Taylor & Francis
Taylor & Francis Group

∂ OPEN ACCESS

Robust Maximum Association Estimators

Andreas Alfons[a], Christophe Croux[b], and Peter Filzmoser[c]

[a]Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands; [b]Faculty of Economics and Business, Katholieke Universiteit Leuven, Leuven, Belgium; [c]Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria

ABSTRACT

The maximum association between two multivariate variables X and Y is defined as the maximal value that a bivariate association measure between one-dimensional projections $\alpha^t X$ and $\beta^t Y$ can attain. Taking the Pearson correlation as projection index results in the first canonical correlation coefficient. We propose to use more robust association measures, such as Spearman's or Kendall's rank correlation, or association measures derived from bivariate scatter matrices. We study the robustness of the proposed maximum association measures and the corresponding estimators of the coefficients yielding the maximum association. In the important special case of Y being univariate, maximum rank correlation estimators yield regression estimators that are invariant against monotonic transformations of the response. We obtain asymptotic variances for this special case. It turns out that maximum rank correlation estimators combine good efficiency and robustness properties. Simulations and a real data example illustrate the robustness and the power for handling nonlinear relationships of these estimators. Supplementary materials for this article are available online.

1. Introduction

Association between two univariate variables U and V can be measured in several ways. The correlation coefficients of Pearson, Spearman, and Kendall are standard tools in statistical practice. For measuring the degree of association between two multivariate variables X and Y, much less literature is existing. An overview of earlier work is given in Ramsay, Ten Berge, and Styan (1984). Recently, Smilde et al. (2009) proposed a matrix association measure for high-dimensional datasets motivated by applications in biology. Nevertheless, the association measures discussed in those articles are based on condensing information from the full covariance matrices and are not very intuitive.

We introduce a class of measures of association between multivariate variables based on the idea of *projection pursuit*. Suppose that X is a p-dimensional random vector and Y is a q-dimensional random vector, with $p \geq q$. A measure of multivariate association between X and Y can be defined by looking for linear combinations $\alpha^t X$ and $\beta^t Y$ of the original variables having maximal association. Expressed in mathematical terms, we seek a measure

$$\rho_R(X, Y) = \max_{\alpha, \beta} R(\alpha^t X, \beta^t Y), \qquad (1)$$

where R is a measure of association between univariate variables. Using the projection pursuit terminology, R is the *projection index* to maximize. Depending on the choice of R used in the above definition, different measures of association between X and Y are obtained. Taking the classical Pearson correlation for R results in the first *canonical correlation coefficient* (see, e.g.,

Johnson and Wichern 2002). Other choices of R yield measures ρ_R having different properties. The bivariate association measures considered in this article are Spearman's rank correlation, Kendall's τ, and an M-estimator (Huber 1981).

For identifying the vectors $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$ in (1), we impose a unit norm restriction such that

$$(\alpha_R(X, Y), \beta_R(X, Y)) = \operatorname*{argmax}_{\|\alpha\|=1, \|\beta\|=1} R(\alpha^t X, \beta^t Y). \qquad (2)$$

We refer to α_R and β_R as the *weighting vectors*. They indicate the contribution of every single component of X and Y in the construction of the linear combinations $\alpha_R{}^t X$ and $\beta_R{}^t Y$ yielding maximal association. This article studies the robustness and efficiency of the maximum association estimators defined in (1) and (2).

An important special case is given for univariate Y, where the weighting vector α_R can be viewed as a normalized coefficient vector in a linear regression model. Using a rank correlation measure R gives an α_R invariant against monotonic transformations of Y. Consider the regression model

$$F(Y) = a^t X + e, \qquad (3)$$

where F is an unspecified strictly monotonic function and e is a random error term. Han (1987) used the Kendall correlation and proved that the resulting estimator yields consistent estimates of the normalized regression coefficients in (3). Furthermore, asymptotic normality of this estimator was shown by Sherman (1993). Nevertheless, those authors only consider the Kendall correlation and do not study robustness.

A related, but different concept is that of *maximal correlation*. There the aim is to find optimal measurable transformations of the variables in X and Y such that the first canonical correlation coefficient is maximized. This problem is already relevant and nontrivial for $p = q = 1$, where one searches measurable functions F and G such that the Pearson correlation between $F(X)$ and $G(Y)$ is maximal. See Papadatos and Xifara (2013) and López Blázquez and Salamanca Miño (2014) for recent references. If (X, Y) follows a bivariate normal distribution, the maximal correlation equals the absolute value of the Pearson correlation (see Yu 2008, for an elegant proof). For $p > q = 1$, this problem was addressed by Breiman and Friedman (1985), who proposed an algorithm for finding optimal transformations for multiple least-square regression. The general case of both $p > 1$ and $q > 1$ does not seem to have received much attention in the statistical literature. The maximum association measure we define in (1) is different from maximal correlation, since (i) we search for optimal linear combinations of the measured variables without transforming them, and (ii) we gain robustness by taking other choices for R than the Pearson correlation coefficient.

We emphasize that we propose a measure of multivariate association, not a measure of dependence. For instance, if there is perfect association between one component of X and one component of Y, there is also a perfect multivariate association, even if the other components are independent of each other (see Example 1 of the supplementary report Alfons, Croux, and Filzmoser 2016, p. 1). This should not necessarily be seen as a disadvantage of the proposed measure, as we aim at finding linear projections of multivariate random vectors that are highly associated with each other. In addition, $\rho_R(X, Y) = 0$ does not imply independence of X and Y (see Example 2 of the supplementary report Alfons, Croux, and Filzmoser 2016, p. 1). For comparison, the standard Pearson correlation measure is not sufficient as an index of independence either (for nonnormal distributions), since zero correlation does not imply independence. The projection pursuit approach that we follow is convenient, but not sufficient for describing the full set of possible dependencies. We provide a one-number summary—together with weighting vectors—of the most important association between linear projections of two sets of random variables.

This article studies a multivariate association measure and is not aiming to provide a fully robustified version of canonical correlation analysis (CCA). Note that several articles on robust CCA can be found in the literature. One stream of research is devoted to robustly estimating the covariance matrices involved in solving the CCA problem and investigating the properties of this approach (e.g., Taskinen et al. 2006). Robust estimators for the multivariate linear model are considered in Kudraszow and Maronna (2011) to obtain robust canonical coordinates. A robust alternating regression technique has been used in Branco et al. (2005). They also used a projection pursuit-based algorithm to estimate the canonical variates, and they compared the different approaches by means of simulation studies.

In this article, we go much further and study the theoretical robustness properties of the estimators. In addition, we emphasize on the special case of regression, where asymptotic variances are computed. As a further contribution, we use an extension of the *grid algorithm*, which was developed by Croux,

Filzmoser, and Oliveira (2007) for projection pursuit principal component analysis. A fast implementation of this algorithm is made available for the statistical computing environment R in package ccaPP.

2. Definitions and Basic Properties

Denote R the projection index to be maximized in (1). The association measure R should verify the following properties, where (U, V) stands for any pair of univariate random variables:

(i) $R(U, V) = R(V, U)$

(ii) $R(aU + b, cV + d) = \text{sign}(ac)R(U, V)$ for all $a, b, c, d \in \mathbb{R}$

(iii) $-1 \leq R \leq 1$

Note that condition (ii) gives $R(-U, V) = -R(U, V)$, therefore the proposed measure defined in (1) verifies $\rho_R \geq 0$. If (U, V) follows a distribution F, we denote $R(F) \equiv R(U, V)$.

The equivariance property (ii) ensures the association measures to be invariant under affine transformations. Indeed, for any nonsingular matrices A and B and vectors a and b, it holds that

$$\rho_R(AX + a, BY + b) = \rho_R(X, Y).$$

The weighting vectors are affine equivariant in the sense that

$$\alpha_R(AX + a, BY + b) = (A^t)^{-1}\alpha_R(X, Y)/\|(A^t)^{-1}\alpha_R(X, Y)\|$$

and similarly for β_R. Now we briefly review the definitions of several bivariate association measures R.

2.1. Pearson Correlation

This classical measure for linear association is defined as

$$R_P(U, V) = \frac{\text{cov}(U, V)}{\sqrt{\text{var}(U)\text{var}(V)}}. \tag{4}$$

The maximization problem in (1) can now be solved explicitly, since it corresponds to the definition of the first canonical correlation coefficient. We have that ρ_{R_P} is given by the square root of the largest eigenvalue of the matrix

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}, \tag{5}$$

where $\Sigma_{XX} = \text{cov}(X)$, $\Sigma_{YY} = \text{cov}(Y)$, $\Sigma_{XY} = \text{cov}(X, Y)$, and $\Sigma_{YX} = \Sigma_{XY}^t$. Existence of ρ_{R_P} requires existence of second moments, while the other measures to be discussed do not require any existence of moments.

We do not aim at constructing robust maximum association measures by plugging robust scatter matrices into (5). The projection pursuit approach that we follow does not require to compute the full scatter matrices and can also be applied if the number of variables in any of the two datasets exceeds the number of observations. However, this article is not focused on such high-dimensional applications.

2.2. Spearman and Kendall Correlation

These well-known measures are based on ranks and signs. The Spearman rank correlation is defined as

$$R_S(U, V) = R_P(\text{rank}(U), \text{rank}(V)),$$

where $\text{rank}(u) = F_U(u)$, with F_U the cumulative distribution function of U, stands for the population rank of u. The definition of Kendall's τ is

$$R_K(U, V) = E[\text{sign}((U_1 - U_2)(V_1 - V_2))],$$

where (U_1, V_1) and (U_2, V_2) are two independent copies of (U, V). Estimators of the population measures are simply given by the sample counterparts. For example, from an iid sample $(u_1, v_1), \ldots, (u_n, v_n)$, we can compute the sample version of $R_K(U, V)$ as

$$\hat{R}_K = \frac{1}{\binom{n}{2}} \sum_{i < j} \text{sign}((u_i - u_j)(v_i - v_j)).$$

2.3. M-Association Derived From a Bivariate M-Scatter Matrix

A robust scatter matrix C is as an alternative to the classical covariance matrix. We use M-estimators of Maronna (1976), since they are quite efficient, but also robust in the bivariate case. Indeed, their breakdown point is $1/(k + 1)$, where k denotes the number of variables. Given a two-dimensional variable $Z = (U, V)^t$, the M-location $\boldsymbol{\mu}(Z)$ and M-scatter matrix $C(Z)$ are implicitly defined as solutions of the equations

$$\begin{aligned}
\boldsymbol{\mu} &= E\left[w_1\left((Z - \boldsymbol{\mu})^t C^{-1}(Z - \boldsymbol{\mu})\right) Z\right] / \\
&\quad E\left[w_1\left((Z - \boldsymbol{\mu})^t C^{-1}(Z - \boldsymbol{\mu})\right)\right] \\
C &= E\left[w_2\left((Z - \boldsymbol{\mu})^t C^{-1}(Z - \boldsymbol{\mu})\right) (Z - \boldsymbol{\mu})(Z - \boldsymbol{\mu})^t\right],
\end{aligned}$$

where $\boldsymbol{\mu}$ is a bivariate vector and C is a symmetric positive definite two-by-two matrix. Furthermore, w_1 and w_2 are specified weight functions. We focus on Huber's M-estimator, where $w_1(d) = \max(1, \delta/d)$ and $w_2(d) = c \max(1, (\delta/d)^2)$, with $\delta = \chi^2_{2,0.9}$ the 10% upper quantile of a chi-squared distribution with 2 degrees of freedom and c selected to obtain a consistent estimator of the covariance matrix at normal distributions (Huber 1981). M-estimators of scatter can be considered as (iteratively) reweighted covariance matrices, and are easy to compute. The association measure implied by a bivariate scatter matrix $C(Z) \equiv C(U, V)$ is then simply given by

$$R_C(U, V) = \frac{C_{12}(U, V)}{\sqrt{C_{11}(U, V) C_{22}(U, V)}},$$

where $C_{11}(U, V)$ and $C_{22}(U, V)$ are the diagonal elements of $C(U, V)$, and $C_{12}(U, V) = C_{21}(U, V)$ are the off-diagonal elements. The association measure based on Huber's M-scatter matrix is denoted by R_M.

It is important to realize that different measures of association R represent different population quantities in (1) and (2). Consider a bivariate elliptical distribution F_ρ with location zero and scatter matrix

$$\Sigma_\rho = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \tag{6}$$

For studying association measures verifying the equivariance property (ii), we can take $\sigma_1 = \sigma_2 = 1$ without loss of generality. The density of F_ρ can be written as

$$f_\rho(\boldsymbol{x}) = g_2(\boldsymbol{x}^t \Sigma_\rho^{-1} \boldsymbol{x}), \tag{7}$$

with $\boldsymbol{x} \in \mathbb{R}^2$ and g_2 a nonnegative function. For the important case of the bivariate normal, we have $g_2(z) = \exp(-z/2)/(2\pi\sqrt{1 - \rho^2})$ and we denote $F_\rho = \Phi_\rho$. Let the function $\kappa_R : [-1, 1] \to \mathbb{R}$ be

$$\kappa_R(\rho) = R(F_\rho) \quad \text{for any } -1 \le \rho \le 1. \tag{8}$$

The function κ_R depends on the underlying distribution. At the bivariate normal, it is known that

$$\kappa_S(\rho) = \frac{6}{\pi} \sin^{-1}(\rho/2) \quad \text{and} \quad \kappa_K(\rho) = \frac{2}{\pi} \sin^{-1}(\rho)$$

for the Spearman and Kendall correlation. M-estimators are consistently estimating the shape of normal distributions and therefore estimate the same quantity as Pearson's correlation, so $\kappa_M(\rho) = \rho$.

Recall that ρ_R is a measure of multivariate association, not an index of dependence. For instance, it is possible that $\rho_R = 0$ even if X and Y are not independent. This is to be expected, since it is known that a zero Pearson correlation does not imply independence (for nonnormal distributions), and the same holds for the rank correlation measures. Furthermore, it is possible that $\rho_R = 1$ while Y is not fully dependent on X. Take the example where X and Y have the same first component, but all other components are independent of each other. Then the weighting vectors are $\alpha_R = (1, 0, \ldots, 0)^t$ and $\beta_R = (1, 0, \ldots, 0)^t$, and $\rho_R = 1$. The projection pursuit approach is convenient, but it cannot explore the full dependence structure between X and Y. The multivariate maximum association measure ρ_R represents the strongest association between linear projections of two sets of random variables. Together with the weighting vectors, it provides a useful tool for analyzing association between multivariate random vectors.

3. Fisher Consistency and Influence Functions

Take $(X, Y) \sim H$, a $(p + q)$-dimensional distribution function. By convention, $Q(H) \equiv Q(X, Y)$ when $(X, Y) \sim H$, for any statistical functional Q. The statistical functionals of interest are

$$(\boldsymbol{\alpha}_R(H), \boldsymbol{\beta}_R(H)) = \underset{\|\boldsymbol{\alpha}\|=1, \|\boldsymbol{\beta}\|=1}{\text{argmax}} \ R(\boldsymbol{\alpha}^t X, \boldsymbol{\beta}^t Y) \tag{9}$$

and

$$\rho_R(H) = R(\boldsymbol{\alpha}_R(H)^t X, \boldsymbol{\beta}_R(H)^t Y). \tag{10}$$

Let H_0 be elliptically symmetric with scatter matrix

$$\Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix},$$

where due to the translation invariance of the functionals, the location is without loss of generality taken to be zero. We assume that Σ has full rank. It holds that the distribution of the pairs $(\boldsymbol{\alpha}^t X, \boldsymbol{\beta}^t Y)$ all belong to the same bivariate elliptical family with density of the form (7). Denote now

$$r(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\boldsymbol{\alpha}^t \Sigma_{XY} \boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}^t \Sigma_{XX} \boldsymbol{\alpha}} \sqrt{\boldsymbol{\beta}^t \Sigma_{YY} \boldsymbol{\beta}}} = R_P(\boldsymbol{\alpha}^t X, \boldsymbol{\beta}^t Y). \tag{11}$$

Then

$$R(\boldsymbol{\alpha}^t X, \boldsymbol{\beta}^t Y) = R(F_{r(\boldsymbol{\alpha}, \boldsymbol{\beta})}) = \kappa_R \left(r(\boldsymbol{\alpha}, \boldsymbol{\beta}) \right) \qquad (12)$$

with F_ρ defined by (7) for any $-1 < \rho < 1$. We need the following condition:

(iv) $\rho \to \kappa_R(\rho) = R(F_\rho)$ is a strictly increasing and differentiable function.

Here F_ρ is the distribution of $(\boldsymbol{\alpha}^t X, \boldsymbol{\beta}^t Y)$ with $\rho = r(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $(X, Y) \sim H_0$. Assumption (iv) holds for all considered association measures at the normal distribution.

Since κ_R is supposed to be strictly increasing, the functionals $\boldsymbol{\alpha}_R(H_0)$ and $\boldsymbol{\beta}_R(H_0)$ defined in (9) are the same for all association measures verifying condition (iv). Taking $R = R_P$ immediately yields the Fisher consistency property

$$\boldsymbol{\alpha}_R(H_0) = \boldsymbol{\alpha}_1 / \|\boldsymbol{\alpha}_1\| \quad \text{and} \quad \boldsymbol{\beta}_R(H_0) = \boldsymbol{\beta}_1 / \|\boldsymbol{\beta}_1\|, \qquad (13)$$

with $\boldsymbol{\alpha}_1$ (and similarly for $\boldsymbol{\beta}_1$) being the eigenvector corresponding to the largest eigenvalue of (5), where $\boldsymbol{\Sigma}$ is the scatter matrix of the model distribution (the covariance matrix if the second moment exists). The vectors $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ are the first canonical vectors. It follows that

$$\rho_R(H_0) = R(\boldsymbol{\alpha}_R(H_0)^t X, \boldsymbol{\beta}_R(H_0)^t Y) = R(\boldsymbol{\alpha}_1^t X, \boldsymbol{\beta}_1^t Y)$$
$$= \kappa_R \left(r(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1) \right) = \kappa_R(\rho_1), \qquad (14)$$

where ρ_1 stands for the first population canonical correlation.

A fairly simple expression for the influence functions can now be derived. The influence function (IF) gives the influence that an observation $\mathbf{x}$ has on a functional Q at a distribution H. If we denote a point mass distribution at $\mathbf{x}$ by $\Delta_{\mathbf{x}}$ and write $H_\varepsilon = (1 - \varepsilon)H + \varepsilon \Delta_{\mathbf{x}}$, the IF is given by

$$\text{IF}(\mathbf{x}, Q, H) = \frac{\partial}{\partial \varepsilon} Q(H_\varepsilon)|_{\varepsilon=0}$$

(see Hampel et al. 1986). The proof of the following theorem is rather lengthy and is therefore given in a supplementary report (Alfons, Croux, and Filzmoser 2016).

Theorem 1. Let H_0 be an elliptically symmetric distribution and let R be an association measure satisfying conditions (i)–(iv). Denote $\rho_1 > \cdots > \rho_q > 0$ be the population canonical correlations and $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q$ the population canonical vectors. Furthermore, set $\rho_k := 0$ for $k > q$. Denote F_ρ the bivariate elliptical distribution as in condition (iv). The influence function of the association measure ρ_R is then given by

$$\text{IF}((\mathbf{x}, \mathbf{y}), \rho_R, H_0) = \text{IF}((u_1, v_1), R, F_\rho), \qquad (15)$$

with $u_j = \boldsymbol{\alpha}_j^t \mathbf{x}$, $j = 1, \ldots, p$, and $v_j = \boldsymbol{\beta}_j^t \mathbf{y}$, $j = 1, \ldots, q$, being the canonical variates for any $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{p+q}$. The influence functions for the weighting vectors are given by

$$\text{IF}\left((\mathbf{x}, \mathbf{y}), \boldsymbol{\alpha}_R, H_0 \right) = \sum_{k=2}^{p} \frac{1}{\rho_1^2 - \rho_k^2} \Big\{ \text{IF}_1((u_1, v_1), R, F_\rho) \rho_1 u_k$$
$$+ \text{IF}_2((u_1, v_1), R, F_\rho) \rho_k v_k \Big\}$$
$$\times \left(I - \frac{\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_1^t}{\|\boldsymbol{\alpha}_1\| \|\boldsymbol{\alpha}_1\|} \right) \frac{\boldsymbol{\alpha}_k}{\|\boldsymbol{\alpha}_1\| \kappa_R'(\rho_1)}, \qquad (16)$$

$$\text{IF}\left((\mathbf{x}, \mathbf{y}), \boldsymbol{\beta}_R, H_0 \right) = \sum_{k=2}^{q} \frac{1}{\rho_1^2 - \rho_k^2} \Big\{ \text{IF}_1((u_1, v_1), R, F_\rho) \rho_k u_k$$
$$+ \text{IF}_2((u_1, v_1), R, F_\rho) \rho_1 v_k \Big\}$$
$$\times \left(I - \frac{\boldsymbol{\beta}_1 \boldsymbol{\beta}_1^t}{\|\boldsymbol{\beta}_1\| \|\boldsymbol{\beta}_1\|} \right) \frac{\boldsymbol{\beta}_k}{\|\boldsymbol{\beta}_1\| \kappa_R'(\rho_1)}, \qquad (17)$$

where IF_1 and IF_2 denote the partial derivatives with respect to the first and second component of $\text{IF}((u_1, v_1), R, F_\rho)$, respectively.

Theorem 1 shows that the IF of the projection index R determines the shape of the IF for the multivariate association measure ρ_R and the weighting vectors. While a bounded $\text{IF}((u_1, v_1), R, F_\rho)$ ensures a bounded IF for ρ_R, this is no longer true for the weighting vectors. Indeed, if u_k or v_k tend to infinity (for a $k \geq 2$), the IF for any of the weighting vectors goes beyond all bounds. Note that an unbounded IF means that if there is a small amount ε of contamination, the change in the value of the functional will be disproportionally large with respect to the level of contamination. It does not mean that the functional breaks down or explodes in the presence of small amounts of outliers. Furthermore, since the partial derivatives of $\text{IF}((u_1, v_1), R, F_\rho)$ appear in the influence functions for the weighting vectors, it is necessary to take a projection index R having a smooth influence function. In particular, discontinuities in $\text{IF}((u_1, v_1), R, F_\rho)$ yield unstable estimates of the weighting vectors.

The influence functions of the association measures R considered in this article are known (e.g., Croux and Dehon 2010). Using for R the Kendall and Spearman rank correlation, as well as the M-estimator, yields bounded influence, whereas the Pearson correlation results in unbounded influence (see the plots in the supplementary report Alfons, Croux, and Filzmoser 2016).

4. Asymptotic Variances

We confine ourselves to the case $p > 1$ and $q = 1$ with H_0 an elliptically symmetric distribution. Due to affine equivariance, set $\boldsymbol{\Sigma}_{XX} = I$, $\Sigma_{YY} = 1$, and $\boldsymbol{\Sigma}_{XY} = (\rho, 0, \ldots, 0)^t$ without loss of generality. In this case, we have

$$\boldsymbol{\alpha}_R = \underset{\|\boldsymbol{\alpha}\|=1}{\text{argmax}} \, R(\boldsymbol{\alpha}^t X, Y).$$

Theorem 1 then reduces to

$$\text{IF}\left((\mathbf{x}, y), \boldsymbol{\alpha}_R, H_0 \right) = \frac{1}{\rho_R \kappa_R'(\rho_R)} \text{IF}_1 \left((\boldsymbol{\alpha}_R^t \mathbf{x}, y), R, F_\rho \right)$$
$$\times \left(I - \boldsymbol{\alpha}_R \boldsymbol{\alpha}_R^t \right) \mathbf{x}. \qquad (18)$$

It follows that the asymptotic variance of the sample version of $\boldsymbol{\alpha}_R$ is given by

$$\text{ASV}(\boldsymbol{\alpha}_R, H_0) = E \left[\text{IF} \left((X, Y), \boldsymbol{\alpha}_R, H_0 \right) \text{IF} \left((X, Y), \boldsymbol{\alpha}_R, H_0 \right)^t \right]$$
$$= \left(I - \boldsymbol{\alpha}_R \boldsymbol{\alpha}_R^t \right) E \left[\text{IF}_1^2 \left((\boldsymbol{\alpha}_R^t X, Y), R, F_\rho \right) X X^t \right]$$
$$\times \left(I - \boldsymbol{\alpha}_R \boldsymbol{\alpha}_R^t \right) \frac{1}{\rho_R^2 \kappa_R'(\rho_R)^2}, \qquad (19)$$

using the relation between the influence function and the asymptotic variance given in Hampel et al. (1986).

Building further upon an earlier version of this manuscript, correctly acknowledged in their article, Jin and Cui (2010) proved asymptotic normality of the joint distribution of the estimators of the weighting vectors and the maximum association under a set of regularity conditions that encompass elliptically symmetric distributions. They did not derive explicit expressions of the asymptotic variances of the estimators, as we do below.

Since $\boldsymbol{\Sigma}_{XX} = \boldsymbol{I}$, we have $\boldsymbol{\alpha}_R = \boldsymbol{\Sigma}_{XY}/\|\boldsymbol{\Sigma}_{XY}\| = (1, 0, \ldots, 0)^t$. All elements of the matrix ASV$(\boldsymbol{\alpha}_R, H_0)$ are then equal to zero, except the diagonal elements

$$\text{ASV}(\boldsymbol{\alpha}_R, H_0)_{jj} = \frac{1}{\rho_R^2 \kappa_R'(\rho_R)^2} E\left[X_j^2 \text{IF}_1^2\left((X_1, Y), R, F_\rho\right)\right]$$

(20)

for $j > 1$. For $j = 1$, the speed of convergence of the estimator is faster than $\sqrt{n}$, corresponding to an asymptotic variance of 0.

For $H_0 = N(\boldsymbol{0}, \boldsymbol{\Sigma})$, the following expressions can be obtained for the asymptotic variance of the weighting vector $\boldsymbol{\alpha}_R$. The proofs follow fairly standard (but tedious) calculus and are omitted.

Proposition 1. Consider the normal model $H_0 = N(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}_{XX} = \boldsymbol{I}$, $\boldsymbol{\Sigma}_{YY} = 1$, and $\boldsymbol{\Sigma}_{XY} = (\rho, 0, \ldots, 0)^t$ without loss of generality, and take $j > 1$. The asymptotic variance of $\boldsymbol{\alpha}_R$ for the Spearman correlation $R = R_S$ is given by

$$\text{ASV}(\boldsymbol{\alpha}_R, H_0)_{jj} = \frac{1 - \rho^2/4}{\rho^2} 16\pi^2 E\left[\varphi^2(X_1) \text{var}\left(\Phi\left(\rho X_1\right.\right.\right.$$
$$\left.\left.\left. + \sqrt{1-\rho^2}Z\right) \Big| X_1\right)\right],$$

(21)

where $Z \sim N(0, 1)$ is independent of X_1, and φ is the probability density function for the standard normal distribution. For the Kendall correlation $R = R_K$, we have

$$\text{ASV}(\boldsymbol{\alpha}_R, H_0)_{jj} = \frac{1 - \rho^2}{\rho^2} \frac{2\pi}{3\sqrt{3}}.$$

(22)

For R derived from a scatter matrix $\boldsymbol{C}$, we have

$$\text{ASV}(\boldsymbol{\alpha}_R, H_0)_{jj} = \frac{1 - \rho^2}{2\rho^2} \left\{ E\left[\gamma^2(d)d^2\right] + \frac{1}{2}E\left[\gamma(d)\gamma'(d)d^3\right] \right.$$
$$\left. + \frac{1}{8}E\left[\gamma'(d)^2 d^4\right] \right\},$$

(23)

with γ such that $\text{IF}((u, v), R_C, F_\rho) = \gamma(d)\text{IF}((u, v), R_P, F_\rho)$ and $d^2 = d^2(u, v) = (u, v)\boldsymbol{\Sigma}_\rho^{-1}(u, v)^t$.

The asymptotic relative efficiency of a maximum association measure with projection index R compared to the Pearson-based approach is now defined as

$$\text{ARE}(\boldsymbol{\alpha}_R, H_0) = \frac{\text{ASV}(\boldsymbol{\alpha}_{R_P}, H_0)_{jj}}{\text{ASV}(\boldsymbol{\alpha}_R, H_0)_{jj}}.$$

(24)

For the Pearson correlation, $\gamma(d) \equiv 1$ and the asymptotic variance in (23) reduces to $\text{ASV}(\boldsymbol{\alpha}_{R_P}, H_0)_{jj} = (1 - \rho^2)/\rho^2$.

Figure 1 (left) displays the asymptotic efficiencies for different choices of R at the multivariate normal distribution with varying maximum correlation ρ. The rank correlation measures and the M-association estimator yield good asymptotic efficiencies of about 80%. The asymptotic efficiencies of the Kendall correlation and the M-estimator are constant, with the former being slightly higher than the latter. For the Spearman correlation, the asymptotic efficiency decreases with increasing ρ. While it is similarly high as for the Kendall correlation initially, it eventually drops below the asymptotic efficiency of the M-estimator.

The good efficiency properties of the Kendall correlation carry over to heavy tailed distributions. Using Equations (20) and (24), we derived asymptotic relative efficiencies at the multivariate t-distribution of the weighting vectors based on the Kendall correlation. Calculation of the ASV is still possible in this case, using the fact that the function κ_K is the same for
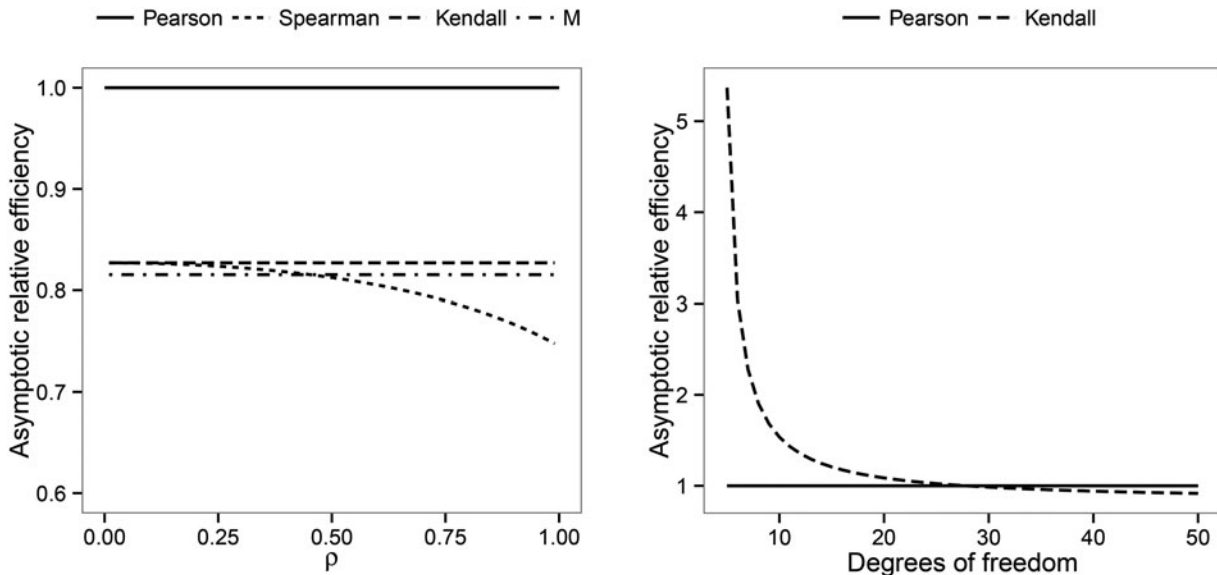
Figure 1. Asymptotic efficiencies of the weighting vector $\boldsymbol{\alpha}_R$ for different choices of the projection index R: at the multivariate normal distribution, as function of the true maximum correlation ρ (left); at the multivariate t-distribution, as function of the degrees of freedom of this distribution (right).

any elliptically symmetric distribution (see Hult and Lindskog 2002). It turns out that the relative efficiencies, as in the normal case, are neither depending on the dimension p nor on the value ρ. Figure 1 (right) plots the asymptotic relative efficiencies as a function of the degrees of freedom of the multivariate t-distribution. We see that the Kendall correlation outperforms the Pearson correlation for smaller degrees of freedom, while the latter only becomes more efficient for degrees of freedom larger than 27.

5. Alternate Grid Algorithm

Let $x_1, \ldots, x_n \in \mathbb{R}^p$ and $y_1, \ldots, y_n \in \mathbb{R}^q$ be the observed data with $p \geq q$. To simplify notation, let $\hat{R}(a, b) := \hat{R}((a^t x_1, b^t y_1), \ldots, (a^t x_n, b^t y_n))$ denote the sample version of the association measure R with projection directions a and b. The aim is to maximize the projection index

$$f(a, b) = |\hat{R}(a, b)|. \tag{25}$$

If $p = 2$ and b is fixed, the problem of finding the maximum association reduces to maximizing the function $\theta \to f((\cos(\theta), \sin(\theta))^t, b)$ over the interval $[-\pi/2, \pi/2]$. This can easily be solved with a grid search, that is, by dividing the interval in $(N_g - 1)$ equal parts and evaluating the function at the grid points $(-1/2 + j/N_g)\pi$ for $j = 0, \ldots, N_g - 1$. The advantages of the grid search are that differentiability of the function is not required and that a global maximum can be distinguished from a local one. With a large enough number of grid points N_g, the found maximum will be close enough to the real solution.

If $p > 2$ and b being kept fixed, a sequence of optimizations in two-dimensional subspaces of the X space is performed. Given a weighting vector a, its kth component a_k can be updated by a grid search in the subspace spanned by a and e_k^p, the kth canonical basis vector of $\mathbb{R}^p$. For a grid of values for the angle θ as described above, the candidate directions are then given by

$$t_k^p(a, \theta) = \frac{\cos(\theta)a + \sin(\theta)e_k^p}{\sqrt{1 + \sin(2\theta)a_k}}, \tag{26}$$

where the denominator ensures unit norm of the candidate directions. This is done for $k = 1, \ldots, p$.

Hence, the idea of the algorithm is to perform series of grid searches, alternating between searching for a with a given b (as described above) and searching for b with a given a. For details, see Algorithm 1. In the first cycle, the whole plane is scanned in each grid search. In the subsequent cycles, the search is limited to a more and more narrow interval of angles while keeping the number of grid points N_g constant. Moreover, the alternate series of grid searches in each cycle is stopped if the improvement is smaller than a certain threshold. Typically only very few iterations are necessary in each cycle, thus keeping computation time low.

If $p > 1$ and $q = 1$, the algorithm reduces to the algorithm of Croux, Filzmoser, and Oliveira (2007) with projection index

$$f_y(a) := |\hat{R}((a^t x_1, y_1), \ldots, (a^t x_n, y_n))|. \tag{27}$$

The algorithm can be extended to higher-order (robust) canonical correlations by transforming the data into suitable subspaces along the lines of Branco et al. (2005). Our implementation

---

Algorithm 1 Alternate grid algorithm

Start with $\hat{\alpha} = e_1^p$ and $\hat{\beta} = e_1^q$.
For $i = 1, \ldots, N_c$ **do** the following cycle.

    For $j = 1, \ldots, N_s$ **do** the following alternate series of grid searches.

    (a) Perform the following series of grid searches in the X space keeping $\hat{\beta}$ fixed.
        For $k = 1, \ldots, p$
        (i) Maximize the objective function in the plane spanned by $\hat{\alpha}$ and e_k^p by a grid search of the function $\theta \to |\hat{R}(t_k^p(\hat{\alpha}, \theta), \hat{\beta})|$, where the angle θ is restricted to the interval $[-\pi/(2^i), \pi/(2^i))$. Let θ_0 denote the angle where the maximum is attained over all grid points.
        (ii) Update $\hat{\alpha} \leftarrow t_k^p(\hat{\alpha}, \theta_0)$.
        End for

    (b) Perform the following series of grid searches in the Y space keeping $\hat{\alpha}$ fixed.
        For $l = 1, \ldots, q$
        (i) Maximize the objective function in the plane spanned by $\hat{\beta}$ and e_l^q by a grid search of the function $\theta \to |\hat{R}(\hat{\alpha}, t_l^q(\hat{\beta}, \theta))|$, where the angle θ is restricted to the interval $[-\pi/(2^i), \pi/(2^i))$. Let θ_0 denote the angle where the maximum is attained over all grid points.
        (ii) Update $\hat{\beta} \leftarrow t_l^q(\hat{\beta}, \theta_0)$.
        End for
    End for
End for
**If** $R(\hat{\alpha}, \hat{\beta}) < 0$ **then** set $\hat{\beta} \leftarrow -\hat{\beta}$.

---

for the statistical computing environment R is available in package *ccaPP*, which can be downloaded from *http://CRAN.R-project.org/package=ccaPP*.

6. Example

We present an application for the maximum rank correlation estimators in the regression setting. We use the *movies* data, which are available from *http://had.co.nz/data/movies/* and contain movie information from the internet movie database (IMDb; *http://imdb.com*). The response variable is given by the average IMDb user rating, and we use the following $p = 11$ predictors: year of release, total budget in U.S. dollars, length in minutes, number of IMDb users who gave a rating, as well as a set of binary variables representing if the movie belongs to the genre action, animation, comedy, drama, documentary, romance, or short film. We limit the analysis to movies with known budget, leaving $n = 5215$ observations in the dataset.

We compare the four estimators based on maximum association measures with least-square (LS) regression and the robust MM-regression estimator tuned for 85% efficiency (e.g., Maronna, Martin, and Yohai 2006). Theoretically, LS regression gives the same results as the maximum Pearson correlation and is included to verify our numeric algorithm. We estimate the prediction performance by randomly splitting the data into
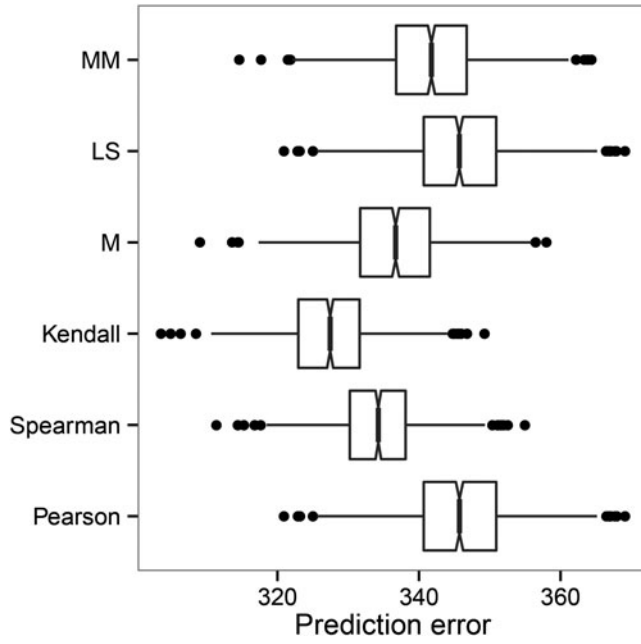
Figure 2. Distribution of the prediction loss for the ranking according to the average IMDB user rating in the *movies* data, estimated via 1000 random splits into training and test data.

training and test sets. We repeat this process $L = 1000$ times and use $m = \lfloor n/3 \rfloor$ observations as test data in each replication. As prediction loss, we use Spearman's footrule distance

$$d = \frac{1}{m} \sum_{i=1}^{m} |r_i - \hat{r}_i|, \quad (28)$$

where $r = (r_1, \ldots, r_m)^t$ are the ranks of the movies according to the average user rating and $\hat{r} = (\hat{r}_1, \ldots, \hat{r}_m)^t$ the predicted ranks.

Figure 2 contains boxplots of the distribution of the prediction error for the considered methods. The maximum rank correlation estimators perform better than their competitors, with the Kendall correlation yielding the lowest prediction error. This hints at a nonlinear relationship between the response and the predictors, which we confirmed by inspecting plots of the response against the fitted values (not shown). The M-association estimator still results in quite good prediction performance, followed by the MM-regression estimator, the Pearson correlation, and LS. The latter two thereby give practically identical results, being in line with the theory. All prediction methods that we compare are linear in the predictor variables. Nonlinear methods as random forests might give even better prediction results.

7. Simulation Experiments

The results presented here are a representative selection from an extensive simulation study. We focus on the case of $p > 1$ and $q = 1$. The sample counterparts of the maximum association measures and the weighting vectors are denoted by $\hat{\rho}_R$ and $\hat{\alpha}_R$. We compare the mean squared error (MSE) of the estimators $\kappa_R^{-1}(\hat{\rho}_R)$ and $\hat{\alpha}_R$ of ρ and α, respectively, for different projection indices R.

Additional experiments with different data configurations, sample sizes and dimensions are discussed in a supplementary

report (Alfons, Croux, and Filzmoser 2016). In particular, we investigate the case of $q > 1$, yielding similar conclusions on the behavior of the estimators. The other additional experiments include an empirical bias study and a comparison of the power of permutation tests for detecting nonlinear relationships.

In each of the $L = 1000$ simulation runs, observations $(x_1, y_1), \ldots, (x_n, y_n)$ with $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$ are generated from the model distribution $N(0, \Sigma)$. We take

$$\Sigma = \begin{pmatrix} I & \Sigma_{XY} \\ \Sigma_{YX} & 1 \end{pmatrix} \quad (29)$$

with $\Sigma_{XY} = \Sigma_{YX}{}^t = (\rho, 0, \ldots, 0)^t$. Thus, the true weighting vector is given by $\alpha = (1, 0, \ldots, 0)^t$, yielding the true maximum correlation ρ. We selected the sample size as $n = 100$ and the dimension of the x_i as $p = 5$.

For the estimators of the maximum correlation ρ, the mean squared error (MSE) is computed as

$$\mathrm{MSE}_R(\rho) = \frac{1}{L} \sum_{l=1}^{L} \left(\phi(\kappa_R^{-1}(\hat{\rho}_R^l)) - \phi(\rho) \right)^2, \quad (30)$$

where $\phi(\rho) = \tanh^{-1}(\rho)$ is the Fisher transformation of ρ, which is often applied to render the finite sample distribution of correlation coefficients more toward normality, and $\hat{\rho}_R^l$ is the estimated maximum association from the lth simulation run. For the weighting vectors, the MSE is computed as

$$\mathrm{MSE}_R(\alpha) = \frac{1}{L} \sum_{l=1}^{L} \left\{ \cos^{-1} \left(\frac{|\alpha^t \hat{\alpha}_R^l|}{||\alpha|| \, ||\hat{\alpha}_R^l||} \right) \right\}^2, \quad (31)$$

where $\hat{\alpha}_R^l$ is the estimated weighting vector from the lth simulation run. The measure (31) is the average of the squared angles between the vectors $\hat{\alpha}_R^l$ and α, making the MSE invariant to the choice of the normalization constraint for the weighting vectors.

For $q = 1$, an alternative procedure is the rank transformation least-square estimator of Garnham and Prendergast (2013). There the response variable Y is replaced by the corresponding ranks, and the weighting vector is then obtained by least-square regression of those ranks on the variables in X. In the supplementary report (Alfons, Croux, and Filzmoser 2016), we add this estimator to the simulation study and show that this approach suffers from leverage points, that is, outliers in the space of the X-variables. In absence of leverage points, it performs well.

7.1 Effect of a Nonlinear Monotonic Transformation

The true regression model is

$$y_i = a^t x_i + e_i, \quad (32)$$

where the regression coefficients a are related to the weighting vector α through rescaling, that is, $\alpha = a/||a||$. We transform the response variable to $\tilde{y}_i = \exp(y_i)$. Then we compute the different estimators once for the data (x_i, y_i) that follow the model and once for the data $(x_i, \tilde{y}_i)$ that do not follow the model. We thereby compare the results for varying maximum correlation $\rho \in [0, 1)$.

(a)

(b)

(c)

(d)
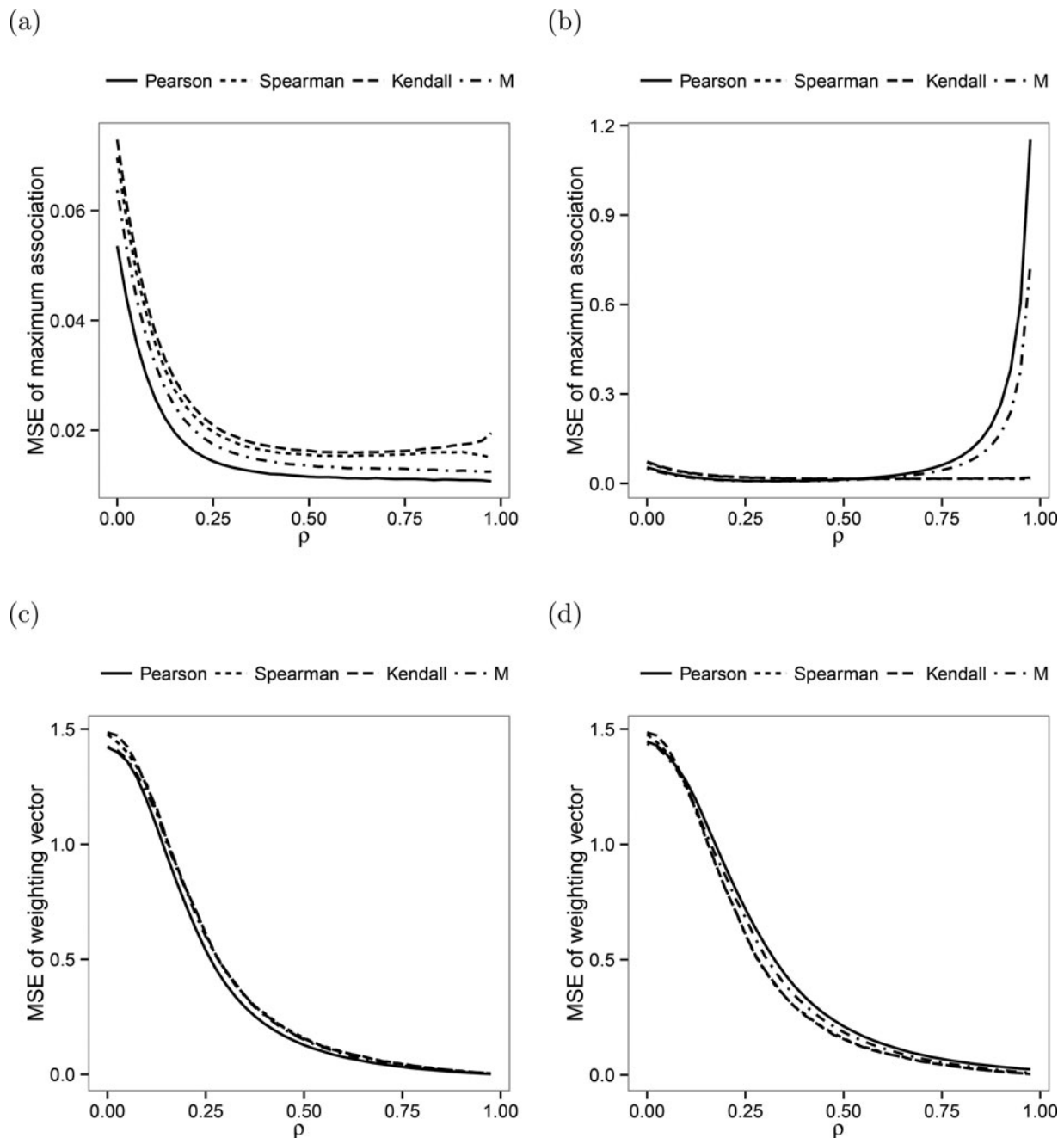
Figure 3. MSEs of the maximum association measures (top) and the weighting vectors (bottom) for the correctly specified model (left) and for the misspecified model (right). In (b) and (d), the lines for the Spearman and Kendall correlation almost coincide. In (c), the lines for all methods except the Pearson correlation almost coincide.

Figure 3(a) and 3(b) shows the MSEs of the maximum association measures for the correctly specified model and the misspecified model, respectively. For the correctly specified model, the methods behave as expected: the Pearson correlation performs best, followed by the M-association estimator and the Spearman and Kendall correlation. However, the results are very different for the misspecified model. As long as the true maximum correlation ρ is small to moderate, the nonlinearity of the relationship is masked by the noise and all methods yield similar results. But as ρ increases, the MSEs of the Pearson correlation and (to a lesser extent) the M-association estimator drift away from 0. The rank correlation measures, on the other hand, are unaffected by the monotonic transformation.

In Figure 3(c) and 3(d), the MSEs of the weighting vectors are displayed for the correctly specified model and the misspecified model, respectively. The nonlinear transformation does not have a very large effect on the MSE of the estimators. Still, while the Pearson correlation yields the best performance under the correctly specified model, the rank correlation measures perform better in the case of the misspecified model. Since the weighting vectors can be seen as normalized regression coefficients, we also computed two regression estimators: least squares (LS) and the robust MM-regression estimator tuned for 85% efficiency (e.g., Maronna, Martin, and Yohai 2006). We did not include those estimators in the plots to keep them from being overloaded. As expected from theory, LS gave the same results as the Pearson correlation, which underlines that our numerical algorithm
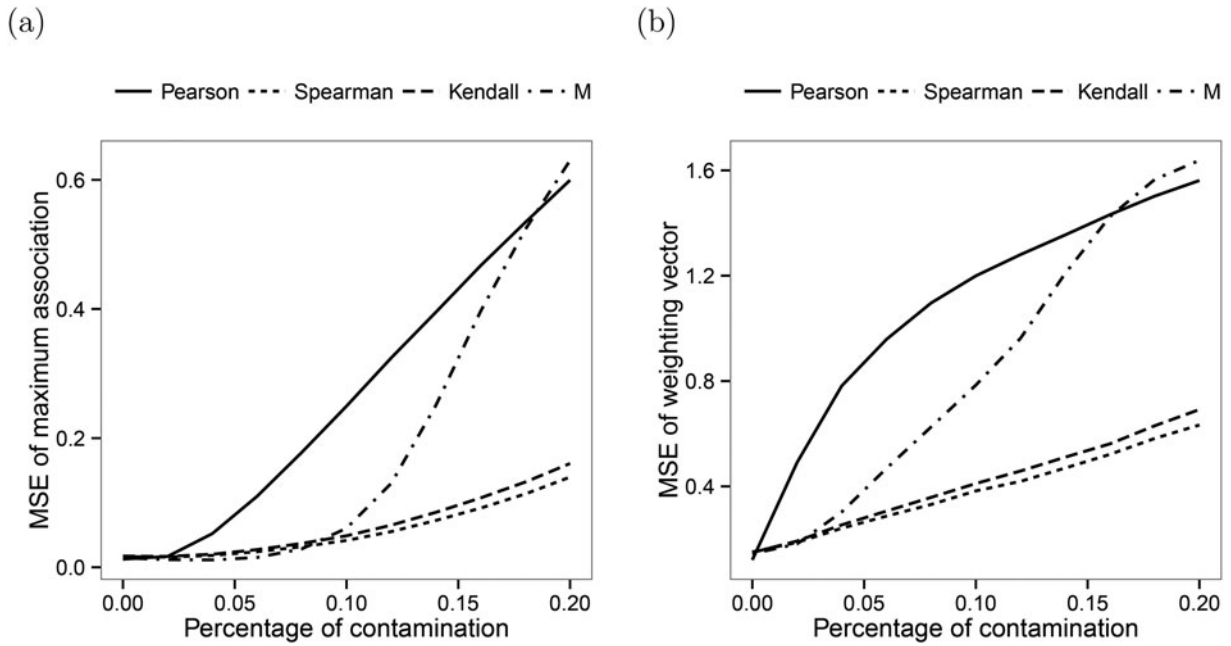
(a)

(b)

Figure 4. MSEs of the maximum association measures (left) and the weighting vectors (right) for varying percentage of contamination ε.

works very well. In addition, the MM-estimator resulted in a slightly larger MSE than the Pearson correlation, yielding the largest MSE among all methods for the misspecified model.

7.2. Effect of Contamination

In the second simulation experiment, we investigate the effect of contamination on the estimators. We fix the true maximum correlation at $\rho = 0.5$. A fraction ε of the data is then replaced by outliers $(\mathbf{x}_i^*, y_i^*)$ coming from an $N(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ distribution, where $\boldsymbol{\mu}^* = (0, -5, 0, 0, 5)^t$ and $\boldsymbol{\Sigma}^*$ is obtained from $\boldsymbol{\Sigma}$ in (29) by replacing $\boldsymbol{\Sigma}_{XY} = \boldsymbol{\Sigma}_{YX}{}^t$ with $\boldsymbol{\Sigma}_{XY}^* = \boldsymbol{\Sigma}_{YX}^*{}^t = (0, -1, 0, 0, 0)^t$. We vary the contamination level ε from 0% to 20% in steps of 2%.

Figure 4(a) shows the resulting MSEs of the maximum association measures. Clearly, the outliers have a strong influence on the Pearson correlation. For the M-estimator, the MSE remains low for small contamination levels but increases dramatically afterward. The rank correlation measures, on the other hand, are much less influenced by the outliers.

The MSEs for the corresponding weighting vectors are depicted in Figure 4(b). Again, the Pearson correlation is strongly influenced by the outliers. Concerning the M-association estimator, the outliers have a more immediate effect on the weighting vector than they have on the maximum association. As before, the maximum rank correlation estimators are more robust against the outliers. Furthermore, the results for the Spearman and Kendall correlation are very similar, confirming the theoretical results for the asymptotic variances.

8. Conclusions

This article studies a measure of association between two multivariate random vectors based on projection pursuit. Its definition is intuitively appealing: it is the highest possible association R between any two projections $\boldsymbol{\alpha}^t X$ and $\boldsymbol{\beta}^t Y$ constructed

from the variables X and Y. Using different projection indices R, different measures of association are obtained. For studying the robustness of the maximum association measures, we carry out influence calculations, showing that the projection index R should have a bounded and smooth influence function. In addition, we emphasize the important special case of univariate Y, in which case the problem of finding the projection yielding maximum association can be interpreted as a regression problem. Using maximum rank correlation estimators then has the advantage that they are invariant against monotonic transformations of the response. To study the robustness of the maximum association estimators against model misspecification and contamination, we present a simulation study and real data applications. Both the theoretical and the numerical results favor maximum rank correlation estimators, as they combine good robustness properties with good efficiency.

The derived influence functions and asymptotic variances can be used to obtain estimates of the standard errors of the coefficients of the weighting vectors. We obtain tractable expressions of the influence functions and asymptotic variances assuming an elliptically symmetric distribution. When the normality assumption does not hold, we suggest to use a bootstrap procedure. Note that Sherman (1993) developed standard errors for the Kendall-based weighting vector in the regression case $q = 1$, using numerical derivation.

We present an approximative algorithm to compute the proposed estimators via projection pursuit. Note that the rank correlation measures are fast to compute. Computation time of the Spearman correlation is dominated by computing the ranks of the two variables, which requires $O(n \log n)$ time. While the definition of the Kendall rank correlation would suggest a computation time of $O(n^2)$, Knight (1966) introduced an $O(n \log n)$ algorithm. The fast algorithm for the Kendall rank correlation is implemented in the R package *ccaPP*, together with the maximum association estimators.

Finally, as a topic for future research, other robustness measures such as breakdown point or maximum asymptotic bias could be derived as a complement to the influence function, which only considers infinitesimal contamination values.

Supplementary Materials

A supplementary report contains further technical details, a proof of Theorem 1, as well as an extensive collection of numerical results.

Acknowledgment

References

Alfons, A., Croux, C., and Filzmoser, P. (2016), "Robust Maximum Association Estimators: Technical Supplement," Supplementary Report. [437,439,442]

Branco, J. A., Croux, C., Filzmoser, P., and Oliveira, M. R. (2005), "Robust Canonical Correlations: A Comparative Study," *Computational Statistics*, 20, 203–229. [437,441]

Breiman, L., and Friedman, J. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association*, 80, 580–619. [437]

Croux, C., and Dehon, C. (2010), "Influence Functions of the Spearman and Kendall Correlation Measures," *Statistical Methods & Applications*, 19, 497–515. [439]

Croux, C., Filzmoser, P., and Oliveira, M. (2007), "Algorithms for Projection-Pursuit Robust Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems*, 87, 218–225. [437,441]

Garnham, A. L., and Prendergast, L. A. (2013), "A Note on Least Squares Sensitivity in Single-index Model Estimation and the Benefits of Response Transformations," *Electronic Journal of Statistics*, 7, 1983–2004. [442]

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley. [439]

Han, A. K. (1987), "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 35, 303–316. [436]

Huber, P. J. (1981), *Robust Statistics*, New York: Wiley. [436,438]

Hult, H., and Lindskog, F. (2002), "Multivariate Extremes, Aggregation and Dependence in Elliptical Distributions," *Advances in Applied Probability*, 34, 587–608. [441]

Jin, J., and Cui, H. J. (2010), "Asymptotic Distributions in the Projection Pursuit Based Canonical Correlation Analysis," *Science China–Mathematics*, 53, 485–498. [440]

Johnson, R. A., and Wichern, D. W. (2002), *Applied Multivariate Statistical Analysis* (5th ed.), Upper Saddle River, NJ: Prentice Hall. [436]

Knight, W. R. (1966), "A Computer Method for Calculating Kendall's Tau With Ungrouped Data," *Journal of the American Statistical Association*, 61, 436–439. [444]

Kudraszow, N. L., and Maronna, R. A. (2011), "Robust Canonical Correlation Analysis: A Predictive Approach," available at *http://www.mate.unlp.edu.ar/publicaciones/06062011164242.pdf* [437]

López Blázquez, F., and Salamanca Miño, B. (2014), "Maximal Correlation in a Non-Diagonal Case," *Journal of Multivariate Analysis*, 131, 265–278. [437]

Maronna, R., Martin, D., and Yohai, V. (2006), *Robust Statistics: Theory and Methods*, Chichester, UK: Wiley. [441,443]

Maronna, R. A. (1976), "Robust M-estimators of Multivariate Location and Scatter," *The Annals of Statistics*, 4, 51–67. [438]

Papadatos, N., and Xifara, T. (2013), "A Simple Method for Obtaining the Maximal Correlation Coefficient and Related Characterizations," *Journal of Multivariate Analysis*, 118, 102–114. [437]

Ramsay, J. O., Ten Berge, J., and Styan, G. P. H. (1984), "Matrix Correlation," *Psychometrika*, 49, 403–423. [436]

Sherman, R. P. (1993), "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123–137. [436,444]

Smilde, A. K., Kiers, H. A. L., Bijlsma, S., Rubingh, C. M., and van Erk, M. J. (2009), "Matrix Correlations for High-Dimensional Data: The Modified RV-Coefficient," *Bioinformatics*, 25, 401–405. [436]

Taskinen, S., Croux, C., Kankainen, A., Ollila, E., and Oja, H. (2006), "Influence Functions and Efficiencies of the Canonical Correlation and Vector Estimates Based on Scatter and Shape Matrices," *Journal of Multivariate Analysis*, 97, 359–384. [437]

Yu, Y. (2008), "On the Maximal Correlation Coefficient," *Statistics and Probability Letters*, 78, 1072–1075. [437]