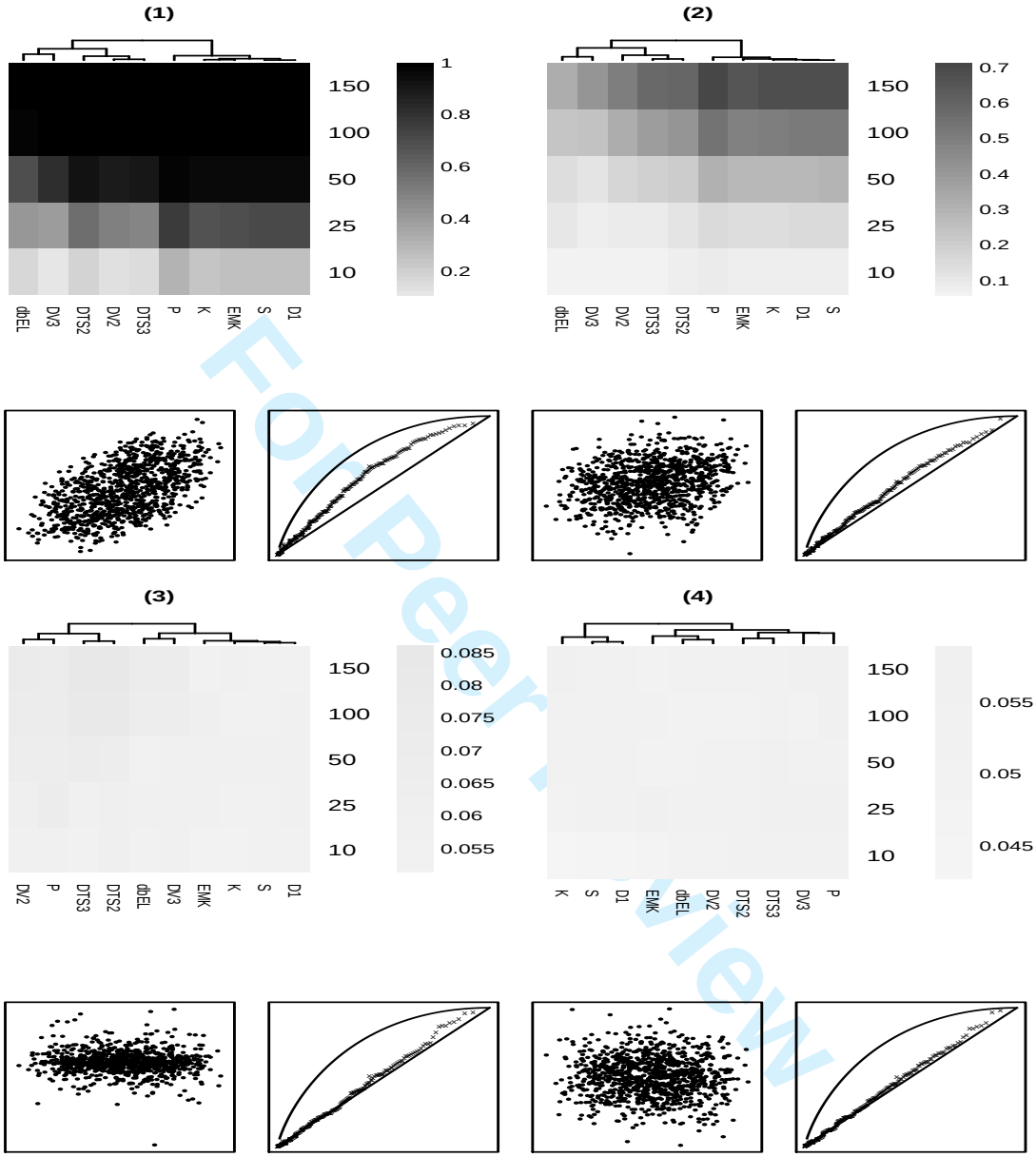


Figure 23: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M19 (EP) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.37(0.6); (2) 0.256(0.545); (3) 0.237(0.505); and (4) 0.237(0.504).

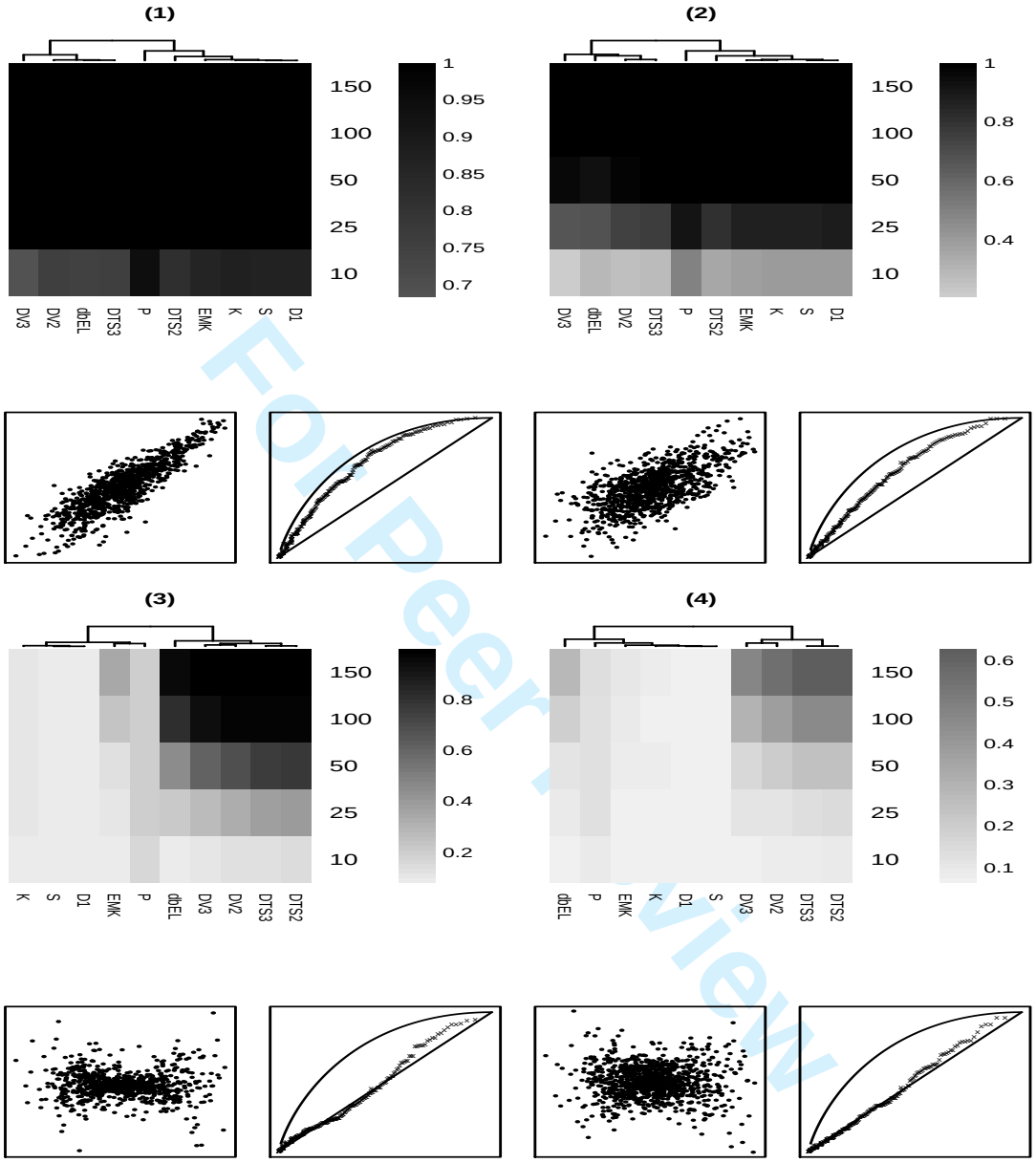


Figure 24: Graphical summarization and the comparison of powers of considered tests at the 0.05 significance level, based on data with the dependence structure of M20 (Gumbel) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left) and Kendall plot (in the right). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.693(0.692); (2) 0.426(0.629); (3) 0.242(0.52); and (4) 0.237(0.513).

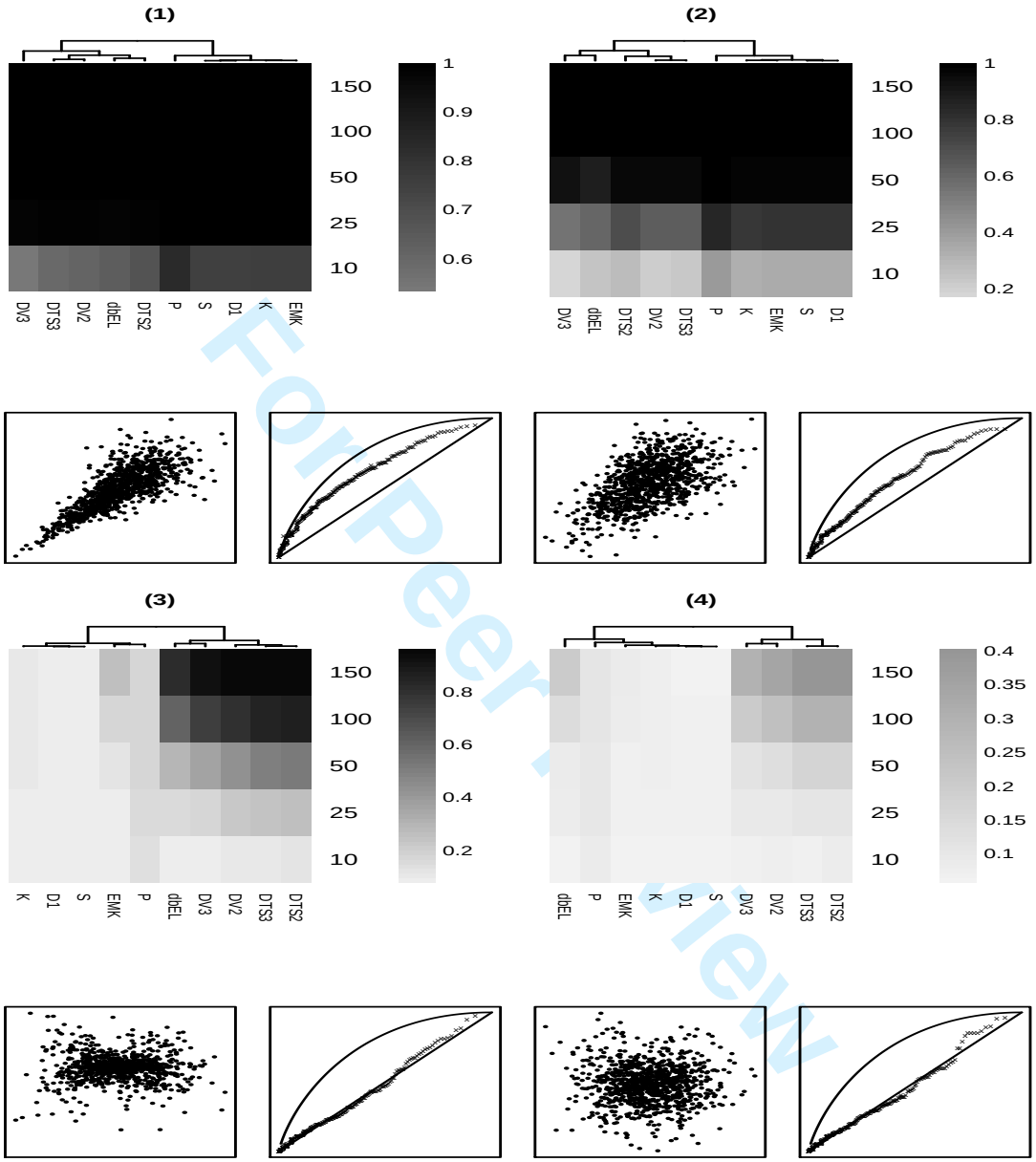


Figure 25: Heat maps of different tests at the 0.05 significance level under the dependence structure of M21 (Clayton) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and Kendall plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.642(0.646); (2) 0.403(0.602); (3) 0.241(0.506); and (4) 0.238(0.505).

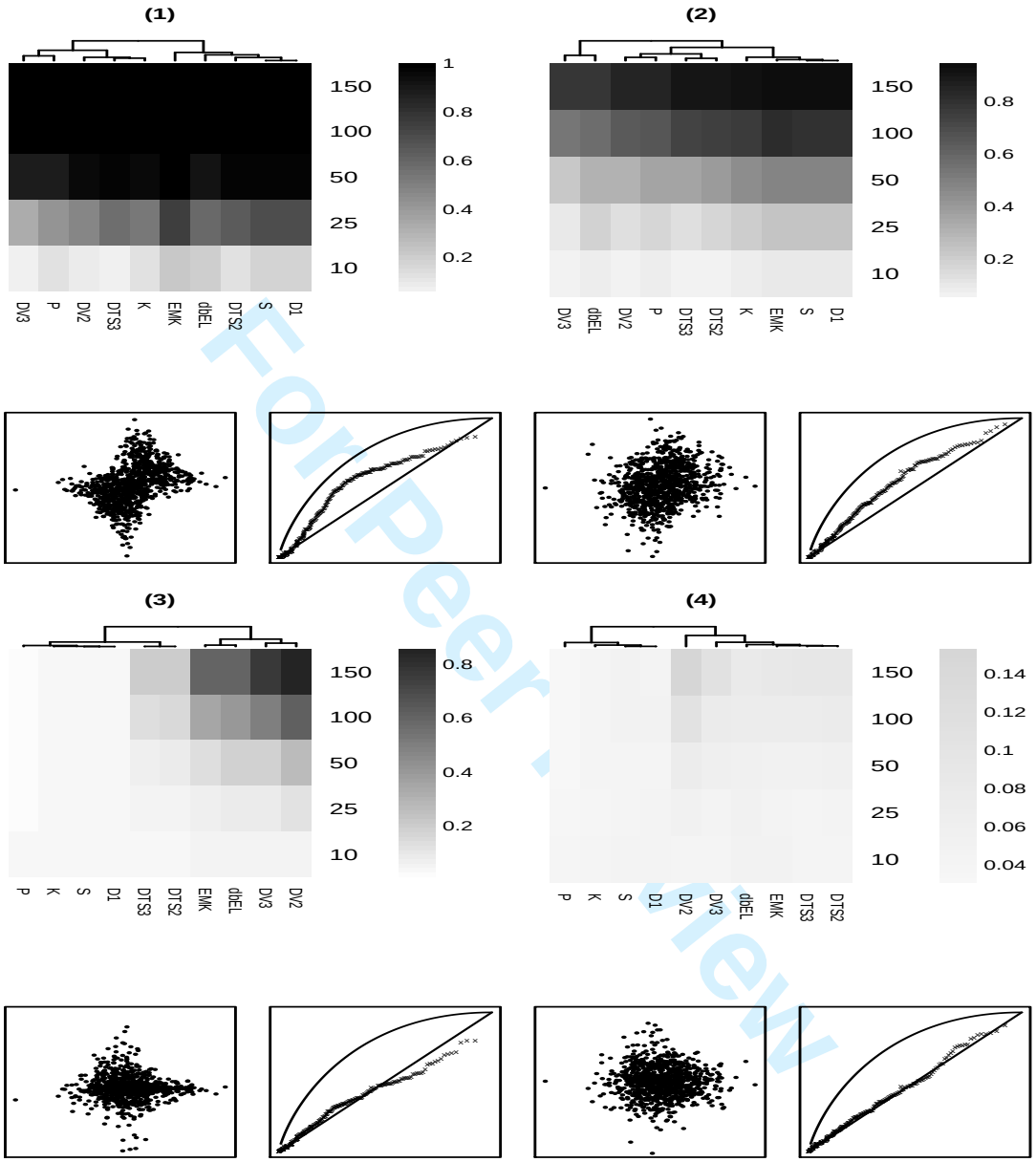


Figure 26: Heat maps of different tests at the 0.05 significance level under the dependence structure of M25 (CondN) and various types of random effects which correspond to one of the sub-panels. In each sub-panel (1-4), the top level is the heat map and the bottom level shows the scatterplot (in the left panel) and Kendall plot (in the right panel). The following random-effect schemes are considered: (1) $X = X_0$, $Y = Y_0$; (2) $X = X_0$, $Y = Y_0 + \varepsilon_A$; (3) $X = X_0$, $Y = \varepsilon_M Y_0$; (4) $X = X_0$, $Y = \varepsilon_M Y_0 + \varepsilon_A$, where the random effects ε_A and $\varepsilon_M \sim N(0, 1)$. The measures of MIC and (AUK) for each type of random effects are (1) 0.425(0.567); (2) 0.292(0.55); (3) 0.268(0.478); and (4) 0.242(0.494).

References

1. Cox DR and Hinkley DV. *Theoretical Statistics*. CRC Press, 1979.

2. Cont R. *Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues*. Taylor & Francis, 2001.

3. Wheeler B. *SuppDists: Supplementary Distributions*, 2013. R package version 1.1-9.1.

4. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.

5. Kallenberg WC and Ledwina T. Data-driven rank tests for independence. *J Amer Statist Assoc* 1999; **94**(445): 285–301.

6. Klauer K. Non-exponential families of distributions. *Metrika* 1986; **33**(1): 299–305.

7. Vexler A, Tsai WM and Hutson AD. A simple density-based empirical likelihood ratio test for independence. *Amer Statist* 2014; **68**(3): 158–169.

8. Vexler A, Schisterman EF and Liu A. Estimation of roc curves based on stably distributed biomarkers subject to measurement error and pooling mixtures. *Stat Med* 2008; **27**(2): 280–296.

9. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psych* 1975; **12**(4): 387–415.

10. Breiman L, Friedman J, Stone CJ and Olshen RA. *Classification and Regression Trees*. CRC press, 1984.

11. Schriever BF et al. An ordering for positive dependence. *The Annals of Statistics* 1987; **15**(3): 1208–1214.

12. Johnson ME. *Multivariate Statistical Simulation: A Guide to Selecting and Generating Continuous Multivariate Distributions*. John Wiley & Sons, 2013.

13. Castillo E and Galambos J. Conditional distributions and the bivariate normal distribution. *Metrika* 1989; **36**(1): 209–214.

14. Kotz S, Balakrishnan N and Johnson NL. *Continuous Multivariate Distributions, Models and Applications*, **1**. John Wiley & Sons, 2004.

15. Jewell EL, Darcy KM, Hutson A, Lee PS, Havrilesky LJ, Grace LA, Berchuck A and Secord AA. Association between the n-terminally truncated (δn) p63α (δnp63α) isoform and debulking status, vegf expression and progression-free survival in previously untreated, advanced stage epithelial ovarian cancer: A gynecologic oncology group study. *Gynecologic Oncology* 2009; **115**(3): 424–429.