



# Canonical correlation analysis for elliptical copulas

Benjamin W. Langworthy<sup>a,\*</sup>, Rebecca L. Stephens<sup>b</sup>, John H. Gilmore<sup>b</sup>,  
Jason P. Fine<sup>a</sup>

<sup>a</sup> Department of Biostatistics, University of North Carolina - Chapel Hill,  
135 Dauer Drive, 3101 McGavran-Greenberg Hall, Chapel Hill, NC 27599, United States of America

<sup>b</sup> Department of Psychiatry, University of North Carolina - Chapel Hill, 101 Manning Dr 1, Chapel Hill, NC 27514,  
United States of America

## ARTICLE INFO

### Article history:

Received 3 May 2020

Received in revised form 10 December 2020

Accepted 10 December 2020

Available online 23 December 2020

### AMS 2010 subject classifications:

primary 62H25

secondary 62F40

### Keywords:

Canonical correlation analysis

Kendall's tau

Resampling methods

Robust methods

Transelliptical distributions

## ABSTRACT

Canonical correlation analysis (CCA) is a common method used to estimate the associations between two different sets of variables by maximizing the Pearson correlation between linear combinations of the two sets of variables. We propose a version of CCA for transelliptical distributions with an elliptical copula using pairwise Kendall's tau to estimate a latent scatter matrix. Because Kendall's tau relies only on the ranks of the data this method does not make any assumptions about the marginal distributions of the variables, and is valid when moments do not exist. We establish consistency and asymptotic normality for canonical directions and correlations estimated using Kendall's tau. Simulations indicate that this estimator outperforms standard CCA for data generated from heavy tailed elliptical distributions. Our method also identifies more meaningful relationships when the marginal distributions are skewed. We also propose a method for testing for non-zero canonical correlations using bootstrap methods. This testing procedure does not require any assumptions on the joint distribution of the variables and works for all elliptical copulas. This is in contrast to permutation tests which are only valid when data are generated from a distribution with a Gaussian copula. This method's practical utility is shown in an analysis of the association between radial diffusivity in white matter tracts and cognitive tests scores for six-year-old children from the Early Brain Development Study at UNC-Chapel Hill. An R package implementing this method is available at [github.com/blangworthy/transCCA](https://github.com/blangworthy/transCCA).

© 2020 Elsevier Inc. All rights reserved.

## 1. Introduction

Canonical correlation analysis (CCA), first introduced by Hotelling [21], is a useful dimension reduction technique for exploring the relationship between two sets of variables. CCA finds the linear combinations of the two sets of variables that have maximal Pearson correlation. After the first direction, further directions are defined as the linear combinations that are maximally correlated subject to the constraint that they are uncorrelated with all previous directions. A small number of directions may be used to summarize the relationship between the two sets of variables.

In Section 4 we present an example where CCA is useful in understanding the relationship between the structure of white matter brain tracts and executive function in six-year-old children. Many of the variables show excess skewness or kurtosis relative to the normal distribution. This suggests transformations may be needed for CCA using Pearson's

\* Corresponding author.

E-mail address: [blangworthy@hsph.harvard.edu](mailto:blangworthy@hsph.harvard.edu) (B.W. Langworthy).

correlation to fully capture the association between the two sets of variables. However it is not clear how to optimally transform the data, especially for heavy tailed distributions where transforming may weaken linear associations. In such settings standard CCA may be problematic, and alternative approaches are valuable.

In the finite dimensional setting when all second moments exist, CCA is valid based on an eigendecomposition involving the sample covariance matrix. In settings where the empirical covariance estimator is either inconsistent or inefficient, including when second moments do not exist or when there are outliers contaminating the observed data, the CCA estimates based on the empirical covariance matrix will also be either inconsistent or inefficient. There is a rich literature on robust estimators of the covariance matrix that are insensitive to outliers and heavy tailed distributions, and may improve the performance of standard CCA based on Pearson correlation. Examples of these are the minimum covariance determinant (MCD) [40], the S-estimator [31], and Tyler's M-estimator [46]. There have been studies examining the performance of CCA using robust estimators of the covariance matrix or by maximizing other robust correlation measures [4,8,44,47]. Many of these robust methods emphasize eigendecompositions employing robust estimates of the covariance or Pearson correlation matrix, which do not exist in the absence of finite moments. Further assumptions are needed to interpret robust CCA in these settings.

We explicitly define a version of CCA for distributions with elliptical copulas that does not require the existence of moments using properties of Kendall's tau for elliptical distributions. For elliptical distributions there is a known monotone relationship between Pearson's correlation and Kendall's tau rank correlation. We utilize this relationship to define CCA using Kendall's tau instead of Pearson correlation such that it is well defined when moments do not exist and has the same canonical directions and correlations as standard CCA for elliptical distributions when moments do exist. Perhaps most importantly this definition of CCA does not make any assumptions about the marginal distributions of the variables, so it can be easily extended to a family of distributions known as transelliptical distributions. The transelliptical family consists of all multivariate distributions which can be transformed into an elliptical distribution using monotone marginal transformations, or equivalently all multivariate distributions with a copula from an elliptical distribution [2,11,12,27,30]. Standard CCA is inadequate to describe the relationship between two sets of variables which are transelliptically distributed and have potentially non-linear associations. CCA using Kendall's tau identifies the linear relationships in the elliptical distribution which characterizes the transelliptical distribution. This is desirable because within elliptical distributions linear relationships describe meaningful association between the variables. We show that CCA for transelliptical distributions can be estimated without transforming the variables to an elliptical distribution, by estimating the scatter matrix based on transformations of Kendall's tau for all pairs of variables [30]. We establish that the resulting estimates for CCA directions and non-zero correlations are consistent and asymptotically normal. This result is more general than previous results which require affine equivariant estimators of the scatter matrix for data generated from elliptical distributions [5,44]. Interestingly, the estimate based on transformations of Kendall's tau for all pairs of variables is not affine equivariant. Simulations indicate that these results can be used to construct confidence intervals that perform similar to bootstrap confidence intervals with close to the desired coverage for the first canonical directions. Confidence intervals for higher order canonical directions and correlations do not perform as well whether using bootstrap or asymptotic results to construct the confidence intervals. This highlights the difficulty in accounting for variability in the estimates due to added constraints for finite samples.

We also develop a testing procedure to identify non-zero canonical correlations using bootstrap bias and standard error estimates. This is necessary because although the asymptotic results for non-zero canonical correlations can be used to construct confidence intervals, asymptotic results for zero canonical correlations are not as straightforward. However based on previous results [6] it can be expected that the zero canonical correlations will converge at rate  $n$  rather than  $\sqrt{n}$ . Therefore by inverting a normal bootstrap confidence interval we derive a test that is consistent and conservative for large sample sizes. This testing procedure can be used for CCA estimated using Kendall's tau or standard methods. This testing procedure is necessary because previously derived asymptotic tests assume the data are generated from a multivariate normal distribution [37,38,48]. Even permutation based tests assume that zero correlation implies independence, which is not true for non-Gaussian elliptical copulas. In non-Gaussian elliptical copulas the canonical directions may not be independent even when they are not informative of any associations between the two sets of variables, and therefore permutation tests which test for independence are not useful in determining which canonical directions capture meaningful associations between the two sets of variables. Our bootstrap based testing procedure makes minimal assumptions, and can even be useful even when data are not generated from a distribution with an elliptical copula.

The rest of the paper is structured as follows. Section 2 overviews the theoretical framework for rank estimation of CCA in the elliptical and transelliptical distributions and provides theoretical results for consistency and asymptotic normality of the estimates. Section 3 reports the results of simulation studies under elliptical and transelliptical distributions. Section 4 provides an analysis of associations between white matter structure and executive function in six-year-old children. Section 5 overviews the paper and concludes with remarks.

## 2. Rank correlation methodology

Assume  $X$  is a  $p \times 1$  dimensional random vector and  $Y$  is a  $q \times 1$  dimensional random vector. The first canonical directions for  $X$  and  $Y$  are the  $p \times 1$  vector,  $a_1$ , and the  $q \times 1$  vector,  $b_1$ , for which the correlation between  $U_1 = a_1^T X$  and  $V_1 = b_1^T Y$  is

maximized. The first canonical correlation is defined as the Pearson's correlation between  $U_1$  and  $V_1$ . In order to uniquely define  $a_1$  and  $b_1$ , it is necessary to add the constraints that  $\text{Var}(U_1) = \text{Var}(V_1) = 1$  [21]. After the first canonical direction and correlation, higher directions are a sequence of  $p \times 1$  vectors,  $a_j$ , and  $q \times 1$  vectors,  $b_j$ , such that  $U_j = a_j^\top X$  and  $V_j = b_j^\top Y$  are maximally correlated, subject to the constraints that  $\text{Cor}(U_j, U_{j'}) = \text{Cor}(U_j, V_{j'}) = \text{Cor}(V_j, U_{j'}) = \text{Cor}(V_j, V_{j'}) = 0$  for all  $j' < j$ , and  $\text{Var}(U_j) = \text{Var}(V_j) = 1$  for all  $j$ . This uniquely defines the canonical directions corresponding to a non-zero canonical correlation except for multiplication of both  $a_j$  and  $b_j$  by  $-1$ . There are at most  $\min(p, q)$  non-zero canonical correlations assuming both  $X$  and  $Y$  are full rank.

The canonical directions and correlations for  $X$  and  $Y$  can be shown to be the solutions to an eigendecomposition based on the covariance matrix between  $X$  and  $Y$ . Estimates of the canonical directions and correlations are commonly based on the same eigendecomposition involving the sample covariance matrix. If we define the joint covariance matrix of  $X$  and  $Y$  as

$$\text{Cov}\{(X^\top, Y^\top)^\top\} = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}$$

then the canonical correlations and directions may be derived from:

$$C = \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1/2}, D = \Sigma_{YY}^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1/2}.$$

The matrices  $C$  and  $D$  share the same first  $\min(p, q)$  eigenvalues, and the canonical correlations are the square root of these eigenvalues [21]. If  $v_{ci}$  is the  $i$ th eigenvector of  $C$ , then  $v_{ci} \Sigma_{XX}^{-1/2} = a_i$ , and if  $v_{di}$  is the  $i$ th eigenvector of  $D$ , then  $v_{di} \Sigma_{YY}^{-1/2} = b_i$  [21].

CCA can be made robust via robust estimation of the covariance matrix [8,44]. Many robust estimates of the covariance matrix are consistent under the elliptical family of distributions. The elliptical family of definitions are commonly defined through their characteristic functions in the following way [9],

**Definition 2.1** (Elliptical Distributions). A  $d \times 1$  random vector  $Z$  is considered to be elliptical if for some  $d \times 1$  vector  $\mu_Z$ , some  $d \times d$  positive semi-definite matrix  $\Sigma_Z$ , and a function  $\psi_Z[0, \infty) \rightarrow \mathbb{R}$ , the characteristic function,  $\Phi$ , satisfies  $\Phi_{Z-\mu_Z}(t) = \psi(t^\top \Sigma_Z t)$  for all  $d \times 1$  vectors  $t$ . In this case we would say that  $Z$  is a  $d \times 1$  dimensional elliptically distributed random variable, which we can note as  $Z \sim \mathcal{ED}_d(\mu_Z, \Sigma_Z, \psi_Z)$

We use  $\Sigma_Z$  in Definition 2.1 because in the elliptical distribution  $\Sigma_Z$  can be viewed as a generalization of the covariance matrix for  $Z$ . When second moments exist  $\Sigma_Z$  equals the covariance matrix up to a scaling factor, and  $\psi_Z$  can be chosen such that it is equal to the covariance matrix. We will refer to  $\Sigma_Z$  as the scatter matrix of  $Z$ , which exists even if second moments do not exist. The following proposition shows that for linear combinations of  $Z$ , the scatter matrix,  $\Sigma_Z$ , and location vector  $\mu_Z$ , are affine equivariant. To be precise, linear combinations of elliptical random variables are also elliptically distributed with a scatter matrix which is a quadratic form in  $\Sigma_Z$ .

**Proposition 2.1** (Linear Combinations of Elliptically Distributed Random Variables). Assume  $Z \sim \mathcal{ED}_d(\mu_Z, \Sigma_Z, \psi)$ . Define  $B$  to be a  $k \times d$  dimensional matrix of rank  $k \leq d$ . Then  $W = BZ$  is a  $k \times 1$  dimensional random vector where  $W \sim \mathcal{ED}_k(B\mu_Z, B\Sigma_Z B^\top, \psi)$

A proof of 2.1 can be found in Owen and Rabinovitch [36].

Letting  $Z = (X^\top, Y^\top)^\top$ , the scatter matrix of  $Z$  can be decomposed as

$$\Sigma_Z = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}.$$

Next we introduce the concept of the scale-invariant scatter matrix of  $Z$ ,  $P_Z$ , which will be equivalent to the correlation matrix of  $Z$  when second moments exist. Analogously to  $\Sigma_Z$ ,  $P_Z$  may be written as,

$$P_Z = \begin{pmatrix} P_{XX} & P_{XY} \\ P_{YX} & P_{YY} \end{pmatrix}.$$

The elements of  $P_Z$ ,  $\rho_{ij}$ , are related to the elements of  $\Sigma_Z$ ,  $\sigma_{ij}$ , through the following equality,  $\rho_{ij} = \sigma_{ij}/(\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}})$ . In general we will assume that  $\Sigma_Z$  and  $P_Z$  are positive-definite in order to guarantee existence of unique solutions for canonical correlation analysis.

A useful extension of elliptical distributions is the transelliptical family of distributions, whose definition is given below,

**Definition 2.2** (Transelliptical Distributions). A  $d \times 1$  dimensional random vector  $Z$  has a transelliptical distribution if there exists a positive-semidefinite matrix  $P_Z$  with all ones along the diagonal, a function  $\psi_Z : [0, \infty) \rightarrow \mathbb{R}$ , and a set of functions  $h_{Z1}, \dots, h_{Zd}$  where  $h_{Zi} : \mathbb{R} \rightarrow \mathbb{R}$  is a monotone increasing function for  $i = 1, \dots, d$ , such that  $\{h_{Z1}(Z_1), \dots, h_{Zd}(Z_d)\}^\top \sim \mathcal{ED}_d(0, P_Z, \psi_Z)$ . The random variable  $Z$  is a  $d \times 1$  dimensional transelliptically distributed random variable, denoted as  $Z \sim \mathcal{TE}_d(h_Z, 0, P_Z, \psi_Z)$ .

The elliptical distribution used in [Definition 2.2](#) is scale invariant and has a scatter matrix with all ones along the diagonal as well as centrality parameter zero in order to uniquely identify the transformations,  $h_Z$ . This definition was given by Liu et al. [30], but an equivalent definition is any multivariate distribution with continuous marginal distributions and a copula from an elliptical distribution [2,11,12,27].

For the elliptical and transelliptical distributions we propose an alternative definition of CCA using a rank correlation measure. This version of CCA has the same true canonical directions and correlations as standard CCA based on Pearson correlation in the elliptical family when second moments exist and still well defined if they do not exist. This construction uses properties of the rank correlation measure, Kendall's tau. For two univariate random variables  $Z_i$  and  $Z_j$  with joint CDF  $F(Z_i, Z_j)$ , Kendall's tau is

$$\tau^{(Z_i, Z_j)} = E\{\text{sign}(Z_i - \tilde{Z}_i)(Z_j - \tilde{Z}_j)\}$$

where  $(\tilde{Z}_i, \tilde{Z}_j)^\top$  is an identically distributed copy of  $(Z_i, Z_j)^\top$  [25]. This quantity exists for all bivariate continuous distributions, and does not require the existence of moments. A consistent estimator of Kendall's tau based on  $n$  iid copies of  $Z_i$  and  $Z_j$ ,  $(z_{i1}, z_{j1})^\top, \dots, (z_{in}, z_{jn})^\top$ , is

$$\hat{\tau}_n^{(Z_i, Z_j)} = \frac{1}{\binom{n}{2}} \sum_{1 \leq k < l \leq n} \text{sign}(z_{ik} - z_{il}) \text{sign}(z_{jk} - z_{jl})$$

This estimator is a U-statistic with consistency and asymptotic normality coming from U-statistic theory [20].

Within the transelliptical family the following proposition, which is equivalent to Theorem 3.2 in Han and Liu [18] and Theorem 3.1 in Fang et al. [12], gives the correspondence between Kendall's tau and the elements of the transelliptical scatter matrix.

**Proposition 2.2** (Kendall's Tau for Transelliptically Distributed Random Variables). Assume  $Z \sim \mathcal{TE}_d(h, 0, P_Z, \psi_Z)$ . If  $p_{ij}$  is the  $ij$ th entry of  $P_Z$  and  $\tau^{(Z_i, Z_j)}$  is the Kendall correlation between the  $i$ th and  $j$ th entries of  $Z$  then  $\tau^{(Z_i, Z_j)} = (2/\pi) \arcsin(\rho_{ij})$

Because the function connecting Kendall's tau and the scale invariant scatter matrix is a monotone increasing function between zero and one that takes the value zero only at zero, maximizing Pearson's correlation is equivalent to maximizing Kendall's tau within the elliptical family, and constraining Pearson's correlation to zero is equivalent to constraining Kendall's tau to zero. Importantly this relationship still holds between elements of the scale invariant scatter matrix and Kendall's tau for transelliptical distributions when moments do not exist.

Given [Propositions 2.2](#) and [2.1](#) we define CCA for transelliptical distributions as follows,

**Definition 2.3** (Canonical Correlation Analysis for Transelliptical Distributions). Assume  $X$  is a  $p \times 1$  dimensional random vector and  $Y$  is a  $q \times 1$  dimensional random vector, and that the random vector  $(X^\top, Y^\top)^\top = Z \sim \mathcal{TE}_{p+q}(h_Z, 0, P_Z, \psi_Z)$ . Define  $h_X$  to be the elementwise functions of  $h_Z$  corresponding to  $X$  and  $h_Y$  to be the elementwise functions of  $h_Z$  corresponding to  $Y$ . The first canonical direction vectors, the  $p \times 1$  vector,  $a_1$ , and the  $q \times 1$  vector,  $b_1$ , are the vectors that maximize  $\tau^{(U_1, V_1)}$  where  $U_1 = a_1^\top h_X(X)$  and  $V_1 = b_1^\top h_Y(Y)$ , subject to the constraint that  $U_1$  and  $V_1$  have scale parameter equal to one. The  $j$ th canonical direction vectors are the  $p \times 1$  vector  $a_j$  and the  $q \times 1$  vector  $b_j$  that maximize  $\tau^{(U_j, V_j)}$  where  $U_j = a_j^\top h_X(X)$   $V_j = b_j^\top h_Y(Y)$ , subject to the constraints that  $\tau^{(U_j, U_{j'})} = \tau^{(U_j, V_{j'})} = \tau^{(V_j, U_{j'})} = \tau^{(V_j, V_{j'})} = 0$  for all  $j' < j$ , and the scale parameter for  $U_j$  and  $V_j$  are equal to one for all  $j$ . The  $j$ th canonical correlation can be defined as  $\sin\{(\pi/2)\tau^{(U_j, V_j)}\}$ .

When second moments exist and  $(X^\top, Y^\top)^\top$  has an elliptical distribution this definition is equivalent to performing CCA based on the correlation matrix. When  $(X^\top, Y^\top)^\top$  has an elliptical distribution but moments do not exist CCA for the transelliptical family uses the same eigendecomposition of the scatter matrix as standard CCA. A large advantage of this definition is when  $(X^\top, Y^\top)^\top$  is transelliptically, but not elliptically distributed. In this setting standard CCA depends heavily on the marginal distributions of the variables in  $X$  and  $Y$ , which depends on  $h_X$  and  $h_Y$ . In many cases  $h_X$  and  $h_Y$  can act to obscure potential linear relationships between the variables. [Definition 2.3](#) is based on  $P_Z$ , which does not depend on the marginal distributions of the variables. In this sense CCA using [Definition 2.3](#) can be thought of as first transforming the variables to elliptical symmetry and then performing CCA. As shown in [Proposition 2.1](#) linear combinations of elliptical distributions meaningfully describe the associations within the variables.

It is important to note that when  $(X^\top, Y^\top)^\top = Z$  is transelliptically distributed and  $\psi_Z$  is not the generating function of a Gaussian distribution  $U_i$  and  $U_j$  for  $i \neq j$  are rank uncorrelated, but not independent. The same is true of  $V_i$  and  $V_j$ , as well as  $U_i$  and  $V_j$ . This is in contrast to CCA when  $Z$  has a multivariate normal distribution, where the different canonical variates are not only uncorrelated, but also independent. However, for all elliptical distributions it will still be the case that  $U_i$  is mean independent of  $U_j$ , in the sense that  $E(U_i|U_j) = E(U_i)$  [36], and likewise for  $V_i$  and  $V_j$  as well as  $U_i$  and  $V_j$ . In addition, if  $\{h_X(X)^\top, h_Y(Y)^\top\}^\top = h_Z(Z)$  has a non-Gaussian elliptical distribution it is not possible to find linear combinations of  $h_X(X)$  and  $h_Y(Y)$  that are independent. This is because any linear combination of  $h_X(X)$  and any linear combination of  $h_Y(Y)$  will jointly have a non-Gaussian elliptical distribution, which cannot be independent [2,24]. Therefore we believe that the constraints in transelliptical CCA requiring different canonical directions to be rank uncorrelated, and therefore mean independent, is a useful way to find a meaningful low-rank representation of the association between the variables.

Requiring fully independent, rather than just uncorrelated canonical variates, for all transelliptical distributions would in some cases require non-linear combinations of the data resulting in difficulty in interpretation.

An issue with estimating CCA for the transelliptical family is estimation of a scatter matrix of transformed versions of  $X$  and  $Y$ . If  $(X^\top, Y^\top)^\top = Z$  is transelliptically distributed and  $h_X$ ,  $h_Y$ , and  $\psi_Z$  are all unknown, then all three must be estimated to transform  $Z$  to its underlying elliptical distribution. Many methods assume that  $\psi_Z$  is the generating function from a Gaussian distribution, which can introduce bias if this assumption is not met. In order to avoid estimation of  $h_X$ ,  $h_Y$  and  $\psi_Z$  we directly estimate the scatter matrix in the transelliptical distribution as follows [30],

**Definition 2.4** (Transelliptical Scatter Matrix Estimate). Assume that  $Z \sim \mathcal{TE}_d(h_Z, 0, P_Z, \psi_Z)$ . Assume that  $\rho_{ij}$  is the element of  $P_Z$  corresponding to the  $i$ th and  $j$ th elements of  $Z$ . Then we can estimate  $\rho_{ij}$  as  $\hat{\rho}_{n,ij} = \sin \left\{ (\pi/2) \hat{\tau}_n^{(Z_i, Z_j)} \right\}$ , and  $P_Z$  by estimating all individual entries in this manner. We will refer to this estimator of the scatter matrix,  $\hat{P}_{Zn}$ , as the transformed Kendall's scatter matrix estimator.

This estimator has also been referred to as the latent generalized correlation matrix, and its statistical properties including convergence rate in high dimensions have been previously studied [17,19]. To obtain estimates for the canonical directions and correlations for the transelliptical family, we simply decompose the transformed Kendall's scatter matrix estimator as we would any correlation matrix estimate when conducting CCA. We note that the transformed Kendall's scatter matrix estimator does not require estimation of the transformations  $h_Z$ , or the generator  $\psi_Z$ , for all transelliptical distributions.

There are other rank based methods that can be used to estimate the scatter matrix for transellipticals when  $\psi_Z$  is assumed to be the generating function for the Gaussian distribution. One such estimator uses transformations of Spearman's correlation. For the bivariate normal distribution Spearman's correlation,  $s$ , and Pearson correlation,  $\rho$ , have the following relationship,  $s = (6/\pi) \arcsin(\rho/2)$ , although this relationship does not extend to other elliptical distributions in the same way that the relationship between Kendall's tau and Pearson's correlation does [22]. Another rank based method is to transform all marginals to be normal using an inverse CDF transformation and then using the standard sample Pearson correlation estimator. When data are generated from a transelliptical distribution and the generating function,  $\psi_Z$ , is from an elliptical distribution other than a Gaussian this method results in biased estimates of the transelliptical scatter matrix.

In addition there is another rank based method, Blomqvist's beta, that can be used to estimate the scatter matrix across all transelliptical distributions. Blomqvist's beta between two variables,  $Z_i$  and  $Z_j$ ,  $\beta_B^{(Z_i, Z_j)}$ , is defined as,  $\beta_B^{(Z_i, Z_j)} = E\{\text{sign}(Z_i - Z_{imed})\text{sign}(Z_j - Z_{jmed})\}$ , where  $Z_{imed}$  and  $Z_{jmed}$  denote the population medians of  $Z_i$  and  $Z_j$  respectively. For elliptical copulas  $\beta_B^{(Z_i, Z_j)} = \tau^{(Z_i, Z_j)}$  [2,23,43]. This means that the correspondence between Blomqvist's beta and Pearson correlation within transelliptical distributions is the same as the correspondence between Kendall's tau and Pearson correlation. Therefore the sample estimate of Blomqvist's beta can be used in a similar fashion to the sample estimate of Kendall's tau in order to estimate transelliptical canonical directions and correlations. However, simulation results in Section 3 indicate that estimates using the sample version of Blomqvist's beta perform much worse than estimates using the sample version of Kendall's tau in the finite sample setting, and for this reason we primarily focus on the transformed Kendall's scatter matrix estimator. One area in which we do consider estimation of Blomqvist's beta is in testing whether data are generated using an elliptical copula. Jaser et al. [23] create a test for the null hypothesis that data are generated from a transelliptical distribution that uses the equivalence between Kendall's tau and Blomqvist's beta for elliptical copulas. We revisit this in Section 4.

A potential issue with the transformed Kendall's scatter matrix estimator is that it is not guaranteed to be positive-definite even when the true scatter matrix,  $P_Z$ , is positive-definite. As discussed by Rousseeuw and Molenberghs [41] various methods are available to adjust  $\hat{P}_{Zn}$  so that it is positive-definite. For simplicity we define  $\tilde{P}_{Zn}$  to be the matrix with the same eigenvectors and positive eigenvalues as  $\hat{P}_{Zn}$  but with all negative eigenvalues set to some small positive constant.  $\tilde{P}_{Zn}$  will have the same asymptotic behavior as  $\hat{P}_{Zn}$  based on the following theorem:

**Theorem 2.1** (Transformed Kendall's Scatter Matrix Estimator Eigenvalues). Assume  $z_1, \dots, z_n$  are  $d$ -dimensional iid realizations of transelliptically distributed vector,  $Z$ , with positive-definite scale invariant scatter matrix  $P_Z$ . Define the ordered eigenvalues of the transformed Kendall's scatter matrix,  $\hat{P}_{Zn}$  to be  $\hat{\lambda}_{n1}, \dots, \hat{\lambda}_{nd}$ , where  $\hat{\lambda}_{nd}$  is the minimum eigenvalue of  $\hat{P}_{Zn}$ . Then  $\Pr(\hat{\lambda}_{nd} > 0) \rightarrow_p 1$

A proof of Theorem 2.1 is presented in Appendix A. Theorem 2.1 gives that the probability of  $\tilde{P}_{Zn}$  being equal to  $\hat{P}_{Zn}$  converges to one for transelliptically distributed  $Z$  with positive-definite  $P_Z$ . This means for transelliptical  $Z$  when  $P_Z$  is positive-definite  $\sqrt{n}(\tilde{P}_{Zn} - P_Z)$  and  $\sqrt{n}(\hat{P}_{Zn} - P_Z)$  will have the same limiting distribution. The limiting distribution of  $\sqrt{n}(\tilde{P}_{Zn} - P_Z)$  can be shown to be asymptotically normal with mean zero and finite variance based on U-statistic theory [20,42] and the delta method.

Next we will show asymptotic properties for estimates of transelliptical canonical correlation using an eigendecomposition based off of a consistent and asymptotically normal estimate of the scatter matrix. Specifically we will focus on the unique non-zero transelliptical canonical correlations and there corresponding directions. As before we will assume



$(X^T, Y^T)^T = Z \sim \mathcal{TE}_{p+q}(h_Z, 0, P_Z, \psi_Z)$ , and that there are  $r \leq \min(p, q)$  unique non-zero transelliptical canonical correlations. We will denote these as  $\lambda_1, \dots, \lambda_r$ , with  $\lambda_1 > \dots > \lambda_r > 0$ . Define  $\Lambda_r = \text{diag}(\lambda_1, \dots, \lambda_r)$  to be the diagonal matrix with the ordered non-zero canonical correlations on the diagonal. Let  $A_r = (a_1, \dots, a_r)$  be the  $p \times r$  matrix where the  $i$ th column is the  $i$ th transelliptical canonical direction for  $X$ , and  $B_r = (b_1, \dots, b_r^T)$  be the  $q \times r$  matrix where the  $i$ th column is the  $i$ th transelliptical canonical direction for  $Y$ . Define  $A_{r+} = (a_{r+1}, \dots, a_p)$  and  $B_{r+} = (b_{r+1}, \dots, b_q)$  to be a solution to the canonical directions corresponding to the zero canonical correlations. This means for  $A = (A_r, A_{r+})$  and  $B = (B_r, B_{r+})$ ,  $A^T P_{XX} A = I_p$ ,  $B^T P_{YY} B = I_q$ , and  $A^T P_{XY} B = \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix}$ . Note that  $A_r$  and  $B_r$  are well defined up to a sign change and  $A_{r+}$  and  $B_{r+}$  are well defined up to multiplication by an orthogonal matrix on the right.  $A_{r+}$  and  $B_{r+}$  can be made unique by imposing suitable constraints.

$P_{Zn}^* = \begin{pmatrix} P_{XXn}^* & P_{XYn}^* \\ P_{YXn}^* & P_{YYn}^* \end{pmatrix}$  will be used to denote an arbitrary consistent and asymptotically normal estimator of  $P_Z$  based on  $n$  iid realizations of  $Z$ .  $A_r$ ,  $A_{r+}$ , and  $B_r$  can all be estimated by the eigendecomposition of the relevant function of  $P_{Zn}^*$ . Denote these estimates as  $\Lambda_{rn}^*$ ,  $A_{rn}^*$ , and  $B_{rn}^*$  respectively. For notational simplicity the subscript  $n$  may be dropped in future references. For theoretical purposes we will define  $P_{UU}^* = A^T P_{XX}^* A$ ,  $P_{VV}^* = B^T P_{YY}^* B$ ,  $P_{UV}^* = A^T P_{XY}^* B$ , and  $P_{VU}^* = P_{UV}^{*\top}$ .  $P_{UUij}^*$  will denote the entry for the  $i$ th row and  $j$ th column of  $P_{UU}^*$ , with similar notation used for  $P_{VV}^*$ ,  $P_{UV}^*$  and  $P_{VU}^*$ . Further define  $G = (g_1, \dots, g_r)$  and  $H = (h_1, \dots, h_r)$  to be the solutions to the system of equations

$$\begin{pmatrix} -\lambda_i^* P_{UU}^* & P_{UV}^* \\ P_{VU}^* & \lambda_i^* P_{VV}^* \end{pmatrix} \begin{pmatrix} g_i \\ h_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (1)$$

where

$$\begin{vmatrix} -\lambda_i^* P_{UU}^* & P_{UV}^* \\ P_{VU}^* & -\lambda_i^* P_{VV}^* \end{vmatrix} = 0. \quad (2)$$

In order to uniquely define  $G$  and  $H$  we will assume that

$$G^T P_{UU}^* G = I_r, \quad H^T P_{VV}^* H = I_r, \quad (3)$$

and  $g_{ii} > 0$  where  $g_{ii}$  is the  $i$ th entry of  $g_i$ . Further define  $g_{ij}$  to be the  $i$ th entry of  $g_j$ , and likewise for  $h_{ij}$ . [Theorem 2.2](#) establishes conditions under which the estimates of transelliptical CCA directions and correlations will be consistent and asymptotically normal and gives results on the form of the limiting variances.

**Theorem 2.2** (Asymptotic Results for Transelliptical CCA). Assume  $(x_l^T, y_l^T)^T$  for  $l \in \{1, \dots, n\}$  are iid realizations of the  $(p+q) \times 1$  dimensional random vector  $(X^T, Y^T)^T = Z \sim \mathcal{TE}_{p+q}(h_Z, 0, P_Z, \psi_Z)$ , with positive-definite  $P_Z$ . Further assume that  $p \geq q$  and there are  $r \leq \min(p, q)$  unique non-zero transelliptical canonical correlations for  $X$  and  $Y$ . If  $P_Z^*$  is guaranteed to be positive-definite and

$$\sqrt{n} \begin{Bmatrix} \text{vec}(P_{XX}^*) & \text{vec}(P_{XX}) \\ \text{vec}(P_{XY}^*) - \text{vec}(P_{XY}) \\ \text{vec}(P_{YY}^*) & \text{vec}(P_{YY}) \end{Bmatrix} \rightarrow_d N_{p^3 \times q^3}(0, \Theta),$$

then,

$$\sqrt{n} \begin{Bmatrix} \text{vec}(P_{UU}^*) & \text{vec}(I_p) \\ \text{vec}(P_{UV}^*) - \text{vec}(\Lambda_{r,pq}) \\ \text{vec}(P_{VV}^*) & \text{vec}(I_q) \end{Bmatrix} \rightarrow_d N_{p^3 \times q^3}(0, J_Z \Theta J_Z^T), \quad (4)$$

where  $\Lambda_{r,m_1 m_2}$  is an  $m_1 \times m_2$  matrix with  $m_1, m_2 \geq r$  and the upper left hand corner equal to  $\Lambda_r$  and all other entries equal to zero, and  $J_Z$  is the  $(p^3 \times q^3) \times (p^3 \times q^3)$  block matrix,

$$J_Z = \begin{pmatrix} A^T \otimes A^T & 0 & 0 \\ 0 & B^T \otimes A^T & 0 \\ 0 & 0 & B^T \otimes B^T \end{pmatrix}.$$

For  $i \in \{1, \dots, r\}$

$$\sqrt{n}(\lambda_i^* - \lambda_i) = \frac{\sqrt{n} \left\{ \lambda_i(P_{UVii}^* - \lambda_i) + \lambda_i(P_{VUii}^* - \lambda_i) - \lambda_i(P_{VVii}^* - 1)\lambda_i(P_{UUii}^* - 1) \right\}}{4\lambda_i^2} + o_p(1), \quad (5)$$

$$\sqrt{n}(g_{ii} - 1) = \frac{\sqrt{n}(P_{UUii} - 1)}{4} + o_p(1), \quad (6)$$

$$\sqrt{n}(h_{ii} - 1) = \frac{\sqrt{n}(P_{W_{ii}} - 1)}{4} + o_p(1), \quad (7)$$

where  $\lambda_i = 0$  if  $i > r$ . For  $i \in \{1, \dots, p\}$ ,  $j \in \{1, \dots, r\}$  and  $i \neq j$

$$\sqrt{n}(g_{ij}) = \frac{\sqrt{n} \{P_{UV_{ij}}^* \lambda_j + P_{VU_{ij}}^* \lambda_i - P_{VV_{ij}}^* \lambda_i \lambda_j - P_{UU_{ij}}^* \lambda_j^2\}}{(\lambda_i^2 - \lambda_j^2)^2} + o_p(1), \quad (8)$$

and for  $i \in \{1, \dots, q\}$ ,  $j \in \{1, \dots, r\}$  and  $i \neq j$

$$\sqrt{n}(h_{ij}) = \frac{\sqrt{n} \{P_{VU_{ij}}^* \lambda_j + P_{UV_{ij}}^* \lambda_i - P_{UU_{ij}}^* \lambda_i \lambda_j - P_{VV_{ij}}^* \lambda_j^2\}}{(\lambda_i^2 - \lambda_j^2)^2} + o_p(1), \quad (9)$$

and finally for  $j \in \{1, \dots, r\}$

$$\sqrt{n}(a_j^* - a_j) = \sum_{j=1}^p a_j \sqrt{n}\{g_{ij} - 1(i=j)\} + o_p(1), \quad (10)$$

$$\sqrt{n}(b_j^* - b_j) = \sum_{i=1}^q b_i \sqrt{n}\{h_{ij} - 1(i=j)\} + o_p(1). \quad (11)$$

$\sqrt{n}\{\text{vec}(\Lambda_r^*) - \text{vec}(\Lambda_r)\}$ ,  $\sqrt{n}\{\text{vec}(A_r^*) - \text{vec}(A_r)\}$  and  $\sqrt{n}\{\text{vec}(B_r^*) - \text{vec}(B_r)\}$  jointly have a multivariate normal limiting distribution with mean zero and finite variance that is a function of  $\Theta$ ,  $A$ ,  $B$ , and  $\Lambda_r$  and can be solved for by using Eqs. (4)–(11).

The proof for Theorem 2.2 can be found in Appendix A. A consistent estimate of the relevant limiting variances can be found by plugging in consistent estimates of  $\Theta$ ,  $A$ ,  $B$ , and  $\Lambda_r$ . This result, and the limiting variance of the estimates, is more general than previous results from Anderson [5] and Taskinen et al. [44], and requires only that the estimate of the covariance matrix be asymptotically normal and positive-definite. Anderson [5] shows the asymptotic results for standard CCA directions and correlations when  $Z$  has a multivariate normal distribution and CCA is estimated using the sample covariance matrix. Taskinen et al. [44] expanded this result to CCA for elliptical distributions when using positive-definite and affine equivariant estimators of the covariance matrix. Because we make minimal assumptions about the form of  $\Theta$  we do not get a concise form of the limiting variances as in previous results. Because  $\tilde{P}_{Zn}$  is not affine equivariant our more general result is needed.

We have already shown that  $\tilde{P}_{Zn}$  is positive-definite, consistent, and asymptotically normal, which leads directly to Corollary 2.2.1.

**Corollary 2.2.1** (Asymptotic Results for Transformed Kendall's Scatter Matrix Estimator). Assume the same set up as in Theorem 2.2 and that  $P_Z$  is estimated using  $\tilde{P}_{Zn}$ . Define  $\tilde{\Lambda}_m$ ,  $\tilde{A}_m$ , and  $\tilde{B}_m$  as the corresponding estimates for the non-zero transelliptical canonical correlations and their corresponding transelliptical canonical directions. Then  $\sqrt{n}\{\text{vec}(\tilde{\Lambda}_m) - \text{vec}(\Lambda_r)\}$ ,  $\sqrt{n}\{\text{vec}(\tilde{A}_m) - \text{vec}(A_r)\}$ , and  $\sqrt{n}\{\text{vec}(\tilde{B}_m) - \text{vec}(B_r)\}$  jointly have a multivariate normal limiting distribution with mean zero and finite variance. The form of the variances can be found using Theorem 2.2 by substituting the limiting variance of  $\tilde{P}_{Zn}$  for  $\Theta$ .

This result follows directly from Theorem 2.2. Methods from Rublík [42] can be used to obtain estimators for the limiting covariance matrix for all pairwise estimates of Kendall's tau. An estimate of the limiting variance of  $\tilde{P}_{Zn}$  can then be found using the delta method. This can be used as a consistent estimate of  $\Theta$  in Theorem 2.2. This allows for the limiting variances of  $\tilde{\Lambda}_m$ ,  $\tilde{A}_m$ , and  $\tilde{B}_m$  to be estimated by a “plug-in” estimator using Eqs. (4)–(11) in Theorem 2.2. Section 3 and the supplementary materials include simulations studies that compare the coverages of confidence intervals using this method to bootstrapped confidence intervals.

These results show that the transelliptical CCA estimates using the transformed Kendall's estimator are consistent and asymptotically unbiased. For finite samples the estimates of the transelliptical canonical correlations have a positive bias that is also present in the estimation of canonical correlations using standard methods. Because of this bias we recommend using a jackknife bias correction for the estimates of both transelliptical canonical correlations and standard canonical correlations.

It is important to note that Theorem 2.2 and Corollary 2.2.1 only apply to non-zero canonical correlations and cannot be used for hypothesis testing for zero correlations. Anderson [6] gives the asymptotic distribution for the zero canonical correlations for standard CCA when  $X$  and  $Y$  are jointly multivariate normal and show that in this case the estimates of the correlations converge at rate  $n$ . A number of asymptotic tests have been derived for the specific case where standard CCA is used and  $X$  and  $Y$  have a multivariate normal joint distribution [37,38,48]. In addition Muirhead and Waternaux [33] show how test statistics used to test for a true canonical correlation of zero when  $X$  and  $Y$  are multivariate normal can be modified for elliptical distributions. These results exploit special properties of elliptical distributions and sample covariance matrix, but it is unclear how to generalize these results to transelliptical CCA using the transformed Kendall's

scatter matrix estimator. Because of this we propose a testing procedure based on bootstrapped replicates. To control the type I error at  $\alpha$  simply invert a  $(1-2\alpha)$ -bootstrapped confidence interval using the normal approximation with bias correction. A  $(1-2\alpha)$ -confidence interval is used because this test is only one sided, so using a  $(1 - \alpha)$  interval will unnecessarily reduce power. Other bootstrap confidence intervals may be used, although it is important not to use the simple percentile method. This is because within each bootstrap sample the estimated canonical correlation will be above zero. This means some type of bias correction is necessary. Although the asymptotic distribution for true correlations of zero is not normal, the fact that the correlations converge at rate  $n$  as opposed to  $\sqrt{n}$  implies that this will have conservative type I error as sample size increases. Simulation results in Section 3 indicate that this is the case.

Given the conservative nature of this test, particularly as sample size increases, it is important to point out why it this bootstrapping procedure is preferred to other testing procedures, including permutation based testing. Permutation or randomization testing assumes that under the null hypothesis observations are exchangeable. For transelliptical distributions this assumption is only met when data are generated from a distribution with a Gaussian copula where having a true correlation of zero implies independence. For all other elliptical copulas this is not the case, so permutation tests will lead to inflated type I error. Even for CCA estimated using the sample correlation or covariance matrix a permutation test will lead to inflated type I error if the data are not generated from a distribution with a Gaussian copula, and asymptotic testing procedures assume the data are generated from a multivariate Gaussian distribution. Importantly this means that even if all the marginal distributions are Gaussian, permutation and asymptotic tests will result in inflated type I error if the copula defining the joint distribution is not a Gaussian copula. The permutation test can be thought of as a test of the null hypothesis that the canonical variates are independent, rather than a test for a true canonical correlation of zero. However for transelliptical CCA a test of independence can be misleading if data are not generated from a Gaussian copula. As an example consider  $(X^T, Y^T)^T = Z$  with a multivariate Cauchy distribution and a scatter matrix equal to the identity matrix. In this case there are no well defined unique canonical directions for transelliptical CCA, as any linear combinations of  $X$  and  $Y$  which meet the relevant constraints will be rank uncorrelated. Any estimated directions will be purely due to random variability within the particular sample, and will not be informative of associations between  $X$  and  $Y$ . However any linear combinations of  $X$  and  $Y$  will not be independent. Therefore even when transelliptical CCA variables are not independent, the estimated directions may still be due purely to random noise and not be informative of any true associations of the variables. For this reason we recommend testing the null that the true canonical correlation is zero, which can be done using the inverted bootstrap procedure. This is particularly the case when data are not generated from a distribution with a Gaussian copula.

The inverted bootstrap procedure does not even need the transelliptical assumption, just the assumption that the estimated correlation or covariance matrix is asymptotically normal. For the transformed Kendall's estimator this only requires that the data from different subjects be independent and identically distributed, and for the sample correlation or covariance matrix this only requires that the data be independent and identically distributed and fourth moments exist. When using the transformed Kendall's estimator this bootstrap procedure will test the null hypothesis that for all variables in  $X$  the true pairwise Kendall's tau coefficient with all variables in  $Y$  is zero. Therefore even when data are not generated from a distribution with an elliptical copula this provides a meaningful test for association between the two sets of variables. Simulation results comparing the bootstrap testing procedures with other testing procedures are presented in Section 3.3.

### 3. Simulation results

#### 3.1. Empirical bias and variance of CCA with robust covariance estimation

Simulations are conducted to compare transelliptical CCA using the transformed Kendall's estimator and standard CCA under both elliptical and transelliptical settings. In addition CCA based on two robust covariance matrix estimators are considered, the re-weighted MCD estimator from the R package *robustbase* [45] and the S estimator from the R package *rrcov* [45]. For the re-weighted MCD estimator a maximum proportion of 0.75 and 0.5 of the observations were considered. When using a cutoff of 0.75 the bias and standard deviation of the direction and correlation of estimates are improved relative to those using a cutoff of 0.5, so only those results using a cutoff of 0.75 are reported. For the S estimator a breakdown point of 0.75 and 0.5 was considered. Results were very similar across both breakdown points so only the breakdown point of 0.75 is presented. In addition two different rank based correlation estimators are considered, one based on Spearman's correlation and one based on Blomqvist's beta. The Spearman correlation estimator estimates all pairwise Spearman correlation values and uses the inverse of the equivalence relationship between Spearman's correlation and Pearson correlation,  $s = (6/\pi) \arcsin(\rho/2)$ , to get an estimate of the Pearson correlation matrix. Because this equivalence only holds for the multivariate normal distribution and not elliptical distributions in general this estimator was only considered when data were generated with a Gaussian copula. Likewise the Blomqvist Beta estimator first estimates all pairwise Blomqvist's beta values and then uses the inverse of the equivalence relationship between Blomqvist's beta and Pearson correlation,  $\beta = (2/\pi) \arcsin(\rho)$ , to get an estimate of the Pearson correlation matrix. Standard CCA is calculated using the R package *CCA* [16].

The distributions of the simulated data sets are multivariate normal, multivariate Cauchy, multivariate t with five and ten degrees of freedom, and the multivariate lognormal. The first four distributions satisfy the elliptical assumptions,



while the latter is a member of the transelliptical but not elliptical family. The sample size of the simulated data sets are  $n = 200$  and  $n = 1000$ , and the dimension of  $X$  and  $Y$  are  $p = q = 4, 8$  and  $16$ . Results for  $p = q = 8$  and  $n = 200$  are presented below, with the other results given in the supplementary material. The relative performance of the different methods is generally consistent across the different dimensions and sample sizes, with methods improving as sample size increases or dimension decreases as long as they have been shown to be consistent for a given distribution. The true scatter matrix for  $X$  and  $Y$  is  $\Sigma_{XX} = \Sigma_{YY} = I_p$ , and  $\Sigma_{XY}$  is a diagonal matrix where the first four diagonal entries are  $0.9, 0.5, 0.4$ , and  $1/3$ , and all remaining diagonal entries are zero. The structure of these scatter matrices is similar to those in Branco et al. [8]. To define  $\hat{P}_{Zn}$  all negative eigenvalues are set to  $0.001$ . For each simulation setting, at most  $0.2\%$  of simulations resulted in  $\hat{P}_{Zn}$  not being positive-definite. The total number of simulated data sets for each simulation setting is  $1000$ .

Based on the  $1000$  simulated data sets the empirical bias and standard deviation are calculated for the canonical correlations and directions for each of CCA methods. For the canonical correlation estimates the bias and variance are calculated after a Fisher inverse hyperbolic tangent transformation. For the  $i$ th canonical direction the angle between the true direction for  $X$ ,  $a_i$ , and estimated direction for  $X$  for the  $j$ th simulation,  $\hat{a}_i^{(j)}$ , is calculated as

$$\cos^{-1} \left( \frac{|\hat{a}_i^{(j)\top} a_i|}{\|\hat{a}_i^{(j)}\| \cdot \|a_i\|} \right).$$

The bias for the canonical directions is estimated as the average angle across all simulated data sets and the standard deviation is estimated as the empirical standard deviation of the angles across all simulated data sets. Table 1 gives the output for the canonical correlation and canonical direction. Because  $\Sigma_{XY}$  is symmetric and  $p = q$  only the bias and standard deviation for the  $X$  direction are presented, with the results for  $Y$  being nearly identical.

Based on results in Table 1 standard CCA has the smallest bias and standard deviation when simulating under the multivariate normal distribution, with the  $S$  estimator outperforming the other robust methods. The transformed Kendall's estimator performs similarly to the re-weighted MCD estimator and estimator based on Spearman's correlation, while the Blomqvist's beta estimator is by far the worst estimator when data are simulated from a multivariate normal distribution. As noted previously, we can see in the multivariate normal setting that there is evidence of a positive finite sample bias in the estimates of the canonical correlations for all methods considered. This is also true for all simulations from all other distributions considered, except for the lognormal distribution when methods are not invariant to monotone transformations of the data. For simulations from the Cauchy distribution the  $S$  estimator performs the best and the transformed Kendall's estimator performs slightly worse than the re-weighted MCD estimator for the first direction, and better than the re-weighted MCD estimator higher directions. The estimator using Blomqvist's beta is again the worst performing robust estimator. Standard CCA has high bias and variance under this setting because of the lack of moments. When data are generated from a multivariate  $t$  distribution with five or ten degrees of freedom the  $S$  estimator is again the best performing robust method, followed by the transformed Kendall's estimator. When data are simulated from a  $t$  distribution with ten degrees of freedom the standard CCA estimates perform similarly to the  $S$  estimator, while when data are simulated using a  $t$  distribution with five degrees of freedom the standard estimator performs worse than the  $S$  estimator and transformed Kendall's estimator. This suggests that even when standard CCA is well defined, robust estimators such as the  $S$  estimator or transformed Kendall's estimator outperforms the sample covariance matrix for heavy tailed elliptical distributions. Under the lognormal setting the standard estimator, the re-weighted MCD estimator, and the  $S$  estimator all underestimate the transelliptical canonical correlations, while the transformed Kendall's estimator and estimator using Spearman's correlation has small positive bias that reduces with sample size. This is particularly evident for  $n = 1000$  presented in the supplementary material. In addition the estimated canonical directions using the transformed Kendall's estimator have lower bias than standard CCA or the estimates using the  $S$  or re-weighted MCD estimators, particularly in the second and higher directions. These findings illustrate the advantages of the transelliptical CCA with data that are transelliptically but not elliptically distributed. Even without transforming potentially skewed marginal distributions the transformed Kendall's estimator can estimate the strongest linear relationships based on the underlying copula. The transformed Kendall's estimator also has advantages over the other rank based methods. Unlike the method based on transformed Spearman's correlation it can be used for non-Gaussian elliptical copulas. It also greatly outperforms the method based on transformations of Blomqvist's beta in all simulation settings considered.

### 3.2. Confidence intervals for non-zero canonical correlations

Simulations are run to compare coverages for transelliptical canonical correlations estimated using the transformed Kendall's estimator using normal bootstrapped confidence intervals as well as asymptotic confidence intervals using "plug-in" estimators of the asymptotic variance for the estimates from Theorem 2.2. Details on the form of the variance estimates are in Theorem 2.2 and Corollary 2.2.1, as well as in Appendix A. The "plug-in" variance estimator is calculated using estimates of transelliptical canonical correlations and directions, based on the transformed Kendall's estimator. An estimate of the variance of the transformed Kendall's scatter matrix is obtained using methods from Rublík [42] and the delta method. Because canonical correlations must be between zero and one the asymptotic confidence intervals are truncated at zero or one as necessary. For the bootstrap confidence intervals  $1000$  bootstrap replicates are used. Confidence

**Table 1**

Average empirical bias (standard deviation) from 1000 simulations of the estimates of the first four canonical correlations and directions. Six different estimation techniques used with data simulated from five different elliptical and transelliptical distributions with  $p = q = 8$  and  $n = 200$ .

		Canonical correlations				
		Normal	Cauchy	Lognormal	t5	t10
Cor 1	Standard	0.05 (0.07)	1.94 (1.13)	−0.12 (0.19)	0.09 (0.11)	0.05 (0.08)
	Kendall	0.06 (0.08)	0.17 (0.14)	0.05 (0.08)	0.07 (0.09)	0.06 (0.08)
	S	0.05 (0.08)	0.09 (0.11)	−0.16 (0.12)	0.06 (0.08)	0.04 (0.08)
	MCD	0.06 (0.08)	0.13 (0.12)	−0.15 (0.13)	0.08 (0.10)	0.07 (0.09)
	Spearman	0.05 (0.09)	–	0.04 (0.08)	–	–
	Blomqvist	0.60 (0.60)	0.63 (0.65)	0.60 (0.60)	0.60 (0.57)	0.57 (0.58)
Cor 2	Standard	0.09 (0.06)	1.70 (0.69)	0.01 (0.12)	0.17 (0.09)	0.11 (0.07)
	Kendall	0.10 (0.07)	0.19 (0.09)	0.10 (0.07)	0.12 (0.07)	0.11 (0.07)
	S	0.09 (0.06)	0.17 (0.08)	−0.02 (0.07)	0.11 (0.07)	0.10 (0.06)
	MCD	0.11 (0.07)	0.23 (0.09)	0.02 (0.08)	0.15 (0.08)	0.15 (0.08)
	Spearman	0.10 (0.07)	–	0.10 (0.07)	–	–
	Blomqvist	0.29 (0.13)	0.29 (0.13)	0.30 (0.12)	0.29 (0.12)	0.30 (0.12)
Cor 3	Standard	0.07 (0.05)	1.22 (0.46)	−0.02 (0.07)	0.13 (0.07)	0.09 (0.06)
	Kendall	0.08 (0.06)	0.14 (0.07)	0.08 (0.06)	0.09 (0.06)	0.09 (0.06)
	S	0.07 (0.06)	0.13 (0.07)	−0.02 (0.06)	0.08 (0.06)	0.08 (0.06)
	MCD	0.09 (0.06)	0.17 (0.08)	0.01 (0.06)	0.12 (0.07)	0.11 (0.06)
	Spearman	0.08 (0.06)	–	0.08 (0.06)	–	–
	Blomqvist	0.19 (0.09)	0.20 (0.09)	0.20 (0.09)	0.19 (0.09)	0.20 (0.09)
Cor 4	Standard	0.03 (0.06)	0.86 (0.34)	−0.05 (0.06)	0.07 (0.06)	0.04 (0.06)
	Kendall	0.03 (0.06)	0.07 (0.06)	0.03 (0.06)	0.04 (0.06)	0.04 (0.06)
	S	0.03 (0.06)	0.07 (0.06)	−0.05 (0.05)	0.04 (0.06)	0.03 (0.06)
	MCD	0.04 (0.06)	0.11 (0.07)	−0.03 (0.06)	0.06 (0.06)	0.06 (0.06)
	Spearman	0.03 (0.06)	–	0.03 (0.06)	–	–
	Blomqvist	0.11 (0.07)	0.11 (0.07)	0.11 (0.07)	0.11 (0.07)	0.12 (0.07)
		Canonical directions				
Dir 1	Standard	0.10 (0.03)	0.69 (0.43)	0.13 (0.07)	0.13 (0.04)	0.11 (0.03)
	Kendall	0.11 (0.03)	0.19 (0.06)	0.11 (0.03)	0.13 (0.04)	0.12 (0.04)
	S	0.10 (0.03)	0.14 (0.04)	0.12 (0.04)	0.11 (0.03)	0.11 (0.03)
	MCD	0.11 (0.03)	0.16 (0.05)	0.15 (0.05)	0.13 (0.04)	0.13 (0.04)
	Spearman	0.11 (0.03)	–	0.11 (0.03)	–	–
	Blomqvist	0.31 (0.10)	0.30 (0.10)	0.31 (0.10)	0.30 (0.10)	0.30 (0.10)
Dir 2	Standard	0.57 (0.30)	1.26 (0.26)	0.80 (0.42)	0.72 (0.34)	0.63 (0.32)
	Kendall	0.60 (0.31)	0.75 (0.32)	0.60 (0.32)	0.63 (0.31)	0.63 (0.31)
	S	0.58 (0.31)	0.74 (0.33)	0.73 (0.37)	0.62 (0.31)	0.59 (0.30)
	MCD	0.63 (0.32)	0.82 (0.33)	0.81 (0.36)	0.71 (0.33)	0.69 (0.32)
	Spearman	0.60 (0.31)	–	0.60 (0.32)	–	–
	Blomqvist	0.87 (0.32)	0.88 (0.32)	0.87 (0.32)	0.85 (0.31)	0.88 (0.32)
Dir 3	Standard	0.82 (0.35)	1.25 (0.24)	1.04 (0.36)	0.97 (0.34)	0.89 (0.35)
	Kendall	0.84 (0.34)	0.99 (0.32)	0.85 (0.34)	0.88 (0.34)	0.88 (0.34)
	S	0.83 (0.35)	0.99 (0.33)	0.99 (0.34)	0.87 (0.35)	0.85 (0.34)
	MCD	0.87 (0.33)	1.06 (0.30)	1.06 (0.32)	0.96 (0.34)	0.93 (0.33)
	Spearman	0.84 (0.34)	–	0.86 (0.35)	–	–
	Blomqvist	1.06 (0.29)	1.07 (0.30)	1.07 (0.30)	1.08 (0.30)	1.05 (0.30)
Dir 4	Standard	0.79 (0.31)	1.27 (0.24)	1.07 (0.33)	1.00 (0.31)	0.88 (0.32)
	Kendall	0.83 (0.31)	1.03 (0.31)	0.84 (0.32)	0.89 (0.32)	0.86 (0.32)
	S	0.80 (0.30)	1.01 (0.31)	1.02 (0.32)	0.87 (0.32)	0.84 (0.32)
	MCD	0.85 (0.30)	1.08 (0.30)	1.12 (0.29)	0.99 (0.31)	0.96 (0.31)
	Spearman	0.83 (0.31)	–	0.84 (0.32)	–	–
	Blomqvist	1.11 (0.29)	1.12 (0.28)	1.11 (0.28)	1.10 (0.29)	1.10 (0.29)

intervals using 2000 bootstrap replicates were also considered and had similar coverages. The bootstrap confidence intervals for the canonical correlations use the normal approximation method including a bias correction. The confidence intervals are constructed from the square of the canonical correlations, and then transformed using the square root to give the bounds for the transelliptical canonical correlations. Because we are using the normal approximation bootstrap confidence intervals may need to be truncated at zero or one, similar to the asymptotic confidence intervals. This can be done before applying the square root transformation. We use the square of the canonical correlations because bootstrap confidence intervals constructed in this way have better coverage than using the canonical correlations themselves in our simulations. We believe this is related to the fact that the canonical correlations are themselves the square root of the eigenvalues for the eigensystem used to solve for CCA directions and correlations.

The simulation set-ups are the same as Section 3.1. Table 2 reports the coverages for the canonical correlations when  $p = q = 8$ , with other dimensions found in the supplementary materials. As with the results for the empirical bias and variance, the relative performance of different methods for constructing confidence intervals is similar across

**Table 2**

Proportion of 1000 simulations in which the estimated confidence interval for the transelliptical canonical correlation contains the true transelliptical canonical correlation. Calculated for bootstrap and asymptotic confidence intervals using the transformed Kendall's estimator for first four canonical correlations with data simulated from five different elliptical and transelliptical distributions with  $p = q = 8$  and  $n = 200$  and  $1000$ .

Canonical correlation	Bootstrap coverages					Asymptotic coverages				
	Normal	Cauchy	Lognormal	t5	t10	Normal	Cauchy	Lognormal	t5	t10
<b>n = 200</b>										
1	0.90	0.81	0.92	0.88	0.90	0.84	0.60	0.84	0.79	0.82
2	0.90	0.90	0.89	0.91	0.89	0.73	0.51	0.74	0.70	0.70
3	0.87	0.88	0.89	0.86	0.88	0.85	0.75	0.84	0.82	0.83
4	0.85	0.86	0.83	0.84	0.81	0.97	0.96	0.97	0.97	0.98
<b>n = 1000</b>										
1	0.94	0.90	0.93	0.93	0.92	0.93	0.89	0.93	0.93	0.93
2	0.93	0.92	0.93	0.91	0.94	0.93	0.89	0.92	0.90	0.91
3	0.89	0.89	0.91	0.91	0.90	0.94	0.93	0.95	0.93	0.93
4	0.91	0.91	0.91	0.89	0.93	0.98	0.97	0.97	0.96	0.97

different dimensions. For the non-zero canonical correlations both the asymptotic and bootstrap confidence intervals tend to have undercoverage in our simulations when  $n = 200$ . This is particularly the case for asymptotic confidence intervals as dimension increases, likely due to the lack of bias correction. Coverage for both the asymptotic and bootstrap confidence intervals improves in our simulations as sample size increases. The bootstrap confidence intervals still have undercoverage in our simulations for higher directions even in simulations when  $n = 1000$  across all distributions. This may be due to difficulty with accounting for variance due to additional constraints. The asymptotic confidence intervals for the canonical correlations have close to the desired coverage for the higher directions, but have undercoverage for the first two directions in our simulations, particularly when data are simulated from a multivariate Cauchy distribution. This is likely due to the same positive finite sample bias that causes severe undercoverage when  $n = 200$ .

For the transelliptical canonical directions using transformed Kendall's estimator, bootstrap and asymptotic confidence intervals are calculated for the loading of each variable in directions corresponding to non-zero canonical correlations. For each bootstrap replicate the estimates of both transelliptical canonical directions are flipped if necessary in order to minimize the sum of the angles between the estimated direction within the bootstrap replicate and the original sample. Table 3 reports the coverages for  $p = q = 8$  and data simulated from a multivariate normal distribution. Results for  $p = q = 4$  and  $16$  as well as data simulated from the multivariate Cauchy distribution are reported in the supplementary materials. Some of the issues with undercoverage, particularly with  $n = 200$  that are apparent in Table 3 are more severe for  $p = q = 16$  and less severe for  $p = q = 4$  in the tables presented in supplementary materials. The performance when data are generated from a multivariate Cauchy distribution is similar to the performance when data are generated from a multivariate normal distribution. The coverages are close to 95% for the first canonical direction, with overcoverage for the variable with a non-zero loading for the first direction. For both the bootstrap confidence intervals and asymptotic confidence intervals there is undercoverage for some loadings in the second, third, and fourth directions. This is likely due to the added complexity of additional constraints for higher order canonical directions, similar to what is seen in the bootstrap confidence intervals for the canonical correlations. We recommend interpreting any confidence intervals for higher order directions with caution. In finite samples it is difficult to fully quantify the uncertainty that arises as the number of constraints increases.

### 3.3. Testing procedures to identify non-zero canonical correlations

In addition to constructing confidence intervals for the non-zero canonical correlations and the associated directions, testing the null hypothesis that the true canonical correlation equals zero is also of interest. As noted in Section 2 we propose testing for a true canonical correlation of zero at the 0.05 significance level by inverting a 90% normal bootstrap confidence interval for the transelliptical canonical correlation, and rejecting the null hypothesis if the lower bound for the confidence interval is above zero. We use a 90% confidence interval because the alternative for this test is one sided. If the test for the  $i$ th transelliptical canonical correlation fails to reject the null hypothesis of a true transelliptical canonical correlation of zero then we will also fail to reject the null hypothesis for all higher order transelliptical canonical correlations. This procedure can be done iteratively, starting with the first transelliptical canonical correlation and moving on to higher order correlations, stopping when the test fails to reject the null hypothesis of a true correlation of zero.

We compare the type I error and power for the bootstrapped testing procedure using the transformed Kendall's estimator with a permutation test also using the transformed Kendall's estimator and the asymptotic Wilk's Lambda [48] from the R package CCP [32] using standard CCA based on the sample correlation matrix. In addition the bootstrap and permutation testing procedures using the sample correlation matrix estimator are presented. We consider  $p = q = 8$  and  $\Sigma_{XX} = \Sigma_{YY} = I$  where  $\Sigma_{XY}$  has either all zeros or a single non-zero entry ranging from 0.2 to 0.8 in increments of 0.2.

**Table 3**

Proportion of 1000 simulations in which the estimated confidence interval for the transelliptical canonical direction loading contains the true canonical direction loading. Calculated for bootstrap and asymptotic confidence intervals using the transformed Kendall's estimator for the first four canonical directions with data simulated from the multivariate normal distribution with  $p = q = 8$  and  $n = 200$  and 1000.

Variable	Bootstrap coverages				Asymptotic coverages			
	Dir 1	Dir 2	Dir 3	Dir 4	Dir 1	Dir 2	Dir 3	Dir 4
<b>n = 200</b>								
1	1.00	0.96	0.95	0.94	0.99	0.94	0.95	0.95
2	0.97	0.98	0.76	0.86	0.96	0.94	0.88	0.80
3	0.96	0.76	0.94	0.65	0.99	0.92	0.94	0.77
4	0.96	0.85	0.67	0.95	0.98	0.86	0.78	0.93
5	0.96	0.94	0.92	0.91	0.99	0.99	0.97	0.99
6	0.96	0.92	0.92	0.92	0.99	0.99	0.99	0.99
7	0.96	0.93	0.92	0.91	0.99	0.99	0.99	0.99
8	0.97	0.94	0.93	0.91	0.99	0.99	0.98	0.98
<b>n = 1000</b>								
1	1.00	0.94	0.96	0.94	1.00	0.99	1.00	1.00
2	0.94	1.00	0.92	0.94	0.94	0.99	0.95	0.93
3	0.94	0.92	0.99	0.86	0.96	0.95	0.98	0.86
4	0.95	0.94	0.85	0.99	0.95	0.94	0.87	0.98
5	0.94	0.94	0.94	0.96	0.98	0.98	0.99	0.99
6	0.96	0.95	0.95	0.96	0.99	1.00	0.99	0.99
7	0.94	0.94	0.95	0.94	0.99	1.00	0.99	0.99
8	0.96	0.95	0.95	0.96	0.99	0.99	0.99	0.99

This set up is employed for multivariate normal, multivariate Cauchy, multivariate  $t$  with five and ten degrees of freedom, and multivariate lognormal distributions for both  $n = 200$  and  $n = 1000$ . 1000 data sets are simulated for each setting. Table 4 gives the proportion of simulated data sets for which the null hypothesis that the first transelliptical canonical correlation is zero is rejected for each testing procedure. The bootstrap test using Kendall's transformed estimator controls for type I error, being conservative in all settings when  $n = 1000$ . For  $n = 200$  type I error is not controlled for using the bootstrap method when data are simulated from a multivariate Cauchy or multivariate  $t$  with five degrees of freedom, but is closer to the nominal level than asymptotic or permutation tests. The type I error not being controlled for heavy tailed elliptical distributions is likely due to the fact that the asymptotic distribution of the canonical correlations with a true value of zero is not normal and bootstrap procedures may not give the desired coverage, particularly for smaller sample sizes. However, given that for standard CCA the zero canonical correlations converge at rate  $n$  rather than  $\sqrt{n}$  we would expect overcoverage for the bootstrap confidence intervals for zero canonical correlations as sample size increases, which would result in the conservative type I errors we see for  $n = 1000$ . For the multivariate Cauchy distribution the transelliptical CCA bootstrap method is the only procedure that does not have a type I error of at least 0.77 for both sample sizes. The permutation based test only controls for type I error when data are simulated from a normal or lognormal distribution for both transelliptical CCA estimated using the transformed Kendall's estimator and standard CCA. This is because permutation based tests for CCA are only valid when zero correlation also implies independence, which is only true for the Gaussian copula. When the sample size is 1000 the type I error rate for the permutation test is even higher than for a sample size of 200. Also as expected the asymptotic Wilk's Lambda test only controls type I error when the data are simulated from a multivariate normal distribution. The type I error rate for this test is inflated even for the lognormal, because changes to the marginal distributions also affect this testing procedure. Power for the bootstrap method is generally comparable to the other testing methods, particularly as sample size increases.

#### 4. White matter tractography data and executive function in six year old children

We provide a comparison of transelliptical CCA estimated with the transformed Kendall's estimator and standard CCA estimated with the sample correlation matrix, using diffusion tensor imaging (DTI) and executive function (EF) data from six-year olds. The data come from an ongoing longitudinal study at the University of North Carolina investigating behavior and brain development from birth through adolescence [13,28,29]. The data include some sibling and twin pairs in addition to singletons. In our analysis, the data from one randomly selected child per family is used.

For DTI, we focus on 20 white matter tracts previously associated with cognitive function [15]. The 20 tracts included in the analysis can be found in Table 5. Imaging measures of diffusion rate and direction are available on these tracts including fractional anisotropy (FA), radial diffusivity (RD) and axial diffusivity (AD). We employ a single value for each tract, calculated by averaging measurements across all locations in the tract. Additional information on these measures and their interpretations can be found at Alexander et al. [3]. Results for RD are presented in the main text with those for FA and AD given in the supplementary material.

EF measures are an executive composite score from the Behavior Rating Inventory of Executive Function (BRIEF) [14], Cambridge Neuropsychological Test Automated Battery (CANTAB) Spatial Span (SSP), CANTAB Stockings of Cambridge

**Table 4**

Proportion of 1000 simulations in which the null hypothesis of a true first canonical correlation of zero is rejected. Calculated for the Wilk's Lambda testing procedure as well as bootstrap and permutation testing procedure for both transelliptical canonical correlation estimates using transformed Kendall's estimator and standard canonical correlation estimates. Simulated for five different elliptical and transelliptical distributions with  $p = q = 8$  and  $n = 200$  and  $1000$ .

		True correlation	0	0.2	0.4	0.6	0.8
<b>n = 200</b>							
Normal	Kendall Bootstrap	0.05	0.17	0.85	1	1	1
	Kendall Permutation	0.06	0.16	0.86	1	1	1
	Wilk's Lambda	0.05	0.17	0.80	1	1	1
	Standard Bootstrap	0.06	0.19	0.92	1	1	1
	Standard Permutation	0.05	0.17	0.90	1	1	1
Cauchy	Kendall Bootstrap	0.14	0.19	0.64	0.99	1	1
	Kendall Permutation	0.77	0.85	0.98	1	1	1
	Wilk's Lambda	1	1	1	1	1	1
	Standard Bootstrap	1	1	1	1	1	1
	Standard Permutation	1	1	1	1	1	1
Lognormal	Kendall Bootstrap	0.05	0.15	0.87	1	1	1
	Kendall Permutation	0.05	0.15	0.87	1	1	1
	Wilk's Lambda	0.10	0.17	0.43	0.88	1	1
	Standard Bootstrap	0.03	0.06	0.30	0.87	0.99	1
	Standard Permutation	0.05	0.08	0.30	0.83	1	1
t5	Kendall Bootstrap	0.08	0.16	0.78	1	1	1
	Kendall Permutation	0.17	0.33	0.90	1	1	1
	Wilk's Lambda	0.94	0.96	1	1	1	1
	Standard Bootstrap	0.30	0.40	0.84	1	1	1
	Standard Permutation	0.76	0.85	0.99	1	1	1
t10	Kendall Bootstrap	0.06	0.15	0.84	1	1	1
	Kendall Permutation	0.10	0.22	0.88	1	1	1
	Wilk's Lambda	0.43	0.65	0.96	1	1	1
	Standard Bootstrap	0.11	0.23	0.86	1	1	1
	Standard Permutation	0.28	0.45	0.96	1	1	1
<b>n = 1000</b>							
Normal	Kendall Bootstrap	0.02	0.85	1	1	1	1
	Kendall Permutation	0.06	0.94	1	1	1	1
	Wilk's Lambda	0.05	0.88	1	1	1	1
	Standard Bootstrap	0.02	0.91	1	1	1	1
	Standard Permutation	0.05	0.97	1	1	1	1
Cauchy	Kendall Bootstrap	0.02	0.45	1	1	1	1
	Kendall Permutation	0.80	1	1	1	1	1
	Wilk's Lambda	1	1	1	1	1	1
	Standard Bootstrap	1	1	1	1	1	1
	Standard Permutation	1	1	1	1	1	1
Lognormal	Kendall Bootstrap	0.02	0.84	1	1	1	1
	Kendall Permutation	0.06	0.92	1	1	1	1
	Wilk's Lambda	0.10	0.45	0.99	1	1	1
	Standard Bootstrap	0.00	0.09	0.92	0.99	1	1
	Standard Permutation	0.06	0.29	0.99	1	1	1
t5	Kendall Bootstrap	0.02	0.67	1	1	1	1
	Kendall Permutation	0.18	0.95	1	1	1	1
	Wilk's Lambda	0.98	1.00	1	1	1	1
	Standard Bootstrap	0.01	0.28	0.99	1	1	1
	Standard Permutation	0.94	1	1	1	1	1
t10	Kendall Bootstrap	0.01	0.79	1	1	1	1
	Kendall Permutation	0.12	0.97	1	1	1	1
	Wilk's Lambda	0.48	0.99	1	1	1	1
	Standard Bootstrap	0.01	0.70	1	1	1	1
	Standard Permutation	0.37	0.99	1	1	1	1

(SOC) [1], Stanford-Binet Verbal Fluid Reasoning (SB VFR), and Stanford-Binet Non-verbal Fluid Reasoning (SB NVFR) [39]. The BRIEF is a parent report measure whereas, CANTAB and Stanford-Binet are child assessments. For all EF variables except BRIEF a higher score indicates better EF, while for BRIEF a lower score indicates better EF. A total of 214 children have data for all EF measures plus all of the white matter tracts, and 216 have data for all EF measures plus all the bilateral tracts.

For each method p-values testing whether the true canonical correlation is zero are based on the bootstrap testing procedure using 1000 replicates. For both methods bootstrap confidence intervals for the direction loadings are reported using 1000 bootstrap replicates and the normal approximation bootstrap method. Confidence intervals based on a "plug-in" variance estimator for transelliptical CCA directions using the transformed Kendall's estimator are also reported.



**Table 5**

List of 20 white matter tracts and their abbreviations used in transelliptical and standard canonical correlation analysis estimating association between white matter and executive function tests in six-year-old children.

Tract name	Abbreviation
Arcuate fasciculus direct pathway left/right	ARC FT Left/Right
Arcuate fasciculus indirect anterior pathway left/right	ARC FP Left/Right
Arcuate fasciculus indirect posterior pathway left/right	ARC TP Left/Right
Anterior cingulum left/right	CGC Left/Right
Corticothalamic prefrontal projections left/right	CTPF Left/Right
Inferior fronto-occipital fasciculus left/right	IFOF Left/Right
Inferior longitudinal fasciculus left/right	ILF Left/Right
Superior longitudinal fasciculus left/right	SLF Left/Right
Uncinate Left/Right	UNC Left/Right
Splenium of the corpus callosum	Splenium
Genu of the corpus callosum	Genu

The marginal distributions for each of the variables to be included in the CCA analysis are tested for violations of normality which would indicate that transelliptical CCA may be more effective at summarizing the associations between the variables than standard CCA and that the transformed Kendall's estimator may be more efficient than the sample correlation estimator. Specifically all variables are tested for excess kurtosis using the Anscombe test [7], and skewness using the D'Agostino test [10]. The average RD values for a number of white matter tracts show excess kurtosis relative to a normal distribution including ARC FT Right, ARC FP Left, ARC TP Right, CTPF Left, CTPF Right, ILF Left, SLF Left, and Splenium. The ARC FP Left, ARC TP right, CTPF Left, CTPF Right, and Splenium also have positive skewness. In addition the SB VFR scores also have excess kurtosis and negative skewness, while the BRIEF scores show positive skewness.

Transelliptical CCA assumes the data come from a transelliptical distribution which can be tested using the methods from Jaser et al. [23]. This test is based on the equivalence between Kendall's tau and Blomqvist's beta for elliptical copulas. After testing for the equivalence between Blomqvist's beta and Kendall's tau for all pairs of variables and applying a false discovery rate (FDR) correction there were no significant differences between Blomqvist's beta and Kendall's tau at the 0.05 level. This suggests that deviations from the transelliptical assumption are relatively minor.

Table 6 gives the first canonical directions and correlations for both transelliptical CCA and standard CCA. The jackknife corrected estimate for the first transelliptical canonical correlation is 0.49, compared to 0.32 for standard CCA. In both cases, the first canonical correlation has  $p$ -value less than 0.05 using the bootstrap testing procedure. No other canonical correlations are significant. The DTI variable loadings are similar for the two methods, with the largest differences arising from tracts such as CTPF, ARC FT, ARC FP, ARC TP, and Splenium that show excess kurtosis or skewness. For all direction loadings the confidence intervals overlap between transelliptical CCA and standard CCA. The asymptotic confidence intervals for the transelliptical CCA direction loadings are narrower than the bootstrap confidence intervals for the DTI variables and similar to the bootstrap confidence intervals for the EF variables. When interpreting the direction loadings for RD values for the white matter tracts we note that lower RD is indicative of higher myelination, which would result in faster transmission of electrical impulses.

For all of the bilateral DTI tracts except the CTPF, SLF, and UNC the loading for the left hemisphere is larger than that for the right hemisphere. This is particularly noticeable in the ARC FT and IFOF tracts. Further analysis is done to examine the association between lateralization of RD among the bilateral tracts and EF tests. In order to do this we employ the lateralization measure from Niogi and McCandliss [35]. For the  $i$ th bilateral tract the lateralization measure,  $RD_{LATi}$ , is defined as

$$RD_{LATi} = \frac{RD_{Li} - RD_{Ri}}{(RD_{Li} + RD_{Ri})/2},$$

where  $RD_{Li}$  is the RD measure from the  $i$ th bilateral tract on the left hemisphere and  $RD_{Ri}$  is the RD measure from the tract on the right hemisphere. The lateralization measure for CTPF shows both excess kurtosis and positive skewness as measured by the Anscombe and D'Agostino tests. The test from Jaser et al. [23] is again used to test for potential violations of the transelliptical assumption, and again none of the pairwise tests show a significant difference between Kendall's tau and Blomqvist's beta after an FDR correction.

Table 7 reports the estimated first direction and correlation for transelliptical CCA using the transformed Kendall's estimator and standard CCA. The jackknife corrected estimate for the first canonical correlation using the transformed Kendall's estimator is 0.33, compared to 0.23 for standard CCA. In this case only CCA using the transformed Kendall's estimator has a  $p$ -value less than 0.05 for the first direction based on the bootstrap testing procedure. A higher lateralization measure for ARC FT and IFOF is correlated with higher SB VFR scores. A higher lateralization score means that RD is lower on the right hemisphere, indicating higher myelination for the right hemisphere tract. This gives evidence that for ARC FT and IFOF tracts greater development of the right hemisphere relative to the left hemisphere is associated with greater fluid reasoning. To the authors' knowledge, this is a novel finding.

This analysis illustrates the benefits of CCA based on the transformed Kendall's estimator. Even with moderate violations of normality as measured by kurtosis and skewness we uncover stronger associations than with standard CCA.

**Table 6**

Estimates for first canonical correlation direction loadings and correlations along with jackknife corrected correlation and bootstrap p-values for canonical correlation analysis between radial diffusivity measures for 20 white matter tracts and five executive function tests in six-year-old children. Estimated for transelliptical canonical correlation analysis using the transformed Kendall's estimator as well as standard canonical correlation analysis.

DTI vars	Transelliptical CCA loadings	Boot CI	Asymp CI	Standard CCA loadings	Boot CI
ARCFT Left	0.93	(0.30, 2.06)	(0.39, 1.47)	0.21	(−0.60, 1.09)
ARCFT Right	−0.83	(−2.53, 0.45)	(−1.87, 0.22)	0.48	(−0.38, 1.77)
ARCFP Left	−0.26	(−1.18, 0.47)	(−0.84, 0.32)	0.12	(−0.52, 0.86)
ARCFP Right	0.24	(−0.38, 0.98)	(−0.32, 0.81)	0.09	(−0.46, 0.67)
ARCTP Left	0.05	(−0.62, 0.79)	(−0.45, 0.56)	0.10	(−0.55, 0.85)
ARCTP Right	−0.05	(−1.06, 0.95)	(−0.76, 0.65)	−0.73	(−1.70, −0.27)
CGC Left	0.25	(−0.51, 1.22)	(−0.32, 0.82)	0.10	(−0.62, 0.83)
CGC Right	−0.02	(−0.84, 0.64)	(−0.50, 0.47)	−0.21	(−0.85, 0.30)
CTPF Left	0.27	(−0.13, 0.84)	(−0.12, 0.66)	0.11	(−0.27, 0.51)
CTPF Right	0.17	(−0.28, 0.83)	(−0.21, 0.56)	0.19	(−0.16, 0.69)
Genu	0.02	(−0.64, 0.68)	(−0.47, 0.51)	−0.01	(−0.56, 0.53)
ILF Left	−0.04	(−0.92, 0.80)	(−0.65, 0.57)	−0.02	(−0.75, 0.68)
ILF Right	0.26	(−0.45, 1.04)	(−0.23, 0.76)	−0.17	(−0.81, 0.35)
IFOF Left	0.89	(0.12, 2.10)	(0.19, 1.58)	0.75	(0.16, 1.84)
IFOF Right	−1.31	(−2.73, −0.52)	(−1.89, −0.73)	−1.10	(−2.27, −0.52)
SLF Left	−0.50	(−1.42, 0.26)	(−1.06, 0.07)	−0.05	(−0.73, 0.62)
SLF Right	0.06	(−0.64, 0.72)	(−0.47, 0.60)	−0.12	(−0.89, 0.54)
Splenium	−0.66	(−1.39, −0.27)	(−1.04, −0.28)	−0.65	(−1.39, −0.33)
UNC Left	−0.26	(−1.09, 0.48)	(−0.83, 0.32)	−0.02	(−0.75, 0.70)
UNC Right	0.64	(0.01, 1.54)	(0.14, 1.14)	0.68	(0.23, 1.50)
EF Vars					
SB V	1.02	(0.71, 1.77)	(0.86, 1.19)	0.87	(0.55, 1.57)
SB NV	−0.41	(−1.09, 0.22)	(−0.94, 0.11)	−0.50	(−1.30, 0.08)
Brief	−0.26	(−0.77, 0.10)	(−0.59, 0.07)	−0.54	(−1.18, −0.17)
SOC	−0.01	(−0.64, 0.58)	(−0.44, 0.41)	0.14	(−0.40, 0.73)
SSPSpan	0.17	(−0.55, 0.91)	(−0.24, 0.57)	−0.18	(−0.79, 0.31)
Cor	0.63			0.48	
Jackknife Cor	0.49			0.32	
Pval	4.80E−04			4.138E−03	
N	214			214	

**Table 7**

Estimates for first canonical correlation direction loadings and correlations along with jackknife corrected correlation and bootstrap p-values for canonical correlation analysis between radial diffusivity lateralization measures for seven bilateral white matter tracts and five executive function tests in six-year-old children. Estimated for transelliptical canonical correlation analysis using the transformed Kendall's estimator as well as standard canonical correlation analysis.

Lat vars	Transelliptical CCA loadings	Boot CI	Asymp CI	Standard CCA loadings	Boot CI
ARC FT	0.71	(0.17, 1.51)	(0.14, 1.29)	0.18	(−0.60, 1.09)
ARC FP	−0.36	(−0.92, 0.05)	(−0.81, 0.08)	−0.04	(−0.55, 0.47)
ARC TP	−0.13	(−0.74, 0.49)	(−0.65, 0.40)	0.25	(−0.28, 0.88)
CGC	0.29	(−0.04, 0.77)	(−0.03, 0.61)	0.40	(0.12, 0.88)
CTPF	0.05	(−0.35, 0.48)	(−0.33, 0.43)	0.07	(−0.35, 0.54)
ILF	−0.09	(−0.54, 0.36)	(−0.49, 0.32)	0.04	(−0.42, 0.55)
IFOF	0.59	(0.20, 1.25)	(0.17, 1.01)	0.65	(0.28, 1.34)
SLF	0.28	(−0.18, 0.87)	(−0.16, 0.71)	0.30	(−0.19, 1.04)
UNC	−0.38	(−0.91, 0.00)	(−0.76, 0.01)	−0.35	(−0.94, 0.10)
EF Vars					
SB V	0.82	(0.39, 1.58)	(0.43, 1.21)	0.80	(0.41, 1.50)
SB NV	−0.13	(−0.76, 0.59)	(−0.72, 0.45)	0.02	(−0.67, 0.80)
Brief	−0.28	(−0.75, 0.06)	(−0.60, 0.04)	−0.41	(−1.03, −0.02)
SOC	−0.41	(−1.14, 0.17)	(−0.93, 0.12)	−0.10	(−0.88, 0.53)
SSPSpan	0.43	(−0.05, 1.00)	(0.07, 0.80)	0.20	(−0.24, 0.73)
Cor	0.44			0.35	
Jackknife Cor	0.33			0.23	
Pval	0.03			0.11	
N	216			216	

## 5. Discussion

In this paper we define a version of CCA for transelliptical distributions using Kendall's tau. Consistent and asymptotically normal estimates of canonical directions and correlations can be obtained using a transformed Kendall's scatter matrix estimator. Simulation studies suggest that CCA estimates using the transformed Kendall's estimator perform well relative to other robust CCA methods in finite sample settings. The transformed Kendall's scatter matrix estimator can be used to consistently estimate transelliptical CCA directions and correlations for all transelliptical distributions, unlike other commonly used robust covariance matrix estimators such as MCD or S estimators which are only consistent for elliptical distributions. An R package implementing transelliptical CCA using the transformed Kendall's scatter matrix estimator as well as the data for the example in Section 4 is available at [github.com/blangworthy/transCCA](https://github.com/blangworthy/transCCA).

Confidence intervals for the canonical directions and nonzero correlations can be obtained using bootstrap methods or based on the asymptotic variances of the estimator. In simulation studies the bootstrap confidence intervals are superior for the canonical correlations when sample size is small, and the two methods perform similarly for the canonical directions. Confidence intervals for higher level canonical directions do not perform as well in terms of coverage as confidence intervals for the first canonical directions for both bootstrap and asymptotic based methods. This is likely due to difficulty in accounting for variability due to additional constraints for finite samples. For this reason we suggest interpreting confidence intervals for higher order directions with caution. In addition we propose a bootstrap procedure for testing if the true canonical correlation is equal to zero based on inverting bootstrap confidence intervals. This is necessary because both asymptotic and permutation based methods are not useful for identifying informative canonical directions if data are not generated from a distribution with a Gaussian copula. One area for future research is finding an improved testing procedure based on the asymptotic distribution for canonical correlations with true value zero using either the sample correlation or transformed Kendall's estimator when data are not generated from a multivariate normal distribution. These results have been found for standard CCA under a multivariate normality, but it is not straightforward to extend these results to transelliptical CCA.

CCA using the transformed Kendall's estimator shows promise for use in high dimensions or for more than two sets of variables. A number of formulations for sparse CCA have been proposed [49–51]. The methods from Yoon et al. [51] also use Kendall's tau to estimate the correlation matrix, however they only considered data generated from a Gaussian copula and neither establish the large sample properties of their estimators nor consider testing and confidence interval construction. Extensions of CCA to more than two sets of variables have been proposed which can be adapted to the transelliptical setting [26,34].

## CRedit authorship contribution statement

**Benjamin W. Langworthy:** Writing - original draft, Methodology, Formal analysis, Conceptualization. **Rebecca L. Stephens:** Writing - review & editing. **John H. Gilmore:** Writing - review & editing, Project administration, Funding acquisition. **Jason P. Fine:** Conceptualization, Writing - review & editing.

## Acknowledgments

This work was supported by the National Institute of Mental Health, United States of America under Grant NIMH R01 MH111944, Grant NIMH U01 MH070890, and Grant T32-MH106440. We would also like to thank the Editor, Associate Editor and referees.

## Appendix A. Proof of theorems

**Proof of Theorem 2.1.** Define  $\lambda_d$  to be the minimum eigenvalue of  $P_Z$ . By assumption  $\lambda_d > 0$ . Given that  $\hat{\tau}^{(Z_i, Z_j)} \rightarrow_p \tau^{(Z_i, Z_j)}$  for all  $1 \leq i, j \leq d$ , by continuous mapping theorem  $\hat{\rho}_{n,ij} \rightarrow_p \rho_{ij}$ . Define  $\delta_{n,ij} = \hat{\rho}_{n,ij} - \rho_{ij}$ . Then for every  $\gamma > 0$  there exists an  $N_\gamma$  such that for every  $n > N_\gamma$ ,

$$\Pr \left( \sum_{1 \leq i, j \leq d} |\delta_{n,ij}| \geq \frac{\lambda_d}{2} \right) \leq \gamma \quad (12)$$

Define  $\Delta_n = \hat{P}_{Zn} - P_Z$ , and  $\omega_{ni}$  to be the  $i$ th eigenvalue of  $\Delta_n$ . Because  $|\omega_{ni}| \leq \sum_{1 \leq i, j \leq n} |\delta_{n,ij}|$ , Eq. (12) implies that for  $1 \leq i \leq d$  and  $n > N_\gamma$

$$\Pr \left( |\omega_{ni}| \geq \frac{\lambda_d}{2} \right) \leq \gamma \quad (13)$$

Because  $\hat{P}_n = P_Z + \Delta_n$  Weyl's inequality is used to put a bound on  $\hat{\lambda}_{nd}$ . Specifically  $\hat{\lambda}_{nd} \geq \lambda_d + \omega_{nd}$ . Combining this with Eq. (13) the result  $\Pr(\hat{\lambda}_{nd} > 0) \rightarrow_p 1$  is obtained.

**Proof of Theorem 2.2.** This proof is a generalization of the work in Anderson [5] and uses similar algebraic steps. The true canonical directions,  $a_i$  and  $b_i$ , corresponding to the canonical correlation  $\lambda_i$  are solutions to the system of equations

$$\begin{pmatrix} -\lambda_i P_{XX} & P_{XY} \\ P_{YX} & -\lambda_i P_{YY} \end{pmatrix} \begin{pmatrix} a_i \\ b_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The  $r$  non-zero canonical correlations,  $\lambda_1 > \dots > \lambda_r > 0$ , are the non-zero values such that

$$\begin{vmatrix} -\lambda_i P_{XX} & P_{XY} \\ P_{YX} & -\lambda_i P_{YY} \end{vmatrix} = 0.$$

$A_r = (a_1, \dots, a_r)$  and  $B_r = (b_1, \dots, b_r)$  will be uniquely determined up to a change in sign. In order to uniquely define  $A_r$  and  $B_r$  it is possible to order the rows of  $X$  such that  $a_{ii} > 0$ . Define  $A_{r+} = (a_{r+1}, \dots, a_p)$  and  $B_{r+} = (b_{r+1}, \dots, b_q)$  to be solutions such that for  $A = (A_r, A_{r+})$  and  $B = (B_r, B_{r+})$  the equalities  $A^\top P_{XX} A = I_p$ ,  $B^\top P_{YY} B = I_q$ , and  $A^\top P_{XY} B = \begin{pmatrix} A_r & 0 \\ 0 & 0 \end{pmatrix}$  hold. The solutions for  $A_{r+}$  and  $B_{r+}$  are only unique up to multiplication by an orthogonal matrix on the right hand side, but the following results will hold for any unique value of  $A_{r+}$  and  $B_{r+}$  obtained by imposing suitable constraints. For simplicity the subscript  $n$  will be dropped from the notation for the estimates of canonical correlations and directions. The estimates of  $a_i$  and  $b_i$ ,  $a_i^*$  and  $b_i^*$ , are solutions to the system of equations

$$\begin{pmatrix} -\lambda_i^* P_{XX}^* & P_{XY}^* \\ P_{YX}^* & \lambda_i^* P_{YY}^* \end{pmatrix} \begin{pmatrix} a_i^* \\ b_i^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (14)$$

The estimate of  $\lambda_i$ ,  $\lambda_i^*$ , is the  $i$ th solution to

$$\begin{vmatrix} -\lambda_i^* P_{XX}^* & P_{XY}^* \\ P_{YX}^* & -\lambda_i^* P_{YY}^* \end{vmatrix} = 0. \quad (15)$$

Recall the transformations  $P_{UU}^* = A^\top P_{XX}^* A$ ,  $P_{VV}^* = B^\top P_{YY}^* B$ ,  $P_{UV}^* = A^\top P_{XY}^* B$ , and  $P_{VU}^* = P_{UV}^{*T}$ , with limiting distribution defined in Eq. (4). It is straightforward to show that the limiting distribution in Eq. (4) holds by the delta method. Further recall  $G$  and  $H$  which are defined by the system of equations in Eqs. (1)–(3). Because the determinant of the product of two square matrices equals product of the determinants, if  $\lambda_i^*$  is a solution to (2) it is also a solution to (15). The solutions to (14) and (1) are related through the identities  $A g_i = a_i^*$  and  $B h_i = b_i^*$ . Eq. (1) implies

$$P_{UV}^* H = P_{UU}^* G \Lambda_r^*, \quad P_{VU}^* G = P_{VV}^* H \Lambda_r^*. \quad (16)$$

Define the matrix  $I_{r,m_1 m_2}$  to be the  $m_1 \times m_2$  matrix with  $m_1, m_2 \geq r$  and the upper left hand corner equal to  $I_r$  and all other entries equal to zero. Based on (4)  $P_{UU} \rightarrow_p I_p$ ,  $P_{VV} \rightarrow_p I_q$ , and  $P_{UV} \rightarrow_p \Lambda_{r,pq}$ . It follows that  $G \rightarrow_p I_{r,pr}$  and  $H \rightarrow_p I_{r,qr}$ . Because  $G, H$ , and  $\Lambda_r^*$  are single valued functions of  $P_{UU}^*, P_{VV}^*,$  and  $P_{UV}^*$  that are differentiable in the neighborhood of  $I_{r,pr}$  and  $I_{r,qr}$ , by the Delta method if we define  $\tilde{G} = \sqrt{n}(G - I_{r,pr})$ ,  $\tilde{H} = \sqrt{n}(H - I_{r,qr})$ , and  $\tilde{\Lambda}_r = \sqrt{n}(\Lambda_r^* - \Lambda_r)$ , then  $\text{vec}(\tilde{G})$ ,  $\text{vec}(\tilde{H})$ , and  $\text{vec}(\tilde{\Lambda}_r)$  all have a normal limiting distribution with mean zero and finite variances. Note the following equalities obtained by expanding and rearranging the terms of (16),

$$\sqrt{n}\{(P_{UV}^* I_{r,qr} - \Lambda_{r,pr}) - (P_{UU}^* - I_p) \Lambda_{r,pr}\} = \tilde{G} \Lambda_r + \tilde{\Lambda}_{r,pr} - \Lambda_{r,pq} \tilde{H} + o_p(1), \quad (17)$$

$$\sqrt{n}\{(P_{VU}^* I_{r,pr} - \Lambda_{r,qr}) - (P_{VV}^* - I_q) \Lambda_{r,qr}\} = \tilde{H} \Lambda_r + \tilde{\Lambda}_{r,qr} - \Lambda_{r,qp} \tilde{G} + o_p(1). \quad (18)$$

Multiplying (17) by  $\Lambda_r$  on the right hand side and (18) by  $\Lambda_{r,pq}$  on the left hand side and taking the sum, and then multiplying (17) by  $\Lambda_{r,qp}$  on the left hand side and (18) by  $\Lambda_r$  on the right hand side and taking the sum results in the following two equalities:

$$\begin{aligned} & \sqrt{n}\{(P_{UV}^* I_{r,qr} - \Lambda_{r,pr}) \Lambda_r + \Lambda_{r,pq} (P_{VU}^* I_{r,pr} - \Lambda_{r,qr}) - \Lambda_{r,pq} (P_{VV}^* - I_q) \Lambda_{r,qr} - (P_{UU}^* - I_p) \Lambda_{r,pr} \Lambda_r\} \\ & = 2 \Lambda_{r,pr} \tilde{\Lambda}_r + \tilde{G} \Lambda_{r,pr}^2 - \Lambda_{r,qp}^2 \tilde{G} + o_p(1), \end{aligned} \quad (19)$$

$$\begin{aligned} & \sqrt{n}\{(P_{VU}^* I_{r,pr} - \Lambda_{r,qr}) \Lambda_r + \Lambda_{r,qp} (P_{UV}^* I_{r,qr} - \Lambda_{r,pr}) - \Lambda_{r,qp} (P_{UU}^* - I_p) \Lambda_{r,pr} - (P_{VV}^* - I_q) \Lambda_{r,qr} \Lambda_r\} \\ & = 2 \Lambda_{r,qr} \tilde{\Lambda}_r + \tilde{H} \Lambda_{r,qr}^2 - \Lambda_{r,qq}^2 \tilde{H} + o_p(1). \end{aligned} \quad (20)$$

The  $i$ th diagonal term where  $1 \leq i \leq r$  of the right hand side of both (19) and (20) will be equal to  $2\lambda_i \tilde{\lambda}_i + o_p(1)$ , which can be used to find Eq. (5). The  $i$ th row and  $j$ th column where  $1 \leq i \leq p$ ,  $1 \leq j \leq r$  and  $j \neq i$  of the right hand side of (19) is equal to  $\tilde{g}_{ij}(\lambda_j^2 - \lambda_i^2)$ , while the  $i$ th row and  $j$ th column where  $1 \leq i \leq q$ ,  $1 \leq j \leq r$  and  $i \neq j$  from the right hand side of (20) is equal to  $\tilde{h}_{ij}(\lambda_j^2 - \lambda_i^2)$ , where  $\tilde{g}_{ij}$  and  $\tilde{h}_{ij}$  are the entry from the  $i$ th row and  $j$ th column of  $\tilde{G}$  and  $\tilde{H}$  respectively. These can be used to show Eqs. (8) and (9). In order to solve for the variances and covariances of  $\tilde{g}_{ii}$  and  $\tilde{h}_{ii}$  the following equalities are obtained by substituting  $\tilde{G}$  and  $\tilde{H}$  into (3)

$$\sqrt{n}(I_{r,rp} P_{UU}^* I_{r,pr} - I_r) = -(\tilde{G}^\top I_{r,pr} + I_{r,rp} \tilde{G}) + o_p(1), \quad (21)$$

$$\sqrt{n}(I_{r,rq} P_{VV}^* I_{r,qr} - I_r) = -(\tilde{H}^\top I_{r,qr} + I_{r,rq} \tilde{H}) + o_p(1). \quad (22)$$

The  $i$ th diagonal term for  $1 \leq i \leq r$  of the right hand side of (21) is  $-2\tilde{g}_{ii} + o_p(1)$  and the  $i$ th diagonal term for  $1 \leq i \leq r$  of the right hand side of (22) is  $-2\tilde{h}_{ii} + o_p(1)$ . The variances and covariances for each term on the left hand side of (19), (20), (21), and (22) can be solved for using (4) and can be used to show Eqs. (6) and (7). Given the variances of  $\tilde{G}$  and  $\tilde{H}$  the asymptotic variances of  $\sqrt{n}(a_j^* - a_j)$  and  $\sqrt{n}(b_j^* - b_j)$  for  $1 \leq j \leq r$  are solved for using the following equalities,

$$\sqrt{n}(a_j^* - a_j) = \sqrt{n}A(g_j - \iota_{pj}) = \sum_{i=1}^p a_i \tilde{g}_{ij} + o_p(1), \quad (23)$$

$$\sqrt{n}(b_j^* - b_j) = \sqrt{n}B(h_j - \iota_{qj}) = \sum_{i=1}^q b_i \tilde{h}_{ij} + o_p(1), \quad (24)$$

where  $\iota_{pj}$  is a  $p \times 1$  vector where the  $j$ th element is one and the rest are zero. Eqs. (23) and (24) can be used to show Eqs. (10) and (11). The limiting covariances on the left hand side can be found by solving for the covariances of the right hand side. Consistent estimates for the covariance matrix for  $\sqrt{n}(A_r^* - A_r)$ ,  $\sqrt{n}(B_r^* - B_r)$ , and  $\sqrt{n}(\Lambda_r^* - \Lambda_r)$  can be obtained by plugging in consistent estimates of  $\Theta$ ,  $A$ ,  $B$ , and  $\Lambda$ .

## Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jmva.2020.104715>.

## References

- [1] CANTAB, Cantab cognitive assessment software, 2017, [www.cantab.com](http://www.cantab.com).
- [2] B. Abdous, C. Genest, B. Rémillard, Dependence properties of meta-elliptical distributions, in: *Statistical Modeling and Analysis for Complex Data Problems*, Springer, 2005, pp. 1–15.
- [3] A.L. Alexander, J.E. Lee, M. Lazar, A.S. Field, Diffusion tensor imaging of the brain, *Neurotherapeutics* 4 (3) (2007) 316–329.
- [4] A. Alfons, C. Croux, P. Filzmoser, Robust maximum association estimators, *J. Amer. Statist. Assoc.* 112 (517) (2017) 436–445.
- [5] T.W. Anderson, Asymptotic theory for canonical correlation analysis, *J. Multivariate Anal.* 70 (1) (1999) 1–29.
- [6] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, third ed., Wiley, 2003. New York.
- [7] F.J. Anscombe, W.J. Glynn, Distribution of the kurtosis statistic  $b_2$  for normal samples, *Biometrika* 70 (1) (1983) 227–234.
- [8] J.A. Branco, C. Croux, P. Filzmoser, M.R. Oliveira, Robust canonical correlations: A comparative study, *Comput. Statist.* 20 (2) (2005) 203–229.
- [9] S. Cambanis, S. Huang, G. Simons, On the theory of elliptically contoured distributions, *J. Multivariate Anal.* 11 (3) (1981) 368–385.
- [10] R.B. D'Agostino, Transformation to normality of the null distribution of  $g_1$ , *Biometrika* (1970) 679–681.
- [11] P. Embrechts, A. McNeil, D. Straumann, Correlation and dependence in risk management: properties and pitfalls, in: *Risk Management: Value at Risk and Beyond*, 2002, pp. 176–223.
- [12] H.-B. Fang, K.-T. Fang, S. Kotz, The meta-elliptical distributions with given marginals, *J. multivariate Anal.* 82 (1) (2002) 1–16.
- [13] J.H. Gilmore, J.E. Schmitt, R.C. Knickmeyer, J.K. Smith, W. Lin, M. Styner, G. Gerig, M.C. Neale, Genetic and environmental contributions to neonatal brain structure: a twin study, *Human Brain Mapp.* 31 (8) (2010) 1174–1182.
- [14] G.A. Gioia, P.K. Isquith, S.C. Guy, L. Kenworthy, Test review behavior rating inventory of executive function, *Child Neuropsychol.* 6 (3) (2000) 235–238.
- [15] J.B. Girault, E. Cornea, B.D. Goldman, R.C. Knickmeyer, M. Styner, J.H. Gilmore, White matter microstructural development and cognitive ability in the first 2 years of life, *Human Brain Mapp.* 40 (4) (2019) 1195–1210.
- [16] I. González, S. Déjean, P. Martin, A. Baccini, CCA: An R package to extend canonical correlation analysis, *J. Stat. Softw.* 23 (12) (2008) 1–14.
- [17] F. Han, H. Liu, Optimal rates of convergence for latent generalized correlation matrix estimation in transelliptical distribution, 2013, arXiv preprint arXiv:1305.6916v3.
- [18] F. Han, H. Liu, Scale-invariant sparse PCA on high-dimensional meta-elliptical data, *J. Amer. Statist. Assoc.* 109 (505) (2014) 275–287.
- [19] F. Han, H. Liu, Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution, *Bernoulli* 23 (1) (2017) 23.
- [20] W. Hoeffding, The strong law of large numbers for  $u$ -statistics, Technical Report, North Carolina State University. Dept. of Statistics, 1961.
- [21] H. Hotelling, Relations between two sets of variates, *Biometrika* 28 (3/4) (1936) 321–377.
- [22] H. Hult, F. Lindskog, Multivariate extremes, aggregation and dependence in elliptical distributions, *Adv. Appl. Probab.* (2002) 587–608.
- [23] M. Jaser, S. Haug, A. Min, A simple non-parametric goodness-of-fit test for elliptical copulas, *Depend. Model.* 5 (1) (2017) 330–353.
- [24] D. Kelker, Distribution theory of spherical distributions and a location-scale parameter generalization, *Sankhyā* (1970) 419–430.
- [25] M.G. Kendall, A new measure of rank correlation, *Biometrika* 30 (1/2) (1938) 81–93.
- [26] J.R. Kettenring, Canonical analysis of several sets of variables, *Biometrika* 58 (3) (1971) 433–451.
- [27] C. Klüppelberg, G. Kuhn, Copula structure analysis, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 71 (3) (2009) 737–753.
- [28] R.C. Knickmeyer, S. Gouttard, C. Kang, D. Evans, K. Wilber, J.K. Smith, R.M. Hamer, W. Lin, G. Gerig, J.H. Gilmore, A structural MRI study of human brain development from birth to 2 years, *J. Neurosci.* 28 (47) (2008) 12176–12182.
- [29] R.C. Knickmeyer, K. Xia, Z. Lu, M. Ahn, S.C. Jha, F. Zou, H. Zhu, M. Styner, J.H. Gilmore, Impact of demographic and obstetric factors on infant brain volumes: a population neuroscience study, *Cerebral Cortex* 27 (12) (2016) 5616–5625.
- [30] H. Liu, F. Han, C. Zhang, Transelliptical graphical models, in: *Proceedings in the 25th International Conference on Neural Information Processing Systems*, (2012), 2012, pp. 800–808.
- [31] H.P. Lopuhaa, On the relation between  $s$ -estimators and  $m$ -estimators of multivariate location and covariance, *Ann. Statist.* (1989) 1662–1683.
- [32] U. Menzel, Ccp r package, 2009, <https://CRAN.R-project.org/package=CCP>.
- [33] R. Muirhead, C. Waternaux, Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika* 67 (1) (1980) 31–43.
- [34] A.A. Nielsen, Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data, *IEEE Trans. Image Process.* 11 (3) (2002) 293–305.
- [35] S. Niogi, B. McCandless, Left lateralized white matter microstructure accounts for individual differences in reading ability and disability, *Neuropsychologia* 44 (11) (2006) 2178–2188.



- [36] J. Owen, R. Rabinovitch, On the class of elliptical distributions and their applications to the theory of portfolio choice, *J. Finance* 38 (3) (1983) 745–752.
- [37] K. Pillai, On the distribution of the largest or the smallest root of a matrix in multivariate analysis, *Biometrika* 43 (1/2) (1956) 122–127.
- [38] C.R. Rao, C.R. Rao, M. Statistiker, C.R. Rao, C.R. Rao, *Linear Statistical Inference and its Applications*, second ed., Wiley, New York, 1973.
- [39] G.H. Roid, *Stanford-Binet intelligence scales*, Riverside Pub., 2003.
- [40] P.J. Rousseeuw, Least median of squares regression, *J. Amer. Statist. Assoc.* 79 (388) (1984) 871–880.
- [41] P.J. Rousseeuw, G. Molenberghs, Transformation of non positive semidefinite correlation matrices, *Commun. Statist. Theory Methods* 22 (4) (1993) 965–984.
- [42] F. Rublík, Estimates of the covariance matrix of vectors of u-statistics and confidence regions for vectors of Kendall's tau, *Kybernetika* 52 (2) (2016) 280–293.
- [43] F. Schmid, R. Schmidt, Nonparametric inference on multivariate versions of blomqvist's beta and related measures of tail dependence, *Metrika* 66 (3) (2007) 323–354.
- [44] S. Taskinen, C. Croux, A. Kankainen, E. Ollila, H. Oja, Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices, *J. Multivariate Anal.* 97 (2) (2006) 359–384.
- [45] V. Todorov, P. Filzmoser, An object-oriented framework for robust multivariate analysis, *J. Stat. Softw.* 32 (3) (2009) 1–47.
- [46] D.E. Tyler, A distribution-free M-estimator of multivariate scatter, *Ann. Statist.* (1987) 234–251.
- [47] S. Visuri, E. Ollila, V. Koivunen, J. Möttönen, H. Oja, Affine equivariant multivariate rank methods, *J. Statist. Plann. Inference* 114 (1–2) (2003) 161–185.
- [48] S. Wilks, On the independence of k sets of normally distributed statistical variables, *Econometrica* (1935) 309–326.
- [49] I. Wilms, C. Croux, Sparse canonical correlation analysis from a predictive point of view, *Biom. J.* 57 (5) (2015) 834–851.
- [50] D. Witten, R. Tibshirani, Extensions of sparse canonical correlation analysis with applications to genomic data, *Stat. Appl. Genet. Mol. Biol.* 8 (1) (2009) 1–27.
- [51] G. Yoon, R.J. Carroll, I. Gaynanova, Sparse semiparametric canonical correlation analysis for data of mixed types, *Biometrika* In press (2019).