EXTENSIONS TO CANONICAL CORRELATION ANALYSIS AND PRINCIPAL COMPONENTS ANALYSIS WITH APPLICATIONS TO SURVIVAL AND BRAIN IMAGING DATA

Benjamin W. Langworthy

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2020

Approved by:

Jianwen Cai

Michael R. Kosorok

John H. Gilmore

J. S. Marron

Fei Zou

ABSTRACT

Benjamin W. Langworthy: Extensions to Canonical Correlation Analysis and Principal Components Analysis with Applications to Survival and Brain Imaging Data
(Under the direction of Jianwen Cai and Michael R. Kosorok)

Canonical correlation analysis (CCA) is a method for finding a low dimension representation of the linear associations between two sets of variables. Likewise principal components analysis (PCA) is a tool for finding a low dimensional representation of a single set of variables. The solution to CCA is an eigendecomposition involving the joint covariance or correlation matrix of both sets of variables and the solution to PCA is an eigendecomposition involving the covariance or correlation matrix of the single set of variables. We extend CCA and PCA using robust or non-standard estimators of the covariance or correlation matrix.

First we extend CCA using a robust correlation estimator based on transformations of Kendall's tau rank correlation coefficient. We show that the CCA estimates using this robust correlation estimator are consistent and asymptotically normal. We also define a bootstrap based testing procedure for identifying informative canonical directions. Simulations show that this robust estimator performs better than standard CCA for data from heavy tailed and skewed distributions. We apply this method to brain white matter structure data from diffusion tensor imaging (DTI) and executive function (EF) test scores in six-year-old children to show that lateralization of white matter brain structure is correlated with higher EF scores. Next we define PCA for the multivariate survival setting where failure time data can be right censored. We estimate the covariance and correlation matrices of the counting processes defined by the failure times and their associated martingales. We use eigendecomposition of these covariance and correlation matrix estimates to obtain principal direction estimates. These estimates are consistent and asymptotically normal. We apply this method to a data set from a clinical trial for

patients with pancreatic cancer and are able to define medically relevant groupings of adverse events. Finally we extend robust CCA to the multi-set and high dimensional setting in which there are more than two sets of variables and one or more of the sets is high-dimensional. We use the same robust correlation estimate using transformations of Kendall's tau. We also use cross-validation for dimension reduction and testing procedures. Unlike existing methods this cross-validation testing procedure is valid when data come from a heavy tailed elliptical distribution. We extend our analysis of DTI and EF data to include brain gray matter volume data from 88 different brain regions to further investigate the association between brain structure and EF test scores.

To Phoebe.

ACKNOWLEDGEMENTS

This work was supported by the National Institute of Mental Health under Grant NIMH R01 MH111944, Grant NIMH U01 MH070890, and Grant T32-MH106440. In addition to my adviser, Jianwen Cai, and my committee members, I would like to thank my collaborators on these projects including: John Gilmore, Rebecca Stephens, Jason Fine, Michael Kosorok, and Robert Corty. I would also like to thank the members of the Early Brain Development Study project here at UNC who have been integral to my development as a researcher. I would also like to thank the staff, students and faculty at the Department of Biostatistics who have helped me in so many ways during my time at UNC. There are too many names to mention here, but know that I have appreciated all that I have learned here at UNC. Finally I would like to thank my family, in particular my parents and siblings. Your support throughout my life has helped me to get to where I am.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	Axial Diffusivity
ARC FP	Arcuate fasciculus indirect anterior pathway
ARC FT	Arcuate fasciculus direct pathway
ARC TP	Arcuate fasciculus indirect posterior pathway
BASC	Behavioral Assessment System for Children
BRIEF	Behavior Ratinv Inventory of Executive Function
CANTAB	Cambridge Neuropsychological Test Automated Battery
CCA	Canonical Correlation Analysis
CGC	Anterior cingulum
CIF	Cumulative Incidence Function
CTPF	Corticothalamic prefrontal projections
CV	Cross Validation
DTI	Diffusion Tensor Imaging
EF	Executve Function
FA	Fractal Anisotropy
GM	Grey Matter
IFOF	Inferior fronto-occipital fasciculus
ILF	Inferior longitudinal fasciculus
LOESS	Locally Estimated Scatterplot Smoothing
MCD	Minimum Covariance Determinant
mCCA	Multiset Canonical Correlation Analysis
PCA	Principal Components Analysis
PH	Proportional Hazards
RD	Radial Diffusivity
SB NV FR	Stanford-Binet Fluid Reasoning Non-verbal
SB V FR	Stanford-Binet Fluid Reasoning Verbal

SLF	Superior longitudinal fasciculus
SOC	Stockings of Cambridge
SSP	Spatial Span
SVD	Singular Value Decomposition
UNC	Uncinate

CHAPTER 1: INTRODUCTION

Many medical or public health data sets have a large number of variables that are collected across the same subjects. Methods such as canonical correlation analysis (CCA) and principal components analysis (PCA) are useful for summarizing entire data sets into a small number of linear combinations of the data. These low dimension representations of the data can make it easier for medical and scientific researchers to understand the structure of complex data sets which can lead to new scientific hypotheses or discoveries.

Standard CCA is estimated with an eigendecomposition involving the joint sample covariance or correlation matrix of two sets of variables. This can be shown to work well when the data come from a multivariate normal distribution. However when data have extreme outliers, or come from heavy tailed or skewed distributions the sample covariance or correlation matrix may not be suitable for estimating CCA. We propose using a robust correlation estimator using properties of Kendall's tau rank correlation that performs better than the sample correlation matrix for data coming from heavy-tailed distributions. This estimator also outperforms other robust correlation estimators, particularly when data come from a skewed distribution. A number of methods have been proposed to identify informative CCA directions that explain true correlation between the two sets of variables and not just random noise. All of these methods can be shown to fail when assumptions related to multivariate normality are not met. For this reason we propose a bootstrap based method for identifying informative CCA directions that makes minimal distributional assumptions. Our robust CCA method is studied in a data set with brain imaging data and executive function (EF) test scores for six-year-old children where many of the variables have more heavy tailed and skewed distributions than would be expected if the data came from a multivariate normal distribution.

Next we consider the multivariate survival setting, where we observe time to event data for many different types of adverse events across the same subjects. The time to event for all of these events may be right censored. In the presence of censoring the covariance between the failure times cannot be estimated without strong parametric assumptions. We instead focus on the counting processes or martingales defined by the failure times. The full covariance or correlation matrices of these counting processes and martingales can be estimated non-parametrically. We use the estimates of the counting process and martingale covariance or correlation matrices to estimate the corresponding principal directions. We also show how this can be extended to the semi-competing risk setting where each of the different event types is subject to a competing risk such as death. We apply our martingale PCA method to data from a clinical trial for patients with pancreatic cancer and are able to define medically relevant groupings of adverse events.

Finally we extend CCA to the multi-set and high-dimensional setting. In this case there are more than two sets of variables, and we use multi-set CCA (mCCA) to get a low dimensional representation of the linear correlation between all the sets of variables. As with standard CCA, previous methods for mCCA work best when data come from a multivariate normal distribution. We combine previously developed methods for mCCA in the high-dimensional setting with our robust correlation techniques to define a robust version of mCCA. We propose a method based on cross-validation for reducing the dimension of the data and identifying informative directions for multi-set CCA. This allows us to include additional brain volume measures across 88 different regions in our analysis of brain imaging and EF data for six-year-old children, and provide added insight for the connection between brain structure and EF ability in children.

The remainder of this document is organized as follows. Chapter 2 gives a review of the relevant literature, chapter 3 presents our robust version of CCA, chapter 4 defines and investigates PCA for survival data, and chapter 5 extends robust CCA to the multi-set and high-dimensional setting, chapter 6 gives concluding remarks and potential areas of future research, and is followed by technical details and references.

CHAPTER 2: LITERATURE REVIEW

Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) first introduced by Hotelling (1936) is a useful technique for summarizing the linear association between two sets of variables. CCA finds the linear combinations of the two sets of variables that are maximally correlated. Subsequent canonical directions and correlations are found in the same manner subject to the constraint that they are uncorrelated with all previous directions. The solution to the canonical correlations and directions are based on an eigendecomposition involving the joint covariance matrix of the two sets of variables. This is a powerful data reduction technique that allows researchers to look at a much smaller set of correlations than all possible pairwise correlations.

Robust CCA

Standard CCA estimates are based on an eigendecomposition of the sample covariance or correlation matrix. In cases where the sample covariance matrix is inefficient or inconsistent, robust estimates of the covariance or correlation matrix can be used as a substitute. Examples of these are the minimum covariance determinant (MCD) (Rousseeuw, 1984) and Tyler's M-estimator (Tyler, 1987). Taskinen et al. (2006) and Visuri et al. (2003) study the asymptotic properties of CCA using robust covariance estimators. Branco et al. (2005) uses simulation studies to investigate the finite sample performance of robust CCA methods including robust estimation of the covariance matrix, a robust alternating regression technique and a projection pursuit approach. Alfons et al. (2017) proposes a projection pursuit approach that directly searches for the directions that maximize a robust correlation measure such as Kendall's tau or Spearman's correlation. Wilms and Croux (2016) proposed a robust alternating regression approach to CCA that also works for high-dimensional data. One issue with many of these

estimators is that they primarily focus on data from heavy tailed distributions and do not work well for data from skewed distributions.

Kendall's tau

One alternative to Pearson's correlation is Kendall's tau (Kendall, 1938) which uses the ranks of the data within a sample. For two random variables, $X^{(1)}$ and $X^{(2)}$, Kendall's tau is defined as

$$\tau^{(X^{(1)},X^{(2)})} = E[\text{sign}(X^{(1)} - \tilde{X}^{(1)})(X^{(2)} - \tilde{X}^{(2)})],$$

where $[\tilde{X}^{(1)}, \tilde{X}^{(2)}]$ is an identically distributed copy of $[X^{(1)}, X^{(2)}]$. If $[x_1^{(1)}, x_1^{(2)}]^T, \ldots,$ $[x_n^{(1)}, x_n^{(2)}]^T$ are iid realizations of $[X_1, X_2]^T$ then Kendall's tau can be consistently estimated by

$$\hat{\tau}_n(X^{(1)}, X^{(2)}) = \frac{1}{\binom{n}{2}} \sum_{1 \le k < l \le n} \sum \text{sign}(x_k^{(1)} - x_l^{(1)})\text{sign}(x_k^{(2)} - z_l^{(2)})$$

Consistency and asymptotically normality follow from U-statistic theoryHoeffding (1961), which makes minimal distributional assumptions on $[X^{(1)}, X^{(2)}]^T$.

Elliptical and transelliptical families of distributions

Any $p \times 1$ dimensional random vector, Z, is elliptically distributed if it has a characteristic function $\Phi_{Z-\mu}(t) = \psi(t^T \Sigma t)$ where μ is a p-dimensional vector, Σ is a $p \times p$ positive semi-definite matrix and ψ is a function from $[0, \infty) \to \mathbb{R}$. This family of distributions includes a number of well known distributions including the multivariate normal and multivariate t distributions. There are a number of theoretical results that can be extended to the entire family of elliptical distributions. One such result is the relationship between the population Kendall's tau, τ, and Pearson correlation, ρ, for any bivariate elliptical distribution where Pearson correlation exists. Lindskog et al. (2003) shows the following equivalence holds for all such distributions:

$$\tau = \frac{2}{\pi} \arcsin(\rho).$$

This relationship is also useful for the transelliptical family of distributions, defined by Liu et al. (2012), which includes any multivarate distribution created by transforming an elliptical distribution with a set of univariate monotonic transformations. This is equivalent to the family of multivariate distributions with an elliptical copula (Embrechts et al., 2002; Klüppelberg and Kuhn, 2009). Because Kendall's tau is invariant to monotone increasing marginal transformations of the data it is very useful for transelliptical distributions. Recent studies have investigated using transformations of Kendall's tau to estimate the correlation matrix for elliptical and transelliptical distributions (Han and Liu, 2012). This estimator has been shown to work well in high dimensions (Han and Liu, 2013, 2017).

Distribution of canonical correlations and directions

Anderson (2003) shows that CCA estimates corresponding to unique non-zero canonical correlations are asymptotically normal when using the sample covaraince matrix and the data come from multivariate normal distribution. Taskinen et al. (2006) extends this result to elliptical distributions and covariance or correlation estimates that are positive definite functions of the data.

High dimensional Canonical Correlation Analysis

For many modern data sets the number of variables can be larger than the number of observations. Standard CCA estimation techniques do not work in this setting. A number of methods have been proposed for this purpose. One approach for high dimensional CCA is to add a penalty term to the canonical directions (Suo et al., 2017; Witten et al., 2009; Parkhomenko et al., 2007; Waaijenborg et al., 2008; Wilms and Croux, 2016). Vinod (1976) proposes applying a ridge penalty directly to the estimates of the covariance matrix, which can then be decomposed to get the CCA estimates.

Multiset Canonical Correlation Analysis

If instead of two sets of variables, we are interested in the connections between three or more sets of variables there are a number of ways to extend CCA. Nielsen (2002) and Kettenring

(1971) give an overview of the different ways to extend CCA into multiple sets of variables. In this case we have m sets of variables, $X_1, \ldots, X_m$, where X_i is an $d_i \times 1$ random vector. For the ith set of variables we can define the jth canonical variate as $U_i^{(j)} = a_i^{(j)T}\mathbf{X_i}$. Kettenring (1971) presents five different ways to think of maximizing the correlation between the canonical variates

1. **SUMCOR**: Maximize $\sum_{i=1}^{m}\sum_{k=1}^{m} \text{Cov}(U_i^{(j)}, U_k^{(j)})$
2. **MAXVAR**: Maximize the largest eigenvalue for the covariance matrix of $[U_1^{(j)}, \ldots, U_m^{(j)}]^T$
3. **MINVAR**: Minimize the smallest eigenvalue for the covariance matrix of $[U_1^{(j)}, \ldots, U_m^{(j)}]^T$
4. **SSQCOR**: Maximize $\sum_{i=1}^{m}\sum_{k=1}^{m} \text{Cov}(U_i^{(j)}, U_k^{(j)})^2$
5. **GENVAR**: Minimize the determinant of the covariance matrix of $[U_1^{(j)}, \ldots, U_m^{(j)}]^T$

Nielsen (2002) extended this by allowing for four different constraints on the direction vectors, $a_i^{(j)}$.

1. **NORM**: $a_i^{(j)T} a_i^{(j)} = 1$ for $i = 1, \ldots, d$.
2. **AVGNORM**: $\sum_{i=1}^{d} a_i^{(j)T} a_i^{(j)} = 1$
3. **VAR**: $\text{Var}(U_1^{(j)}) = \ldots = \text{Var}(U_d^{(j)}) = 1$.
4. **AVGVAR**: $\sum_{i=1}^{d} \text{Var}(U_i^{(j)}) = 1$.

This allows for 20 different formulations of the multiset CCA problem. Asendorf (2015) gives a useful overview of the 20 different formulations and shows which have closed form solutions.

Principal Components Analysis

Principal components analysis (PCA) is a technique for transforming a potentially correlated set of variables into an uncorrelated set of variables (Pearson, 1901; Hotelling, 1933). The first principal direction is the unit length linear combination of the variables that explains the maximum possible variance. All further directions are unit length linear combinations of the

variables that explain the maximum variance subject to the constraint that they are uncorrelated with all previous directions. The principal directions can be shown to be the eigenvectors of the joint covariance matrix of the variables. Estimates of the principal directions are typically the eigenvectors of the sample covariance matrix of the observed data. PCA can be used for both dimension reduction and reducing multicollinearity between the variables. Frequently a large proportion of the variance within a set of variables can be explained by a small number of principal components. Collinearity among predictors can create instability in estimation for various statistical methods including linear regression. Principal component regression deals with this issue by including the uncorrelated principal component scores instead of the original data. Anderson (2003) shows that if the data come from a multivariate normal distribution the estimation of the eigenvalues and eigenvectors of the sample covariance matrix are asymptotically normal when properly standardized.

Estimates of bivariate survival function

Bivariate or multivariate survival data can come in the context of paired or clustered subjects all with a single survival time, or a single person with multiple non-competing failure times. In the bivariate case the ith subject or cluster will have two failure times $T^{(1)}$ and $T^{(2)}$ as well as two censoring times $C^{(1)}$ and $C^{(2)}$. The bivariate survival function for the failure times is

$$S^{(1,2)}(t,s) = P(T^{(1)} > t, T^{(2)} > s).$$

If we define $X^{(j)} = \min(T^{(j)}, C^{(1)})$ and $\delta^{(j)} = I(X^{(j)} = C^{(j)})$, the observed data will be $\{X^{(1)}, X^{(2)}, \delta^{(1)}, \delta^{(2)}\}$. The Kaplan-Meier estimator is most commonly used to estimate the univariate survival function, but it is not straightforward to extend this to the bivariate survival function. Common estimators for the bivariate survival function include those by Dabrowska et al. (1988), Lin and Ying (1993) and Prentice and Cai (1992). It was shown in Gill et al. (1995) that the Dabrowska and Prentice & Cai estimators are consistent and converge weakly to a Guassian process.

Survival Analysis with Semi-Competing Risks

In many settings failure times can be subject to both independent censoring and competing risks. This includes time to event data for certain diseases in which death acts as a competing risk. This means that once death is observed the disease of interest will never be observed. This competing risk set up can extend both the univariate and bivariate survival set up. In the bivariate competing risk set up there are still two failure times of interest, $T_1^{(1)}$ and $T_1^{(2)}$. In addition, for each failure time of interest there is also a competing risk time, $T_2^{(1)}$ and $T_2^{(2)}$. Without censoring the observed data is $T^{(j)} = T_1^{(j)} \wedge T_2^{(j)}$ and $\epsilon^{(j)} = 2 - I(T_2^{(j)} > T_1^{(j)})$ for $j = 1, 2$. If we introduce an independent censoring time for each failure time of interest, $C^{(1)}$ and $C^{(2)}$, the observed data will be $X^{(j)} = T^{(j)} \wedge C^{(j)}$, $\delta^{(j)} = I(T^{(j)} \leq C^{(j)})$, and $\eta^{(j)} = \delta^{(j)}\epsilon^{(j)}$ for $j = 1, 2$. For this competing risk set up it has been shown that the cause specific cumulative incidence functions (CIF) can be estimated non-parametrically. The univariate cause specific CIF is $F_k^{(j)}(t_j) = P(T^{(j)} \leq t_j, \ddot{\epsilon}^{(j)} = k)$ for $j = 1, 2$ and $k = 1, 2$. Details on estimation of the univariate cause specific CIF can be found in Kalbfleisch and Prentice (2011). The bivariate cause specific CIF is defined as $F_{kl}^{(j,j')}(t_j, t_{j'}) = P(\ddot{T}^{(j)} \leq t_j, \ddot{\gamma}^{(j)} = k, \ddot{T}^{(j')} \leq t_{j'}, \ddot{\gamma}^{(j')} = l)$, for $j, j' = 1, 2$ and $k, l = 1, 2$. Details on estimation of the bivariate cause specific CIF can be found in Cheng et al. (2007).

Functional delta method

The delta method is a well known result that can be used to derive the distribution of functions of random variables. When an estimator is a functional the functional delta method (Kosorok, 2008) can be used for the same purpose. In order to give the functional delta method we must first define the concept of Hadamard differentiability. For two complete normed spaces $\mathbb{D}$ and $\mathbb{E}$ a map $\phi : \mathbb{D} \to \mathbb{E}$ is Hadamard differentiable at h if there exists a map $\phi'_\theta : \mathbb{D} \to \mathbb{E}$ such that

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{tn} \to \phi'_\theta(h)$$

then ϕ is Hadamard differentiable with derivative ϕ'_θ. With this definition we can give the functional delta method which is theorem 2.8 from Kosorok (2008).

Functional delta method *For normed spaces $\mathbb{D}$ and $\mathbb{E}$, let $\phi : D_\phi \subset \mathbb{D} \to \mathbb{E}$ be Hadamard differentiable at θ tangentially to $D_0 \subset \mathbb{D}$. Assume that $r_n(X_n - \theta) \rightsquigarrow X$ for some sequence of constants $r_n \to \infty$, where X_n takes it's values in D_ϕ and X is a tight process taking it's values in D_0. Then $r_n(\phi(X_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(X)$*

CHAPTER 3: CANONICAL CORRELATION ANALYSIS FOR ELLIPTICAL COPULAS

3.1 Introduction

Canonical correlation analysis (CCA), first introduced by Hotelling (1936), 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. Typically a small number of directions may be used to summarize the relationship between the two sets of variables.

In Section 3.4 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 have excess skewness or kurtosis relative to the normal distribution. This suggests transformations may be needed for CCA using Pearson's 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) (Rousseeuw, 1984) and Tyler's (1987) M-estimator. There have been studies examining the performance of CCA using robust estimators of the covariance matrix or by maximizing other robust correlation measures (Taskinen et al., 2006; Visuri et al., 2003; Branco et al., 2005; Alfons et al., 2017). 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 data that have a multivariate distribution defined by an elliptical copula that does not require the existence of moments using properties of Kendall's tau for elliptically distributed data. 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 is equivalent to standard CCA for elliptically distributed data when moments do exist. Perhaps most importantly this definition of CCA does not make any assumptions about the marginal distributions of the data, 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 (Embrechts et al., 2002; Klüppelberg and Kuhn, 2009; Liu et al., 2012). 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 data to an elliptical

distribution by estimating the scatter matrix based on transformations of Kendall's tau for all pairs of variables (Liu et al., 2012). 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 elliptically distributed data (Anderson, 1999; Taskinen et al., 2006). Interestingly, the estimate based on transformations of Kendall's tau for all pairs of variables is not affine equivariant. Simulations show that these results can be used to construct confidence intervals with close to the desired coverage.

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 (Anderson, 2003) 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 come from a multivariate normal distribution. Even permutation based tests assume that zero correlation implies independence, which is not true for non-Gaussian elliptical copulas. Both of these types of tests will have inflated type I error when their assumptions are not met. Our bootstrap based testing procedure makes no such assumptions, and is useful even when data do not have an elliptical copula.

The rest of the paper is structured as follows. Section 3.2 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 3.3 reports the results of simulation studies under elliptical and transelliptical distributions. Section 3.4 provides an analysis of associations between white matter structure and executive function in six-year-old children. Section 3.5 overviews the paper and concludes with remarks.

3.2 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 $\text{Var}(U_1) = \text{Var}(V_1) = 1$ (Hotelling, 1936). 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 $U_j = a_j^T X$ and $V_j = b_j^T Y$ 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 $\text{Var}(U_j) = \text{Var}(V_1) = 1$ for all j. This uniquely defines the canonical directions corresponding to a non-zero canonical correlation except for multiplication of both a_i and b_i 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 of the sample covariance matrix. If we define the joint covariance matrix of X and Y as

$$\text{Cov}([X^T, Y^T]^T) = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$$

then the canonical correlations and directions may be derived from:

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

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

CCA can be made robust via robust estimation of the covariance matrix. Many robust estimates of the covariance matrix are consistent under the elliptical family of distributions, defined as,

Definition 3.1 (Elliptical Distributions)**.** *A $d \times 1$ random vector Z is considered to be elliptical if for some $d \times 1$ vector μ_Z, some $d \times d$ positive semi-definite matrix Σ_Z, and a function $\psi_Z[0, \infty) \to \mathbb{R}$, the characteristic function, Φ, satisfies $\Phi_{Z-\mu_Z}(t) = \psi(t^T\Sigma_Z t)$ 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 $Z \sim ED_d(\mu_Z, \Sigma_Z, \psi_Z)$*

We use Σ_Z in definition 3.1 because in the elliptical distribution Σ_Z can be viewed as a generalization of the covariance matrix for Z. When second moments exist Σ_Z equals the covariance matrix up to a scaling factor, and ψ_Z can be chosen such that it is equal to the covariance matrix. We will refer to Σ_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, Σ_Z behaves in the same way as a covariance matrix. To be precise, linear combinations of elliptical random variables are also elliptically distributed with a scatter matrix which is a quadratic form in Σ_Z.

Proposition 3.1 (Linear combinations of elliptically distributed random variables)**.** *Assume $Z \sim ED_d(\mu_Z, \Sigma_Z, \psi_Z)$. Define B to be a $k \times d$ dimensional matrix. Then $W = BZ$ is a $k \times 1$ dimensional random vector where $W \sim ED_k(B\mu_Z, B\Sigma_Z B^T, \psi_W)$*

Proposition 3.1 can be shown using the form of characteristic functions for elliptical distributions given in definition 3.1.

Letting $Z = [X^T, Y^T]^T$, the scatter matrix of Z can be decomposed as

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

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 Σ_Z, P_Z may be written as,

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

The elements of P_Z, ρ_{ij}, are related to the elements of Σ_Z, σ_{ij}, through the following equality, $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$. In general Σ_Z and P_Z will be assumed to be 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 3.2 (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 $\psi_Z : [0, \infty) \to \mathbb{R}$, and a set of functions $h_{Z1}, \ldots, h_{Zd}$ where $h_{Zi} : \mathbb{R} \to \mathbb{R}$ is a monotone increasing function for $i = 1, 2, \ldots, d$ such that $[h_{Z1}(Z_1), \ldots, h_{Zd}(Z_d)]^T \sim ED_d(0, P_Z, \psi_Z)$. The random variable Z is a $d \times 1$ dimensional transelliptically distributed random variable, denoted as $Z \sim TE_d(h_Z, 0, P_Z, \psi_Z)$.*

The elliptical distribution used in Definition 3.2 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. (2012), but an equivalent definition is any multivariate distribution with continuous marginal distributions and a copula from an elliptical distribution (Embrechts et al., 2002; Klüppelberg and Kuhn, 2009).

For the elliptical and transelliptical distributions we propose an alternative definition of CCA using a rank correlation measure. This version of CCA is equivalent to 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 $[\tilde{Z}_i, \tilde{Z}_j]^T$ is an identically distributed copy of $[Z_i, Z_j]^T$. 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, $[z_{i1}, z_{j1}]^T, \ldots, [z_{in}, z_{jn}]^T$, is

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

This estimator is a U-statistic with consistency and asymptotic normality coming from established U-statistic theory (Hoeffding, 1961).

Within the transelliptical family the following property gives the correspondence between Kendall's tau and the elements of the transelliptical scatter matrix.

Proposition 3.2 (Kendall's tau for transelliptically distributed random variables)**.** *Assume* $Z \sim TE_d(h, 0, P_Z, \psi_Z)$*. If* p_{ij} *is the* $i-jth$ *entry of* P_Z *and* $\tau^{(Z_i,Z_j)}$ *is the Kendall correlation between the* ith *and* jth *entries of* Z *then* $\tau^{(Z_i,Z_j)} = \frac{2}{\pi} \arcsin(\rho_{ij})$

The proof for Proposition 3.2 follows from the proof of the same result for elliptical distributions in Lindskog et al. (2003) and the fact that Kendall's tau is invariant to monotone increasing transformations of the data. 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 3.2 and 3.1 we define CCA for transelliptically distributed data as follows,

Definition 3.3 (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 $[X^T, Y^T]^T = Z \sim TE_{p+q}(h_Z, 0, P_Z, \psi_Z)$. Define h_X to be the element-wise functions of h_Z corresponding to X and h_Y to be the element-wise 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 $U_1 = a_1^T h_X(X)$ and $V_1 = b_1^T h_Y(Y)$, subject to the constraint that U_1 and V_1 have scale parameter equal to one. The jth 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 $U_j = a_j^T h_X(X)$ $V_j = b_j^T h_Y(Y)$, 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 jth canonical correlation can be defined as* $\sin\left(\frac{\pi}{2}\tau^{(U_j,V_j)}\right)$.

When second moments exist and $[X^T, Y^T]^T$ has an elliptical distribution this definition is equivalent to performing CCA based on the correlation matrix. When data are elliptically distributed 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 is definition is when $[X^T, Y^T]^T$ 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 3.3 is based on P_Z, which does not depend on the marginal distributions of the data. In this sense CCA using Definition 3.3 can be thought of as first transforming the variables to elliptical symmetry and then performing CCA. As shown in proposition 3.1 linear combinations of elliptical distributions meaningfully describe the associations within the data.

An issue with estimating CCA for the transelliptical family is estimation of a scatter matrix of transformed versions of X and Y. If $[X^T, Y^T]^T = Z$ is transelliptically distributed and h_X, h_Y, and ψ_Z are all unknown, then all three must be estimated to transform Z to it's underlying elliptical distribution. Many methods assume that ψ_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 ψ_Z we directly estimate the scatter matrix in the transelliptical distribution as follows (Liu et al., 2012),

Definition 3.4 (Transelliptical scatter matrix estimate)**.** *Assume that $Z \sim TE_d(h_Z, 0, P_Z, \psi_Z)$. Assume that ρ_{ij} is the element of P_Z corresponding to the ith and jth elements of Z. Then we can estimate ρ_{ij} as $\hat{\rho}_{n,ij} = \sin\left\{\frac{\pi}{2}\hat{\tau}_n^{(Z_i,Z_j)}\right\}$, 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.*

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 is the only known estimate of the scatter matrix that can be used for CCA for all members of the transelliptical family, without having to estimate the transformations h_Z, or the generator ψ_Z.

There are other rank based methods that can be used to estimate the scatter matrix for transellipticals when ψ_Z is assumed to be the generating function for the Guassian distribution. One such estimator uses transformations of Spearman's correlation. For the bivariate normal distribution Spearman's correlation, s, and Pearson correlation, ρ, have the following relationship, $s = \frac{6}{\pi}\arcsin\left(\frac{\rho}{2}\right)$. 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. 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 transelliptically

distributed and the generating function, ψ_Z, is from an elliptical distribution other than a Guassian this method results in biased estimates of the transelliptical scatter matrix.

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 (1993) 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 positive eigenvalues as $\hat{P}_{Zn}$ but with with all negative eigenvalues set 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 3.1 (Transformed Kendall's scatter matrix estimator eigenvalues)**.** *Assume $z_1, \ldots, z_n$ 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 $\hat{\lambda}_{n1}, \ldots, \hat{\lambda}_{nd}$, 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 3.1 is presented in Appendix A. Theorem 3.1 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 $\sqrt{n}(\hat{P}_{Zn} - P_Z)$ can be shown to be asymptotically normal with mean zero and finite variance based on U-statistic theory and the delta method. Theorem 3.2 establishes conditions under which the estimates of transelliptical CCA directions and correlations will be consistent and asymptotically normal.

Theorem 3.2 (Asymptotic results for transelliptical CCA)**.** *Assume $[x_i^T, y_i^T]^T$ for $i = 1, \ldots, n$ are iid realizations of the $(p + q) \times 1$ dimensional random vector $[X^T, Y^T]^T = Z \sim TE_{p+q}(h_Z, 0, P_Z, \psi_Z)$, with positive-definite P_Z. Further assume that $p \geq q$ and there are $r \leq q$ unique non-zero transelliptical canonical correlations for X and Y. Let $\lambda_1, \ldots, \lambda_r$ be the non-zero transelliptical canonical correlations with $\lambda_1 > \ldots > \lambda_r > 0$. Define*

$\Lambda_r = diag(\lambda_1, \ldots, \lambda_r)$ to be the diagonal matrix with the ordered non-zero canonical correlations on the diagonal. Let $A_r = [a_1, \ldots, a_r]$ be the $p \times r$ matrix where the ith column is the ith transelliptical canonical direction for X, and $B_r = [b_1, \ldots, b_r^T]$ be the $q \times r$ matrix where the ith column is the ith transelliptical canonical direction for Y. Define $A_{r^+} = [a_{r+1}, \ldots, a_p]$ and $B_{r^+} = [a_{r+1}, \ldots, b_q]$ to be a solution to the canonical directions corresponding to the zero canonical correlations. This means for $A = [A_r, A_{r^+}]$ and $B = [B_r, B_{r^+}]$, $A^T P_{XX} A = I_p$, $B^T P_{XX} B = I_q$, and $A^T P_{XY} B = \begin{bmatrix} \Lambda_q & 0 \\ 0 & 0 \end{bmatrix}$. 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. We will assume that the estimates of Λ_r, A_r and B_r, which we will denote as Λ^*_{nr}, A^*_{nr}, and B^*_{nr} are based on the eigendecomposition of a function of a consistent estimate of the scatter matrix, P_Z, which is denoted by P^*_{Zn}. For an arbitrary matrix, M, define $vec(M)$ to be the column vector made by stacking the columns of M on top of each other. The following holds: if P^*_{Zn} is guaranteed to be positive-definite and

$$\sqrt{n} \begin{bmatrix} vec(P^*_{XXn}) & & vec(P_{XX}) \\ vec(P^*_{XYn}) & - & vec(P_{XY}) \\ vec(P^*_{YYn}) & & vec(P_{YY}) \end{bmatrix} \rightarrow_d N_{p^3 \times q^3}(0, \Theta),$$

then $\sqrt{n}[vec(\Lambda^*_{nr}) - vec(\Lambda_r)]$, $\sqrt{n}[vec(A^*_{nr}) - vec(A_r)]$ and $\sqrt{n}[vec(B^*_{nr}) - vec(B_r)]$ jointly have multivariate normal limiting distributions with mean zero and a finite limiting variance that is a function of Θ, A, B, and Λ_q. The form of the limiting variance can be found in Appendix A.

A proof of Theorem 3.2 is presented in Appendix A. This result, and the limiting variance of the estimates, is more general than previous results from Anderson (1999) and Taskinen et al. (2006), and requires only that the estimate of the covariance matrix be asymptotically normal and positive-definite. Anderson (1999) show the asymptotic results for standard CCA directions and correlations when Z has a multivariate normal distribution and the sample covariance matrix

is used. Taskinen et al. (2006) expanded this result to CCA in the elliptical distribution when using positive-definite and affine equivariant estimators of the covariance matrix. Because we make minimal assumptions about the form of Θ 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. To our knowledge there are no other estimators of the scatter matrix that work for all transelliptical distributions that do not require the estimation of h_Z and ψ_Z.

We have already shown that $\tilde{P}_{Zn}$ is positive-definite, consistent, and asymptotically normal, which leads directly to corollary 3.1.

Corollary 3.1 (Asymptotic results for transformed Kendall's scatter matrix estimator)**.** *Assume the same set up as in Theorem 3.2 and that P_Z is estimated using $\tilde{P}_{Zn}$. Define $\tilde{\Lambda}_{nr}$, $\tilde{A}_{nr}$, and $\tilde{B}_{nr}$ as the corresponding estimates for the non-zero transelliptical canonical correlations and their corresponding transelliptical canonical directions. Then $\sqrt{n}[vec(\tilde{\Lambda}_{nr}) - vec(\Lambda)]$, $\sqrt{n}[vec(\tilde{A}_{nr}) - vec(A)]$, and $\sqrt{n}[vec(\tilde{B}_{nr}) - vec(B)]$ jointly multivariate normal limiting distributions with mean 0 and finite variances. The form of the variances can be found using methods from the proof for Theorem 3.2 in Appendix A.*

Methods from Rublík (2016) can be used to obtain estimators for the covariance matrix for all pairwise estimates of Kendall's tau. An estimate of the variance of $\tilde{P}_{Zn}$ can then be found using the delta method, which allows for estimates of the limiting variances for $\tilde{\Lambda}_n$, $\tilde{A}_n$, and $\tilde{B}_n$ to be estimated by a "plug-in" estimator using the form of the variance found in Appendix A. Section 3.3 and Appendix A include simulations studies that compare the coverage of confidence intervals using this method to bootstrapped confidence intervals.

These results show that the transelliptical CCA estimates using the transformed Kendall's scatter matrix 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 correlation.

It is important to note that Theorem 3.2 and Corollary 3.1 only apply to non-zero canonical correlations and cannot be used for hypothesis testing for zero correlations. Anderson (2003) 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. In addition Muirhead and Waternaux (1980) shows 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 the elliptical distribution 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 α simply invert a (1-2α) bootstrapped confidence interval using the normal approximation with bias correction. A (1-2α) 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 as sample size and dimension increase the probability that for each bootstrap sample the estimated canonical correlation will be above zero converges to one. 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 bootstrap will have conservative type I error as sample size increases. This is shown in simulation results found in Section 3.3.

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 data this assumption is only met when data have 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 do not have a Gaussian copula, and asymptotic testing procedures assume the data come 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. For this reason we recommend using the inverted bootstrap procedure if there is any reason to believe data do not follow a multivariate normal distribution. 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 that for all variables in X the true pairwise Kendall's tau coefficient with all variables in Y is 0. Therefore even when data do not have 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.

3.3 Simulation Results

3.3.1 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 minimum covariance determinant (MCD) estimator from the R package *robustbase* (Todorov and Filzmoser, 2009a) and the M estimator from the R package *rrcov* (Todorov and Filzmoser, 2009b). Standard CCA is calculated using the R package *CCA*.

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=1,000, 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 Appendix A. The true scatter matrix for X and Y is $\Sigma_{XX} = \Sigma_{YY} = I_p$, and $\Sigma_{XY} = \text{diag}(0.9, 0.5, 0.4, 1/3, 0, \ldots)$. The structure of these scatter matrices is similar to those in Branco et al. (2005). To define $\tilde{P}_{Zn}$ 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 1,000.

Based on the 1,000 simulated data sets the empirical bias and standard deviation is 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 ith canonical direction the angle between the true direction for X, a_i, and estimated direction for X for the jth simulation, $\hat{a}_{ji}$, is calculated as $\cos^{-1}\left(\frac{|\hat{a}_{ji}^T, a_i|}{||\hat{a}_{ji}^T, \hat{a}_{ji}|| \cdot ||a_i^T, a_i||}\right)$. 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 3.1 gives the output for the canonical correlation and canonical direction. Because Σ_{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 3.1 standard CCA has the smallest bias and standard deviation when simulating under the multivariate normal distribution, with transelliptical CCA using the transformed Kendall's matrix outperforming the other two robust methods. For simulations from the Cauchy distribution the transformed Kendall's scatter matrix estimator performs similarly to the MCD and M estimators for the first direction and is superior for higher directions. Standard CCA completely "breaks down" under this setting because of the lack of moments. When data are generated from a multivariate t distribution with five degrees of freedom transelliptical CCA using the transformed Kendall's estimator outperforms all other methods. This shows that even when standard CCA is well defined, the transformed Kendall's scatter estimator outperforms the sample covariance matrix for heavy tailed elliptical distributions. When data come from a

multivariate t with ten degrees of freedom standard CCA using the sample covariance matrix and the transformed Kendall's scatter estimator, both of them outperforming the other two methods.. Under the lognormal setting the standard estimator, the MCD estimator, and the M estimator all underestimate the transelliptical canonical correlations, while the transformed Kendall estimator provides consistent estimates. This is particularly evident for $n = 1,000$ presented in Appendix A. 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 scatter estimator can consistently estimate the strongest linear relationships based on the underlying copula. As noted previously, the finite sample bias is positive for both standard and transelliptical canonical correlation estimates.

3.3.2 Confidence intervals for non-zero canonical correlations

Simulations are run to compare coverages for normal bootstrapped confidence intervals as well as asymptotic confidence intervals using "plug-in" estimators of the asymptotic variance for the estimates from Theorem 3.2. Details on the form of the variance estimates are in Appendix A. The "plug-in" variance estimator is calculated using estimates of transelliptical canonical correlations, and directions based on the transformed Kendall's correlation estimate. An estimate of the variance of the transformed Kendall's scatter matrix is obtained using methods from Rublík (2016). For the bootstrap confidence intervals 1,000 bootstrap replicates are used. The bootstrap confidence intervals for the canonical correlations are constructed from the square of the estimated canonical correlations, and then transformed using the square root to give the bounds for the transelliptical canonical correlations. Bounds may be truncated at one or zero as necessary for both the bootstrap and asymptotic confidence intervals.

The simulation set-ups are the same as Section 3.3.1, but simulations for the multivariate lognormal distribution are not included because they are similar to results using the multivariate normal distribution when using the transformed Kendall's estimator. Tables reporting the coverages can be found in Appendix A. For the non-zero canonical correlations both the the

Table 3.1: Bias (SD) of canonical correlation and direction estimates, p=q=8, n=200

		Normal	Cauchy	Lognormal	t5	t10
Canonical Correlations						
Cor 1	Standard	0.05 (0.07)	1.93 (1.10)	-0.13 (0.19)	0.09 (0.10)	0.06 (0.08)
	Kendall	0.06 (0.08)	0.17 (0.14)	0.06 (0.08)	0.07 (0.09)	0.06 (0.09)
	MCD	0.07 (0.09)	0.14 (0.13)	-0.12 (0.15)	0.10 (0.11)	0.08 (0.10)
	M	0.06 (0.09)	0.13 (0.12)	-0.14 (0.14)	0.09 (0.10)	0.07 (0.09)
Cor 2	Standard	0.08 (0.06)	1.69 (0.63)	0.01 (0.11)	0.17 (0.09)	0.12 (0.07)
	Kendall	0.10 (0.07)	0.19 (0.09)	0.10 (0.06)	0.12 (0.07)	0.11 (0.07)
	MCD	0.13 (0.08)	0.27 (0.10)	0.09 (0.09)	0.19 (0.08)	0.17 (0.08)
	M	0.12 (0.07)	0.25 (0.09)	0.05 (0.08)	0.17 (0.08)	0.15 (0.08)
Cor 3	Standard	0.07 (0.05)	1.21 (0.44)	-0.02 (0.07)	0.13 (0.07)	0.09 (0.06)
	Kendall	0.08 (0.06)	0.14 (0.07)	0.08 (0.05)	0.09 (0.06)	0.09 (0.06)
	MCD	0.10 (0.06)	0.20 (0.08)	0.06 (0.07)	0.15 (0.07)	0.13 (0.07)
	M	0.09 (0.06)	0.19 (0.08)	0.02 (0.06)	0.13 (0.07)	0.12 (0.06)
Cor 4	Standard	0.03 (0.05)	0.86 (0.34)	-0.05 (0.06)	0.07 (0.06)	0.04 (0.06)
	Kendall	0.03 (0.06)	0.08 (0.06)	0.03 (0.05)	0.04 (0.06)	0.04 (0.06)
	MCD	0.05 (0.06)	0.13 (0.07)	0.01 (0.06)	0.08 (0.06)	0.07 (0.06)
	M	0.04 (0.06)	0.12 (0.07)	-0.01 (0.05)	0.07 (0.06)	0.06 (0.06)
Canonical Directions						
Dir 1	Standard	0.10 (0.03)	0.65 (0.43)	0.12 (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 (0.03)
	MCD	0.12 (0.04)	0.19 (0.06)	0.19 (0.07)	0.15 (0.05)	0.14 (0.04)
	M	0.12 (0.04)	0.18 (0.06)	0.17 (0.06)	0.14 (0.04)	0.13 (0.04)
Dir 2	Standard	0.58 (0.30)	1.26 (0.26)	0.83 (0.43)	0.73 (0.34)	0.63 (0.32)
	Kendall	0.60 (0.31)	0.75 (0.32)	0.60 (0.31)	0.65 (0.32)	0.63 (0.32)
	MCD	0.67 (0.32)	0.88 (0.34)	0.89 (0.35)	0.79 (0.34)	0.76 (0.34)
	M	0.64 (0.31)	0.85 (0.33)	0.83 (0.36)	0.74 (0.33)	0.73 (0.33)
Dir 3	Standard	0.83 (0.34)	1.26 (0.25)	1.05 (0.36)	0.97 (0.33)	0.89 (0.35)
	Kendall	0.85 (0.34)	0.98 (0.33)	0.85 (0.34)	0.91 (0.34)	0.89 (0.35)
	MCD	0.92 (0.33)	1.08 (0.30)	1.12 (0.29)	1.02 (0.32)	0.98 (0.33)
	M	0.89 (0.33)	1.05 (0.31)	1.09 (0.32)	0.99 (0.32)	0.96 (0.33)
Dir 4	Standard	0.80 (0.31)	1.27 (0.23)	1.08 (0.33)	0.99 (0.31)	0.89 (0.32)
	Kendall	0.84 (0.30)	1.02 (0.31)	0.84 (0.31)	0.89 (0.31)	0.87 (0.31)
	MCD	0.92 (0.31)	1.13 (0.28)	1.17 (0.27)	1.05 (0.30)	1.02 (0.31)
	M	0.89 (0.31)	1.12 (0.28)	1.16 (0.27)	1.00 (0.31)	0.99 (0.31)

asymptotic and bootstrap confidence intervals tend show undercoverage 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 as sample size increases.

For the transelliptical canonical directions 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. 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. 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.

3.3.3 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 3.2 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 ith 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 from the R package *CCP* (Menzel, 2009) 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 in Appendix A. We consider $p = q = 8$ and $\Sigma_{XX} = \Sigma_{YY} = I$ where Σ_{XY} has either all zeros or a single non-zero entry ranging from 0.2 to 0.8 in increments of 0.2. 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 $n = 1,000$. 1,000 data sets are simulated for each setting. Table 3.2 gives the proportion of simulated data sets for which the null hypothesis that the first transelliptical canonical correlation is zero is rejected for the for each testing procedure. The bootstrap test using Kendall's transformed estimator controls for type I error, being conservative in all settings when n = 1,000. For n = 200 type I error is not controlled for the bootstrap method when data come from multivariate Cauchy or multivariate t distribution with five degrees of freedom, but is closer to the nominal level than asymptotic or permutation tests tests. For the multivariate Cauchy distribution the transelliptical CCA bootstrap method is the only procedure that doesn't have a type I error of one. The permutation based test only controls for type I error when data come from a normal or lognormal distribution. This is because permutation based tests for CCA are only valid when zero correlation also implies independence, which is not true for non-Gaussian elliptical copulas. When the sample size is 1,000 the type I error rate for the permutation test is even higher than for a sample size of 200 for the multivariate Cauchy and multivariate t distributions. Also as expected the asymptotic Wilk's Lambda test only works when the data come 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.

Table 3.2: Power and type I error for asymptotic, permutation and bootstrap testing procedures

	True correlation	0	0.2	0.4	0.6	0.8
		n=200				
Normal	Kendall Bootstrap	0.05	0.17	0.85	1.00	1.00
	Wilk's Lambda	0.05	0.17	0.80	1.00	1.00
	Kendall Perm	0.06	0.16	0.86	1.00	1.00
Cauchy	Kendall Bootstrap	0.14	0.19	0.64	0.99	1.00
	Wilk's Lambda	1.00	1.00	1.00	1.00	1.00
	Kendall Perm	0.77	0.85	0.98	1.00	1.00
Lognormal	Kendall Bootstrap	0.05	0.15	0.87	1.00	1.00
	Wilk's Lambda	0.10	0.17	0.43	0.88	1.00
	Kendall Perm	0.05	0.15	0.87	1.00	1.00
t5	Kendall Bootstrap	0.08	0.16	0.78	1.00	1.00
	Wilk's Lambda	0.94	0.96	1.00	1.00	1.00
	Kendall Perm	0.17	0.33	0.90	1.00	1.00
t10	Kendall Bootstrap	0.06	0.15	0.84	1.00	1.00
	Wilk's Lambda	0.43	0.65	0.96	1.00	1.00
	Kendall Perm	0.10	0.22	0.88	1.00	1.00
		n=1,000				
Normal	Kendall Bootstrap	0.02	0.85	1.00	1.00	1.00
	Wilk's Lambda	0.05	0.88	1.00	1.00	1.00
	Kendall Perm	0.06	0.94	1.00	1.00	1.00
Cauchy	Kendall Bootstrap	0.02	0.45	1.00	1.00	1.00
	Wilk's Lambda	1.00	1.00	1.00	1.00	1.00
	Kendall Perm	0.80	1.00	1.00	1.00	1.00
Lognormal	Kendall Bootstrap	0.02	0.84	1.00	1.00	1.00
	Wilk's Lambda	0.10	0.45	0.99	1.00	1.00
	Kendall Perm	0.06	0.92	1.00	1.00	1.00
t5	Kendall Bootstrap	0.02	0.67	1.00	1.00	1.00
	Wilk's Lambda	0.98	1.00	1.00	1.00	1.00
	Kendall Perm	0.18	0.95	1.00	1.00	1.00
t10	Kendall Bootstrap	0.01	0.79	1.00	1.00	1.00
	Wilk's Lambda	0.48	0.99	1.00	1.00	1.00
	Kendall Perm	0.12	0.97	1.00	1.00	1.00

3.4 White matter tractography and executive function in six year old children

We provide a comparison of transelliptical CCA estimated with the transformed Kendall's 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 (Gilmore et al., 2010; Knickmeyer et al., 2008, 2016). 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 (Girault et al., 2019). The 20 tracts included in the analysis can be found in Table 3.3. Imaging measures of diffusion rate and direction are available on these tracts including fractal 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. (2007). Results for RD are presented in the main text with those for FA and AD given in Appendix A.

Table 3.3: List of white matter tracts used in CCA analysis

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

EF measures are an executive composite score from the Behavior Rating Inventory of Executive Function (BRIEF) (Gioia et al., 2000), Cambridge Neuropsychological Test Automated Battery (CANTAB) Spatial Span (SSP), CANTAB Stockings of Cambridge (SOC) (CANTAB, 2017) , Stanford-Binet Fluid Reasoning Verbal (SB V FR), and Stanford-Binet Fluid Reasoning Non-verbal (SB NV FR) (Roid, 2003). 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 1,000 replicates. For both methods bootstrap confidence intervals for the direction loadings are reported using 1,000 bootstrap replicates and the normal approximation bootstrap method. Confidence intervals base on a "plug-in" variance estimator for transelliptical CCA directions using the transformed Kendall's estimator are also reported.

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, and skewness using the Agostino test. The average RD values for a number of white matter tracts shows 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 Stanford Binet verbal fluid reasoning scores also have excess kurtosis and negative skewness, while the BRIEF scores show positive skewness.

Transelliptical CCA assumes the data are transelliptically distributed which can be tested using the methods from Jaser et al. (2017). This test is based on the equivalence between Kendall's tau and Blomqvist's beta for elliptical copulas. Blomqvist's beta between two variables, Z_1 and Z_2, $\beta_B^{(Z_1,Z_2)}$, is defined as, $\beta_B^{(Z_1,Z_2)} = E[\text{sign}(Z_1 - Z_{1med})\text{sign}(Z_2 - Z_{2med})]$, where Z_{1med} and Z_{2med} denote the population medians of Z_1 and Z_2. 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 any deviations from the transelliptical assumption are relatively minor.

Table 3.4 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, 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 through the white matter tracts.

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 (2006). For the ith bilateral tract the lateralization measure, $RD_{\text{LAT}i}$, is defined as $RD_{\text{LAT}i} = \frac{RD_{Li}-RD_{Ri}}{(RD_{Li}+RD_{Ri})/2}$, where RD_{Li} is the RD measure from the ith 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 Agostino tests. The test from Jaser et al. (2017) is again used to test for potential violations of the transelliptical assumption, and again none of the pairwise tests show a significant difference at the 0.05 level between Kendall's tau and Blomqvist's beta after an FDR correction.

Table 3.4: Estimates for first canonical correlation and directions for white matter tract RD values and EF tests for transelliptical and standard CCA

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 FR	1.02	(0.71, 1.77)	(0.86, 1.19)	0.87	(0.55, 1.57)
SB NV FR	-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)
SSP	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 3.5 reports the estimated first direction and correlation for transformed Kendall's CCA and standard CCA. The jackknife corrected estimate for the first canonical correlation using transformed Kendall's CCA is 0.33, compared to 0.23 for standard CCA. In this case only CCA using the transformed Kendall estimator has a p-value less than 0.05 for the first direction based on the bootstrap testing procedure. The first direction for the DTI lateralization measures is driven by positive loadings for the ARC FT and IFOF tracts. The first direction for the EF test cores is driven by a positive loading for the SB V FR scores. This indicates that higher lateralization score for the ARC FT and IFOF tracts is associated with higher SB V FR 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 3.5: Estimates for first canonical correlation and direction for RD lateralization measure and EF tests for transformed Kendall's CCA and standard CCA

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 FR	0.82	(0.39 ,1.58)	(0.43 ,1.21)	0.80	(0.41, 1.50)
SB NV FR	-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)
SSP	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	

3.5 Discussion

In this paper we define a version of CCA for transelliptically distributed data using Kendall's tau. Consistent estimates of canonical directions and correlations can be obtained using a transformed Kendall's scatter matrix estimator. These estimates are shown to be consistent and asymptotically normal. Simulation studies show that CCA estimates using transformed Kendall's tau perform well relative to other robust CCA methods in finite sample settings. The transformed Kendall's scatter matrix estimator is the only consistent estimator of CCA for all transelliptical distributions.

Confidence intervals for the canonical directions and nonzero correlations can be obtained using bootstrap methods or based on the asymptotic variances of the estimator. 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. 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 fail if data do not come from 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 estimate when data do not come 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 general transelliptical distributions.

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 (Witten and Tibshirani, 2009; Wilms and Croux, 2015; Yoon et al., 2018). The methods from Yoon et al. (2018) 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 (Kettenring, 1971; Nielsen, 2002).

CHAPTER 4: PRINCIPAL COMPONENTS ANALYSIS FOR RIGHT CENSORED DATA

4.1 Introduction

Principal Component Analysis (PCA) (Pearson, 1901; Hotelling, 1933), transforms a set of potentially linearly correlated variables into a set of linearly uncorrelated variables called the principal components. The first principal component is defined to be the unit length linear combination of the variables with maximum possible variance. All subsequent directions are the unit length linear combinations of the variables with maximal variance, subject to the constraint that they are orthogonal to all previous principal components. PCA is frequently used for dimension reduction and removing collinearity within a set of variables. PCA is generally estimated using eigendecomposition of the estimated covariance matrix or singular value decomposition (SVD) of the mean centered data.

PCA is used across many different settings, but to our knowledge one area where PCA has not been used is multivariate time to event data in the presence of right censoring. In the multivariate survival setting with p different event times for each subject we can define $T^{(j)}$ to be the failure time for the jth event time. If we assume that each subject also has an independent censoring time, C, the observed data for the jth event type consists of $Y^{(j)} = T^{(j)} \wedge C$ as well as an indicator for whether the observed time is a censoring time or failure time. In the presence of censoring the full covariance matrix for $T = [T^{(1)}, \ldots, T^{(p)}]$ cannot be estimated without strong parametric assumptions. This means that PCA cannot be estimated for T. In order to overcome this issue we propose two versions of PCA for multivariate survival data and show how both can be estimated. One version uses the counting processes for each event type, defined as $N^{(j)}(t) = I(T^{(j)} \leq t)$. The other version uses the martingales based on the decomposition

of these counting processes. Even though these counting processes are not observed for all time points if subjects are censored, we show that the covariance between the counting processes and martingales for two different event types can both be estimated non-parametrically. This allows for the principal component directions for the counting processes and martingales to be estimated consistently even in the presence of independent censoring. The corresponding component scores can be estimated for each subject up until the time they are censored. In order to make these methods more flexible we also allow for semi-competing risks, and show how to extend both survival PCA methods to this setting.

The estimation of the covariance between counting processes and martingales uses existing methods for estimation of bivariate and univariate survival functions, and univariate hazard functions. Prentice and Cai (1992) shows that at a fixed timepoint the covariance for the counting process martingales for two different event types can be written as a function of bivariate and univariate survival functions, as well as univariate hazard functions. Similarly, the covariance between the counting processes for two different event types can be written as a function of the bivariate and univariate survival functions. In this paper we use the Kaplan-Meier and Nelson-Aalen estimators for the univariate survival and cumulative hazard functions. Estimators for the bivariate survival function include those by Dabrowska et al. (1988), Prentice and Cai (1992), van der Laan (1993), and Lin and Ying (1993). If there are more than two event types estimates of the full covariance and correlation matrices for the martingales or counting processes are found by estimating all the elements individually using bivariate methods. This allows for estimation of principal component directions for the martingales and counting processes using eigendecomposition of their estimated covariance or correlation matrices.

In the presence of competing risks we use previous methods for estimation of bivariate and univariate cause specific hazard functions and cause specific cumulative incidence functions (CIFs). Prentice et al. (1978) gives an overview of cause specific hazards and incidence functions in the univariate setting, and Kalbfleisch and Prentice (2011) gives details on estimating univariate cause specific CIFs. Details on estimating bivariate cause specific hazard functions and CIFs

can be found in Cheng et al. (2007). In order to extend our methods to the competing risk setting we use the cause specific counting processes and the martingales based on their decomposition. We are able to show that the covariance between the cause specific martingales or counting processes for two different events can be written as a function of the bivariate and univariate cause specific CIFs as well as the univariate cause specific hazard functions. Using these results we show that the full covariance matrix for cause specific martingales and counting processes can be estimated in the presence of competing risks, which makes it possible to estimate the corresponding principal component directions using eigendecomposition.

Being able to estimate covariance and principal component directions for martingales and counting processes in the presence of competing risks is of particular interest in sick populations where death acts as a competing risk for many different adverse events. We present one such example using data from the MPACT study for subjects with pancreatic cancer. In this study subjects are followed and have many different types of adverse events due to both cancer and treatment. Using PCA of the martingales we are able to define medically relevant groupings of the event types. We also show how the principal component scores can be estimated and used as predictors in a Cox Proportional Hazards (PH) model. This can be used to remove multicollinearity among predictors and is analogous to principal component linear regression.

The rest of this paper is structured as follows. Section 4.2 studies estimation of covariance and correlation matrices for counting processes and martingales. Section 4.3 defines the estimators for survival PCA and shows that they are consistent and asymptotically normal. Section 4.4 provides results of simulation studies for survival PCA methods. Section 4.5 provides an analysis of adverse events among subjects in the MPACT pancreatic cancer clinical trial. Proofs for theorems can be found in Appendix B.

4.2 Covariance estimation for bivariate counting processes and counting process martingales

4.2.1 Estimation of covariance in presence of right censoring

As in Section 4.1 assume that for each subject in a population there are p event types of interest, and let $T^{(j)}$ denote the failure time for the jth event type. We will assume that $T^{(j)}$ is a continuous random variable. The full vector of failure times for a subject can be written as $T = [T^{(1)}, \ldots, T^{(p)}]^T$. If $\mathbf{t} = [t_1, \ldots, t_p]^T$ is a vector of fixed timepoints, the joint distribution for T is defined as $F_T(t) = P(T^{(1)} \leq t_1, \ldots, T^{(p)} \leq t_p)$, and the univariate distribution functions are defined as $F^{(j)}(t_j) = P(T^{(j)} \leq t_j)$ for $j = 1, \ldots, p$. We will also assume that there is an independent censoring time, C, with distribution $F_C(c) = P(C \leq c)$. This censoring time is the same for all event types. This assumption is reasonable in the case where all event types are measured for the same subject, and end of study or loss to follow up will make it impossible to get any additional information from that subject for any of the event types. The observed data for the jth event type is the observed time, $Y^{(j)} = T^{(j)} \wedge C$, and the censoring indicator, $\eta^{(j)} = I(T^{(j)} \leq C)$. Define the counting process associated with the jth event type as $N^{(j)}(t) = I(T^{(j)} \leq t)$. Note that the value of $N^{(j)}(t)$ is not always observed, unlike $N^{(j1)}(t) = N^{(j)}(t)\eta^{(j)}$, which we will use in later sections for deriving asymptotic properties of estimates.

We can define the cumulative hazard function for the jth event type at time t to be $\Lambda^{(j)}(t) = \int_0^t \lambda^{(j)}(s)ds$, where

$$\lambda^{(j)}(t) = \lim_{\delta \to 0} \frac{1}{\delta} P(t \leq T^{(j)} < t + \delta | T^{(j)} \geq t).$$

The martingale defined by the decomposition of $N^{(j)}(t)$ is $M^{(j)}(t) = N^{(j)}(t) - \Lambda^{(j)}(t \wedge T^{(j)})$.

Define the covariance between $N^{(j)}(t_j)$ and $N^{(j')}(t_{j'})$ to be $CN^{(j,j')}(t_j, t_{j'})$, and similarly the covariance between $M^{(j)}(t_j)$ and $M^{(j')}(t_{j'})$ to be $CM^{(j,j')}(t_j, t_{j'})$. Because $N^{(j)}(t)$ may not be observed for every subject in the presence of censoring, $CN^{(j,j')}(t_j, t_{j'})$ and $CM^{(j,j')}(t_j, t_{j'})$ cannot be calculated using standard methods. The following equality can be used to estimate

$CN^{(j,j')}(t_j, t_{j'})$,

$$CN^{(j,j')}(t_j, t_{j'}) = S^{(j,j')}(t_j, t_{j'}) - S^{(j)}(t_j)S^{(j')}(t_{j'}), \tag{4.1}$$

where $S^{(j,j')}(t_j, t_{j'}) = P(T^{(j)} > t_j, T^{(j')} > t_{j'})$ is the bivariate survival function and $S^{(j)}(t_j) = P(T^{(j)} > t_j)$ is the univariate survival function. Similarly, the following equality from Prentice and Cai (1992) is obtained through Stieltjes integration and repeated integration by parts, and can be used to estimate $CM^{(j,j')}(t_j, t_{j'})$,

$$CM^{(j,j')}(t_j, t_{j'}) = S^{(j,j')}(t_j, t_{j'}) - 1 + \int_0^{t_j} S^{(j,j')}(s_j^-, t_{j'})\lambda^{(j)}(s_j)ds_+ \\ \int_0^{t_{j'}} S^{(j,j')}(t_j, s_{j'}^-)\lambda^{(j')}(s_{j'})ds_{j'} + \int_0^{t_j}\int_0^{t_{j'}} S(s_j^-, s_{j'}^-)\lambda^{(j)}(s_j)\lambda^{(j')}(s_{j'})ds_j ds_{j'}. \tag{4.2}$$

This allows for the consistent estimation of $CN^{(j,j')}(t_j, t_{j'})$ and $CM^{(j,j')}(t_j, t_{j'})$ by plugging in consistent estimates of all of the quantities on the right hand side of Equations (4.1) and (4.2) respectively. The variance of $N^{(j)}(t)$ and $M^{(j)}(t)$ can also be written as a function of the univariate survival functions,

$$CN^{(j,j)}(t_j) = S^{(j)}(t_j)\{1 - S^{(j)}(t_j)\} \tag{4.3}$$

$$CM^{(j,j)}(t_j) = 1 - S^{(j)}(t_j). \tag{4.4}$$

We use the Kaplan-Meier estimator for the univariate survival functions and the Nelson-Aalen estimator for the univariate cumulative hazard function. Potential estimators for the bivariate survival function are discussed in Section 4.1. We use the estimator from Dabrowska et al. (1988) because it is shown to converge weakly to a Gaussian process (Gill et al., 1995), and has good performance in simulations (Cheng et al., 2007).

In many cases the correlation between the counting processes or martingales of two failure times may be more useful for describing their relationship than the covariance. Define

$RN^{(j,j')}(t_j, t_{j'}) = Cor\{N^{(j)}(t_j), N^{(j')}(t_{j'})\}$ and $RM^{(j,j')}(t_j, t_{j'}) = Cor\{M^{(j)}(t_j), M^{(j')}(t_{j'})\}$. The following equality can be used to estimate $RN^{(j,j')}(t_j, t_{j'})$,

$$RN^{(j,j')}(t_j, t_{j'}) = \frac{CN^{(j,j')}(t_j, t_{j'})}{\sqrt{S^{(j)}(t_j)\{1 - S^{(j)}(t_j)\}}\sqrt{S^{(j')}(t_{j'})\{1 - S^{(j')}(t_{j'})\}}}. \tag{4.5}$$

Similarly the following equality can be used to estimate $RM^{(j,j')}(t_j, t_{j'})$,

$$RM^{(j,j')}(t_j, t_{j'}) = \frac{CM^{(j,j')}(t_j, t_{j'})}{\sqrt{1 - S^{(j)}(t_j)}\sqrt{1 - S^{(j')}(t_{j'})}}. \tag{4.6}$$

Note that Equation (4.5) requires that $0 < S^{(j)}(t_j) < 1$ for all j and Equation (4.6) requires that $S^{(j)}(t_j) < 1$ for all j in order to be well defined. The right hand side of Equations (4.5) and (4.6) can both be consistently estimated using the Kaplan-Meier estimator and estimates for $CN^{(j,j')}(t_j, t_{j'})$ and $CM^{(j,j')}(t_j, t_{j'})$ respectively.

For a fixed set of timepoints, $\mathbf{t} = [t_1, \ldots, t_p]^T$ we define the full covariance matrix for all p counting processes as

$$CN(\mathbf{t}) = \begin{bmatrix} CN^{(1,1)}(t_1, t_1) & \ldots & CN^{(1,p)}(t_1, t_p) \\ \vdots & \ddots & \vdots \\ CN^{(p,1)}(t_p, t_1) & \vdots & CN^{(p,p)}(t_p, t_p) \end{bmatrix}, \tag{4.7}$$

and the full correlation matrix as

$$RN(\mathbf{t}) = \begin{bmatrix} RN^{(1,1)}(t_1, t_1) & \ldots & RN^{(1,p)}(t_1, t_p) \\ \vdots & \ddots & \vdots \\ RN^{(p,1)}(t_p, t_1) & \vdots & RN^{(p,p)}(t_p, t_p) \end{bmatrix}. \tag{4.8}$$

Similarly at $\mathbf{t}$ we define the full covariance matrix for the martingales as

$$CM(\mathbf{t}) = \begin{bmatrix} CM^{(1,1)}(t_1,t_1) & \dots & CM^{(1,p)}(t_1,t_p) \\ \vdots & \ddots & \vdots \\ CM^{(p,1)}(t_p,t_1) & \vdots & CM^{(p,p)}(t_p,t_p) \end{bmatrix}, \tag{4.9}$$

and the full correlation matrix as

$$RM(\mathbf{t}) = \begin{bmatrix} RM^{(1,1)}(t_1,t_1) & \dots & RM^{(1,p)}(t_1,t_p) \\ \vdots & \ddots & \vdots \\ RM^{(p,1)}(t_p,t_1) & \vdots & RM^{(p,p)}(t_p,t_p) \end{bmatrix}. \tag{4.10}$$

We will denote the estimators of each of these matrices created by using the estimator for each element of the matrix as $\widehat{CN}(\mathbf{t})$, $\widehat{RN}(\mathbf{t})$, $\widehat{CM}(\mathbf{t})$, and $\widehat{RM}(\mathbf{t})$ respectively. Note that these estimates may not be positive semi-definite. In this case there are a number of methods that can be used to transform these matrices to be positive semi-definite or positive-definite (Rousseeuw and Molenberghs, 1993). We define $\widetilde{CN}(\mathbf{t})$ to be the matrix with the same eigenvectors as $\widehat{CN}(\mathbf{t})$, and all negative eigenvalues set to some small constant. $\widetilde{RN}(\mathbf{t})$, $\widetilde{CM}(\mathbf{t})$, and $\widetilde{RM}(\mathbf{t})$ are defined analogously. If we assume that the true covariance or correlation matrices are positive-definite then it can be shown that transforming these matrices to be positive-definite will not change the limiting behavior at a fixed timepoint using results from Section 4.2.3 and Weyl's inequality.

4.2.2 Estimation of covariance in presence of right censoring and competing risks

In this section we will allow for the introduction of a single competing risk for all of the non-competing event types of interest. As a motivating example consider a study where subjects are followed over time for a number of non-competing adverse events and death acts as a competing risk for each of the adverse events. As before we will assume that each subject has p non-competing failure times, $T = [T^{(1)}, \dots, T^{(p)}]^T$. Define $\ddot{T}$ to be the failure time for the competing risk. In the case where there is more than one competing risk these can be combined

into a single competing event time. In the absence of censoring for the jth event type we will observe $\ddot{T}^{(j)} = T^{(j)} \wedge \ddot{T}$ and $\ddot{\gamma}^{(j)} = 2 - I(\ddot{T} > T^{(j)})$. If we again assume an independent censoring time, C, we will observe $\ddot{Y}^{(j)} = \ddot{T}^{(j)} \wedge C$, failure indicator $\ddot{\eta}^{(j)} = I(\ddot{T}^{(j)} \leq C)$, and event type indicator $\ddot{\epsilon}^{(j)} = \ddot{\eta}^{(j)}\ddot{\gamma}^{(j)}$. The cause specific counting process for the jth event type and lth cause is defined as $\ddot{N}_l^{(j)}(t) = I(\ddot{T}^{(j)} \leq t, \ddot{\gamma}^{(j)} = l)$ for $j = 1, \ldots, p$ and $l = 1, 2$. Similar to the previous section $\ddot{N}_l^{(j)}$ is not always observed, so we will also define $\ddot{N}_l^{(j1)}(t) = \ddot{N}_l^{(j)}(t)\ddot{\eta}^{(j)}$ to be used for estimation purposes. The cumulative cause specific hazard for the jth event type and lth cause evaluated at time t is $\ddot{\Lambda}_l^{(j)}(t) = \int_0^t \ddot{\lambda}_l^{(j)}(s)ds$ where

$$\ddot{\lambda}_l^{(j)}(t) = \lim_{\delta \to 0} \frac{1}{\delta} P(t \leq \ddot{T}^{(j)} < t + \delta, \ddot{\gamma}^{(j)} = l | \ddot{T}^{(j)} \geq t).$$

The martingale based on the decomposition of $\ddot{N}_l^{(j)}(t)$ is $\ddot{M}_l^{(j)}(t) = \ddot{N}_l^{(j)}(t) - \ddot{\Lambda}_l^{(j)}(t \wedge \ddot{T}^{(j)})$.

We will focus on the covariance between counting processes and martingales for the non-competing adverse events. That is $\ddot{CN}^{(j,j')}(t_j, t_{j'}) = Cov\{\ddot{N}_1^{(j)}(t_j), \ddot{N}_1^{(j')}(t_{j'})\}$ and $\ddot{CM}^{(j,j')}(t_j, t_{j'}) = Cov\{\ddot{M}_1^{(j)}(t_j), \ddot{M}_1^{(j')}(t_{j'})\}$. These quantities cannot be estimated using standard methods in the presence of censoring. The following equality can be used to estimate $\ddot{CN}^{(j,j')}(t_j, t_{j'})$,

$$\ddot{CN}^{(j,j')}(t_j, t_{j'}) = F_{11}^{(j,j')}(t_j, t_{j'}) - F_1^{(j)}(t_j)F_1^{(j')}(t_{j'}, \tag{4.11}$$

where $F_{kl}^{(j,j')}(t_j, t_{j'}) = P(\ddot{T}^{(j)} \leq t_j, \ddot{\gamma}^{(j)} = k, \ddot{T}^{(j')} \leq t_{j'}, \ddot{\gamma}^{(j')} = l)$ is the bivariate cause specific CIF and $F_k^{(j)}(t_j) = P(\ddot{T}^{(j)} \leq t_j, \ddot{\gamma}^{(j)} = k)$ is the univariate cause specific CIF. For notational purposes, for a bivariate function, $G(t_i, t_2)$ define

$$G(dt_1, dt_2) = G^{(1,1)}(t_1, t_2)dt_1dt_2$$

$$G(dt_1, t_2) = G^{(1,0)}(t_1, t_2)dt_1$$

$$G(t_1, dt_2) = G^{(0,1)}(t_1, t_2)dt_2$$

where $G^{(1,1)}(t_1, t_2)$ is the second partial derivative of $G(t_1, t_2)$ with respect to t_1 and t_2, $G^{(1,0)}(t_1, t_2)$ is the first partial derivative of $G(t_1, t_2)$ with respect to t_1, and $G^{(0,1)}(t_1, t_2)$ is the first partial derivative of $G(t_1, t_2)$ with respect to t_2. Using the fact that

$$P(\ddot{T}_j \leq s, \ddot{\gamma}_j = k, \ddot{T}_{j'} > t) = F_k^{(j)}(s) - F_{k,1}^{(j,j')}(s,t) - F_{k,2}^{(j,j')}(s,t),$$

the following result is obtained and can be used to estimate $C\ddot{M}^{(j,j')}(t,t)$,

$$\begin{aligned}
C\ddot{M}^{(j,j')}(t,t) = {} & \ddot{\Lambda}_1^{(j)}(t)\ddot{\Lambda}_1^{(j')}(t)\ddot{S}^{(j,j')}(t,t) + \int_0^t \ddot{\Lambda}_1^{(j)}(s)\ddot{\Lambda}_1^{(j')}(s)F_{22}^{(j,j')}(ds) \\
& + \int_0^t \int_0^t \{1 - \ddot{\Lambda}_1^{(j)}(s_1)\}\{1 - \ddot{\Lambda}_1^{(j')}(s_2)\}F_{11}^{(j,j')}(ds_1, ds_2) \\
& + \int_0^t \{1 - \ddot{\Lambda}_1^{(j)}(s)\}\{-\ddot{\Lambda}_1^{(j')}(t)\}\{F_1^{(j)}(ds) - F_{11}^{(j,j')}(ds,t) - F_{12}^{(j,j')}(ds,t)\} \\
& + \int_0^t \{1 - \ddot{\Lambda}_1^{(j')}(s)\}\{-\ddot{\Lambda}_1^{(j)}(t)\}\{F_1^{(j')}(ds) - F_{11}^{(j,j')}(t,ds) - F_{21}^{(j,j')}(t,ds)\} \\
& + \int_0^t \int_{s_1}^t \{1 - \ddot{\Lambda}_1^{(j)}(s_1)\}\{-\ddot{\Lambda}_1^{(j')}(s_2)\}F_{12}^{(j,j')}(ds_1, ds_2) \\
& + \int_0^t \int_{s_1}^t \{1 - \ddot{\Lambda}_1^{(j')}(s_1)\}\{-\ddot{\Lambda}_1^{(j)}(s_2)\}F_{12}^{(j',j)}(ds_1, ds_2),
\end{aligned} \tag{4.12}$$

where $\ddot{S}^{(j,j')}(t_j, t_{j'}) = P(\ddot{T}^{(j)} > t_j, \ddot{T}^{(j')} > t_{j'})$. The bivariate cause specific CIFs can be estimated using methods from Cheng et al. (2007), while the univariate cause specific CIFs and cumulative hazards can be estimated using methods from Kalbfleisch and Prentice (2011). Therefore estimators for $C\ddot{N}^{(j,j')}(t_j, t_{j'})$ and $C\ddot{M}^{(j,j')}(t,t)$ are obtained by consistently estimating all the elements on the right hand side of Equations (4.11) and (4.12) respectively. The variance of $\ddot{N}^{(j)}(t_j)$ and $\ddot{M}^{(j)}(t_j)$ can also be written as functions of the bivariate and univariate cause specific CIFs,

$$C\ddot{N}^{(j,j)}(t_j) = F_1^{(j)}(t_j)\{1 - F_1^{(j)}(t_j)\} \tag{4.13}$$

$$C\ddot{M}^{(j,j)}(t_j) = F_1^{(j)}(t_j). \tag{4.14}$$

As in the previous section we can extend these methods to estimate the correlation between the counting processes or martingales. Define $\ddot{RN}^{(j,j)}(t_j, t_{j'}) = Cor\{\ddot{N}_1^{(j)}(t_j), \ddot{N}_1^{(j')}(t_{j'})\}$ and $\ddot{RM}^{(j,j)}(t_j, t_{j'}) = Cor\}\ddot{M}_1^{(j)}(t_j), \ddot{M}_1^{(j')}(t_{j'})\}$. The following equality is used to estimate $\ddot{RN}^{(j,j)}(t_j, t_{j'})$,

$$\ddot{RN}^{(j,j)}(t_j, t_{j'}) = \frac{\ddot{CN}^{(j,j')}(t_j, t_{j'})}{\sqrt{F_1^{(j)}(t_j)\}1 - F_1^{(j)}(t_j)\}}\sqrt{F_1^{(j')}(t_{j'})\{1 - F_1^{(j')}(t_{j'})\}}}, \tag{4.15}$$

and

$$\ddot{RM}^{(j,j)}(t) = \frac{\ddot{CM}^{(j,j')}(t)}{\sqrt{F_1^{(j)}(t)}\sqrt{F_1^{(j')}(t)}} \tag{4.16}$$

can be used to estimate $\ddot{RM}^{(j,j)}(t_j, t_{j'})$. Equation (4.15) requires $0 < F_1^{(j)}(t_j) < 1$ for all j and Equation (4.16) requires $0 < F_1^{(j)}(t)$ for all j in order to be well defined. The full covariance and correlation matrices at the vector of timepoints $\mathbf{t}$, $\ddot{CN}(\mathbf{t})$, $\ddot{RN}(\mathbf{t})$, $\ddot{CM}(\mathbf{t})$, and $\ddot{RM}(\mathbf{t})$ can all be defined analogously to $CN(\mathbf{t})$, $RN(\mathbf{t})$, $CM(\mathbf{t})$, and $RM(\mathbf{t})$ from the previous section. The estimators of these matrices created by estimating each element of the matrix are defined as $\widehat{\ddot{CN}}(\mathbf{t})$, $\widehat{\ddot{RN}}(\mathbf{t})$, $\widehat{\ddot{CM}}(\mathbf{t})$, and $\widehat{\ddot{RM}}(\mathbf{t})$. These matrix estimates may not be positive-definite. In this case we can define $\widetilde{\ddot{CN}}(\mathbf{t})$, $\widetilde{\ddot{RN}}(\mathbf{t})$, $\widetilde{\ddot{CM}}(\mathbf{t})$, and $\widetilde{\ddot{RM}}(\mathbf{t})$ using the same techniques as in Section 4.2.1.

4.2.3 Weak convergence of covariance and correlation estimates

In this section we show weak convergence properties for the estimates of the covariance and correlation matrices for martingales and counting processes. Define PF to be the expectation of a random function, F, and $H^{(j)}(t) = I(Y^{(j)} \geq t)$. Further define $\mathbb{P}_n N^{(j1)}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i^{(j)} \leq t, \eta_i^{(j)} = 1)$ and $\mathbb{P}_n H^{(j)}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i^{(j)} \geq t)$, and for an arbitrary $q \times r$ matrix M define the function Vec(M) to be the column vector created by stacking the columns of M on top of each other. In general we will assume that $\Lambda^{(j)}$ is estimated with the Nelson-Aalen estimator and $S^{(j)}$ is estimated with the Kaplan-Meier estimator. Theorem 4.3 shows the weak convergence of the estimates of the elements of CM, CN, RM and RN:

Theorem 4.3. *Assume that the estimator of $S^{(j,j')}$ converges weakly such that*

$$\sqrt{n}\begin{bmatrix} \mathbb{P}_n N^{(11)} - PN^{(11)} \\ \vdots \\ \mathbb{P}_n N^{(p1)} - PN^{(p1)} \\ \mathbb{P}_n H^{(1)} - PH^{(1)} \\ \vdots \\ \mathbb{P}_n H^{(p)} - PH^{(p)} \\ \hat{S}^{(1,2)} - S^{(1,2)} \\ \vdots \\ \hat{S}^{(p-1,p)} - S^{(p-1,p)} \end{bmatrix} \rightsquigarrow \begin{bmatrix} Z_{N_1} \\ \vdots \\ Z_{N_p} \\ Z_{H_1} \\ \vdots \\ Z_{H_p} \\ Z_{S_{12}} \\ \vdots \\ Z_{S_{p-1,p}} \end{bmatrix}$$

in $D[0,\tau_1]^2 \times \ldots \times D[0,\tau_p]^2 \times D[0,\tau_{12}] \times \ldots \times D[0,\tau_{p-1,p}]$, where $\tau_{jj'} = (\tau_j, \tau_{j'})$, $(D[0,\tau_j], ||\cdot||_\infty)$ is the space of univariate cadlag functions of bounded variation in $[0,\tau_j]$ equipped with uniform norm, $(D[0,\tau_{jj'}], ||\cdot||_\infty)$ is the space of bivariate cadlag functions of bounded variation in $[0,\tau_{jj'}]$ equipped with uniform norm, and $[Z_{N_1}, \ldots, Z_{N_p}, Z_{H_1}, \ldots, Z_{H_p}, Z_{S_{12}}, \ldots, Z_{S_{p-1,p}}]^T$ is a mean zero tight Gaussian process. Assume that $P(Y^{(j)} > \tau_j, Y^{(j')} > \tau_{j'}) > 0$ for all $1 \leq j, j' \leq p$ and $\Lambda^{(j)} < \infty$ for $j = 1, \ldots, p$, then for any $[0,\mathbf{t}] \subset [0,\tau]$ where $\tau = [\tau_1, \ldots, \tau_p]$,

$$\sqrt{n}[Vec(\widehat{CM}) - Vec(CM)](\mathbf{t}) \rightsquigarrow Z_{CM} \tag{4.17}$$

$$\sqrt{n}[Vec(\widehat{CN}) - Vec(CN)](\mathbf{t}) \rightsquigarrow Z_{CN}, \tag{4.18}$$

where Z_{CM} and Z_{CN} are p^2 dimensional mean zero Gaussian processes. Further if $\omega = [\omega_1, \ldots, \omega_p]$, where $\omega_j < \tau_j$ for $j = 1, \ldots, p$ and $P(Y^{(j)} \leq \omega_j, Y^{(j')} \leq \omega_{j'}) > 0$ for all $1 \leq j, j' \leq p$, then for any $[\omega,\mathbf{t}] \subset [\omega,\tau]$

$$\sqrt{n}[Vec(\widehat{RM}) - Vec(RM)](\mathbf{t}) \rightsquigarrow Z_{RM} \tag{4.19}$$

$$\sqrt{n}[Vec(\widehat{RN}) - Vec(RN)](\mathbf{t}) \rightsquigarrow Z_{RN}, \tag{4.20}$$

where Z_{RM} and Z_{RN} are p^2 dimensional mean zero Gaussian processes.

The proof for this theorem is presented in Appendix B, and is done by showing that the right hand sides of Equations (4.18) and (4.17) are Hadamard differentiable mappings and applying the functional delta method (Theorem 2.8 in Kosorok (2008)). The additional assumption required for Equations (4.19) and (4.20) is needed to ensure that $RM^{(j,j)}$ and $RN^{(j,j')}$ are well defined. In addition the assumption of joint weak convergence can be shown for the Nelson-Aalen, Kaplan-Meier and Dabrowska estimators using methods similar to Lemma A.6 in the appendix of Cheng et al. (2007). Next we will show weak convergence properties for the martingale and counting process covariance and correlation matrices in the presence of competing risks. Define $\ddot{H}^{(j)}(t) = I(\ddot{Y}^{(j)} \geq t)$. Further let $\mathbb{P}_n \ddot{N}_l^{(j1)}$ and $\mathbb{P}_n \ddot{H}^{(j)}$ be defined analogously to $\mathbb{P}_n N^{(j1)}$ and $\mathbb{P}_n H^{(j)}$. We will assume that $\ddot{\Lambda}_1$ is estimated using a Nelson-Aalen style estimator and $F_1^{(j)}$ is estimated using methods from Kalbfleisch and Prentice (2011), and $F_{kl}^{(j,j')}$ is estimated using methods from Cheng et al. (2007). This leads to Theorem 4.4 which shows weak convergence for the estimates of the elements of $\ddot{CM}$, $\ddot{CN}$, $\ddot{RM}$, and $\ddot{RN}$:

Theorem 4.4. *Assume that the estimator of $\ddot{S}^{(j,j')}$ converges weakly such that*

$$\sqrt{n}\begin{bmatrix} \mathbb{P}_n\ddot{N}_1^{(11)} - P\ddot{N}_1^{(11)} \\ \vdots \\ \mathbb{P}_n\ddot{N}_1^{(p1)} - P\ddot{N}_1^{(p1)} \\ \mathbb{P}_n\ddot{H}^{(1)} - P\ddot{H}^{(1)} \\ \vdots \\ \mathbb{P}_n\ddot{H}^{(p)} - P\ddot{H}^{(p)} \\ \hat{\ddot{S}}^{(1,2)} - \ddot{S}^{(1,2)} \\ \vdots \\ \hat{\ddot{S}}^{(p-1,p)} - \ddot{S}^{(p-1,p)} \end{bmatrix} \rightsquigarrow \begin{bmatrix} Z_{\ddot{N}_1} \\ \vdots \\ Z_{\ddot{N}_p} \\ Z_{\ddot{H}_1} \\ \vdots \\ Z_{\ddot{H}_p} \\ Z_{\ddot{S}_{12}} \\ \vdots \\ Z_{\ddot{S}_{p-1,p}} \end{bmatrix}$$

in $D[0,\ddot{\tau}_1]^2\times\ldots\times D[0,\ddot{\tau}_p]^2\times D[0,\ddot{\tau}_{12}]\times\ldots\times D[0,\ddot{\tau}_{p-1,p}]$*, where* $\ddot{\tau}_{jj'}=(\ddot{\tau}_j,\ddot{\tau}_{j'})$*,* $(D[0,\ddot{\tau}_j],||\cdot||_\infty)$ *is the space of univariate cadlag functions of bounded variation in* $[0,\ddot{\tau}_j]$ *equipped with uniform norm,* $(D[0,\ddot{\tau}_{jj'}],||\cdot||_\infty)$ *is the space of bivariate cadlag functions of bounded variation in* $[0,\ddot{\tau}_{jj'}]$ *equipped with uniform norm, and* $[Z_{\ddot{N}_1},\ldots,Z_{\ddot{N}_p},Z_{\ddot{H}_1},\ldots,Z_{\ddot{H}_p},Z_{\ddot{S}_{12}},\ldots,Z_{\ddot{S}_{p-1,p}}]^T$ *is a mean zero tight Gaussian process. Assume that* $P(\ddot{Y}^{(j)}>\tau_j,\ddot{Y}^{(j')}>\tau_{j'})>0$ *for all* $1\leq j,j'\leq p$ *and* $\ddot{\Lambda}_k^{(j)}<\infty$ *for* $j=1,\ldots,p$ *and* $k=1,2$*, then for any* $[0,\ddot{t}]\subset[0,\ddot{\tau}]$ *where* $\ddot{\tau}=[\ddot{\tau}_1,\ldots,\ddot{\tau}_p]$*,*

$$\sqrt{n}[Vec(\widehat{\ddot{CM}})-Vec(\ddot{CM})](t)\rightsquigarrow Z_{\ddot{CM}} \tag{4.21}$$

$$\sqrt{n}[Vec(\widehat{\ddot{CN}})-Vec(\ddot{CN})](t)\rightsquigarrow Z_{\ddot{CN}}, \tag{4.22}$$

where $Z_{\ddot{CM}}$ *and* $Z_{\ddot{CN}}$ *are* p^2 *dimensional mean zero Gaussian processes. Further if* $\ddot{\omega}=[\ddot{\omega}_1,\ldots,\ddot{\omega}_p]$*, where* $\ddot{\omega}_j<\ddot{\tau}_j$ *for* $j=1,\ldots,p$ *and* $P(\ddot{Y}^{(j)}\leq\ddot{\omega}_j,\ddot{Y}^{(j')}\leq\ddot{\omega}_{j'})>0$ *for all* $1\leq j,j'\leq p$*, then for any* $[\ddot{\omega},\ddot{t}]\subset[\ddot{\omega},\ddot{\tau}]$

$$\sqrt{n}[Vec(\widehat{\ddot{RM}})-Vec(\ddot{RM})](t)\rightsquigarrow Z_{\ddot{RM}} \tag{4.23}$$

$$\sqrt{n}[Vec(\widehat{\ddot{RN}})-Vec(\ddot{RN})](t)\rightsquigarrow Z_{\ddot{RN}}, \tag{4.24}$$

where $Z_{\ddot{RM}}$ *and* $Z_{\ddot{RN}}$ *are* p^2 *dimensional mean zero Gaussian processes.*

The proof of Theorem 4.4 is provided in Appendix B, and uses similar methods as the proof for Theorem 4.3.

4.3 PCA methods for right censored data

PCA transforms a set of variables into linearly uncorrelated variables. For a $p\times 1$ dimensional random vector X the first principal component direction, v_1, is the $p\times 1$ vector for which Var(v_1^TX) is maximized subject to the constraint $||v_1||_2=1$. The jth principal component direction is the $p\times 1$ vector for which Var(v_j^TX) is maximized subject to the constraints $||v_j||_2=1$ and $v_j^Tv_{j'}=0$ for $j'<j$. The principal components can be shown to be the eigenvectors of the covariance matrix of X. The solutions will not be unique if there are repeated eigenvalues. The

principal components, $v_j^T X$, are linearly uncorrelated. The proportion of the variance of the data explained by the jth principal component is equal to $\frac{\lambda_j}{\sum_{i=1}^{p} \lambda_i}$, where λ_i is the ith eigenvalue of the covariance matrix of X. Estimates PCA are found through eigendecomposition of the sample covariance matrix or SVD of the mean centered data.

In the presence of right censoring it is not possible to non-parametrically estimate the principal components for T, the $p \times 1$ dimensional vector of failure times. This is because it is not possible to estimate the covariance matrix using standard methods without making strong assumptions on the form of the joint distribution. This is the case with or without the presence of competing risks. When there are no competing risks, instead of estimating the principal components for T we consider the principal components for $N(\mathbf{t}) = [N^{(1)}(t_1), \ldots, N^{(p)}(t_p)]^T$ or $M(\mathbf{t}) = [M^{(1)}(t_1), \ldots, M^{(p)}(t_p)]^T$. The principal directions of $N(\mathbf{t})$ are the eigenvectors of $CN(\mathbf{t})$ and the principal directions of $M(\mathbf{t})$ are the eigenvectors of $CM(\mathbf{t})$. Similarly in the presence of competing risks we consider $\ddot{N}(\mathbf{t}) = [\ddot{N}_1^{(1)}(t_1), \ldots, \ddot{N}_1^{(p)}(t_p)]^T$, which has principal directions equal to the eigenvectors of $\ddot{CN}(\mathbf{t})$ and $\ddot{M}(\mathbf{t}) = [\ddot{M}_1^{(1)}(t_1), \ldots, \ddot{M}_1^{(p)}(t_p)]^T$ which has principal directions equal to the eigenvectors of $\ddot{CM}(\mathbf{t})$. In all cases the correlation matrix can be used instead of the covariance matrix to get a scaled version PCA.

We obtain estimates for principal directions for $N(\mathbf{t}), M(\mathbf{t}), \ddot{N}(\mathbf{t})$, and $\ddot{M}(\mathbf{t})$ using the eigenvectors of consistent estimates of the relevant covariance or correlation matrices, which were derived in Section 4.2. Consistent estimates of the proportion of variance explained are based on the corresponding eigenvalues. If we assume that the eigenvalues of $CM(\mathbf{t}), CN(\mathbf{t}), \ddot{CM}(\mathbf{t})$, or $\ddot{CN}(\mathbf{t})$ are unique then the estimates of the corresponding principal components will be consistent and asymptotically normal based on the results from Theorem 4.5. This is also true if we use the correlation matrices instead of the covariance matrices.

Theorem 4.5. *Assume that $t_1, \ldots, t_n$ are iid realizations of the $p \times 1$ random vector T, with joint distribution F_T. Assume that Σ is a $p \times p$ positive-definite function of F_T and that $\widehat{\Sigma}_n$ is a positive-definite function of $(t_1, \ldots, t_n)$. Define $v_{\Sigma i}$ to be the ith eigenvector of Σ and $\xi_{\Sigma i}$ to be the ith eigenvalue of Σ, and likewise $\hat{v}_{\hat{\Sigma} in}$ to be the ith eigenvector of $\widehat{\Sigma}_n$ and $\hat{\xi}_{\hat{\Sigma} in}$*

to be the ith eigenvalue of $\widehat{\Sigma}_n$. Assume that $\sqrt{n}[Vec(\widehat{\Sigma}_n) - Vec(\Sigma)] \rightarrow_d N(0, \Psi_\Sigma)$, where Ψ_Σ is some positive definite matrix. If $\Xi = diag(\xi_{\Sigma 1}, \ldots, \xi_{\Sigma p})$, $\hat{\Xi}_n = diag(\hat{\xi}_{\Sigma 1n}, \ldots, \hat{\xi}_{\Sigma pn})$, $V = [v_{\Sigma 1}, \ldots, v_{\Sigma p}$, and $\hat{V} = [\hat{v}_{\hat{\Sigma} 1n}, \ldots, \hat{v}_{\hat{\Sigma} pn}]$. Then $\sqrt{n}[Vec(\hat{V}) - Vec(V)] \rightarrow_d N(0, \Psi_V)$ and $\sqrt{n}[Vec(\hat{\Xi}) - Vec(\Xi)] \rightarrow_d N(0, \Psi_\Xi)$, where Ψ_V and Ψ_Ξ are positive-semidefinite matrices

The proof for Theorem 4.5 can be found in Appendix B. When combined with results from Section 4.2 this shows that estimates for the principal component vectors based on $\widetilde{CM}(\mathbf{t}), \widetilde{CN}(\mathbf{t}), \widetilde{C\ddot{M}}(\mathbf{t}), \widetilde{C\ddot{N}}(\mathbf{t})$, or the corresponding correlation matrix estimates are consistent and asymptotically normal.

The principal component scores can also be estimated for those subjects who have not been censored by time point $\mathbf{t}$. In order for this to be the case it must be that either $C > t_j$, or $\eta^{(j)} = 1$ for $j = 1, \ldots, p$. In this case $N(\mathbf{t}) = [N^{(1)}(t_1), \ldots, N^{(p)}(t_p)]^T$, the entire vector of failure counting processes at time point $\mathbf{t}$ will be observed. Define $\hat{v}_{cnj}(\mathbf{t})$ to be the jth eigenvector of $\widehat{CN}(\mathbf{t})$. Then the estimate of the jth principal component score is $\hat{v}_{cnj}(\mathbf{t})^T N(\mathbf{t})$. In the presence of a competing risk a similar estimate can be made using $\ddot{N}_1(\mathbf{t}) = [\ddot{N}_1^{(1)}(t_1), \ldots, \ddot{N}_1^{(p)}(t_p)]^T$ and the jth eigenvector of $\widehat{C\ddot{N}}(\mathbf{t})$, $\hat{v}_{\ddot{cn}j}$. In the case where a subject is not censored by time point $\mathbf{t}$, the entire vector of martingales, $M(\mathbf{t}) = [M^{(1)}(t_1), \ldots, M^{(p)}(t_p)]^T$ can be consistently estimated by plugging in a consistent estimate of $\Lambda^{(j)}(t_j \wedge T^{(j)})$ for $j = 1, \ldots, p$. If $\hat{v}_{cmj}(\mathbf{t})$ is the jth eigenvector of $\widehat{CM}(\mathbf{t})$, then the estimate of the jth principal component score is $\hat{v}_{cmj}(\mathbf{t})^T \hat{M}(\mathbf{t})$. Similar calculations can be done in the case of a competing risk. In this case the full vector of martingales, $\ddot{M}_1(\mathbf{t}) = [\ddot{M}_1^{(1)}(t_1), \ldots, \ddot{M}_1^{(p)}(t_p)]^T$, can be estimated by plugging in consistent estimates of $\ddot{\Lambda}_1^{(j)}(t_j \wedge T^{(j)})$, for $j = 1, \ldots, p$. The jth eigenvector of $\widehat{C\ddot{M}}(\mathbf{t})$ is defined as $\hat{v}_{c\ddot{m}_j}(\mathbf{t})$, and the estimate of the jth principal component score is $\hat{v}_{\widehat{cm}j}(\mathbf{t})^T \hat{M}_1(\mathbf{t})$. All of the estimates for principal component scores based on the scaled counting processes or martingales can be estimated in the same way using the eigenvectors of the correlation matrices. In Section 4.5 we show an example where the principal component scores estimates can be used as covariates in a Cox PH model. This is possible because the form of the partial likelihood for the Cox PH model only requires the covariate values for those subjects who have not been

censored by a given timepoint. This allows for principal component scores for censored time to event variables to be used as covariates in a Cox PH model similar to how principal component scores are used for principal component regression.

4.4 Simulation results

Simulations were conducted to examine the estimation of principal components of $M(\mathbf{t}), N(\mathbf{t}), \ddot{M}_1(\mathbf{t}), \ddot{N}_1(\mathbf{t})$ focusing on results using correlation matrices. Data sets were simulated with $p = 4$ and $p = 8$. First we simulate $W = [W^{(1)}, \ldots, W^{(p)}]$ with a multivariate normal distribution were $W^{(j)}$ has mean zero and standard deviation one for $j = 1, \ldots, p$. When $p = 4$ W has covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0.6 & 0.4 & 0.2 \\ 0.6 & 1 & 0.6 & 0.4 \\ 0.4 & 0.6 & 1 & 0.6 \\ 0.2 & 0.4 & 0.6 & 1 \end{bmatrix}, \tag{4.25}$$

and when $p = 8$, W has covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0.6 & 0.4 & 0.2 & 0 & 0 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.4 & 0.2 & 0 & 0 & 0 \\ 0.4 & 0.6 & 1 & 0.6 & 0.4 & 0.2 & 0 & 0 \\ 0.2 & 0.4 & 0.6 & 1 & 0.6 & 0.4 & 0.2 & 0 \\ 0 & 0.2 & 0.4 & 0.6 & 1 & 0.6 & 0.4 & 0.2 \\ 0 & 0 & 0.2 & 0.4 & 0.6 & 1 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0.2 & 0.4 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0 & 0 & 0.2 & 0.4 & 0.6 & 1 \end{bmatrix}. \tag{4.26}$$

$T = [T^{(1)}, \ldots, T^{(p)}]$ is defined to be a transformation of W, such that $T^{(j)} \sim \text{Exponential}(1)$. Specifically $T^{(j)} = -\ln(1 - \Phi(W^{(j)}))$, where $\Phi(\cdot)$ is the CDF of a standard normal distribution.

The competing risk setting has a similar set up. W has the same distribution as above and $[W^T, \ddot{W}]^T$ has a multivariate normal distribution and the correlation between $\ddot{W}$ and $W^{(j)}$ is 0.1 for $j = 1, \ldots, p$. T is still defined as the same transformation of W and $\ddot{T} = -\ln(1 - \Phi(\ddot{W}))$.

In all settings two different censoring distributions are considered. The first is $C_1 \sim$ Uniform$(0, 4)$, and the second is $[1/4 \cdot C_2] \sim$ Beta$(1.5, 6.5)$. For all settings a sample size of $n = 200$ and $n = 1000$ are simulated. In the setting without competing risks the average censoring rate for each event was 25% for both $p = 4$ and $p = 8$ when using C_1 and 47% for when using C_2. In the semi-competing risk setting the censoring rate was 13% when using C_1 and 34% when using C_2. For both censoring schemes in the competing risk setting the competing and non-competing events were equally likely to be observed. The lower censoring rate in the competing risk setting is due to the fact that the censoring time must come before both the non-competing event time and the competing event time in order for the subject to be censored.

For each data set $\widetilde{RN}(\mathbf{t})$, $\widetilde{RM}(\mathbf{t})$, $\widetilde{\ddot{R}N}(\mathbf{t})$, and $\widetilde{\ddot{R}M}(\mathbf{t})$ were estimated using the Dabrowska estimator for bivariate survival functions, the Kaplan-Meier estimator for univariate survival functions, and the Nelson-Aalen estimator for all cumulative hazard functions. In order to ensure all estimated covariance and correlation matrices are positive-definite a minimum eigenvalue of 0.001 is used. Matrices are estimated at $\mathbf{t} = [1, 1, 1, 1, \ldots]^T$ and $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$. The true covariance matrices are calculated empirically by simulating a single data set with 500,000 subjects and no censoring which allows for standard correlation and covariance estimation methods to be used. In the competing risk setting the cumulative hazard necessary to calculate the martingales is estimated using the Nelson-Aalen estimator based on the 500,000 simulated subjects. In the setting without competing risks the cumulative hazard is known based on the distribution of $T^{(j)}$. The true principal component directions are calculated as the eigenvectors of the true correlation and covariance matrices. More information on the true correlation matrices can be found in Appendix B.

For the ith simulated data and jth principal component direction the angle in radians between the estimated direction, $\hat{v}_{ij}$ and the true direction, v_j, and is calculated as

$$\text{Angle}(\hat{v}_{ij}, v_j) = \cos^{-1}\left[\frac{|\hat{v}_{ij}^T v_j|}{||\hat{v}_{ij}||_2 \cdot ||v_j||_2}\right].$$

For each setting the bias and standard deviation for the jth principal direction is calculated as the empirical mean and standard deviation of the angle between the estimated direction and true direction for all 1000 simulated data sets. Table 4.6 reports the the average angle and standard deviation for the principal component directions using eigendecomposition of $\widetilde{RN}(\mathbf{t})$ and $\widetilde{RM}(\mathbf{t})$ for both censoring distributions and $\mathbf{t} = [1, 1, 1, 1, \ldots]^T$. Results for $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$ are reported in Appendix B. As expected the average angle is larger in higher dimensions. For both $p = 4$ and $p = 8$ the sample size increases the mean and standard deviation of the angles between the estimated direction and true direction across all 1,000 simulated data sets decreases. Specifically for $n = 1000$ the average angle between the estimated first direction and true first direction is less than 0.10 radians, which is just under six degrees, except for when using the counting process correlation matrix when $p = 8$ and using C_2 as the censoring distribution. Unsurprisingly the average angle across all sample sizes and dimensions is higher for C_2 which has a higher censoring rate. However even for this higher censoring rate performance is still reasonable. Even 0.36 radians which is the highest average angle between the true and estimated first principal direction at $n = 200$ is just over 20 degrees. At $\mathbf{t} = [1, 1, 1, 1, \ldots]^T$ the estimates using the martingale correlation matrix outperform the estimates using the estimates using the counting process correlation matrix across all sample sizes, censoring distributions and dimensions. This is also true for when $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$ as shown in Appendix B. Table 4.7 reports the average angle and standard deviation for the principal component directions using eigendecomposition of $\widetilde{\ddot{RN}}(\mathbf{t})$ and $\widetilde{\ddot{RM}}(\mathbf{t})$ for both censoring distributions when $\mathbf{t} = [1, 1, 1, 1, \ldots]^T$. Results for $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$ are presented in Appendix B. The average angle between the true direction and estimated direction in the competing risk setting tends to be

higher than the average angle for between the estimated and true direction in the setting with no competing risks. In addition the relative effect of increases in dimension, sample size, or censoring rate is similar between the competing risk and no competing risk settings. Simulations based on estimates of the martingale and counting process covariance matrices for both the non-competing risk and competing risk setting shown similar results and are not included.

One difference between the competing risk and no competing risk settings is that the average angle for first direction is lower for counting process PCA than martingale PCA when $p = 8$ in the competing risk setting. This is because $\ddot{N}_1^{(j)}(t)$ and $\ddot{N}_1^{(j')}(t)$ will tend to be correlated even when $T^{(j)}$ and $T^{(j')}$ are uncorrelated. Consider the fact that in order for $\ddot{N}_1^{(j)}(t)$ to equal one it must be the case that both $T^{(j)} \leq t$ and $T^{(j)} \leq \ddot{T}$. If $T^{(j)}$ and $T^{(j')}$ are independent and uncorrelated $P(T^{(j')} \leq \ddot{T} | T^{(j)} \leq \ddot{T})$ will be higher than $P(T^{(j')} \leq \ddot{T})$. This in turn means that when $T^{(j)}$ and $T^{(j')}$ are independent $P(\ddot{N}_1^{(j')}(t) = 1 | P(\ddot{N}_1^{(j)}(t) = 1)$ will be larger than $P(\ddot{N}_1^{(j')}(t) = 1)$, leading to a positive correlation between $\ddot{N}_1^{(j)}(t)$ and $\ddot{N}_1^{(j')}(t)$. The effect on $\ddot{M}_1^{(j)}(t)$ and $\ddot{M}_1^{(j')}(t)$ is not as large because it will not introduce correlation between $\ddot{\Lambda}_1^{(j)}(t \wedge \ddot{T}^{(j)})$ and $\ddot{\Lambda}_1^{(j')}(t \wedge \ddot{T}^{(j')})$ to the same degree. Because of this in our simulation set up the separation between the first and second eigenvalues for $\ddot{RN}(\mathbf{t})$ is larger than for $\ddot{RM}(\mathbf{t})$, which leads to more precise estimates of the first principal direction. More information on this issue, and specifically the correlation between $\ddot{N}^{(j)}(t)$ and $\ddot{N}^{(j')}(t)$, even when $T^{(j)}$ and $T^{(j')}$ are uncorrelated can be found in Appendix B.

Table 4.6: Average (SD) of the angle in radians between the true and estimated PCA directions based on counting process and martingale correlation matrices with no competing risks

$\mathbf{t} = [1, 1, 1, 1, \ldots]^T$									
		C_1				C_2			
		Counting Process		Martingale		Counting Process		Martingale	
		n=200	n=1000	n=200	n=1000	n=200	n=1000	n=200	n=1000
4 Dim	PC 1	0.07 (0.04)	0.03 (0.02)	0.05 (0.03)	0.02 (0.01)	0.12 (0.07)	0.05 (0.02)	0.08 (0.04)	0.03 (0.02)
	PC 2	0.21 (0.13)	0.09 (0.05)	0.14 (0.08)	0.06 (0.03)	0.33 (0.22)	0.13 (0.07)	0.21 (0.12)	0.09 (0.05)
	PC 3	0.61 (0.41)	0.36 (0.29)	0.53 (0.39)	0.26 (0.22)	0.73 (0.41)	0.48 (0.37)	0.63 (0.40)	0.37 (0.31)
	PC 4	0.59 (0.41)	0.36 (0.30)	0.52 (0.40)	0.26 (0.23)	0.69 (0.42)	0.48 (0.37)	0.61 (0.41)	0.37 (0.31)
8 Dim	PC 1	0.23 (0.15)	0.09 (0.04)	0.17 (0.10)	0.07 (0.03)	0.36 (0.27)	0.13 (0.08)	0.25 (0.20)	0.09 (0.05)
	PC 2	0.28 (0.16)	0.11 (0.04)	0.21 (0.11)	0.09 (0.03)	0.45 (0.27)	0.17 (0.08)	0.31 (0.20)	0.11 (0.05)
	PC 3	0.33 (0.13)	0.13 (0.04)	0.24 (0.09)	0.10 (0.03)	0.51 (0.24)	0.20 (0.06)	0.33 (0.13)	0.14 (0.04)
	PC 4	0.75 (0.33)	0.32 (0.14)	0.57 (0.29)	0.23 (0.09)	0.94 (0.33)	0.52 (0.27)	0.77 (0.34)	0.34 (0.17)
	PC 5	1.10 (0.31)	0.92 (0.36)	1.00 (0.33)	0.75 (0.36)	1.16 (0.29)	1.02 (0.33)	1.09 (0.32)	0.86 (0.36)
	PC 6	1.14 (0.30)	1.02 (0.34)	1.07 (0.32)	0.88 (0.38)	1.17 (0.28)	1.07 (0.33)	1.13 (0.30)	0.97 (0.36)
	PC 7	1.10 (0.30)	0.96 (0.36)	1.01 (0.33)	0.74 (0.35)	1.14 (0.29)	1.07 (0.33)	1.09 (0.31)	0.91 (0.35)
	PC 8	0.91 (0.35)	0.61 (0.32)	0.79 (0.35)	0.44 (0.26)	1.02 (0.32)	0.80 (0.34)	0.93 (0.35)	0.63 (0.32)

Table 4.7: Average (SD) of the angle in radians between the true and estimated PCA directions based on counting process and martingale correlation matrices in the presence of a competing risk

$\mathbf{t} = [1, 1, 1, 1, \ldots]^T$									
		C_1				C_2			
		Counting Process		Martingale		Counting Process		Martingale	
		n=200	n=1000	n=200	n=1000	n=200	n=1000	n=200	n=1000
4 Dim	PC 1	0.08 (0.05)	0.03 (0.02)	0.07 (0.04)	0.03 (0.01)	0.11 (0.08)	0.05 (0.02)	0.10 (0.06)	0.04 (0.02)
	PC 2	0.33 (0.20)	0.15 (0.08)	0.18 (0.12)	0.08 (0.04)	0.45 (0.27)	0.21 (0.12)	0.27 (0.17)	0.11 (0.06)
	PC 3	0.66 (0.38)	0.40 (0.32)	0.58 (0.39)	0.31 (0.26)	0.77 (0.40)	0.50 (0.35)	0.69 (0.39)	0.44 (0.35)
	PC 4	0.64 (0.38)	0.40 (0.32)	0.57 (0.39)	0.31 (0.26)	0.74 (0.39)	0.49 (0.35)	0.66 (0.40)	0.43 (0.35)
8 Dim	PC 1	0.16 (0.09)	0.07 (0.03)	0.24 (0.20)	0.09 (0.05)	0.23 (0.16)	0.09 (0.03)	0.35 (0.28)	0.13 (0.08)
	PC 2	0.37 (0.19)	0.16 (0.07)	0.29 (0.20)	0.11 (0.05)	0.50 (0.26)	0.21 (0.09)	0.42 (0.28)	0.15 (0.08)
	PC 3	0.44 (0.20)	0.19 (0.07)	0.30 (0.11)	0.12 (0.04)	0.63 (0.29)	0.26 (0.10)	0.43 (0.20)	0.18 (0.06)
	PC 4	0.80 (0.32)	0.42 (0.20)	0.65 (0.31)	0.29 (0.13)	0.98 (0.32)	0.60 (0.29)	0.86 (0.34)	0.44 (0.22)
	PC 5	1.12 (0.31)	0.98 (0.35)	1.07 (0.32)	0.83 (0.37)	1.14 (0.29)	1.05 (0.32)	1.12 (0.30)	0.96 (0.34)
	PC 6	1.14 (0.29)	1.05 (0.35)	1.10 (0.31)	0.97 (0.36)	1.15 (0.29)	1.09 (0.32)	1.13 (0.29)	1.04 (0.33)
	PC 7	1.14 (0.30)	1.01 (0.34)	1.08 (0.32)	0.86 (0.37)	1.16 (0.29)	1.09 (0.32)	1.13 (0.30)	0.99 (0.35)
	PC 8	1.00 (0.32)	0.68 (0.32)	0.88 (0.34)	0.53 (0.29)	1.09 (0.30)	0.84 (0.34)	1.00 (0.33)	0.72 (0.34)

4.5 MPACT trial

The MPACT trial was a clinical trial that ran from 2009 to 2013 in which 861 patients with metastatic pancreatic cancer were cancer were randomized to be treated with either the standard of care, gemcitabine, or a novel medication, paclitaxel. The data for the 430 patients randomized to standard of care is available through Project Data Sphere®. Further details on the MPACT trial have been published at Von Hoff et al. (2013).

During the course of this trial, there were nine adverse events that occurred in at least 50 of the patients: abdominal pain, anemia, constipation, decreased appetite, fatigue, nausea, neutropenia, thrombocytopenia, and vomiting. For each adverse event the failure time is the time from beginning of observation to the first occurrence of that event. Patients who left the study before having a given adverse event due to death or progression of disease were considered to have a competing event. Patients who left the study before having a given adverse event for any other reason were considered to be censored. Overall the censoring percentage for each event type ranged between 28% for neutropenia and 35% for constipation. The percentage of patients who had the event of interest ranged from 13% for constipation to 26% for anemia.

We used the martingale correlation matrix because the martingales contain information on when an event occurred in addition to whether it occurred by a given time. $R\ddot{M}(\mathbf{t})$ was estimated between $\mathbf{t} = [30, 30, \ldots]^T$ and $\mathbf{t} = [360, 360, \ldots]^T$ in the increment of one day. Day 360 was chosen as a final time point because by that time 420 of the 430 patients had left the study. The principal component loadings based on $R\ddot{M}(\mathbf{t})$ at $\mathbf{t} = [360, 360, \ldots]^T$ are presented in Table 4.8. Figure 4.1 shows the directions for the first two principal components plotted over time between day 30 and 360. The line type is based on the following grouping of event types. Constitutional (C): Fatigue; Gastrointestinal (G): Abdominal pain, constipation, decreased appetite, nausea, vomiting; Hematologic (H): Anemia, neutropenia, thrombocytopenia.

Based on Figure 4.1 it is apparent that the largest loadings in the first principal component are gastrointestinal and constitutional events, and the loadings for these events all go in the same direction. The largest loadings for the second principal component are the hematologic events

Table 4.8: Principal component directions at day 360 and proportion of variance explained for each principal component using estimates based on martingale covariance matrix

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Abdominal pain	0.38	-0.21	-0.18	0.03	0.05	0.73	-0.41	-0.26	-0.04
Anemia	0.09	0.42	0.56	0.51	-0.39	0.26	-0.02	0.15	-0.05
Constipation	0.25	0.26	0.36	-0.16	0.80	0.10	0.09	0.21	-0.11
Decreased appetite	0.42	-0.16	-0.42	0.33	-0.03	-0.09	0.11	0.68	-0.16
Fatigue	0.43	0.00	-0.06	0.47	0.17	-0.28	0.28	-0.53	0.34
Nausea	0.46	0.14	0.06	-0.35	-0.29	-0.21	0.12	-0.27	-0.65
Neutropenia	-0.10	0.55	-0.45	-0.14	-0.06	0.40	0.54	-0.04	0.09
Thrombocytopenia	-0.01	0.60	-0.34	0.09	0.10	-0.30	-0.65	-0.04	-0.01
Vomiting	0.44	0.10	0.14	-0.48	-0.27	-0.07	-0.09	0.22	0.64
Proportion Variance	0.26	0.15	0.11	0.10	0.10	0.08	0.08	0.06	0.05

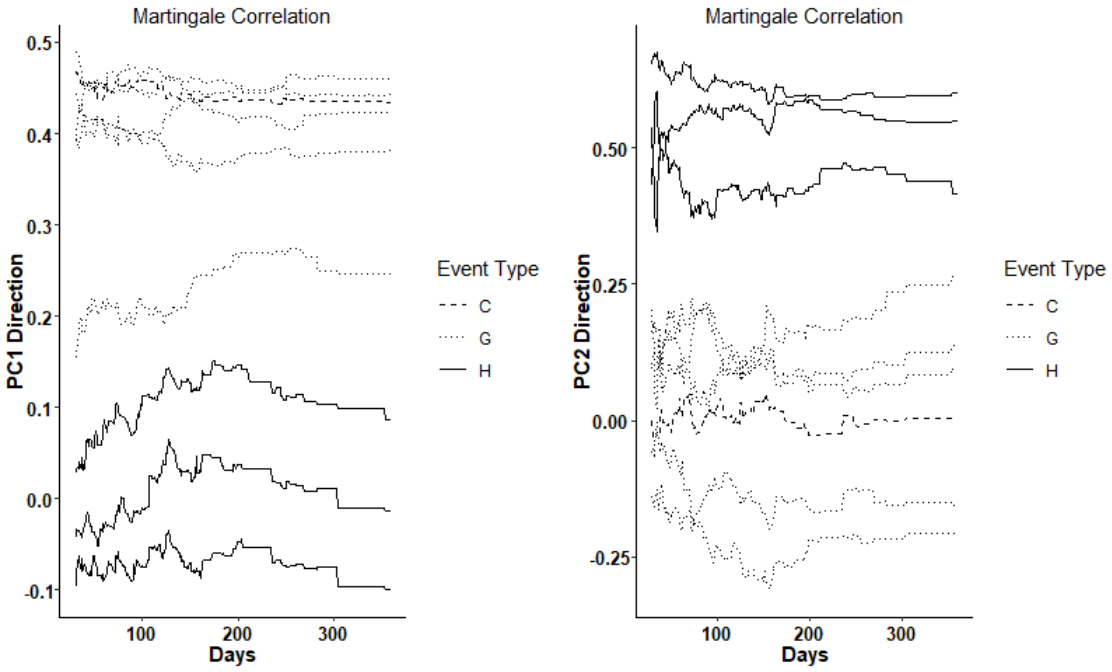

Figure 4.1: Principal component direction loadings from day 30 to 360 for first two principal components using martingale correlation matrix estimates. Line types indicate constitutional, gastrointestinal and hematologic event types.

which also all go in the same direction. Therefore the first principal component is driven by the occurrence of gastrointestinal events and fatigue, and the second principal component is driven by the occurrence of hematologic events. This shows that martingales for gastrointestinal and constitutional events tend to be correlated with martingales for other gastrointestinal and constitutional events, while martingales for hematologic events tend to be correlated with martingales for other hematologic events. Together the first two principal components explain close to 40% of the total variance, which is consistent over time.

Unlike the first two principal components, the third through ninth principal components do not have a straightforward interpretation. In addition, as can be seen in the proportion of variance explained in Table 4.8, the proportion of variance explained for the third through seventh principal components are similar. This is true for all the time points estimated, and caused issues with potentially crossing eigenvalues. In order to make them comparable over time, instead of ordering the principal directions based on proportion of variance explained the order was chosen to minimize the sum of the angles between principal directions across time. First day 360 was chosen as a reference date, and all the principal directions were ordered in descending order of the associated eigenvalues. We can define $\hat{v}_{jt}$ to be the jth principal direction for $\mathbf{t} = [t, t, \ldots]^T$. For all days other than 360 the ordering of the principal directions was chosen in order to minimize $\sum_{i=1}^{9} \text{Angle}(\hat{v}_{jt}, \hat{v}_{j360})$. This meant for some days the ordering of the principal directions would be changed. A figure of the principal directions using this method for ordering can be found in the Appendix B.

In addition to simple analysis of the principal component loadings we used the first two principal component scores in a Cox PH model with death or progression of disease as the outcome. Because of the changing principal component loadings and difficulty of interpreting coefficients for time varying covariates we used a landmark analysis, with landmark times at 30, 60, 90 and 120 days. At each of the landmark times the martingales for each of the nine events were estimated as an indicator of whether the adverse event had happened by the landmark time minus the estimated cumulative hazard evaluated at the minimum of the observed event time

and the landmark time. The cumulative hazard for each of the adverse event types was estimated using the Nelson-Aalen estimator. The principal component scores were calculated as the inner product of the martingale scores and the principal component loadings at the landmark time. The principal component scores for the first two principal components were then used as covariates in a Cox PH model with time until death or progression of disease past the landmark time as the outcome of interest. In addition to the principal component scores, the age in years, sex, and Karnofsky score were included as covariates. The Karnofsky score is a numeric measure of the general well being of cancer patients, and ranged from 70 to 100 within the sample (Karnofsky et al., 1948). One subject was dropped from the model for day 30 due to a missing Karnofsky score.

The estimated hazard ratio for the first principal component score is greater than one, although the associated p-values are greater than 0.05 for all four landmark times. Recall that the major loadings onto the first principal component are gastrointestinal and constitutional events. This implies that among patients who survive up to a given landmark time, those who have experienced more and earlier gastrointestinal and constitutional events have a higher estimated hazard of death or progression of disease after the landmark, holding all other covariates constant. However given that the p-values for the PC 1 coefficients are all above 0.05 further study is necessary to see if this result holds across a larger population or is just due to random variation in the MPACT trial sample.

Similarly the estimated hazard ratio for the second principal component score is below one for all landmark times after day 30, although the associated p-values are greater than 0.05 for all three of these landmark times. Using the same logic as above, this implies that among patients who have survived to day 60, 90 or 120, those patients who have experienced more and earlier hematologic events have a lower hazard for death or progression of disease after the landmark time, holding all other covariates constant. As above, given the p-values for the PC 2 score regression coefficients further study is necessary to see if this result holds across a larger population.

These results are consistent with the idea that many gastrointestinal and constitutional adverse are the result of cancer itself and may be a precursor to progression of disease, while hematologic adverse events are toxicities of the treatment and may indicate either a greater exposure to drug (due to increased absorption or decreased degradation) or a greater sensitivity to its effects.

One potential issue with this analysis is that the principal component loadings are calculated using the entire sample, while the landmark analysis considers only those patients who are still in the study at the landmark date. As a sensitivity analysis the principal component loadings and corresponding scores were estimated at each landmark time using only those patients still in the study at the time. For both the first and second principal component the correlation between the principal component scores using the loadings from the full sample and landmark sample had a correlation above 0.92 and landmark days 30, 60, 90, and 120. Therefore we believe it is reasonable to use the principal loadings using the full sample. In particular at the later times this may give more precise estimates since the sample size for the landmark analysis is less than half the full sample.

Table 4.9: Estimated hazard ratios for landmarked Cox PH models using PC score estimates as covariates

	Day 30		Day 60		Day 90		Day 120	
	HR	P-Value	HR	P-Value	HR	P-Value	HR	P-Value
PC 1	1.12	0.42	1.08	0.61	1.29	0.11	1.20	0.34
PC 2	1.05	0.79	0.79	0.19	0.72	0.07	0.74	0.14
Age	0.99	0.11	0.99	0.21	0.99	0.22	0.99	0.46
SEX:Male	1.00	1.00	1.02	0.90	0.95	0.76	0.94	0.78
Karnofsky Index	0.99	0.09	1.00	0.63	1.00	0.95	1.01	0.58
n	355		249		201		139	

Although landmarked analysis makes the interpretation of the coefficients from the Cox PH model more straightforward, it only analyzes those patients who survived up to the landmark time. For this reason at each landmark time we compared the average PC 1 and PC 2 scores for patients who survived up to the landmark time, and those who died or had their disease progress

before the landmark time. The martingale PC scores were estimated as described in Section 4.3, and the PC loadings used to calculate the PC scores were estimated at the specified landmark time. Patients who were censored before the landmark date were not included in the analysis. We also tested for differences in the distribution of the scores for those who died or progressed before the landmark time and those still alive at the landmark time using a Wilcoxon rank-sum test.

We found that the patients who died or progressed before day 30 have a higher average score on PC 1 than those who are still alive after day 30 (β = 0.73, p=6.43E-8), Table 4.10). The direction of this difference stays the same for the other three landmark times, although the magnitude of the difference is not as large for days 60 (β = 0.36, p=7.39E-8), 90 (β = 0.28,p=1.11E-7) or 120 (β = 0.26,p=1.33E-7). That patients who die earlier have a higher PC 1 score is consistent with the notion that gastrointestinal events and fatigue, the main drivers of PC 1, are caused by cancer per se and are early indicators of death or progression of disease. At day 30 (β = 0.04,p=0.10), 60 (β = 0.09, p = 1.73E-3) and 90 ($\beta = 0.06$, p = 4.80E-3) the average PC 2 score of patients who died before the given landmark time is higher than the patients who survived past the landmark time. However the difference between the two groups is not as large as it is for PC 1 at the same landmark times. The direction of the differences changes at day 120, although it is not statistically significant ($\beta = -0.02, p = 017$). The smaller difference between patients who died before the given landmark times and those who survived up to the landmark times for PC 2 scores relative to PC 1 scores is consistent with hematologic adverse events being at least in part reactions to toxicities of the treatment and are therefore not early indicators of death or progression of disease in the same way as gastrointestinal events or fatigue.

Table 4.10: Comparison of average PC 1 scores and PC 2 scores for patients who died/progressed before landmark date compared to those who survived up to landmark date

	Day 30			Day 60		
	PC 1	PC 2	N	PC 1	PC 2	N
Died/Progressed	0.69	0.03	21	0.26	0.06	89
Survived	-0.04	-0.01	356	-0.10	-0.03	249
P-value	6.43E-08	0.10		7.39E-08	1.73E-03	
	Day 90			Day 120		
	PC 1	PC 2	N	PC 1	PC 2	N
Died/Progressed	0.18	0.03	120	0.12	-0.01	170
Survived	-0.10	-0.03	201	-0.14	0.01	139
P-value	1.11E-07	4.80E-03		1.33E-07	0.17	

4.6 Discussion

In this paper we show how PCA can be defined for multivariate survival data in the presence of censoring by using either the counting processes or the corresponding martingales defined by each event type. We build on previous results for bivariate survival data to show how to estimate the full covariance and correlation matrices for either the counting processes or martingales at a given time point. In addition we are able to extend this to the semi-competing risk setting where each of the different event types are subject to a competing risk as well as an independent censoring time. For both the standard censoring only setting and the semi-competing risk setting we are able to show that the estimators for the martingale and counting process covariance or correlation matrices converge to a mean zero Gaussian process when properly normalized.

We also show that the loadings for the principal components based on the martingales and counting processes can be consistently estimated through eigendecomposition of the corresponding covariance or correlation matrix. The corresponding estimates are shown to be consistent and asymptotically normal. A subject's principal component score can only be estimated up to the time they are censored, however this still allows for principal component scores to be used as coefficients in Cox PH regression.

The usefulness of this method is shown using data from the comparator arm of the MPACT trial for patients with pancreatic cancer. We estimate the principal components based on the martingale correlation matrix for nine different adverse events experienced by patients in the trial, and are able to define medically relevant groupings of these events. One area for future research is to further consider ordering and potential changing of principal components over time. In the case where two or more principal components have similar eigenvalues the ordering may change over time due to random noise. It may be of interest to investigate how to identify when changes in principal component loadings and eigenvalues over time are due to random noise or an actual change in true underlying covariance or correlation matrix.

CHAPTER 5: ROBUST ESTIMATION OF MULTISET CANONICAL CORRELATION ANALYSIS

5.1 Introduction

Many modern data sets have many different modalities or sets of variables all measured across the same subjects. Because of this many methods have been developed to concisely summarize the associations between these different modalities. One of the oldest methods for summarizing the associations between two sets of variables is canonical correlation analysis (CCA) (Hotelling, 1936). CCA finds the linear combinations of each of the two sets of variables, called canonical directions, with maximal Pearson correlation. Estimation of CCA using the sample covariance or correlation matrix typically works well in the low dimension high sample size setting. When there are more than two sets of variables there are a number of ways to extend CCA to multiset CCA (mCCA) (Kettenring, 1971; Nielsen, 2002). Recently a number of extensions of mCCA to the high-dimensional setting have been proposed (Witten and Tibshirani, 2009; Asendorf, 2015; Xu et al., 2012; Fu et al., 2017). There are also a number of methods that try to summarize both the associations between the different sets of variables, as well as the independent structure within each set of variables (Lock et al., 2013; Feng et al., 2018; Gaynanova and Li, 2019). Many of these methods have similarities to different formulations of the mCCA problem.

One drawback to these methods is that they are not robust to extreme outliers or data coming from heavy tailed or skewed distributions. We propose a version of high-dimensional mCCA for these settings that is more robust than existing methods. In order to do this we build on existing mCCA methods by using robust correlation estimation techniques. To make our correlation estimation robust we use properties of Kendall's tau estimator in the transelliptical

family of distributions (Embrechts et al., 2002; Klüppelberg and Kuhn, 2009; Liu et al., 2012). A transformation of Kendall's tau can be used to estimate the correlation matrix for this family of distributions. This estimator has been shown to have good performance in the high-dimensional setting (Han and Liu, 2017). We show that using this correlation estimator to estimate mCCA directions can lead to better performance than the sample Pearson correlation estimator when data come from heavy tailed elliptical or transelliptical distributions. Another advantage is that it is invariant to monotone transformations of the data, so if data have a skewed marginal distributions it is not necessary to transform the data in order to find meaningful linear associations. We also show that if the data do not come from a multivariate normal distribution existing methods to test for informative canonical directions, including permutation testing, do not control for type I error. Because of this we propose a cross-validation based technique for identifying informative mCCA directions.

The rest of this paper is structured as follows. Section 5.2 gives an in-depth review of mCCA techniques. Section 5.3 reviews high-dimensional mCCA and proposes our high-dimensional robust mCCA method. Section 5.4 compares performance of our method to existing methods in simulation studies. Section 5.5 gives an example of how our method can be applied using executive function (EF) test scores and brain imaging data in six-year-old children. Section 5.6 gives concluding remarks.

5.2 Multi-set Canonical Correlation Analysis

CCA is a common technique for looking at the association between two sets of variables. For a $p_1 \times 1$ random vector $X_1 = [X_{11}, \ldots, X_{1p_1}]^T$ and a $p_2 \times 1$ random vector $X_2 = [X_{21}, \ldots, X_{2p_2}]^T$ the first canonical variables, $U_1^{(1)} = a_1^{(1)T} X_1$ and $U_2^{(1)} = a_2^{(1)T} X_2$, are the linear combinations of X_1 and X_2 where $\text{Cor}(U_1^{(1)}, U_2^{(1)})$ is maximized. The canonical directions are the $p_1 \times 1$ vector $a_1^{(1)}$ and the $p_2 \times 1$ vector $a_2^{(1)}$, which can be made unique up to a change in sign of both vectors by requiring that $\text{Var}(U_1^{(1)}) = \text{Var}(U_2^{(1)}) = 1$. Further canonical variables and directions are defined with the additional constraint that they are uncorrelated with all previous canonical variables. It can be shown that the population canonical directions are

solutions to an eigendecomposition based on a function of the joint covariance matrix of X_1 and X_2 (Hotelling, 1936), and estimates of the canonical directions are typically based on the same eigendecomposition based on a consistent estimate of the joint covariance matrix of X_1 and X_2.

Many modern data sets include more than two sets of variables, modalities, or views. For this reason it is desirable to define mCCA for more than two sets of variables. There are a number of ways to extend CCA to more than two sets of variables, both in terms of the optimization and constraint functions. In order to define mCCA we will use the following notation. Define $X_i = [X_{i1}, \ldots, X_{ip_i}]^T$ to be a $p_i \times 1$ dimensional random vector for $i = 1, \ldots, d$. The covariance matrix for $X = [X_1^T, \ldots, X_d^T]^T$ can be defined as

$$\Sigma_X = \begin{bmatrix} \Sigma_{11} & \ldots & \Sigma_{1d} \\ \vdots & \ddots & \vdots \\ \Sigma_{d1} & \ldots & \Sigma_{dd} \end{bmatrix},$$

where Σ_{ii} is the covariance matrix for X_i. The jth canonical variable for the ith set of variables can be defined as $U_i^{(j)} = a_i^{(j)T} X_i$ where $a_i^{(j)}$ is a $p_i \times 1$ vector. There are a number of mCCA formulations based on maximizing a function of the covariance matrix of $U^{(j)} = [U_1^{(j)}, \ldots, U_d^{(j)}]$,

$$\Sigma_{U^{(j)}} = \begin{bmatrix} a_1^{(j)T}\Sigma_{11}a_1^{(j)} & \ldots & a_1^{(j)T}\Sigma_{1d}a_d^{(j)} \\ \vdots & \ddots & \vdots \\ a_d^{(j)T}\Sigma_{d1}a_1^{(j)} & \ldots & a_d^{(j)T}\Sigma_{dd}a_d^{(j)} \end{bmatrix}.$$

The five most common objective functions for mCCA were first summarized by Kettenring (1971) and are:

1. **SUMCOR**: Maximize $\sum_{i=1}^d \sum_{k=1}^d a_i^{(j)T}\Sigma_{ik}a_k^{(j)}$.
2. **SSQCOR**: Maximize $\sum_{i=1}^d \sum_{k=1}^d (a_i^{(j)T}\Sigma_{ik}a_k^{(j)})^2$.
3. **MAXVAR**: Maximize the largest eigenvalue of $\Sigma_{U^{(j)}}$.
4. **MINVAR**: Minimize the smallest eigenvalue of $\Sigma_{U^{(j)}}$.

5. **GENVAR**: Minimize the determinant of $\Sigma_{U^{(j)}}$.

In general it is assumed that all elements of $U^{(j)}$ are uncorrelated with all elements of $U^{(j')}$ for $j > j'$. In order to make the $a_i^{(j)}$'s unique up to a sign change additional constraints are needed. The most obvious extension from standard two set CCA is $\text{Var}(U_1^{(j)}) = \ldots = \text{Var}(U_d^{(j)}) = 1$. However other constraints are possible and sometimes desirable. Nielsen (2002) considers the following constraints:

1. **NORM**: $a_i^{(j)T} a_i^{(j)} = 1$ for $i = 1, \ldots, d$.
2. **AVGNORM**: $\sum_{i=1}^{d} a_i^{(j)T} a_i^{(j)} = 1$
3. **VAR**: $\text{Var}(U_1^{(j)}) = \ldots = \text{Var}(U_d^{(j)}) = 1$.
4. **AVGVAR**: $\sum_{i=1}^{d} \text{Var}(U_i^{(j)}) = 1$.

The combination of objective functions and constraints results in 20 possible combinations that can be used for mCCA. Asendorf (2015) gives a useful overview of the 20 different combinations and shows which have well defined solutions, which of these have closed form solutions, and which need to be solved numerically. The two most commonly used formulations are the SUMCOR objective function subject to the AVGVAR constraint and the MAXVAR objective function subject to the VAR constraint. In both of these cases the canonical directions can be shown to be solutions to a generalized eigensystem. If we define the matrix

$$\Sigma_D = \begin{bmatrix} \Sigma_{11} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \Sigma_{dd} \end{bmatrix},$$

and assume Σ_D is invertible, then the solutions for the canonical directions using the SUMCOR/AVGVAR formulation can be shown to be related to the solutions for the generalized eigensystem,

$$\Sigma_D^{-1} \Sigma_X a = \lambda a. \tag{5.27}$$

If we define $\lambda^{(j)}$ to be the jth eigenvalue for this system, then the corresponding generalized eigenvector, $a^{(j)} = [a_1^{(j)T}, \ldots, a_d^{(j)T}]^T$, is equal to the jth canonical direction vector when properly normalized (Nielsen, 2002; Asendorf, 2015; Parra, 2018). In order to properly normalize this vector it is helpful to consider the transformation $\tilde{a}^{(j)} = \Sigma_D^{(1/2)} a^{(j)}$ which makes Equation (5.27) equal to

$$\Sigma_D^{-1/2} \Sigma_X \Sigma_D^{-1/2} \tilde{a} = \lambda \tilde{a}. \tag{5.28}$$

If $||\tilde{a}^{(j)}||_2 = 1$ then $a_S^{(j)} = \Sigma_D^{-1/2} \tilde{a}^{(j)}$ will be the solution to the SUMCOR/AVGVAR formulation of mCCA. $a_S^{(j)}$ can be written as $a_S^{(j)} = [a_{S1}^{(j)T}, \ldots, a_{Sd}^{(j)T}]^T$ where $a_{Si}^{(j)}$ is the $p_i \times 1$ vector corresponding to the ith set of variables, X_i. $\lambda^{(j)}$ will also be the SUMCOR value for the jth canonical direction for this formulation of mCCA.

The solution to the MAXVAR/VAR formulation of mCCA can also be shown to be based on the solution to Equation (5.28). Define $\tilde{a}_i^{(j)}$ to be a $p_i \times 1$ vector for $i = 1, \ldots, d$ such that $\tilde{a}^{(j)} = [\tilde{a}_1^{(j)T}, \ldots, \tilde{a}_d^{(j)T}]$. Then

$$a_M^{(j)} = \Lambda_{\tilde{a}^{(j)}}^{-1} \Sigma_D^{-1/2} \tilde{a}^{(j)},$$

where $\Lambda_{\tilde{a}^{(j)}} = \text{blkdiag}(||a_i^{(j)}||_2 I_{p_i})$, is the solution to the MAXVAR/VAR formulation of mCCA (Kettenring, 1971; Asendorf, 2015). As with $a_S^{(j)}$, $a_M^{(j)}$ can be written as $a_M^{(j)} = [a_{M1}^{(j)T}, \ldots, a_{Md}^{(j)T}]^T$ where $a_{Mi}^{(j)}$ is the $p_i \times 1$ vector corresponding to the ith set of variables, X_i. The MAXVAR value for the jth canonical direction is equal to $\lambda^{(j)}$ for the MAXVAR/VAR formulation of mCCA.

The covariance matrix for the SUMCOR/AVGVAR canonical variates can be written as

$$\Sigma_{US}^{(j)} = \begin{bmatrix} a_{S1}^{(j)T} \Sigma_{11} a_{S1}^{(j)} & \ldots & a_{S1}^{(j)T} \Sigma_{1d} a_{Sd}^{(j)} \\ \vdots & \ddots & \vdots \\ a_{Sd}^{(j)T} \Sigma_{d1} a_{S1}^{(j)} & \ldots & a_{Sd}^{(j)T} \Sigma_{dd} a_{Sd}^{(j)} \end{bmatrix}.$$

The covariance matrix for the MAXVAR/VAR canonical variates can be written as

$$\Sigma_{UM}^{(j)} = \begin{bmatrix} a_{M1}^{(j)T}\Sigma_{11}a_{M1}^{(j)} & \dots & a_{M1}^{(j)T}\Sigma_{1d}a_{Md}^{(j)} \\ \vdots & \ddots & \vdots \\ a_{Md}^{(j)T}\Sigma_{d1}a_{M1}^{(j)} & \dots & a_{Md}^{(j)T}\Sigma_{dd}a_{Md}^{(j)} \end{bmatrix}.$$

As noted previously the sum of the elements of $\Sigma_{US}^{(j)}$ and the maximum eigenvalue of $\Sigma_{UM}^{(j)}$ are both equal to $\lambda^{(j)}$, which is the jth eigenvalue for the generalized eigensystem in Equation (5.28). In both cases a value of $\lambda^{(j)} > 1$ indicates linear association between the canonical variates, and an informative canonical direction. This means that there are the same number of informative canonical directions using either the SUMCOR/AVGVAR or MAXVAR/VAR formulations of mCCA. This relationship can be very useful because there are instances where the SUMCOR/AVGVAR formulation may be more useful, and other instances where the MAXVAR/VAR solution may be more useful. We will see one example of this in Section 5.3.4 where the SUMCOR/AVGVAR is more useful for identifying informative canonical directions in the high dimensional setting using cross-validation.

5.3 High dimensional robust mCCA

5.3.1 High-dimensional CCA methods

Although both the SUMCOR/AVGVAR and MAXVAR/VAR have simple closed form solutions when the estimate of Σ_D is invertible, in the case where one or more of the sets of variables is of high dimension this may not be the case. Further, even if the estimate of Σ_D is invertible, in the high dimensional setting the solution may be very unstable and lead to imprecise estimates. Further as dimension of the sets of variables increases the estimated canonical correlations will tend to increase due to random noise. For this reason additional methods are needed to estimate mCCA directions in the high-dimensional setting, and to ensure that the correlations in the estimated directions are not due to random noise.

There are a number of methods that can be used for mCCA in the high-dimensional setting, we will highlight two of them. One is to add a penalty term to the entries of the canonical directions, $a_i^{(j)}$. This technique has been well studied in the standard CCA setting with two sets of variables (Parkhomenko et al., 2009; Witten et al., 2009; Wilms and Croux, 2015, 2016; Yoon et al., 2018). Extensions to the mCCA setting using this technique are less common, with Witten and Tibshirani (2009) being the primary example. Two of the drawbacks to the penalized mCCA proposed by Witten and Tibshirani (2009) are that it does not account for correlation within the sets of variables, and it only uses the SUMCOR/NORM formulation.

Another method for high-dimensional mCCA is to first reduce the dimension of each of the sets of variables using principal components analysis (PCA) or equivalently the singular value decomposition (SVD) and then using the closed form solutions from the low dimensional setting. Asendorf (2015) uses this method with the MAXVAR/VAR formulation of mCCA. Another closely related method is angle-based joint and individual variation explained (Feng et al., 2018). This is very closely related to using the SUMCOR/AVGVAR formulation of mCCA after using PCA/SVD to reduce the dimension of each of the sets of variables. Other examples include Xu et al. (2012) and Fu et al. (2017) which combine reducing the rank of each set of variables using SVD and adding a penalty or regularization term to the CCA directions. As noted in Section 5.1, an issue with existing methods is that they are not robust if the data have extreme outliers, or come from heavy tailed or skewed marginal distribution. In addition methods such as Asendorf (2015) and Feng et al. (2018) assume a model in which the random error term for every variable within a set has the same variance. For data sets in which the variables within a set have variances that differ by an order of magnitude or all variables are standardized to have variance equal to one, such an assumption is not reasonable.

In Section 5.5 we analyze data from the Early Brain Development Study (EBDS) in which all of these issues are present. We look at the relationship between cognitive test scores, average radial diffusivity (RD) in 20 white matter tracts using Diffusion Tensor Imaging (DTI), and grey matter (GM) volume for 88 different brain regions as defined by the Automated Anatomic

Labelling (AAL) atlas using structural Magnetic Resonance Imaging. Across all three of these modalites there are variables which show excess skewness or kurtosis relative to a normal distribution. The cognitive tests are not all measured on the same scale so they have different variances unless standardized and the brain regions are different sizes so the GM volume by region also has different variances, sometimes by an order of magnitude. For this reason a robust version of high-dimensional mCCA using the correlation matrix instead of the covariance matrix is necessary. In order to make our method robust we use properties of Kendall's tau in a family of distributions known as the transelliptical distribution (Embrechts et al., 2002; Klüppelberg and Kuhn, 2009; Liu et al., 2012) to estimate a version of the correlation matrix that is robust to both skewness and heavy tailed distributions. For such distributions additional techniques for selecting the number of principal components and identifying informative mCCA directions are also necessary.

5.3.2 Robust correlation estimation in transelliptical family

Using properties of the transelliptical family of distributions we will define a robust estimator for high dimensional mCCA. In order to define the transelliptical family of distributions it is necessary to first define the elliptical family of distributions:

Definition 5.5 (Elliptical Distributions)**.** *A $p \times 1$ random vector Y is considered to be elliptical if for some $p \times 1$ vector μ_Y, some $p \times p$ positive semi-definite matrix Σ_Y, and a function $\psi_Y[0, \infty) \to \mathbb{R}$, the characteristic function, Φ, satisfies $\Phi_{Y-\mu_Y}(t) = \psi(t^T \Sigma_Y t)$ for all $p \times 1$ vectors t. In this case we would say that Y is a $p \times 1$ dimensional elliptically distributed random variable, which we can note as $Y \sim ED_p(\mu_Y, \Sigma_Y, \psi_Y)$*

Common elliptical distributions include multivariate normal, multivariate t, and multivariate logistic distributions. This family of distributions is useful for CCA because all linear combinations of elliptically distributed random variables are still elliptically distributed, and Σ_Y which is equal to the covariance matrix of Y up to a scalar when second moments exist. Even when moments do not exist Σ_Y defines the linear associations between the elements of Y. A useful extension of the elliptical family of distributions is the transelliptical family of distributions

(Embrechts et al., 2002; Klüppelberg and Kuhn, 2009; Liu et al., 2012). A definition of the transelliptical family of distributions is given below:

Definition 5.6 (Transelliptical distributions)**.** *A $p \times 1$ dimensional random vector Z has a transelliptical distribution if there exists a positive-semidefinite matrix Σ_{hZ} with all ones along the diagonal, a function $\psi_{hZ} : [0, \infty) \to \mathbb{R}$, and a set of functions $h_{Z1}, \ldots, h_{Zp}$ where $h_{Zi} : \mathbb{R} \to \mathbb{R}$ is a monotone increasing function for $i = 1, 2, \ldots, p$ such that $h_Z(Z) = [h_{Z1}(Z_1), \ldots, h_{Zp}(Z_p)]^T \sim ED_p(0, \Sigma_{hZ}, \psi_{hZ})$. The random variable Z is a $p \times 1$ dimensional transelliptically distributed random variable, denoted as $Z \sim TE_p(h_Z, 0, \Sigma_{hZ}, \psi_Z)$.*

An equivalent definition is any multivariate distribution with continuous marginal distributions and a copula that comes from a multivariate elliptical distribution. Because transelliptical distributions allow for monotonic marginal transformations of elliptical distributions this can include heavily skewed marginal distributions. When considering methods such as CCA and Z is transelliptically distributed, it can be more useful to consider the elliptically distributed $W = h_Z(Z)$ rather than Z itself. This is because as mentioned above linear combinations of elliptically distributed random variables are still elliptically distributed. Further the parameter Σ_{hZ} describes the linear associations between W, and not Z, if the transformation functions h_Z are nonlinear. In fact the correlation or covariance matrix of Z itself may not be fully informative of the relationship between the elements of Z if the marginal distributions are heavily skewed or otherwise non-elliptical.

For this reason it is desirable to estimate Σ_{hZ} rather than the correlation matrix of Z. As shown in Liu et al. (2012) a consistent estimate of every element of Σ_{hZ} can be obtained through transformations of consistent estimates of Kendall's tau for all pairs of variables in Z. Assume that $Z = [Z_1^T, \ldots, Z_d^T]^T$ is a $\sum_{i=1}^{d} p_i \times 1$ dimensional transelliptically distributed random vector where $Z_i = [Z_{i1}, \ldots, Z_{ip_i}]^T$, is a $p_i \times 1$ random vector, and $\tilde{Z}$ is an identically distributed copy of Z. Then the Kendall's tau between Z_{ij} for $i = 1, \ldots, d$ and $j = 1, \ldots, p_i$ and Z_{kl} for

$k = 1 \ldots, d$ and $l = 1, \ldots, p_d$ is equal to

$$\tau^{(Z_{ij}, Z_{kl})} = E[\text{sign}(Z_{ij} - \tilde{Z}_{ij})(Z_{kl} - \tilde{Z}_{kl})]$$

For n iid copies of Z, $z_1, \ldots, z_n$ a consistent and asymptotically normal estimate of $\tau^{(Z_{ij}, Z_{kl})}$ is

$$\hat{\tau}_n^{(Z_{ij}, Z_{kl})} = \frac{1}{\binom{n}{2}} \sum_{1 \leq r < s \leq n} \sum \text{sign}(z_{rij} - z_{sij}) \text{sign}(z_{rkl} - z_{skl}),$$

where z_{rij} is the rth copy of Z_{ij}. Within the transelliptical family there is a known correspondence between the, $\sigma_{ij,kl}$, the entry of Σ_{hZ} corresponding to Z_{ij} and Z_{kl} and the Kendall's tau coefficient between Z_{ij} and Z_{kl}. Specifically $\sigma_{ij,kl} = \frac{2}{\pi} \arcsin\left(\tau^{(Z_{ij}, Z_{kl})}\right)$. This is a straightforward extension of the same result from Lindskog et al. (2003) for elliptical distributions. Based on this a consistent estimate of $\sigma_{ij,kl}$ is

$$\hat{\sigma}_{ij,kl} = \frac{2}{\pi} \arcsin\left(\hat{\tau}_n^{(Z_{ij}, Z_{kl})}\right).$$

As shown by Liu et al. (2012) an estimate of Σ_{hZ}, $\hat{\Sigma}_{hZ}$, can be obtained by estimating all the elements in this fashion. Importantly this estimator does not require the estimation of h_Z or ϕ_{hZ} because Kendall's tau is invariant to monotone increasing transformations of the data. Other techniques for estimating Σ_{hZ} will require estimation or assumptions for the form of h_Z and ϕ_{hZ}. Because $\hat{\Sigma}_{hZ}$ is not guaranteed to be positive semidefinite it is sometimes necessary to map $\hat{\Sigma}_{hZ}$ to a positive semidefinite matrix which we will denote as $\tilde{\Sigma}_{hZ}$. In the case where $d > 2$ and p_i is large for some i in $1, \ldots, d$ we can use $\tilde{\Sigma}_Z$ to get robust estimates of high-dimensional mCCA directions. This can be considered to be a latent version of high-dimensional mCCA where we look for the most meaningful relationships between $h_Z(Z) = W$, rather than Z itself.

5.3.3 Latent high-dimensional mCCA in the transelliptical family

We will assume that the observed data, $Z = [Z_1^T, \ldots, Z_d^T]^T$ is a $\sum_{i=1}^d p_i \times 1$ dimensional transelliptically distributed random vector. This means that there is a set of $\sum_{i=1}^d p_i$ monotonic

transformations, $h_Z(\cdot)$, such that $h_Z(Z) = W$ is a $\sum_{i=1}^{d} p_i \times 1$ dimensional elliptically distributed random vector with scatter matrix that has all 1's along the diagonal. W can be written as $W = [W_1^T, \ldots, W_d^T]^T$ where W_i is a $p_i \times 1$ dimensional elliptically distributed random vector. The scatter matrix for W can be written as

$$\Sigma_W = \begin{bmatrix} \Sigma_{W11} & \ldots & \Sigma_{W1d} \\ \vdots & \ddots & \vdots \\ \Sigma_{Wd1} & \ldots & \Sigma_{Wdd} \end{bmatrix},$$

and is equivalent to Σ_{hZ}. Note that each element of Σ_W can be consistently estimated using transformations of Kendall's tau between all pairs of variables in Z as described in the previous section. Define B_{ij} to be the eigenvector corresponding to the jth eigenvalue of Σ_{Wii} and $B_i^{(r_i)} = [B_{i1}, \ldots, B_{ir_i}]$ to be the $p_i \times r_i$ matrix that has the first r_i eigenvectors of Σ_{Wii} along the columns. We can then define $V_{ij} = B_{ij}^T W_i$ to be the jth principal component score for W_i and $V_i^{(r_i)} = B_i^{(r_i)T} W_i$ to be the $r_i \times 1$ vector of principal component scores for W_i. Define $V^{(r)} = [V_1^{(r_1)T}, \ldots, V_d^{(r_d)T}]^T$, then the covariance matrix for $V^{(r)}$ can be defined as

$$\Sigma_{V^{(r)}} = \begin{bmatrix} B_1^{(r_1)T}\Sigma_{W11}B_1^{(r_1)} & \ldots & B_1^{(r_1)T}\Sigma_{W1d}B_d^{(r_d)T} \\ \vdots & \ddots & \vdots \\ B_d^{(r_d)T}\Sigma_{Wd1}B_1^{(r_1)} & \ldots & B_d^{(r_d)T}\Sigma_{Wdd}B_d^{(r_d)} \end{bmatrix}$$

and

$$\Sigma_{V^{(r)}D} = \begin{bmatrix} B_1^{(r_1)T}\Sigma_{W11}B_1^{(r_1)} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & B_d^{(r_d)T}\Sigma_{Wdd}B_d^{(r_d)} \end{bmatrix},$$

It is not possible to get estimates of $V^{(r)}$ without knowing or estimating h_Z, but $B_i^{(r_i)}$, $\Sigma_{V^{(r)}}$, and $\Sigma_{V^{(r)}D}$ can all be estimated using eigendecomposition of the estimate of Σ_W, which can be estimated using transformations of pairwise Kendall's tau estimates of Z.

If r_i is sufficiently small for $i = 1, \ldots, d$ then it is reasonable to assume that $\Sigma^{-1}_{V^{(r)}D}$ exists. Therefore we can solve for the SUMCOR/AVGVAR and MAXVAR/VAR canonical directions of $V^{(r)}$ through the generalized eigensystem

$$\Sigma^{-1/2}_{V^{(r)}D}\Sigma_{V^{(r)}}\Sigma^{-1/2}_{V^{(r)}D}\tilde{a}_{V^{(r)}} = \lambda\tilde{a}_{V^{(r)}}. \tag{5.29}$$

The jth SUMCOR/AVGVAR canonical direction for $V^{(r)}$ is $a^{(j)}_{SV^{(r)}} = \Sigma^{-1/2}_{V^{(r)}D}\tilde{a}^{(j)}_{V^{(r)}}$ and the jth MAXVAR/VAR canonical direction for $V^{(r)}$ is $a^{(j)}_{MV^{(r)}} = \Lambda_{\tilde{a}^{(j)}_{V^{(r)}}}\Sigma^{-1/2}_{V^{(r)}D}\tilde{a}^{(j)}_{V^{(r)}}$. These can all by estimated using the transformed Kendall's tau matrix estimator for Σ_W.

If we assume that r_i is large enough for $i = 1, \ldots, d$ such that $\text{Cov}(V_{ij}, V_{i'j'}) = 0$ where $r_i < j \leq p_i$, $1 \leq j' \leq p_{i'}$, $1 \leq i, i' \leq d$ and $i \neq i'$, then the SUMCOR/AVGVAR and MAXVAR/VAR canonical directions of W can be solved from the canonical directions for V. Define $a^{(j)}_{SV^{(r)}} = [a^{(j)T}_{SV^{(r)}1}, \ldots, a^{(j)T}_{SV^{(r)}d}]^T$ where $a^{(j)T}_{SV^{(r)}i}$ is $r_i \times 1$ vector corresponding to $V^{(r_i)}_i$. $a^{(j)T}_{MV^{(r)}i}$ can be defined analogously. In this case the jth SUMCOR/AVGVAR canonical direction for W_i is $a^{(j)}_{SWi} = B^{(r_i)}_i a^{(j)}_{SV^{(r)}i}$, and the jth MAXVAR/VAR canonical direction for W_i is $a^{(j)}_{MWi} = B^{(r_i)}_i a^{(j)}_{MV^{(r)}i}$. These equalities can be used to obtain estimates of $a^{(j)}_{SW}$ and $a^{(j)}_{MW}$, by plugging in estimates of $B^{(r_i)}_1$, $a^{(j)}_{SV}$, and $a^{(j)}_{MV}$.

5.3.4 Selecting number of principal directions for each set and identifying informative canonical correlation vectors

One of the primary challenges in estimating high dimensional mCCA for the transelliptical family is selecting the appropriate dimension of $B^{(r_1)}_i$ for $i = 1, \ldots, d$. If the estimated rank is too small, then the assumption that all principal components greater than r_i in the ith set are uncorrelated with all other sets will not be reasonable. This means that the estimated canonical correlations and vectors for W will not use all the correlations between the different sets of variables. Alternatively if r_i is too large then estimates will be very unstable and have a high variance. The optimal selection would be the smallest possible r_i for $i = 1, \ldots, d$ such that the assumption $\text{Cov}(V_{ij}, V_{i'j'}) = 0$ where $r_i < j \leq p_i$, $1 \leq j' \leq p_{i'}$, $1 \leq i, i' \leq d$, and $i \neq i'$ is met. However, finding the optimal r_i is not a straightforward task. This is related to the

more general problem of selecting the number of principal components when using PCA for dimension reduction. As mentioned previously there are a number of methods for selecting a number of principal components, however these tend to be either heuristic methods such as looking for an "elbow" in the Scree plot, or make assumptions that are not reasonable for all transelliptical distributions. Specifically methods such as those in Choi et al. (2017) and Lock et al. (2020) assume that each variable within a set has an random error term with the same variance. As noted in Section 5.3.1 this assumption is not reasonable for many data sets.

Because of these challenges we recommend a data driven approach to selecting r_i using cross-validation. Rather than focusing on finding some optimal low rank approximation of each data set based, we focus on finding the number of principal components to maximize the cross-validated SUMCOR value based on the SUMCOR/AVGVAR mCCA formulation. Cross-validation is necessary because increasing the number of principal components for each set can only increase the estimated SUMCOR value. Cross-validation allows us to help determine if this is due to increased signal or just added noise due to adding more directions. As a first step a lower bound, l_i, and upper bound, u_i, of possible values for r_i should be chosen for $i = 1, \ldots, d$. This means we will consider all values of r_i such that $l_i \leq r_i \leq u_i$. Each set l_i and u_i can be chosen by looking at scree plots, eigenvalues, and the cumulative percentage of variance explained by each principal component. Define $\mathcal{R} =: \{(r_1, \ldots, r_d) \in \mathbb{N} \times \ldots \times \mathbb{N} | l_1 \leq r_1 \leq u_1, \ldots, l_d \leq r_d \leq u_d\}$. This means there will be $(u_1 - l_1 + 1) \times \ldots \times (u_d - l_d + 1)$ possible combinations in $\mathcal{R}$. For every $r \in \mathcal{R}$ the k-fold cross-validated SUMCOR value using the SUMCOR/AVGVAR canonical direction can be calculated. For a given r, define $\hat{a}_{SWik}^{(jr)}$ to be the jth estimated SUMCOR/AVGVAR direction for the ith set of variables for the kth training fold. Define $\tilde{\Sigma}_{Wii'k}$ to be the estimated correlation matrix between the ith and $i'th$ sets of variables using the kth test set. Note that $\tilde{\Sigma}_{Wii'k}$ depends only on the split of the data and not r. The cross-validated covariance matrix for

the SUMCOR canonical variates is

$$\Sigma_{USk}^{(jr)} = \begin{bmatrix} \hat{a}_{SW1k}^{(jr)T}\tilde{\Sigma}_{W11k}\hat{a}_{SW1k}^{(jr)} & \dots & \hat{a}_{SW1k}^{(jr)T}\tilde{\Sigma}_{W1dk}\hat{a}_{SWdk}^{(jr)} \\ \vdots & \ddots & \vdots \\ \hat{a}_{SWdk}^{(jr)T}\tilde{\Sigma}_{Wd1k}\hat{a}_{SW1k}^{(jr)} & \dots & \hat{a}_{SWdk}^{(jr)T}\tilde{\Sigma}_{Wddk}\hat{a}_{SWdk}^{(jr)} \end{bmatrix} \times \frac{1}{\sum_{l=1}^{d}\hat{a}_{SWlk}^{(jr)T}\tilde{\Sigma}_{Wllk}\hat{a}_{SWlk}^{(jr)}} \quad (5.30)$$

The extra scaling term on the right hand side of Equation (5.30) is to ensure that the AVGVAR constraint is exactly met for the kth test set. The cross-validated SUMCOR value for the jth SUMCOR/AVGVAR canonical correlation for the kth test set is

$$CV_{Sk}^{(jr)} = \frac{\sum_{i=1}^{d}\sum_{i'=1}^{d}\hat{a}_{SWik}^{(jr)T}\tilde{\Sigma}_{Wii'k}\hat{a}_{SWi'k}^{(jr)}}{\sum_{l=1}^{d}\hat{a}_{SWlk}^{(jr)T}\tilde{\Sigma}_{Wllk}\hat{a}_{SWlk}^{(jr)}}. \quad (5.31)$$

The average cross-validation SUMCOR value can then be calculated as

$$CV_{S}^{(jr)} = \sum_{k=1}^{K} CV_{S_k}^{(jr)}. \quad (5.32)$$

The cross-validated SUMCOR values can be used to select the optimal $r \in \mathcal{R}$ by finding the value of r that maximizes the sum of the first J SUMCOR values. Define $CV_S^{(Jr)} = \sum_{j=1}^{J} CV_S^{(jr)}$, to be the sum of the first J cross-validated SUMCOR values for r. In general we recommend focusing on a small number of directions. This is because the goal is to find the best possible low rank representation of the correlations between data sets, which are driven by the first few directions. A simple way to choose the optimal r would simply be $\text{argmax}_{r\in\mathcal{R}} CV_S^{(Jr)}$. However this will largely depend on the particular split of the data for the chosen r. A way to get around this is to use a locally weighted regression technique such as locally estimated scatterplot smoothing (LOESS) (Cleveland, 1979), and find the r that maximizes the expected value of $CV_S^{(Jr)}$. For LOESS regression or any similar technique the expected value of $CV_S^{(Jr)}$, $E(CV_S^{(Jr)})$, can be written as a function of r. Using this we will choose the optimal r as $\hat{r} = \text{argmax}_r E(CV_S^{(Jr)})$. It is important to note that this method is used to try and find the best canonical directions, and

not give the most efficient low-rank approximation of each set of variables. Not including an important principal component for one of the variables will likely cause a bigger drop-off in performance than including an extra principal component that is not correlated with any of the other sets of variables. For this reason the cross-validation technique may tend to slightly overestimate the optimal r_i for each set.

These cross-validated values can also be used to identify the informative canonical directions. The informative canonical directions are those with a SUMCOR value greater than one for the SUMCOR/AVGVAR formulation. One useful property of $CV_{Sk}^{(jr)}$ is that it is not bounded below by one. This is because even though the diagonals of Equation (5.30) are guaranteed to sum to one, the off diagonals may be positive or negative. This is in contrast to performing an analogous cross-validation technique for the MAXVAR/VAR formulation. In this case if we ensure that the diagonal terms are all equal to one, even if the off diagonal terms are negative the maximum eigenvalue will be greater than one. This means even when the true direction is uninformative if we use the MAXVAR/VAR form for cross-validation the cross-validated MAXVAR values will all be above one, which indicates an informative direction, except on a set of measure zero. This is related to a problem in standard CCA in which the estimated canonical correlation will always be above zero except on a set of measure zero, even when the true canonical correlation is zero. However as noted above there are the same number of true informative directions for both the MAXVAR/VAR and SUMCOR/AVGVAR formulations, which means we can use the SUMCOR cross-validation procedure to identify the informative directions for both formulations of the mCCA problem.

There are a number of previous methods to identify informative canonical directions that do not work across all transelliptical distributions. Among these is permutation based testing, which does not work for non-Gaussian elliptical distributions. This is because permutation testing creates a reference distribution assuming that the sets of variables are all independent. However for non-Gaussian elliptical distributions such as the multivariate t distribution having zero Pearson or Kendall's tau correlation does not imply independence. This leads to inflated

type I error for permutation tests if the data are distributed according to one of these distributions, as shown in Section 5.4. The same issue will arise in transelliptical distributions with a non-Gaussian copula.

cross-validation, and specifically using the SUMCOR/AVGVAR formulation of cross-validation, helps to get around both of these issues. cross-validation does not create a null distribution assuming all variables are independent, and instead tries to create a test and training set with the same underlying distribution. Further as mentioned previously the values for $CV_{SK}^{(jr)}$ are not bounded below by one, since the sum of the off diagonals can be less than one. This means that under the null where $\lambda^{(j)} = 1$, $E(CV_S^{(jr)}) = 1$. This is the primary reason that we use the cross-validated SUMCOR value, rather than the cross-validated MAXVAR value.

This also allows for a testing procedure based on the sample mean and standard deviation of the cross-validated SUMCOR values. If we define $m = (u_1 - l_1 + 1) \times \ldots \times (u_d - l_d + 1)$, then we can enumerate $CV_S^{(jr)}$ for all $r \in \mathcal{R}$ as $CV_S^{(j)} = [CV_S^{(j1)}, \ldots, CV_S^{(jm)}]$. A test statistic based on $CV_S^{(j)}$ is

$$T_S^{(j)} = \frac{\text{Mean}(CV_S^{(j)}) - 1}{\text{SD}(CV_S^{(j)})}, \tag{5.33}$$

where Mean$(\cdot)$ is the sample mean of a vector and SD$(\cdot)$ is the sample standard deviation. We can get an approximate level α test for the null that the true SUMCOR value, $\lambda^{(j)} \leq 1$, by rejecting the null if $T_S^{(j)} > z_{1-\alpha}$ where $z_{1-\alpha}$ is the $1-\alpha$ percentile of a standard normal distribution. This test can be done in a step down procedure, starting with $T_S^{(1)}$, and moving through all the test statistics for all higher order canonical correlations in a sequential order. Once we fail to reject the null that $\lambda^{(j)} \leq 1$ then we would also reject the null that $\lambda^{(j')} \leq 1$ for all $j' > j$. Even though the values $CV_S^{(j)}$ are not independent, simulation studies suggest that this works reasonably well and is slightly conservative. Further this does not depend on the selection of $\mathcal{R}$, so long as m is reasonably large. However the power under the alternative will depend on the selection of $\mathcal{R}$. This is because if u_i and l_i are set to be too low then the cross-validated SUMCOR values will tend to be lower than the true value of $\lambda^{(j)}$. Alternatively if u_i and l_i are set to be too high, then the cross-validated SUMCOR values will have a high variance, making

it tougher to detect true signal. Even when selection of the number of principal components is fixed before any cross-validation procedures, a similar testing procedure can be done by doing repeated cross-validation for the fixed value of r. However if cross-validation is used to both select the optimal value of r, only using repeated cross-validation on this selection of r may lead to inflated type I error. This is related to common multiple testing problems. For a given data set, even if there is no correlation between the sets of variables one or more of the many possible choices of r will be likely to have an unusually high SUMCOR value due to random chance. If r is chosen to maximize the cross-validated SUMCOR value it is very likely that one such value of r will be chosen and lead to potential inflated type I error. This is why we focus on a test statistic averaging many different possible values of r. However if r is chosen using another technique that does not focus on maximizing cross-validated SUMCOR values the same multiple comparison type issues may not be an issue. Simulations that investigate identifying informative directions and choosing the optimal number of principal components for each set are presented in Section 5.4.

5.4 Simulations Results

Simulation studies were conducted to study estimation of high-dimensional mCCA for elliptical distributions using the transformed Kendall's tau correlation estimator. In addition we investigate the performance of cross-validation and other testing procedures for identifying informative canonical directions. For simulations studying the estimation of the mCCA directions a correlation structure that has three informative canonical directions was considered. For simulations studying the identification of informative directions the same correlation structure was considered, as well as one that has zero informative directions. For both set ups there are three sets of variables, each with a dimension of 100. For both correlation structures within each set the first 10 variables in the same block are all correlated with each other with a correlation coefficient of 0.47, variables 11 through 20 in the same block are all correlated with each other with a correlation coefficient of 0.41, variables 21 through 30 in the same block are all correlated with each other with a correlation coefficient of 0.33, and variables 31 through 40 in the same

block are all correlated with each other with a correlation coefficient of 0.23. All other within set correlations are zero. For the first correlation structure the first 10 variables of each set are correlated with the first 10 variables of every other set, with a correlation coefficient of 0.11, variables 11 through 20 of one set are correlated with variables 11 through 20 of every other set with a correlation coefficient of 0.06, and variables 21 through 30 of one set are correlated with variables 21 through 30 in every other set with a correlation coefficient of 0.2. All other between set correlations for the first correlation structure are zero. For the second correlation structure all between set correlations are zero. A heatmap of the two correlation structures are presented in Figure 5.2, with the first 100 rows/columns representing the first set, the next 100 rows/columns representing the second set, and the final 100 rows/columns representing the third set. For the first correlation structure the three informative MAXVAR/VAR and SUMCOR/AVGVAR values are equal to 2, 1.43, and 1.27.

For both correlation structure one and two data sets were simulated assuming a multivariate normal distribution and multivariate t distribution with three and 10 degrees of freedom. Sample sizes of n=200 and n=1000 were considered. A total of 1000 data sets were simulated for each setting. For each simulated data set the high dimensional mCCA directions and correlations were estimated as described in Section 5.3.3 using both the transformed Kendall's correlation matrix estimator and the sample correlation matrix. In order to select the optimal r_1 for each set the cross-validation procedure using LOESS regression described in Section 5.3.4 was used with $l_i = 3$, $u_i = 9$ and $K = 5$. The sum of the cross-validated SUMCOR values for the first four directions for each combination of r_1, r_2, and r_3 was used as the outcome, and the values of r_1, r_2, and r_3 as the predictors. The smoothing parameter was set to 0.5. Smoothing parameters of 0.25 and 0.75 were also considered and gave similar results. The optimal values r_1, r_2, and r_3 were selected to be the values that gave the highest expected cross-validated SUMCOR value summed over the first four directions.

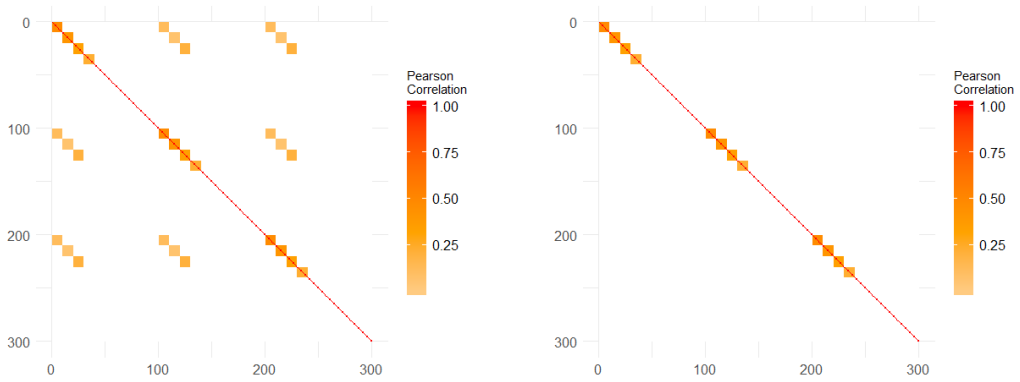

Figure 5.2: Correlation structure one on the left and correlation structure two on the right

5.4.1 Estimation of canonical directions

First we look at how well the directions are estimated using both transformed Kendall's tau and the sample Pearson correlation matrix to estimate the true correlation matrix. Due to the large dimension of each of the sets of variables the angle between the true direction and the estimated direction will tend to be large even when estimation procedures are reasonable. Instead we calculate the "true" SUMCOR and MAXVAR values of the estimated directions $\hat{a}_S^{(j)}$ and $\hat{a}_M^{(j)}$ using the true underlying correlation matrix, which is only possible in simulations when the true correlation matrix is known. Define $\Sigma_{\hat{U}S}^{(j)}$ such that the entry for the ith row and lth column is

$$\frac{\hat{a}_{Si}^{(j)T}\Sigma_{il}\hat{a}_{Sl}^{(j)}}{\sum_{i=1}^{d}\hat{a}_{Si}^{(j)T}\Sigma_{ii}\hat{a}_{Si}^{(j)}}.$$

Similarly define $\Sigma_{\hat{U}M}^{(j)}$ such that the entry for the ith row and lth column is

$$\frac{\hat{a}_{Mi}^{(j)T}\Sigma_{il}\hat{a}_{Ml}^{(j)}}{\sqrt{\hat{a}_{Mi}^{(j)T}\Sigma_{ii}\hat{a}_{Mi}^{(j)}}\sqrt{\hat{a}_{Ml}^{(j)T}\Sigma_{ll}\hat{a}_{Ml}^{(j)}}}.$$

The sum of the entries of $\Sigma_{\hat{U}S}^{(j)}$ is the SUMCOR value in the direction of $\hat{a}_S^{(j)}$, and the maximum eigenvalue of $\Sigma_{\hat{U}M}^{(j)}$ is the MAXVAR value in the direction of $\hat{a}_M^{(j)}$. Because this uses the true correlation matrix instead of the estimated correlation matrix this gives an indicator of how well the estimated directions are able to find true associations between the sets of variables. We

can compare these values to the theoretical maximum in order to examine how well the mCCA directions are estimated.

Table 5.11 reports the estimated bias and standard deviation for these estimated MAXVAR and SUMCOR values relative to the theoretical maximum from 1,000 simulations. For the first three directions the bias is negative, but decreasing in absolute value as sample size increases. This negative bias is to be expected because the estimated directions SUMCOR and MAXVAR values are being compared to the optimal directions. For the first direction the bias can only be negative or zero because there are no directions with a SUMCOR or MAXVAR value better than the optimal direction. Also note that for the fourth direction, which is an uninformative direction with a true MAXVAR and SUMCOR value of one, the bias for the MAXVAR is always positive. This is because the maximum eigenvalue of $\Sigma^{(j)}_{\hat{U}M}$ is bounded below by one. This is related to the reason that SUMCOR is preferred over MAXVAR during cross-validation procedures. The sum of all entries for $\Sigma^{(j)}_{\hat{U}S}$ does not have the same issue, because even though the diagonal will sum to one, the sum of the off diagonal elements may be positive or negative. When data come from a multivariate normal distribution the sample Pearson correlation matrix and transformed Kendall's tau estimator perform nearly identically. For both estimators the bias for the first three directions is between -0.10 and -0.05 for both the SUMCOR and MAXVAR formulations when n=200 and is equal to -0.02 for all three directions when n=1000. For heavier tailed distributions such as a multivariate t with three or 10 degrees of freedom the transformed Kendall's tau estimate performs much better than the sample Pearson correlation matrix. This is particularly noticeable for the first direction when n=200 and for n=1000 when the multivariate t has three degrees of freedom. In addition for the multivariate t with three degrees of freedom the bias when using the sample Pearson correlation matrix is positive for the third direction when n=1000. This is due to imprecision in the estimation of the first two directions which changes the constraints and allows for the average correlation in the direction of the estimated third direction to be higher than the true third direction. This still shows that mCCA using the sample Pearson correlation matrix does not meaningfully summarize the associations between

the three sets of variables as desired due to imprecision in estimating the first two directions. In contrast the mCCA directions using the transformed Kendall's tau estimator perform almost as well when data come from a multivariate t distribution compared to when data come from a multivariate normal distribution. This highlights the robustness of the transformed Kendall's tau estimator for heavy tailed elliptical distributions. Table 5.12 reports the average and standard deviation of the absolute value of the Pearson correlation between $a_M^{(j)}$ and $\hat{a}_M^{(j)}$ as well as $a_S^{(j)}$ and $\hat{a}_S^{(j)}$ over all 1000 simulated data sets. This is an alternative way of looking at the performance of the estimated MAXVAR and SUMCOR directions. This is only reported for the first three directions because only the first three directions are well defined. The absolute value of the Pearson correlation is used because a_M and $\hat{a}_S$ are only defined up to a sign change. This table also shows that using the transformed Kendall's tau estimator outperforms the sample Pearson correlation matrix when data come from a multivariate t distribution with three or 10 degrees of freedom, and performs similarly when data come from a multivariate normal distribution. This shows more clearly that estimates of the first direction are better than estimates of higher order directions across all settings.

5.4.2 Identifying informative canonical directions

As noted in Section 5.3.4, identifying informative mCCA directions is of interest so that purely random mCCA directions are not interpreted as representing meaningful associations between the different sets of variables. For both the SUMCOR/AVGVAR and MAXVAR/VAR formulations this corresponds to directions with a SUMCOR or MAXVAR value greater than one. We compare the CV testing procedure from Section 5.3.4 with two different permutation procedures. In all cases the type I error rate for each test, α, is set to 0.05. For the cross-validation procedure we calculate $T_S^{(j)}$ for $j = 1, 2, \ldots, 4$. We set $l_i = 3$ and $u_i = 9$ for $i = 1, 2, 3$. For the permutation testing procedures the subject labeling is randomly permuted independently for each of the three sets of variables and the estimated SUMCOR values are calculated using this new permuted data set. This is repeated 1,000 times to create a permutation distribution of SUMCOR values when all sets of variables are independent. For the first permutation procedure,

Table 5.11: Bias (SD) of estimated MAXVAR/VAR direction MAXVAR value compared to true direction MAXVAR value and estimated SUMCOR/AVGVAR direction compared to true direction SUMCOR value, selecting between three and nine principal components per set

		n=200		n=1000	
		Transformed Kendall's tau	Pearson Correlation	Transformed Kendall's tau	Pearson Correlation
		MAXVAR			
Normal	Dir 1	-0.10 (0.03)	-0.09 (0.03)	-0.02 (0.01)	-0.02 (0.01)
	Dir 2	-0.09 (0.06)	-0.09 (0.06)	-0.02 (0.02)	-0.02 (0.02)
	Dir 3	-0.06 (0.08)	-0.06 (0.08)	-0.02 (0.02)	-0.02 (0.02)
	Dir 4	0.12 (0.07)	0.11 (0.07)	0.05 (0.05)	0.05 (0.05)
t3	Dir 1	-0.13 (0.05)	-0.83 (0.16)	-0.02 (0.01)	-0.74 (0.28)
	Dir 2	-0.13 (0.09)	-0.22 (0.17)	-0.03 (0.02)	-0.05 (0.28)
	Dir 3	-0.08 (0.09)	-0.03 (0.17)	-0.02 (0.03)	0.08 (0.22)
	Dir 4	0.13 (0.07)	0.22 (0.14)	0.06 (0.05)	0.28 (0.17)
t10	Dir 1	-0.11 (0.04)	-0.25 (0.21)	-0.02 (0.01)	-0.03 (0.03)
	Dir 2	-0.11 (0.07)	-0.10 (0.18)	-0.02 (0.02)	-0.05 (0.07)
	Dir 3	-0.07 (0.08)	-0.05 (0.11)	-0.02 (0.02)	-0.04 (0.07)
	Dir 4	0.12 (0.06)	0.16 (0.09)	0.05 (0.05)	0.10 (0.08)
		SUMCOR			
Normal	Dir 1	-0.10 (0.03)	-0.10 (0.03)	-0.02 (0.01)	-0.02 (0.01)
	Dir 2	-0.10 (0.06)	-0.10 (0.06)	-0.02 (0.02)	-0.02 (0.02)
	Dir 3	-0.08 (0.09)	-0.08 (0.09)	-0.02 (0.02)	-0.02 (0.02)
	Dir 4	0.01 (0.08)	0.01 (0.07)	-0.01 (0.02)	-0.01 (0.02)
t3	Dir 1	-0.14 (0.05)	-0.85 (0.17)	-0.02 (0.01)	-0.74 (0.28)
	Dir 2	-0.14 (0.09)	-0.24 (0.18)	-0.03 (0.02)	-0.06 (0.28)
	Dir 3	-0.10 (0.10)	-0.06 (0.19)	-0.02 (0.03)	0.07 (0.23)
	Dir 4	0.04 (0.09)	0.16 (0.16)	0.00 (0.03)	0.26 (0.18)
t10	Dir 1	-0.11 (0.04)	-0.26 (0.21)	-0.02 (0.01)	-0.03 (0.03)
	Dir 2	-0.11 (0.07)	-0.10 (0.19)	-0.02 (0.02)	-0.05 (0.07)
	Dir 3	-0.09 (0.09)	-0.06 (0.12)	-0.02 (0.02)	-0.04 (0.07)
	Dir 4	0.02 (0.09)	0.11 (0.11)	-0.01 (0.02)	0.06 (0.09)

Table 5.12: Average (SD) of the correlation between estimated and true MAXVAR/VAR directions and SUMCOR/AVGVAR directions, selecting between three and nine principal components per set

		n=200		n=1000	
		Transformed Kendall's tau	Pearson Correlation	Transformed Kendall's tau	Pearson Correlation
			MAXVAR		
Normal	Dir 1	0.87 (0.03)	0.87 (0.02)	0.97 (0.01)	0.97 (0.01)
	Dir 2	0.67 (0.19)	0.68 (0.19)	0.91 (0.06)	0.91 (0.06)
	Dir 3	0.54 (0.23)	0.54 (0.22)	0.86 (0.09)	0.86 (0.08)
t3	Dir 1	0.82 (0.04)	0.19 (0.15)	0.96 (0.01)	0.28 (0.25)
	Dir 2	0.57 (0.22)	0.15 (0.13)	0.89 (0.07)	0.15 (0.14)
	Dir 3	0.42 (0.23)	0.15 (0.12)	0.82 (0.11)	0.18 (0.16)
t10	Dir 1	0.85 (0.03)	0.71 (0.19)	0.97 (0.01)	0.95 (0.04)
	Dir 2	0.64 (0.19)	0.37 (0.25)	0.91 (0.06)	0.80 (0.20)
	Dir 3	0.49 (0.22)	0.32 (0.22)	0.85 (0.09)	0.66 (0.25)
			SUMCOR		
Normal	Dir 1	0.87 (0.03)	0.87 (0.02)	0.97 (0.01)	0.97 (0.01)
	Dir 2	0.67 (0.19)	0.67 (0.19)	0.91 (0.06)	0.91 (0.06)
	Dir 3	0.53 (0.22)	0.53 (0.22)	0.85 (0.09)	0.86 (0.08)
t3	Dir 1	0.82 (0.04)	0.19 (0.15)	0.96 (0.01)	0.28 (0.25)
	Dir 2	0.57 (0.22)	0.15 (0.12)	0.89 (0.07)	0.15 (0.14)
	Dir 3	0.41 (0.23)	0.15 (0.12)	0.82 (0.11)	0.18 (0.15)
t10	Dir 1	0.85 (0.03)	0.70 (0.19)	0.97 (0.01)	0.95 (0.04)
	Dir 2	0.63 (0.19)	0.37 (0.25)	0.90 (0.06)	0.79 (0.21)
	Dir 3	0.48 (0.22)	0.31 (0.22)	0.84 (0.09)	0.65 (0.25)

PERM1, the null of an uninformative direction is rejected if the estimated jth SUMCOR value for the original data set is greater than the (1-α) percentile of the permutation distribution for the jth direction. For the second permutation procedure, PERM2, the null of an uninformative direction is rejected if the estimated jth SUMCOR value for the original data set is greater than the (1-α) percentile of the permutation distribution for the first direction. Similar to the cross-validation procedure both PERM1 and PERM2 are done using a step down procedure, in which the null of a true SUMCOR value of one for higher order directions can only be rejected if it has also been rejected for all previous directions.

Table 5.13 gives the type I error for simulations done using correlation structure two using the transformed Kendall's tau correlation estimator and sample Pearson correlation estimator. In all cases the reported values are the proportion of simulations where the null of a true SUMCOR value of one is rejected. The cross-validation procedure controls for type I error for the first direction across all distributions and sample sizes when using transformed Kendall's correlation estimator. It is in fact slightly conservative, with type the type I error ranging between 0.02 and 0.04. The cross-validation procedure also controls for type I error using the sample Pearson correlation estimator when data come from a multivariate normal distribution. When data come from a t distribution with three degrees of freedom and correlation is estimated using the sample Pearson correlation matrix the cross-validation procedure results in slightly inflated type I error of 0.08 for the first direction when n=200. The low type I error across all settings for the second direction using the cross-validation procedure is not surprising due to the step down procedure of the test. In order for the null to be rejected for the second direction it must also be rejected for the first direction.

For both permutation testing procedures type I error is only controlled when data come from a multivariate normal distribution. When data come from a t distribution with either three or 10 degrees of freedom the type I error is above 0.05. This is true when using the sample Pearson correlation matrix estimator or the transformed Kendall's correlation estimator. However the type I error is particularly large when using the sample Pearson correlation estimator. The inflated

type I error for data with a multivariate t distribution is because permutation testing generates a distribution for the SUMCOR values assuming the sets of variables are all independent of each other, which is is a more strict assumption than the sets of variables being uncorrelated when the data has a multivariate t distribution.

Table 5.13: Type I error for CV and permutation testing procedures using correlation structure two

		n=200			n=1000		
		CV	PERM1	PERM2	CV	PERM1	PERM2
		Transformed Kendall's tau					
Normal	Dir 1	0.03	0.05	0.05	0.04	0.06	0.06
	Dir 2	0.00	0.01	0.00	0.00	0.01	0.00
t3	Dir 1	0.02	0.35	0.35	0.03	0.23	0.23
	Dir 2	0.00	0.17	0.00	0.00	0.08	0.00
t10	Dir 1	0.03	0.12	0.12	0.03	0.11	0.11
	Dir 2	0.00	0.04	0.00	0.00	0.03	0.00
		Pearson Correlation					
Normal	Dir 1	0.03	0.04	0.04	0.04	0.06	0.06
	Dir 2	0.00	0.00	0.00	0.00	0.01	0.00
t3	Dir 1	0.08	1.00	1.00	0.05	1.00	1.00
	Dir 2	0.00	1.00	0.99	0.00	1.00	1.00
t10	Dir 1	0.05	0.72	0.72	0.01	0.43	0.43
	Dir 2	0.00	0.52	0.10	0.00	0.23	0.01

Table 5.14 gives the power and type I error for the cross-validation and permutation testing procedures using correlation structure one and simulating data with a multivariate normal distribution. Table 5.15 gives the same results for the cross-validation testing procedure when data come from a multivariate t distribution. In both cases the first three directions the reported values are the estimated power and the reported values for the fourth direction this are type I error. Based on results from Table 5.13 we do not present the power for permutation testing procedures when data come from a multivariate t distribution because the permutation testing procedures do not control for type I error. For the cross-validation testing procedure when using

the transformed Kendall's tau estimator the type I error is less than 0.01 for all distribution and sample size settings. When using the sample Pearson correlation estimator this is also true when data come from a multivariate normal distribution. The conservative nature of the CV procedure for the fourth direction is likely due to additional constraints when estimating the fourth direction. As a check we performed simulations for a correlation structure that had one true informative direction. In the proportion of simulations where the null was rejected for the second direction was between 0.02 and 0.04, similar to what we see for the first direction in Table 5.13. As the number of constraints increases the cross-validated SUMCOR values may tend to decrease leading to more conservative tests for higher directions.

There is one setting where the cross-validation procedure is not conservative. This is when using the sample Pearson correlation estimator for data with a multivariate t distribution with three degrees of freedom and n=1000. In this case the type I error rate is not controlled and the proportion of simulations where the null is rejected is 0.2. This is because of how poorly the directions based on the sample Pearson correlation matrix estimate the true directions in this setting. Based on Table 5.11 the average "true" SUMCOR value of the estimated fourth direction in this setting is 1.26. In comparison when data come from a multivariate normal distribution or the transformed Kendall's tau estimator is used the average "true" SUMCOR value is between 0.99 and 1.04. It is in the setting where there is a large positive bias in the SUMCOR value of the fourth direction that type I error is inflated.

When data come from a multivariate normal distribution both correlation estimation procedures have similar power for the cross-validation testing procedure for the first three directions, with the sample Pearson correlation matrix having slightly higher power. When data come from a multivariate t distribution with three or 10 degrees of freedom the transformed Kendall's tau estimator has higher power for the first three directions for the cross-validation testing procedure. This difference is largest for a multivariate t with three degrees of freedom. This corresponds to the distributions where the transformed Kendall's tau estimator more efficiently estimates the true mCCA directions.

Correlation matrix structure one is when the difference between the two permutation testing procedures is most noticeable. PERM1 has higher power than PERM2 for the second and third directions, but PERM1 has a rejection proportion above 0.05 for the fourth direction, even when data come from a multivariate normal distribution. This inflated type I error for higher directions using the PERM1 procedure is also present for standard two set CCA (Winkler et al., 2020). Even though the true fourth direction is uninformative for correlation structure one, the estimated SUMCOR value for the fourth direction may tend to be higher than the SUMCOR value for the fourth direction in a data set where all sets of variables are independent. This is because of imprecision in estimating the first three directions, and the corresponding constraints they put on the fourth direction. Alternatively the PERM2 procedure has lower power because it is comparing the second and third directions to the first direction from the permutation sample. Even though the permutation sample is constructed to have no correlation between the sets, the SUMCOR value for the first direction will tend to be reasonably large due to random noise and the fact that there are no constraints on the direction. This makes it harder to reject the null for higher order directions even when they are informative.

Table 5.14: Power and type I error for cross-validation and permutation testing procedures for correlation structure one

		n=200			n=1000		
		CV	PERM1	PERM2	CV	PERM1	PERM2
		Transformed Kendall's tau					
Normal	Dir 1	1.00	1.00	1.00	1.00	1.00	1.00
	Dir 2	0.90	0.99	0.78	1.00	1.00	1.00
	Dir 3	0.44	0.81	0.03	1.00	1.00	0.99
	Dir 4	0.00	0.19	0.00	0.00	0.17	0.00
		Pearson Correlation					
Normal	Dir 1	1.00	1.00	1.00	1.00	1.00	1.00
	Dir 2	0.92	0.99	0.80	1.00	1.00	1.00
	Dir 3	0.47	0.82	0.03	0.99	1.00	0.99
	Dir 4	0.00	0.16	0.00	0.00	0.16	0.00

Table 5.15: Power and type I error for cross-validation testing procedure for correlation structure one when data come from multivariate t distribution

		Dir 1	Dir 2	Dir 3	Dir 4
		Transformed Kendall's tau			
t3	n=200	1.00	0.74	0.28	0.00
	n=1000	1.00	1.00	0.99	0.00
t10	n=200	1.00	0.86	0.36	0.00
	n=1000	1.00	1.00	1.00	0.00
		Pearson Correlation			
t3	n=200	0.34	0.15	0.05	0.01
	n=1000	0.52	0.40	0.33	0.20
t10	n=200	0.95	0.70	0.38	0.06
	n=1000	1.00	0.99	0.95	0.05

5.5 mCCA estimates for executive function and brain structure in six-year-old children

We study the performance of latent high-dimensional multiset CCA using data from the Early Brain Development Study (EBDS) at University of North Carolina (Gilmore et al., 2010; Knickmeyer et al., 2008, 2016). The EBDS is a longitudinal study of children that collects data from birth through early adolescence. We include three sets of variables in our analysis: EF test scores, brain GM volume, and white matter structure as measured using DTI. For this project we use data collected at age six. The primary reason for using this age is that it is the oldest age at which data has been collected for a majority of subjects in the EBDS, and the EF tests are more reliable at age six than at earlier ages.

We include four different EF test scores. These include both the Stanford-Binet intelligence scales verbal fluid reasoning (SB V FR) and non-verbal fluid reasoning (SB NV FR) section scores (Roid, 2003). These both measure fluid reasoning which is known to be associated with working memory and executive function. We also include the Behavior Rating Inventory of Executive Function (BRIEF) (Gioia et al., 2000) composite score, which is a parent reported measure of executive function. Finally we include the executive function scale from the Behavioral Assessment System for Children (BASC) (Reynolds and Kamphaus, 2010).

The white matter structure variables include RD measured using DTI for 20 different white matter tracts. For each tract we used the average of the RD values for all voxels in the tract. RD is primarily driven by myelination in the white matter tracts, with lower RD indicating more myelination. Myelin is a substance that surrounds the white matter axons and increases the speed of electrical impulses through the brain tract. For more information see Alexander et al. (2007). The 20 tracts were chosen based on previous research from Girault et al. (2019). The list of the tracts used can be found in Table 5.16. For each subject the RD values were centered. This was shown to be able to find a stronger canonical correlation signal than using the RD values without centering.

Table 5.16: List of white matter tracts used in CCA analysis

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 lef/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 brain volume variables included total GM volume in 88 of the 90 regions defined by the AAL brain parcellation (Tzourio-Mazoyer et al., 2002). Many subjects had no observed GM in the left and right pallidum so those two regions were not included in the analysis. For each subject we reported the GM volume as a proportion of overall intracranial volume. This was done to control for overall brain/head size for each subject.

The latent high-dimensional multiset canonical correlation directions and cross-validated correlations for the MAXVAR/VAR formulation are presented below. The first step was to identify the optimal number of principal directions for EF test scores, RD values for the DTI

tracts, and GM volume for the AAL regions. We used the cross-validated methods described in Section 5.3.4. We considered between two and four directions (out of four total possible) for the EF, between four and 20 directions (out of 20 total possible) for the DTI variables, and between two and 30 directions (out of 88 total possible) for the GM variables. This was because there was a noticeable drop in cross-validated performance after 30 directions for the GM volume measures. For each of the 1479 possible combinations the cross-validated SUMCOR value was calculated using five fold cross-validation. A LOESS regression was then fit with the cross-validated SUMCOR values for the first canonical direction as the outcome, the number of principal directions for each of the three sets of variables as the predictors, and a smoothing parameter of 0.5. The combination with the highest predicted value for cross-validated SUMCOR was chosen as the optimal number of principal components for each set of variables. A total of three principal directions were chosen for the EF/cognitive variables, 16 principal directions for the DTI variables, and seven principal directions for the GM variables. A smoothing parameter of 0.25 was also considered and selected the same number of principal components.

Based on the cross-validation testing procedure presented in Section 5.3.4 there are two significant canonical directions. In order to look at the strength of the associations for the first two directions five fold cross-validation was repeated 50 times at the chosen number of principal directions for each set of variables. The average MAXVAR/VAR correlation matrices from all 250 cross-validation test sets are presented in Table 5.17. We focus on the MAXVAR/VAR matrices because of the easier interpretability of correlation matrices relative to covariance matrices that has diagonals summing to one. For the first direction all three sets of variables are correlated with each other in similar magnitudes, between 0.29 and 0.35, but for the second and third direction the correlation between GM and DTI is the strongest at 0.28 and 0.18, compared to less than 0.1 for the other two correlations. For both the second and third direction the cross-validated correlation between GM and the EF/cognitive variables is in fact negative. Figure 5.3 shows the first direction loadings for the GM and DTI variables using BrainNet Viewer for the GM variables (Xia et al., 2013) and 3D slicer for the DTI variables (Kikinis et al., 2014). A full

Table 5.17: Cross-validated correlation matrices for MAXVAR/VAR mCCA directions

	Direction 1				Direction 2		
	EF	DTI	GM		EF	DTI	GM
EF	1.00	0.33	0.29	EF	1.00	0.07	-0.08
DTI	0.33	1.00	0.35	DTI	0.07	1.00	0.28
GM	0.29	0.35	1.00	GM	-0.08	0.28	1.00

	Direction 3		
	EF	DTI	GM
EF	1.00	0.09	-0.04
DTI	0.09	1.00	0.18
GM	-0.04	0.18	1.00

listing of the mCCA loadings for the EF, GM, and DTI variables can be found in Appendix C. For the first direction the SB V FR had a positive loading that is more than twice the magnitude of the other three EF test scores for the first direction. For the DTI white matter tracts seven of the nine bilateral tracts (all but ARC FP and SLF) have loadings with the opposite signs between the left and right hemisphere. For the ARC FT, ARC TP, IFOF and ILF the tract on the right hemisphere has a negative loading while the tract on the left has a positive loadings. For the CGC, CTPF, and UNC this pattern is reversed. In contrast only six of the 44 pair of bilateral regions have loadings with opposite signs between the left and right hand side of the brain. Of these only two regions, the supramarginal gyrus and putamen have a loading above 0.02 on either the right or left. The putamen has a loading of -0.09 on the left and 0.03 on the right. The supramarginal has a loading of 0.01 on the left and -0.05 on the right.

For the GM regions one of the most notable variable loadings is for the caudate which has a loading of -0.12 on the left and -0.10 on the right. Of all the GM regions these are two of the three negative loadings with the largest absolute value. Previous research has shown that the caudate is negatively associated with working memory, which is known to be closely associated with fluid reasoning, among subjects at risk of psychosis or with schizotypal personality disorder (Levitt et al., 2002; Hannan et al., 2010). Given that the first direction for the EF variables is largely driven by a fluid reasoning test, the large negative loadings for the caudate gives further

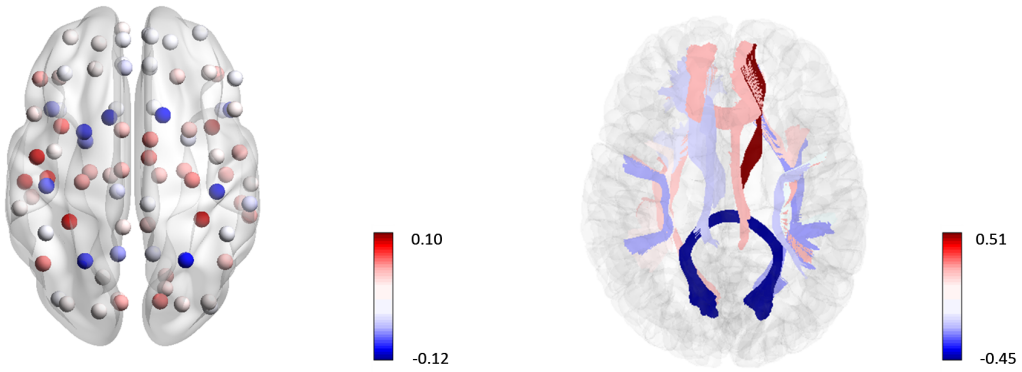

Figure 5.3: GM and DTI loadings for first canonical direction using MAXVAR/VAR formulation

evidence that lower grey matter volume in the caudate is associated with lower working memory and fluid reasoning.

The second direction MAXVAR/VAR loadings for GM and DTI are in Figure 5.4. A full table with the loading values can be found in Appendix C. The second direction loadings for EF/Cognitive variables are not reported because the second direction for EF/Cognitive variables is not strongly correlated with either GM or DTI based on Table 5.17. All of the GM regions with negative loadings with a large absolute value are located near the back of the brain. Late in childhood gray matter starts to decrease, starting with the posterior of the brain (Bray et al., 2015). This indicates that those children with a high score for the second canonical direction for the GM are those who have started this process and may be further along in the development process than those children with low scores. RD in white matter tracts tends to decrease as myelin increases which happens when children mature. This indicates that the DTI tracts with negative loadings for the second directions are those that have relatively more myelin for children further along in the maturation process. This include the Splenium, UNC and ARC FP. Previous research has shown that development of the white matter does not happen in the same posterior to anterior order, but rather based on the fiber orientation of white matter (Bray et al., 2015).

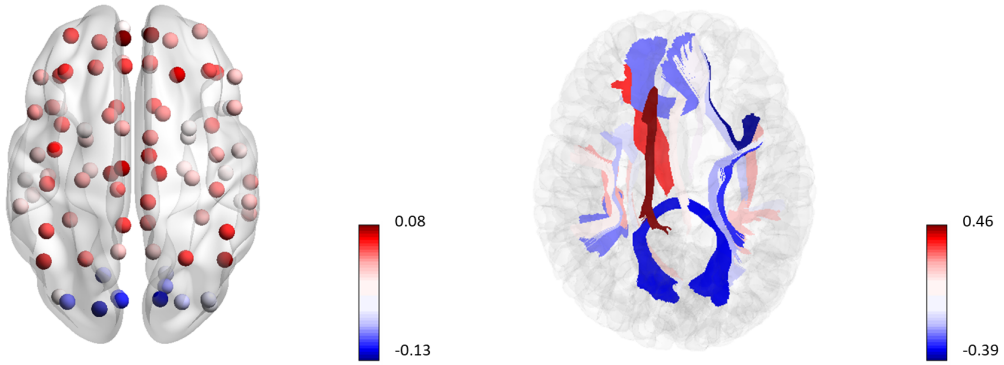

Figure 5.4: GM and DTI loadings for second canonical direction using MAXVAR/VAR formulation

5.6 Discussion

mCCA is a powerful extension of CCA that can find a low rank summary of the associations between multiple sets of variables. There are a number of potential formulations for mCCA, but we have shown how the MAXVAR/VAR and SUMCOR/AVGVAR formulations from Kettenring (1971) and Nielsen (2002) have similar closed form estimates and can be used in tandem with each other. In the high dimensional setting the closed form estimates for both of these formulations can still be useful after first reducing the dimension of each set of variables using PCA, similar to Asendorf (2015) and Feng et al. (2018). We introduce a cross-validation procedure that can be used to select the optimal number of principal components for each set of variables. In this high dimensional setting we show that estimation of the mCCA directions can be greatly improved by using robust estimation of the correlation or covariance matrix. In particular using the transformed Kendall's tau correlation estimator can improve performance when data come from heavy tailed or skewed distributions.

Careful consideration of testing procedures to identify informative mCCA directions is also needed when data do not come from a multivariate normal distribution. In particular permutation testing may lead to inflated type I error regardless of the technique used to estimate the correlation matrix and mCCA directions. For this reason we develop a cross-validation based testing procedure for identifying informative mCCA directions. However future research on both testing procedures and ways to select the optimal number of principal components would be of

value, particularly if they can cut down on computational time. In addition a computationally feasible robust version of penalized mCCA would be useful for high dimensional data sets.

CHAPTER 6: SUMMARY AND FUTURE RESEARCH

In this document we investigate extensions to CCA and PCA. Both CCA and PCA can give useful low rank summaries of high dimensional data sets. The estimates of CCA and PCA typically employ eigendecomposition involving the sample covariance or correlation matrices of the data. For data sets with independent censoring it is not possible to use standard covariance or correlation estimators. For other data sets with extreme outliers or when the data come from heavy tailed or skewed distributions the sample covariance and correlation matrices may not be an efficient estimator of the true covariance and correlation matrices. We investigate alternative ways to estimate covariance and correlation matrices for such data sets in order to improve performance of CCA and PCA.

In Chapters 3 and 5 we extend CCA and mCCA using a robust correlation estimator based on Kendall's tau rank correlation coefficient. In chapter 3 we are able to show that CCA estimates using this robust correlation estimator are consistent and asymptotically normal. Simulation studies show that this robust CCA method using Kendall's tau outperforms standard CCA estimation methods when data come from heavy tailed elliptical or skewed transelliptical distributions. In chapter 5 we show that the same robust correlation estimator can be used in high dimensional and multi-set settings, and improves performance for data coming from heavy tailed elliptical distributions. We also show the need for new methods for testing for informative CCA and mCCA directions for non-Gaussian elliptical copulas. In chapter 3 we use a bootstrap based testing procedure that outperforms existing asymptotic and permutation based testing procedures for heavy tailed elliptical distributions. In chapter 5 we use cross-validation to identify informative directions. Unlike permutation testing this controls for type I error even when the data come from a heavy tailed elliptical distribution.

In chapter 4 we define covariance and correlation measures in the multivariate survival setting with independent censoring, using counting processes and martingales. We show that these covariance and correlation measures can be estimated non-parametrically and these estimates converge to a tight Gaussian process. Both of these covariance and correlation measures can also be extended to the semi-competing risk setting. We show that eigendecomposition of these covariance and correlation matrices can be used to estimate PCA for multivariate survival data. These PCA estimates are consistent and asymptotically normal and perform well in finite sample simulation studies.

For each of the methods proposed in this document there are a number of areas for potential future research. The robust CCA methods in chapter 3 can be extended to the high-dimensional setting by adding a penalty term to the loadings for the canonical directions. A number of studies have shown how this can be done in a computationally efficient way (Wilms and Croux, 2016, 2015; Yoon et al., 2018). Another potential area of research is developing a more powerful method for testing for informative CCA directions when data do not come from a multivariate normal distribution. Many procedures that test for a true canonical correlation of zero assume the data come from multivariate normal distribution. Permutation testing does not make the same assumptions about the true underlying marginal distributions, but still requires a Gaussian copula. An asymptotic test based on the transformed Kendall's tau correlation matrix estimator may be more powerful than the bootstrap testing technique proposed in Chapter 3 and more robust than other existing testing procedures.

In chapter 4 we focus on estimating the martingale and counting process covariance between two failure times that share the same censoring and competing risk times. It may also be of interest to estimate martingale and counting process covariance between two failure times with different censoring or competing risk times. Another potential area of future research is changing of principal components over time. It would be of interest to identify when changes in the estimates of principal components over time are due to random noise or a true change in the underlying martingale or counting process covariance or correlation matrix. Additionally,

although principal component loadings are estimated at the population level and are shared between censored and uncensored subjects, the principal component scores can only be estimated for subjects who are not yet censored. These principal component scores may be of use for clustering or grouping the subjects. It would useful to define an imputation scheme that can be used to help classify or group censored as well as uncensored subjects. In particular, for those subjects censored later in the study there may be enough observed data to get reasonable estimates of principal component scores even after their censoring time.

In chapter 5 we use cross-validation to identify the informative directions, but potential improvements using asymptotic properties of Kendall's tau estimator may be possible and desirable. In addition high dimensional mCCA has also been estimated by adding a penalty term to the canonical direction loadings (Witten and Tibshirani, 2009). To our knowledge no similar methods have been applied using robust covariance or correlation estimation techniques for mCCA with more than two sets of variables.

APPENDIX A: TECHNICAL DETAILS AND ADDITIONAL SIMULATION RESULTS FOR CHAPTER 2

A.1 Proof of theorems

Proof of Theorem 2.1 Define λ_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_γ such that for every $n > N_\gamma$,

$$Pr\left(\sum_{1\leq i,j\leq d} |\delta_{n,ij}| \geq \frac{\lambda_d}{2}\right) \leq \gamma \tag{A.34}$$

Define $\Delta_n = \hat{P}_{Zn} - P_Z$, and $\omega_{ni}i$ to be the ith eigenvalue of Δ_n. Because $|\omega_{ni}| \leq \sum_{1\leq i,j\leq n} |\delta_{n,ij}|$, Equation (A.34) implies that for $1 \leq i \leq d$ and $n > N_\gamma$

$$Pr\left(|\omega_{ni}| \geq \frac{\lambda_d}{2}\right) \leq \gamma \tag{A.35}$$

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 Equation (A.35) the result $Pr(\hat{\lambda}_{nd} > 0) \rightarrow_p 1$ is obtained.

Proof of Theorem 2.2

Using work similar to Anderson (1999), the true canonical directions, a_i and b_i, corresponding to the canonical correlation λ_i are solutions to the system of equations

$$\begin{bmatrix} -\lambda_i P_{XX} & P_{XY} \\ P_{YX} & -\lambda_i P_{YY} \end{bmatrix} \begin{bmatrix} a_i \\ b_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{A.36}$$

The r non-zero canonical correlations, $\lambda_1 > \lambda_2 > \ldots > \lambda_r > 0$, are the non-zero values such that

$$\begin{vmatrix} -\lambda_i P_{XX} & P_{XY} \\ P_{YX} & -\lambda_i P_{YY} \end{vmatrix} = 0. \tag{A.37}$$

$A_r = [a_1, \ldots, a_r]$ and $B_r = [b_1, \ldots, b_r]$ 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 $A_{r^+} = [a_{r+1}, \ldots, a_p]$ and $B_{r^+} = [b_{r+1}, \ldots, a_q]$ to be solutions such that for $A = [A_r, A_{r^+}]$ and $B = [B_r, B_{r^+}]$, $A^T P_{XX} A = I_p$, $B^T P_{XX} B = I_q$, and $A^T P_{XY} B = \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix}$. 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

$$\begin{bmatrix} -\lambda_i^* P_{XX}^* & P_{XY}^* \\ P_{YX}^* & \lambda_i^* P_{YY}^* \end{bmatrix} \begin{bmatrix} a_i^* \\ b_i^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{A.38}$$

The estimate of λ_i, λ_i^*, is the ith solution to

$$\begin{vmatrix} -\lambda_i^* P_{XX}^* & P_{XY}^* \\ P_{YX}^* & -\lambda_i^* P_{YY}^* \end{vmatrix} = 0. \tag{A.39}$$

Consider the transformations $P_{UU}^* = A^T P_{XX}^* A$, $P_{VV}^* = B^T P_{YY}^* B$, $P_{UV}^* = A^T P_{XY}^* B$, and $P_{VU}^* = P_{VU}^{*T}$. Define $\Lambda_{r,m_1 m_2}$ to be an $m_1 \times m_2$ matrix with $m_1, m_2 \geq r$ and the upper left hand corner equal to Λ_r and all other entries equal to zero. Combining the Delta method with the assumption for the asymptotic distribution of $P_{XX}^*, P_{XY}^*, and P_{YY}^*$ gives the result,

$$\sqrt{n} \begin{bmatrix} \text{vec}(P_{UU}^*) & & \text{vec}(I_p) \\ \text{vec}(P_{UV}^*) & - & \text{vec}(\Lambda_{r,pq}) \\ \text{vec}(P_{VV}^*) & & \text{vec}(I_q) \end{bmatrix} \rightarrow_d N_{p^3 \times q^3}(0, J_Z \Theta J_Z^T), \tag{A.40}$$

where J_Z is the $(p^3 \times q^3) \times (p^3 \times q^3)$ block matrix,

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

Consider the system of equations

$$\begin{bmatrix} -\lambda_i^* P_{UU}^* & P_{UV}^* \\ P_{VU}^* & \lambda_i^* P_{VV}^* \end{bmatrix} \begin{bmatrix} g_i \\ h_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{A.41}$$

where

$$\begin{vmatrix} -\lambda_i^* P_{UU}^* & P_{UV}^* \\ P_{VU}^* & -\lambda_i^* P_{VV}^* \end{vmatrix} = 0, \tag{A.42}$$

Because the determinant of the product of two square matrices equals product of the determinants, if λ_i^* is a solution to (A.42) it is also a solution to (A.39). The solutions to (A.38) and (A.41) are related through the identities $Ag_i = a_i^*$ and $Bh_i = b_i^*$. Define $G = [g_1, \ldots, g_r]$, $H = [h_1, \ldots, h_r]$ and $\Lambda_r^* = \text{diag}(\lambda_1^*, \ldots, \lambda_r^*)$. Equation (A.41) implies

$$P_{UV}^* H = P_{UU}^* G \Lambda_r^* \tag{A.43}$$

$$P_{VU}^* G = P_{VV}^* H \Lambda_r^*. \tag{A.44}$$

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

$$G^T P_{UU}^* G = I_r \tag{A.45}$$

$$H^T P_{VV}^* H = I_r, \tag{A.46}$$

and $g_{ii} > 0$. 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 (A.40) $P_{UU} \rightarrow_p I_p$,

$P_{VV} \rightarrow_p I_q$, 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 Λ_r^* are single valued functions of P_{UU}^*, P_{UV}^*, and P_{VV}^* that are differentiable in the neighborhood of $I_{r,pr}$ and Λ_r, by the Delta method the quantities, $\tilde{G} = \sqrt{n}[G - I_{r,pr}]$, $\tilde{H} = \sqrt{n}[H - I_{r,qr}]$, and $\tilde{\Lambda}_r = \sqrt{n}[\Lambda_r^* - \Lambda_r]$ all have a normal limiting distribution with mean zero and finite variances. Note the following equalities obtained by expanding and rearranging the terms of (A.43) and (A.44),

$$\sqrt{n}[(P_{UV}^* I_{r,qr} - \Lambda_{r,pr}) - (P_{UU}^* - I_p)\Lambda_{r,pr}] = \tilde{G}\Lambda_r + \tilde{\Lambda}_{r,pr} - \Lambda_{r,pq}\tilde{H} + o_p(1) \tag{A.47}$$

$$\sqrt{n}[(P_{VU}^* I_{r,pr} - \Lambda_{r,qr}) - (P_{VV}^* - I_q)\Lambda_{r,qr}] = \tilde{H}\Lambda_r + \tilde{\Lambda}_{r,qr} - \Lambda_{r,qp}\tilde{G} + o_p(1). \tag{A.48}$$

Multiplying (A.47) by Λ_r on the right hand side and (A.48) by $\Lambda_{r,pq}$ on the left hand side and taking the sum, and then multiplying (A.47) by $\Lambda_{r,qp}$ on the left hand side and (A.48) by Λ_r on the right hand side and taking the sum results in the following two equalities

$$\begin{aligned}\sqrt{n}[&(P_{UV}^* I_{r,qr} - \Lambda_{r,pr})\Lambda_r + \Lambda_{r,pq}(P_{VU}^* I_{r,pr} - \Lambda_{r,qr}) - \Lambda_{r,pq}(P_{VV}^* - I_q)\Lambda_{r,qr} - \\ &(P_{UU}^* - I_p)\Lambda_{r,pr}\Lambda_r] = 2\Lambda_{r,pr}\tilde{\Lambda}_r + \tilde{G}\Lambda_{r,pr}^2 - \Lambda_{r,pp}^2\tilde{G} + o_p(1).\end{aligned} \tag{A.49}$$

$$\begin{aligned}\sqrt{n}[&(P_{VU}^* I_{r,pr} - \Lambda_{r,qr})\Lambda_r + \Lambda_{r,qp}(P_{UV}^* I_{r,qr} - \Lambda_{r,pr}) - \Lambda_{r,qp}(P_{UU}^* - I_p)\Lambda_{r,pr} - \\ &(P_{VV}^* - I_q)\Lambda_{r,qr}\Lambda_r] = 2\Lambda_{r,qr}\tilde{\Lambda}_r + \tilde{H}\Lambda_{r,qr}^2 - \Lambda_{r,qq}^2\tilde{H} + o_p(1)\end{aligned} \tag{A.50}$$

The ith diagonal term where $1 \leq i \leq r$ of the right hand side of both (A.49) and (A.50) will be equal to $2\lambda_i\tilde{\lambda}_i + o_p(1)$. The ith row and jth column where $1 \leq i \leq p$, $1 \leq j \leq r$ and $j \neq i$ of the right hand side of (A.49) is equal to $\tilde{g}_{ij}(\lambda_j^2 - \lambda_i^2)$, while the ith row and jth column where $1 \leq i \leq q$, $1 \leq j \leq r$ and $i \neq j$ from the right hand side of (A.50) is equal to $\tilde{h}_{ij}(\lambda_j^2 - \lambda_i^2)$. This is used to solve for the variance and covariances of $\tilde{\lambda}_i$, $\tilde{g}_{ij}$ and $\tilde{h}_{ij}$. 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 (A.45) and (A.46)

$$\sqrt{n}[I_{r,rp}P^*_{UU}I_{r,pr} - I_r] = -[\tilde{G}^T I_{r,pr} + I_{r,rp}\tilde{G}] + o_p(1) \tag{A.51}$$

$$\sqrt{n}[I_{r,rq}P^*_{VV}I_{r,qr} - I_r] = -[\tilde{H}^T I_{r,qr} + I_{r,rq}\tilde{G}] + o_p(1) \tag{A.52}$$

The ith diagonal term for $1 \leq i \leq r$ of the right hand side of (A.51) is $-2\tilde{g}_{ii} + o_p(1)$ and the ith diagonal term for $1 \leq i \leq r$ of the right hand side of (A.52) is $-2\tilde{h}_{ii} + o_p(1)$. The variances and covariances for each term on the left hand side of (A.49), (A.50), (A.51), and (A.52) can be solved for using (A.40). 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,

$$\sqrt{n}[a^*_j - a_j] = \sqrt{n}A[g_j - \iota_{pj}] = \sum_{i=1}^{p} a_i\tilde{g}_{ij} + o_p(1) \tag{A.53}$$

$$\sqrt{n}[b^*_j - b_j] = \sqrt{n}B[h_j - \iota_{qj}] = \sum_{i=1}^{q} b_i\tilde{h}_{ij} + o_p(1) \tag{A.54}$$

where ι_{pj} is a $p \times 1$ vector where the jth element is one and the rest are zero. Therefore the limiting covariances on the left hand side of Equations (A.51) and (A.52) can be found by solving for the covariances of the right hand side. Consistent estimates for 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 Θ, A, B, and Λ.

A.2 Additional Simulation Results

A.2.1 Empirical bias and standard deviation of CCA with robust covariance estimation

Tables A1 through A5 report bias and standard deviation for estimates for transelliptical CCA estimated using the transformed Kendall's estimator, standard CCA using the sample covariance estimate, and robust CCA based on the MCD estimator and Tyler's M estimator. The simulation set-ups can be found in section 3.1 of the main text, along with details on how the bias and standard deviation are calculated. The results for p=q=8 and n=200 can be found in Table 1 of the main text. In all cases only the bias and standard deviation for the four

non-zero canonical correlations and the associated directions are reported. As in Table 1 of the main text only the bias and standard deviation for the X direction are reported with the results for the Y direction being nearly identical. When data are multivariate normal CCA based using the standard covariance matrix estimator performs the best, followed by transelliptical CCA. When data are simulated from a multivariate Cauchy distribution transelliptical CCA and the two robust CCA methods based on the MCD and Tyler's M estimator are consistent and perform similarly. Standard CCA estimates are not consistent due to the lack of moments. When data come from a multivariate lognormal distribution only the transelliptical CCA estimator is consistent for canonical correlations and directions. All other estimators underestimate the strength of the canonical correlations due to the non-linear transformation of the marginal distributions. When the data come from a t distribution with 5 degrees of freedom transelliptical CCA using the transformed Kendall's estimator outperforms all other methods. When data come from a t distribution with 10 degrees of freedom the CCA using the standard covariance matrix estimate performs the best, followed by transelliptical CCA using the transformed Kendall's estimator. This shows that the transformed Kendall's estimator tends to be preferred to other robust correlation or covariance estimators for the simulation settings considered. It is also preferred to the standard covariance or correlation estimator for heavier tailed ellipticals. For a fixed sample size as dimension increases the bias and standard deviation of both the correlation and direction estimates increases for all estimators. In particular for the first canonical correlation the positive finite sample bias increases as dimension increases.

Table A1: Bias (SD) of canonical correlation and direction estimates, p=q=4, n=200

		Normal	Cauchy	Lognormal	t5	t10
			Canonical Correlations			
Cor 1	Standard	0.02 (0.07)	1.38 (1.04)	-0.14 (0.18)	0.04 (0.10)	0.03 (0.08)
	Kendall	0.03 (0.08)	0.07 (0.13)	0.02 (0.08)	0.03 (0.09)	0.03 (0.09)
	MCD	0.03 (0.09)	0.07 (0.13)	-0.21 (0.14)	0.04 (0.10)	0.04 (0.10)
	M	0.03 (0.08)	0.06 (0.12)	-0.21 (0.14)	0.04 (0.10)	0.04 (0.10)
Cor 2	Standard	0.05 (0.07)	1.04 (0.58)	-0.05 (0.13)	0.09 (0.09)	0.06 (0.07)
	Kendall	0.05 (0.07)	0.09 (0.09)	0.05 (0.06)	0.06 (0.07)	0.06 (0.07)
	MCD	0.07 (0.08)	0.13 (0.11)	-0.05 (0.10)	0.09 (0.09)	0.08 (0.08)
	M	0.06 (0.08)	0.12 (0.10)	-0.06 (0.09)	0.08 (0.08)	0.08 (0.08)
Cor 3	Standard	0.02 (0.06)	0.42 (0.36)	-0.09 (0.08)	0.03 (0.07)	0.02 (0.06)
	Kendall	0.02 (0.06)	0.03 (0.08)	0.02 (0.06)	0.02 (0.06)	0.02 (0.06)
	MCD	0.02 (0.07)	0.04 (0.10)	-0.10 (0.08)	0.03 (0.07)	0.03 (0.07)
	M	0.02 (0.06)	0.04 (0.09)	-0.10 (0.08)	0.03 (0.07)	0.03 (0.07)
Cor 4	Standard	-0.04 (0.06)	-0.10 (0.20)	-0.16 (0.07)	-0.06 (0.08)	-0.05 (0.07)
	Kendall	-0.04 (0.06)	-0.07 (0.08)	-0.04 (0.07)	-0.05 (0.07)	-0.05 (0.06)
	MCD	-0.06 (0.07)	-0.09 (0.10)	-0.20 (0.08)	-0.07 (0.08)	-0.07 (0.08)
	M	-0.05 (0.07)	-0.09 (0.10)	-0.19 (0.08)	-0.06 (0.08)	-0.06 (0.08)
			Canonical Directions			
Dir 1	Standard	0.07 (0.03)	0.57 (0.44)	0.08 (0.05)	0.09 (0.05)	0.08 (0.03)
	Kendall	0.08 (0.03)	0.13 (0.06)	0.08 (0.03)	0.09 (0.04)	0.08 (0.04)
	MCD	0.08 (0.04)	0.13 (0.06)	0.12 (0.06)	0.10 (0.04)	0.09 (0.04)
	M	0.08 (0.04)	0.12 (0.05)	0.11 (0.06)	0.09 (0.04)	0.09 (0.04)
Dir 2	Standard	0.46 (0.30)	1.09 (0.34)	0.67 (0.46)	0.61 (0.36)	0.53 (0.35)
	Kendall	0.50 (0.33)	0.63 (0.37)	0.49 (0.32)	0.53 (0.33)	0.51 (0.34)
	MCD	0.55 (0.34)	0.72 (0.37)	0.74 (0.41)	0.60 (0.36)	0.58 (0.34)
	M	0.53 (0.35)	0.71 (0.38)	0.69 (0.40)	0.58 (0.35)	0.57 (0.34)
Dir 3	Standard	0.70 (0.38)	1.08 (0.31)	0.90 (0.44)	0.83 (0.39)	0.78 (0.39)
	Kendall	0.73 (0.39)	0.84 (0.40)	0.75 (0.38)	0.78 (0.38)	0.76 (0.39)
	MCD	0.76 (0.38)	0.90 (0.38)	0.93 (0.39)	0.82 (0.38)	0.81 (0.39)
	M	0.76 (0.39)	0.92 (0.38)	0.89 (0.39)	0.82 (0.38)	0.79 (0.38)
Dir 4	Standard	0.59 (0.36)	1.03 (0.30)	0.75 (0.46)	0.70 (0.38)	0.64 (0.37)
	Kendall	0.61 (0.37)	0.71 (0.38)	0.63 (0.38)	0.65 (0.38)	0.63 (0.38)
	MCD	0.64 (0.37)	0.79 (0.38)	0.80 (0.40)	0.70 (0.37)	0.67 (0.38)
	M	0.63 (0.37)	0.79 (0.38)	0.77 (0.40)	0.69 (0.37)	0.66 (0.37)

Table A2: Bias (SD) of canonical correlation and direction estimates, p=q=4, n=1000

		Normal	Cauchy	Lognormal	t5	t10
Canonical Correlations						
Cor 1	Standard	3.9E-03 (0.03)	1.46 (1.17)	-0.20 (0.10)	0.01 (0.05)	0.01 (0.04)
	Kendall	4.6E-03 (0.03)	0.02 (0.06)	0.00 (0.03)	0.01 (0.04)	0.01 (0.04)
	MCD	4.9E-03 (0.03)	0.02 (0.06)	-0.26 (0.06)	0.01 (0.04)	0.01 (0.04)
	M	0.01 (0.04)	0.01 (0.06)	-0.24 (0.06)	0.01 (0.04)	0.01 (0.04)
Cor 2	Standard	0.01 (0.03)	1.04 (0.62)	-0.12 (0.06)	0.02 (0.05)	0.01 (0.04)
	Kendall	0.01 (0.03)	0.02 (0.04)	0.01 (0.03)	0.01 (0.04)	0.01 (0.04)
	MCD	0.01 (0.03)	0.03 (0.05)	-0.14 (0.05)	0.01 (0.04)	0.01 (0.04)
	M	0.01 (0.04)	0.02 (0.05)	-0.13 (0.05)	0.01 (0.04)	0.01 (0.04)
Cor 3	Standard	0.00 (0.03)	0.41 (0.37)	-0.11 (0.05)	0.01 (0.04)	0.00 (0.03)
	Kendall	0.01 (0.03)	0.01 (0.04)	0.01 (0.03)	0.01 (0.03)	0.00 (0.03)
	MCD	0.01 (0.03)	0.01 (0.05)	-0.13 (0.04)	0.01 (0.04)	0.01 (0.04)
	M	0.01 (0.03)	0.01 (0.04)	-0.12 (0.04)	0.01 (0.04)	0.01 (0.04)
Cor 4	Standard	-0.01 (0.03)	-0.09 (0.21)	-0.13 (0.04)	-0.02 (0.04)	-0.01 (0.03)
	Kendall	-0.01 (0.03)	-0.02 (0.04)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)
	MCD	-0.01 (0.03)	-0.03 (0.05)	-0.16 (0.04)	-0.01 (0.04)	-0.02 (0.04)
	M	-0.01 (0.03)	-0.03 (0.05)	-0.14 (0.04)	-0.02 (0.04)	-0.02 (0.04)
Canonical Directions						
Dir 1	Standard	0.03 (0.01)	0.53 (0.44)	0.03 (0.02)	0.05 (0.02)	0.03 (0.01)
	Kendall	0.03 (0.01)	0.06 (0.02)	0.03 (0.01)	0.04 (0.02)	0.03 (0.01)
	MCD	0.03 (0.01)	0.05 (0.02)	0.05 (0.02)	0.04 (0.02)	0.04 (0.02)
	M	0.03 (0.01)	0.05 (0.02)	0.05 (0.02)	0.04 (0.02)	0.04 (0.02)
Dir 2	Standard	0.20 (0.12)	1.08 (0.34)	0.36 (0.36)	0.33 (0.22)	0.23 (0.15)
	Kendall	0.21 (0.14)	0.30 (0.20)	0.22 (0.14)	0.23 (0.14)	0.23 (0.14)
	MCD	0.23 (0.15)	0.38 (0.26)	0.43 (0.28)	0.29 (0.20)	0.26 (0.18)
	M	0.22 (0.15)	0.38 (0.28)	0.39 (0.27)	0.29 (0.20)	0.27 (0.19)
Dir 3	Standard	0.34 (0.23)	1.07 (0.30)	0.64 (0.46)	0.53 (0.32)	0.40 (0.26)
	Kendall	0.36 (0.24)	0.51 (0.33)	0.37 (0.24)	0.41 (0.27)	0.40 (0.26)
	MCD	0.39 (0.25)	0.61 (0.36)	0.67 (0.35)	0.50 (0.33)	0.46 (0.31)
	M	0.40 (0.27)	0.61 (0.36)	0.65 (0.36)	0.50 (0.32)	0.46 (0.30)
Dir 4	Standard	0.31 (0.22)	1.01 (0.31)	0.51 (0.45)	0.46 (0.32)	0.35 (0.26)
	Kendall	0.32 (0.23)	0.44 (0.33)	0.32 (0.23)	0.36 (0.27)	0.35 (0.26)
	MCD	0.35 (0.25)	0.52 (0.34)	0.58 (0.33)	0.43 (0.32)	0.41 (0.30)
	M	0.36 (0.26)	0.51 (0.34)	0.55 (0.35)	0.43 (0.31)	0.40 (0.29)

Table A3: Bias (SD) of canonical correlation and direction estimates, p=q=8, n=1000

		Normal	Cauchy	Lognormal	t5	t10
Canonical Correlations						
Cor 1	Standard	0.01 (0.03)	1.87 (1.16)	-0.19 (0.11)	0.02 (0.05)	0.01 (0.04)
	Kendall	0.01 (0.03)	0.03 (0.06)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)
	MCD	0.01 (0.03)	0.03 (0.06)	-0.22 (0.07)	0.02 (0.05)	0.01 (0.04)
	M	0.01 (0.03)	0.03 (0.06)	-0.22 (0.07)	0.01 (0.04)	0.01 (0.04)
Cor 2	Standard	0.02 (0.03)	1.65 (0.66)	-0.11 (0.06)	0.04 (0.05)	0.02 (0.03)
	Kendall	0.02 (0.03)	0.04 (0.05)	0.02 (0.03)	0.02 (0.04)	0.02 (0.03)
	MCD	0.02 (0.03)	0.05 (0.05)	-0.11 (0.05)	0.03 (0.04)	0.03 (0.04)
	M	0.02 (0.03)	0.05 (0.05)	-0.11 (0.04)	0.03 (0.04)	0.02 (0.04)
Cor 3	Standard	0.01 (0.03)	1.17 (0.47)	-0.10 (0.05)	0.03 (0.04)	0.02 (0.03)
	Kendall	0.02 (0.03)	0.03 (0.04)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)
	MCD	0.02 (0.03)	0.05 (0.05)	-0.09 (0.04)	0.03 (0.04)	0.03 (0.04)
	M	0.02 (0.03)	0.04 (0.04)	-0.09 (0.04)	0.03 (0.04)	0.03 (0.04)
Cor 4	Standard	4.1E-03 (0.03)	0.83 (0.35)	-0.11 (0.04)	0.01 (0.04)	3.2E-03 (0.03)
	Kendall	4.3E-03 (0.03)	0.01 (0.04)	0.00 (0.03)	0.00 (0.03)	3.5E-03 (0.03)
	MCD	4.2E-03 (0.03)	0.01 (0.05)	-0.10 (0.04)	0.01 (0.04)	4.8E-03 (0.04)
	M	4.5E-03 (0.03)	0.01 (0.05)	-0.10 (0.04)	0.01 (0.04)	4.4E-03 (0.04)
Canonical Directions						
Dir 1	Standard	0.04 (0.01)	0.66 (0.44)	0.05 (0.02)	0.06 (0.02)	0.05 (0.01)
	Kendall	0.05 (0.01)	0.08 (0.02)	0.05 (0.01)	0.05 (0.02)	0.05 (0.01)
	MCD	0.05 (0.01)	0.08 (0.02)	0.08 (0.02)	0.06 (0.02)	0.05 (0.01)
	M	0.05 (0.01)	0.07 (0.02)	0.07 (0.02)	0.06 (0.02)	0.05 (0.02)
Dir 2	Standard	0.23 (0.10)	1.26 (0.27)	0.43 (0.37)	0.37 (0.21)	0.27 (0.15)
	Kendall	0.25 (0.12)	0.35 (0.17)	0.25 (0.13)	0.26 (0.12)	0.26 (0.13)
	MCD	0.25 (0.12)	0.46 (0.28)	0.49 (0.29)	0.34 (0.18)	0.31 (0.18)
	M	0.26 (0.12)	0.42 (0.23)	0.44 (0.27)	0.33 (0.18)	0.30 (0.16)
Dir 3	Standard	0.39 (0.21)	1.27 (0.24)	0.70 (0.42)	0.58 (0.32)	0.44 (0.26)
	Kendall	0.41 (0.23)	0.56 (0.29)	0.40 (0.23)	0.45 (0.27)	0.42 (0.25)
	MCD	0.42 (0.23)	0.69 (0.33)	0.74 (0.35)	0.56 (0.31)	0.50 (0.29)
	M	0.42 (0.24)	0.66 (0.32)	0.70 (0.35)	0.55 (0.30)	0.48 (0.27)
Dir 4	Standard	0.38 (0.21)	1.26 (0.24)	0.61 (0.37)	0.55 (0.30)	0.43 (0.24)
	Kendall	0.39 (0.22)	0.53 (0.27)	0.39 (0.22)	0.44 (0.26)	0.41 (0.24)
	MCD	0.40 (0.22)	0.65 (0.29)	0.73 (0.32)	0.53 (0.29)	0.48 (0.26)
	M	0.41 (0.23)	0.62 (0.29)	0.67 (0.31)	0.52 (0.28)	0.46 (0.26)

Table A4: Bias (SD) of canonical correlation and direction estimates, p=q=16, n=200

		Normal	Cauchy	Lognormal	t5	t10
			Canonical Correlations			
Cor 1	Standard	0.10 (0.08)	2.47 (1.13)	-0.07 (0.19)	0.17 (0.10)	0.12 (0.08)
	Kendall	0.13 (0.08)	0.42 (0.20)	0.12 (0.08)	0.16 (0.09)	0.14 (0.09)
	MCD	0.18 (0.10)	0.29 (0.12)	-0.01 (0.15)	0.22 (0.11)	0.21 (0.11)
	M	0.17 (0.10)	0.31 (0.13)	-0.03 (0.15)	0.20 (0.11)	0.18 (0.10)
Cor 2	Standard	0.18 (0.06)	2.36 (0.67)	0.13 (0.10)	0.33 (0.11)	0.23 (0.06)
	Kendall	0.21 (0.06)	0.43 (0.09)	0.21 (0.06)	0.25 (0.07)	0.23 (0.06)
	MCD	0.33 (0.08)	0.50 (0.09)	0.31 (0.08)	0.41 (0.08)	0.39 (0.08)
	M	0.31 (0.08)	0.54 (0.10)	0.26 (0.08)	0.36 (0.08)	0.35 (0.08)
Cor 3	Standard	0.18 (0.05)	1.95 (0.52)	0.13 (0.06)	0.31 (0.07)	0.23 (0.05)
	Kendall	0.21 (0.05)	0.39 (0.07)	0.21 (0.05)	0.25 (0.05)	0.23 (0.05)
	MCD	0.31 (0.06)	0.49 (0.07)	0.31 (0.06)	0.40 (0.06)	0.37 (0.06)
	M	0.30 (0.06)	0.51 (0.08)	0.27 (0.06)	0.36 (0.06)	0.34 (0.06)
Cor 4	Standard	0.17 (0.04)	1.65 (0.41)	0.13 (0.04)	0.29 (0.05)	0.21 (0.04)
	Kendall	0.19 (0.04)	0.36 (0.06)	0.19 (0.04)	0.23 (0.04)	0.21 (0.04)
	MCD	0.29 (0.05)	0.46 (0.06)	0.29 (0.05)	0.37 (0.05)	0.35 (0.05)
	M	0.28 (0.05)	0.48 (0.07)	0.26 (0.05)	0.34 (0.05)	0.32 (0.05)
			Canonical Directions			
Dir 1	Standard	0.15 (0.03)	0.81 (0.44)	0.19 (0.08)	0.19 (0.06)	0.16 (0.03)
	Kendall	0.17 (0.04)	0.29 (0.07)	0.17 (0.04)	0.19 (0.04)	0.18 (0.04)
	MCD	0.21 (0.05)	0.26 (0.06)	0.28 (0.08)	0.24 (0.06)	0.23 (0.06)
	M	0.21 (0.05)	0.26 (0.07)	0.26 (0.07)	0.22 (0.05)	0.22 (0.05)
Dir 2	Standard	0.73 (0.29)	1.34 (0.20)	0.98 (0.38)	0.93 (0.31)	0.79 (0.28)
	Kendall	0.78 (0.29)	0.97 (0.27)	0.77 (0.28)	0.84 (0.29)	0.80 (0.27)
	MCD	0.96 (0.29)	1.07 (0.28)	1.13 (0.29)	1.02 (0.28)	0.99 (0.28)
	M	0.95 (0.29)	1.08 (0.28)	1.07 (0.31)	0.96 (0.27)	0.95 (0.28)
Dir 3	Standard	1.01 (0.29)	1.37 (0.18)	1.23 (0.25)	1.17 (0.25)	1.07 (0.27)
	Kendall	1.05 (0.28)	1.19 (0.24)	1.05 (0.29)	1.11 (0.27)	1.08 (0.26)
	MCD	1.20 (0.24)	1.24 (0.21)	1.29 (0.20)	1.23 (0.22)	1.21 (0.23)
	M	1.18 (0.25)	1.26 (0.22)	1.26 (0.21)	1.21 (0.23)	1.19 (0.24)
Dir 4	Standard	1.12 (0.26)	1.38 (0.17)	1.29 (0.21)	1.25 (0.22)	1.19 (0.24)
	Kendall	1.16 (0.25)	1.27 (0.21)	1.16 (0.25)	1.20 (0.24)	1.18 (0.24)
	MCD	1.26 (0.22)	1.31 (0.19)	1.33 (0.18)	1.28 (0.20)	1.29 (0.20)
	M	1.25 (0.22)	1.31 (0.19)	1.32 (0.19)	1.28 (0.20)	1.27 (0.21)

Table A5: Bias (SD) of canonical correlation and direction estimates, p=q=16, n=1000

		Normal	Cauchy	Lognormal	t5	t10
			Canonical Correlations			
Cor 1	Standard	0.02 (0.03)	2.44 (1.25)	-0.18 (0.11)	0.04 (0.05)	0.02 (0.04)
	Kendall	0.02 (0.03)	0.06 (0.06)	0.02 (0.04)	0.03 (0.04)	0.02 (0.04)
	MCD	0.02 (0.03)	0.05 (0.06)	-0.19 (0.07)	0.04 (0.05)	0.03 (0.04)
	M	0.02 (0.03)	0.05 (0.05)	-0.19 (0.07)	0.03 (0.04)	0.03 (0.04)
Cor 2	Standard	0.03 (0.03)	2.31 (0.76)	-0.09 (0.06)	0.08 (0.06)	0.04 (0.03)
	Kendall	0.04 (0.03)	0.07 (0.05)	0.04 (0.03)	0.05 (0.03)	0.04 (0.03)
	MCD	0.04 (0.03)	0.10 (0.05)	-0.05 (0.05)	0.07 (0.04)	0.06 (0.04)
	M	0.04 (0.03)	0.09 (0.05)	-0.07 (0.05)	0.06 (0.04)	0.05 (0.04)
Cor 3	Standard	0.04 (0.03)	1.85 (0.51)	-0.07 (0.04)	0.08 (0.04)	0.05 (0.03)
	Kendall	0.04 (0.03)	0.07 (0.04)	0.04 (0.03)	0.05 (0.03)	0.04 (0.03)
	MCD	0.04 (0.03)	0.11 (0.04)	-0.03 (0.03)	0.08 (0.04)	0.07 (0.03)
	M	0.04 (0.03)	0.09 (0.04)	-0.04 (0.04)	0.06 (0.03)	0.06 (0.03)
Cor 4	Standard	0.03 (0.03)	1.54 (0.39)	-0.07 (0.03)	0.06 (0.04)	0.04 (0.03)
	Kendall	0.03 (0.03)	0.06 (0.04)	0.03 (0.03)	0.04 (0.03)	0.03 (0.03)
	MCD	0.03 (0.03)	0.09 (0.04)	-0.02 (0.03)	0.06 (0.04)	0.05 (0.03)
	M	0.03 (0.03)	0.08 (0.04)	-0.04 (0.03)	0.05 (0.04)	0.05 (0.03)
			Canonical Directions			
Dir 1	Standard	0.06 (0.01)	0.77 (0.44)	0.08 (0.02)	0.09 (0.02)	0.07 (0.01)
	Kendall	0.07 (0.01)	0.11 (0.02)	0.07 (0.01)	0.08 (0.01)	0.07 (0.01)
	MCD	0.07 (0.01)	0.11 (0.02)	0.11 (0.02)	0.09 (0.02)	0.08 (0.02)
	M	0.07 (0.01)	0.10 (0.02)	0.10 (0.02)	0.08 (0.02)	0.08 (0.02)
Dir 2	Standard	0.29 (0.11)	1.36 (0.20)	0.51 (0.35)	0.47 (0.24)	0.33 (0.12)
	Kendall	0.31 (0.11)	0.44 (0.19)	0.31 (0.12)	0.34 (0.13)	0.32 (0.13)
	MCD	0.31 (0.12)	0.54 (0.23)	0.62 (0.29)	0.46 (0.20)	0.40 (0.16)
	M	0.32 (0.12)	0.50 (0.22)	0.54 (0.27)	0.41 (0.18)	0.37 (0.15)
Dir 3	Standard	0.46 (0.21)	1.37 (0.18)	0.80 (0.39)	0.71 (0.30)	0.51 (0.24)
	Kendall	0.49 (0.22)	0.66 (0.28)	0.48 (0.22)	0.54 (0.24)	0.51 (0.24)
	MCD	0.49 (0.22)	0.81 (0.30)	0.91 (0.32)	0.70 (0.28)	0.62 (0.27)
	M	0.50 (0.22)	0.76 (0.31)	0.83 (0.31)	0.64 (0.27)	0.59 (0.26)
Dir 4	Standard	0.47 (0.20)	1.37 (0.17)	0.81 (0.34)	0.73 (0.28)	0.55 (0.22)
	Kendall	0.50 (0.21)	0.68 (0.24)	0.50 (0.20)	0.55 (0.22)	0.53 (0.22)
	MCD	0.51 (0.21)	0.87 (0.28)	1.00 (0.29)	0.72 (0.26)	0.65 (0.24)
	M	0.51 (0.20)	0.82 (0.27)	0.89 (0.28)	0.66 (0.24)	0.61 (0.24)

A.2.2 Confidence intervals for canonical correlations and canonical direction loadings

This section reports confidence interval coverages for transelliptical CCA correlations and direction loadings. The simulation set-ups are the same as section 3.1 of the main text but results or simulations using the multivariate lognormal distribution are not included because they are similar to results using the multivariate normal distribution for transelliptical CCA. Tables A6, A7, and A8 report the coverages for bootstrap and asymptotic confidence intervals for the canonical correlations. For Table A6 p=q=4, for Table A7 p=q=8, and for Table A8 p=q=16. When p=q=4 the asymptotic confidence intervals perform comparably the bootstrap confidence intervals, while when p=q=8 and p=q=16 the asymptotic confidence intervals perform poorly, particularly when n=200. This poor performance in the higher dimension setting is due to the finite sample bias increasing as dimension increases. Normal approximation bootstrap confidence intervals are able to help control for this bias. Tables A9 through A14 report the bootstrap and asymptotic confidence interval coverages for canonical direction loadings. Only the coverages for the X direction are reported with the coverages for the Y direction being nearly identical. In addition results for t distribution with 5 and 10 degrees of freedom are not reported, and are similar to results for the multivariate Cauchy and multivariate normal distribution. Table A9 reports coverages when p=q=4 and data are simulated from a multivariate normal distribution, and Table A10 reports coverages when p=q=4 and data are simulated from a multivariate Cauchy distribution. In both cases there is undercoverage for loadings higher order directions, which improves as sample size increases. Table A11 reports coverages when p=q=8 and data are simulated from a multivariate normal distribution, and Table A12 reports coverages when p=q=8 and data are simulated from a multivariate Cauchy distribution. Table A13 reports coverages when p=q=16 and data are simulated from a multivariate normal distribution, and Table A14 reports coverages when p=q=16 and data are simulated from a multivariate Cauchy distribution. An increase in dimension does not have a large effect on the confidence interval coverages for the direction loadings, unlike the confidence interval coverages for the canonical correlations which get much worse when p=q=16.

Table A6: Bootstrap and asymptotic confidence interval coverages for transelliptical canonical correlations, p=q=4

	Bootstrap Coverages				Asymptotic Coverages			
	Normal	Cauchy	t5	t10	Normal	Cauchy	t5	t10
				n=200				
Cor 1	0.94	0.88	0.91	0.93	0.94	0.83	0.89	0.90
Cor 2	0.92	0.92	0.91	0.91	0.89	0.86	0.88	0.90
Cor 3	0.90	0.92	0.91	0.91	0.97	0.98	0.97	0.97
Cor 4	0.84	0.76	0.80	0.81	0.94	0.95	0.96	0.95
				n=1000				
Cor 1	0.94	0.94	0.95	0.95	0.95	0.92	0.92	0.95
Cor 2	0.93	0.91	0.93	0.92	0.95	0.94	0.94	0.92
Cor 3	0.90	0.90	0.92	0.90	0.96	0.97	0.97	0.96
Cor 4	0.90	0.90	0.90	0.90	0.96	0.96	0.96	0.96

Table A7: Bootstrap and asymptotic confidence interval coverages for transelliptical canonical correlations, p=q=8

	Bootstrap Coverages				Asymptotic Coverages			
	Normal	Cauchy	t5	t10	Normal	Cauchy	t5	t10
				n=200				
Cor 1	0.90	0.81	0.88	0.90	0.84	0.60	0.79	0.82
Cor 2	0.90	0.90	0.91	0.89	0.73	0.51	0.70	0.70
Cor 3	0.87	0.88	0.86	0.88	0.85	0.75	0.82	0.83
Cor 4	0.85	0.86	0.84	0.81	0.97	0.96	0.97	0.98
				n=1000				
Cor 1	0.94	0.90	0.93	0.92	0.93	0.89	0.93	0.93
Cor 2	0.93	0.92	0.91	0.94	0.93	0.89	0.90	0.91
Cor 3	0.89	0.89	0.91	0.90	0.94	0.93	0.93	0.93
Cor 4	0.91	0.91	0.89	0.93	0.98	0.97	0.96	0.97

Table A8: Bootstrap and asymptotic confidence interval coverages for transelliptical canonical correlations, p=q=16

	Bootstrap Coverages				Asymptotic Coverages			
	Normal	Cauchy	t5	t10	Normal	Cauchy	t5	t10
				n=200				
Cor 1	0.86	0.53	0.80	0.83	0.56	0.18	0.47	0.53
Cor 2	0.84	0.81	0.83	0.82	0.14	0.00	0.05	0.08
Cor 3	0.82	0.87	0.88	0.84	0.07	0.00	0.03	0.06
Cor 4	0.86	0.91	0.90	0.87	0.11	0.00	0.04	0.06
				n=1000				
Cor 1	0.92	0.89	0.92	0.93	0.87	0.73	0.86	0.86
Cor 2	0.90	0.90	0.91	0.91	0.79	0.62	0.77	0.77
Cor 3	0.89	0.86	0.89	0.89	0.77	0.64	0.76	0.78
Cor 4	0.87	0.82	0.88	0.88	0.88	0.79	0.84	0.86

Table A9: Bootstrap and asymptotic confidence interval coverages for transelliptical canonical directions for data with multivariate normal distribution, p=q=4

	Bootstrap Coverages				Asymptotic Coverages			
	Dir 1	Dir 2	Dir 3	Dir 4	Dir 1	Dir 2	Dir 3	Dir 4
				n=200				
Variable 1	1.00	0.95	0.93	0.95	1.00	0.94	0.95	0.95
Variable 2	0.96	0.98	0.80	0.89	0.95	0.96	0.90	0.79
Variable 3	0.96	0.78	0.97	0.69	0.97	0.89	0.94	0.68
Variable 4	0.96	0.89	0.72	0.97	0.97	0.84	0.75	0.94
				n=1000				
Variable 1	1.00	0.95	0.95	0.95	1.00	0.95	0.94	0.95
Variable 2	0.94	1.00	0.92	0.95	0.95	1.00	0.94	0.92
Variable 3	0.96	0.92	0.99	0.85	0.97	0.93	0.99	0.86
Variable 4	0.96	0.94	0.86	0.99	0.97	0.93	0.87	0.99

Table A10: Bootstrap and asymptotic confidence interval coverages for transelliptical canonical directions for data with multivariate Cauchy distribution, p=q=4

	Bootstrap Coverages				Asymptotic Coverages			
	Dir 1	Dir 2	Dir 3	Dir 4	Dir 1	Dir 2	Dir 3	Dir 4
				n=200				
Variable 1	1.00	0.96	0.96	0.96	1.00	0.94	0.94	0.94
Variable 2	0.96	0.96	0.73	0.85	0.97	0.94	0.83	0.77
Variable 3	0.95	0.71	0.95	0.62	0.95	0.88	0.92	0.66
Variable 4	0.95	0.81	0.65	0.95	0.96	0.81	0.76	0.94
				n=1000				
Variable 1	1.00	0.96	0.95	0.94	1.00	0.94	0.94	0.94
Variable 2	0.95	1.00	0.88	0.93	0.96	0.99	0.94	0.89
Variable 3	0.95	0.88	0.99	0.77	0.95	0.94	0.98	0.79
Variable 4	0.95	0.92	0.80	0.98	0.96	0.90	0.83	0.98

Table A11: Bootstrap and asymptotic confidence interval coverages for transelliptical canonical directions for data with multivariate normal distribution, p=q=8

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

Table A12: Bootstrap and asymptotic confidence interval coverages for transelliptical canonical directions for data with multivariate Cauchy distribution, p=q=8

	Bootstrap Coverages				Asymptotic Coverages			
	Dir 1	Dir 2	Dir 3	Dir 4	Dir 1	Dir 2	Dir 3	Dir 4
n=200								
Variable 1	1.00	0.95	0.94	0.91	1.00	0.93	0.93	0.95
Variable 2	0.97	0.93	0.70	0.82	0.97	0.89	0.84	0.74
Variable 3	0.97	0.66	0.89	0.60	0.99	0.93	0.92	0.85
Variable 4	0.97	0.80	0.61	0.90	0.98	0.84	0.81	0.89
Variable 5	0.96	0.92	0.87	0.81	0.99	0.98	0.96	0.97
Variable 6	0.96	0.92	0.86	0.84	0.99	0.99	0.98	0.99
Variable 7	0.96	0.91	0.88	0.85	1.00	0.99	0.98	0.98
Variable 8	0.96	0.89	0.88	0.82	0.99	0.99	0.97	0.97
n=1000								
Variable 1	1.00	0.94	0.95	0.95	1.00	0.95	0.95	0.94
Variable 2	0.95	1.00	0.89	0.89	0.96	0.99	0.93	0.89
Variable 3	0.94	0.88	0.99	0.77	0.97	0.93	0.98	0.80
Variable 4	0.95	0.90	0.78	0.99	0.98	0.92	0.82	0.98
Variable 5	0.95	0.95	0.93	0.96	0.99	0.99	0.99	0.99
Variable 6	0.96	0.95	0.94	0.96	0.99	0.99	0.99	1.00
Variable 7	0.95	0.95	0.95	0.95	0.99	1.00	0.99	1.00
Variable 8	0.94	0.94	0.95	0.96	0.99	0.99	0.99	0.99

Table A13: Bootstrap and asymptotic confidence interval coverages for transelliptical canonical directions for data with multivariate normal distribution, p=q=16

	Bootstrap Coverages				Asymptotic Coverages			
	Dir 1	Dir 2	Dir 3	Dir 4	Dir 1	Dir 2	Dir 3	Dir 4
				n=200				
Variable 1	1.00	0.96	0.94	0.91	0.98	0.94	0.93	0.94
Variable 2	0.98	0.93	0.69	0.83	0.98	0.83	0.88	0.76
Variable 3	0.97	0.67	0.80	0.65	0.99	0.98	0.85	0.90
Variable 4	0.98	0.79	0.59	0.71	0.99	0.87	0.83	0.83
Variable 5	0.98	0.91	0.86	0.80	1.00	0.99	0.97	0.99
Variable 6	0.98	0.92	0.87	0.79	1.00	1.00	0.99	1.00
Variable 7	0.98	0.92	0.84	0.79	1.00	1.00	1.00	1.00
Variable 8	0.97	0.92	0.84	0.80	1.00	1.00	1.00	1.00
Variable 9	0.98	0.94	0.89	0.77	1.00	1.00	1.00	1.00
Variable 10	0.97	0.92	0.86	0.80	1.00	1.00	1.00	1.00
Variable 11	0.98	0.92	0.88	0.78	1.00	1.00	1.00	1.00
Variable 12	0.98	0.93	0.88	0.77	1.00	1.00	1.00	1.00
Variable 13	0.98	0.91	0.87	0.75	1.00	1.00	1.00	1.00
Variable 14	0.97	0.93	0.85	0.80	1.00	1.00	1.00	1.00
Variable 15	0.98	0.92	0.88	0.80	1.00	1.00	1.00	1.00
Variable 16	0.98	0.93	0.87	0.79	1.00	1.00	1.00	1.00
				n=1000				
Variable 1	1.00	0.96	0.95	0.95	0.99	0.95	0.95	0.94
Variable 2	0.96	1.00	0.92	0.93	0.96	0.96	0.96	0.92
Variable 3	0.96	0.92	0.99	0.86	0.96	0.95	0.87	0.97
Variable 4	0.95	0.95	0.85	0.99	0.97	0.96	0.94	0.86
Variable 5	0.95	0.94	0.94	0.95	0.99	0.99	1.00	1.00
Variable 6	0.96	0.95	0.95	0.96	1.00	1.00	1.00	1.00
Variable 7	0.95	0.94	0.95	0.96	1.00	1.00	1.00	1.00
Variable 8	0.95	0.94	0.95	0.94	1.00	1.00	1.00	1.00
Variable 9	0.95	0.95	0.95	0.95	1.00	1.00	1.00	1.00
Variable 10	0.96	0.95	0.95	0.95	1.00	1.00	1.00	1.00
Variable 11	0.95	0.96	0.95	0.95	1.00	1.00	1.00	1.00
Variable 12	0.96	0.95	0.94	0.96	1.00	1.00	1.00	1.00
Variable 13	0.96	0.96	0.94	0.94	1.00	1.00	1.00	1.00
Variable 14	0.95	0.95	0.94	0.95	1.00	1.00	1.00	1.00
Variable 15	0.95	0.95	0.95	0.95	1.00	1.00	1.00	1.00
Variable 16	0.96	0.94	0.96	0.95	1.00	1.00	1.00	1.00

Table A14: Bootstrap and asymptotic confidence interval coverages for transelliptical canonical directions for data with multivariate Cauchy distribution, p=q=16

	Bootstrap Coverages				Asymptotic Coverages			
	Dir 1	Dir 2	Dir 3	Dir 4	Dir 1	Dir 2	Dir 3	Dir 4
				n=200				
Variable 1	1.00	0.99	0.96	0.93	0.98	0.91	0.91	0.91
Variable 2	0.99	0.86	0.66	0.78	0.97	0.68	0.81	0.74
Variable 3	1.00	0.64	0.73	0.66	1.00	0.97	0.77	0.94
Variable 4	1.00	0.72	0.65	0.58	0.99	0.89	0.88	0.77
Variable 5	1.00	0.88	0.82	0.76	1.00	0.99	0.98	0.99
Variable 6	1.00	0.86	0.82	0.77	1.00	1.00	0.98	0.99
Variable 7	1.00	0.89	0.81	0.77	1.00	1.00	0.99	1.00
Variable 8	1.00	0.87	0.84	0.77	1.00	1.00	1.00	1.00
Variable 9	0.99	0.88	0.82	0.78	1.00	1.00	1.00	1.00
Variable 10	1.00	0.90	0.84	0.76	1.00	1.00	1.00	1.00
Variable 11	1.00	0.89	0.82	0.76	1.00	1.00	1.00	1.00
Variable 12	0.99	0.89	0.83	0.77	1.00	1.00	1.00	1.00
Variable 13	1.00	0.88	0.83	0.76	1.00	1.00	1.00	1.00
Variable 14	0.99	0.89	0.83	0.80	1.00	1.00	1.00	1.00
Variable 15	1.00	0.91	0.82	0.77	1.00	1.00	1.00	1.00
Variable 16	1.00	0.88	0.82	0.76	1.00	1.00	1.00	1.00
				n=1000				
Variable 1	1.00	0.95	0.93	0.94	1.00	0.94	0.94	0.93
Variable 2	0.97	0.99	0.87	0.90	0.96	0.96	0.93	0.89
Variable 3	0.95	0.86	0.98	0.71	0.98	0.95	0.92	0.84
Variable 4	0.95	0.92	0.74	0.98	0.98	0.92	0.84	0.94
Variable 5	0.95	0.95	0.92	0.94	0.99	1.00	0.99	1.00
Variable 6	0.94	0.95	0.94	0.93	1.00	1.00	1.00	1.00
Variable 7	0.96	0.94	0.92	0.94	1.00	1.00	1.00	1.00
Variable 8	0.95	0.94	0.93	0.94	1.00	1.00	1.00	1.00
Variable 9	0.95	0.94	0.93	0.94	1.00	1.00	1.00	1.00
Variable 10	0.94	0.94	0.92	0.94	1.00	1.00	1.00	1.00
Variable 11	0.94	0.94	0.94	0.93	1.00	1.00	1.00	1.00
Variable 12	0.93	0.94	0.93	0.92	1.00	1.00	1.00	1.00
Variable 13	0.94	0.94	0.93	0.93	1.00	1.00	1.00	1.00
Variable 14	0.96	0.93	0.95	0.93	1.00	1.00	1.00	1.00
Variable 15	0.95	0.94	0.93	0.93	1.00	1.00	1.00	1.00
Variable 16	0.96	0.93	0.91	0.95	1.00	1.00	1.00	1.00

A.2.3 Testing for non-zero canonical correlations

Table A15 reports the type I error and power of the permutation and bootstrap testing procedures described in Section 3.3 of the main text when using standard covariance matrix estimates. This is in contrast to Table 3.2 in the text where the permutation and bootstrap testing procedure were used with the transformed Kendall's estimator. The simulation set-ups are the same as Section 3.3 of the main text. Just as in Table 3.2 of the main text the permutation test only controls type I error at the desired 0.05 level when data are multivariate normal or multivariate lognormal. For both permutation an bootstrap based testing procedures using the standard covariance matrix estimator results in higher power than the transformed Kendall's estimator when data are multivariate normal. However when data come from a multivariate t distribution with five or ten degrees of freedom this relationship is reversed, with the transformed Kendall's estimator resulting in higher power when n=1000. When n=200 using the standard covariance matrix estimate results in inflated type I error for both multivariate t distributions. Also, unsurprisingly even the bootstrap testing procedure breaks down when data are multivariate Cauchy and the standard covariance matrix estimator is used, with a type I error of 1 for both sample sizes.

Table A15: Power and type I error for permutation and bootstrap testing procedures using standard correlation estimates

	True correlation	0	0.2	0.4	0.6	0.8
		n=200				
Normal	Standard Bootstrap	0.06	0.19	0.92	1.00	1.00
	Standard Perm	0.05	0.17	0.90	1.00	1.00
Cauchy	Standard Bootstrap	1.00	1.00	1.00	1.00	1.00
	Standard Perm	1.00	1.00	1.00	1.00	1.00
Lognormal	Standard Bootstrap	0.03	0.06	0.30	0.87	0.99
	Standard Perm	0.05	0.08	0.30	0.83	1.00
t5	Standard Bootstrap	0.30	0.40	0.84	1.00	1.00
	Standard Perm	0.76	0.85	0.99	1.00	1.00
t10	Standard Bootstrap	0.11	0.23	0.86	1.00	1.00
	Standard Perm	0.28	0.45	0.96	1.00	1.00
		n=1000				
Normal	Standard Bootstrap	0.02	0.91	1.00	1.00	1.00
	Standard Perm	0.05	0.97	1.00	1.00	1.00
Cauchy	Standard Bootstrap	1.00	1.00	1.00	1.00	1.00
	Standard Perm	1.00	1.00	1.00	1.00	1.00
Lognormal	Standard Bootstrap	0.00	0.09	0.92	0.99	1.00
	Standard Perm	0.06	0.29	0.99	1.00	1.00
t5	Standard Bootstrap	0.01	0.28	0.99	1.00	1.00
	Standard Perm	0.94	1.00	1.00	1.00	1.00
t10	Standard Bootstrap	0.01	0.70	1.00	1.00	1.00
	Standard Perm	0.37	0.99	1.00	1.00	1.00

A.3 Additional results for White matter tractography data and executive function in six year old children

This section reports additional results using transelliptical CCA and standard CCA to investigate the relationship between white matter structure and executive function in six year old children. A description of the data set and the white matter tracts and executive function (EF) tests that are used can be found in Section 5 of the main text. Table A16 reports the first canonical direction and correlation using average fractal anisotropy (FA) for each of the white matter tracts. Analysis methods are the same to the results reported in the main text using average radial diffusivity (RD) instead of average FA. The results for transelliptical CCA using average RD values can be found in Table 7 of the main text. The sign of the direction loadings for all of the EF tests except CANTAB Stockings of Cambridge (SOC) are the same between transelliptical CCA using average FA values for the white matter tracts and average RD values. In both cases the Stanford Binet Verbal Fluid Reasoning Score (SB V) has the loading with the highest absolute value. The transelliptical CCA direction loadings for average FA values have the opposite sign of the first direction loadings for average RD values in 11 of the 16 tracts. This is to be expected because RD tends to decrease as myelination increases, while FA tends to increase as myelination increases, so we would expect the effects of FA and RD to go in opposite directions. The loadings for the EF variables and average FA values are similar between transelliptical CCA and standard CCA, and the jackknife corrected correlation for both is close to 0.35. Table A17 reports the results for the first canonical correlation and direction using average AD for each of the white matter tracts. The same methods are used as the analysis using average RD and average FA. Transelliptical CCA did not find a significant direction, and the first direction using standard CCA is only marginally significant (p=0.05).

Table A16: Estimates for first canonical correlation and directions for white matter tract FA values and EF tests for transelliptical and standard CCA

DTI Vars	Transelliptical CCA Loadings	Boot CI	Asymp CI	Standard CCA Loadings	Boot CI
ARC FT Left	-0.57	(-1.50, 0.03)	(-1.15, 0.00)	-0.11	(-0.74, 0.47)
ARC FT Right	-0.05	(-1.04, 0.95)	(-0.99, 0.89)	-0.08	(-0.80, 0.59)
ARC FP Left	0.13	(-0.42, 0.72)	(-0.36, 0.63)	-0.08	(-0.64, 0.42)
ARC FP Right	-0.27	(-1.07, 0.27)	(-0.89, 0.35)	-0.38	(-0.94, -0.08)
ARC TP Left	0.15	(-0.28, 0.64)	(-0.25, 0.55)	-0.10	(-0.57, 0.26)
ARC TP Right	0.10	(-0.41, 0.71)	(-0.39, 0.58)	0.14	(-0.23, 0.61)
CGC Left	-0.09	(-0.84, 0.51)	(-0.71, 0.53)	-0.44	(-1.30, 0.11)
CGC Right	0.05	(-0.50, 0.68)	(-0.44, 0.55)	0.36	(-0.10, 1.07)
CTPF Left	0.01	(-0.62, 0.57)	(-0.50, 0.52)	-0.23	(-0.79, 0.17)
CTPF Right	-0.29	(-1.02, 0.29)	(-0.81, 0.23)	-0.25	(-0.82, 0.16)
Genu	0.16	(-0.55, 1.00)	(-0.50, 0.81)	0.42	(-0.07, 1.26)
ILF Left	-0.06	(-1.01, 0.79)	(-0.85, 0.74)	0.18	(-0.40, 0.86)
ILF Right	1.02	(0.41, 2.26)	(0.28, 1.76)	0.51	(-0.03, 1.36)
IFOF Left	-0.09	(-1.02, 0.76)	(-0.90, 0.71)	-0.52	(-1.38, -0.01)
IFOF Right	-0.55	(-1.79, 0.53)	(-1.66, 0.57)	0.10	(-0.70, 0.91)
SLF Left	0.10	(-0.42, 0.70)	(-0.38, 0.57)	0.06	(-0.39, 0.55)
SLF Right	0.09	(-0.52, 0.85)	(-0.59, 0.77)	0.24	(-0.29, 0.99)
Splenium	0.68	(0.38, 1.51)	(0.29, 1.08)	0.72	(0.47, 1.43)
UNC Left	-0.22	(-0.99, 0.39)	(-0.76, 0.32)	0.07	(-0.48, 0.65)
UNC Right	0.04	(-0.83, 0.78)	(-0.74, 0.81)	-0.25	(-0.89, 0.28)
EF Vars					
SB V	0.98	(0.64, 1.87)	(0.74, 1.22)	0.85	(0.48, 1.55)
SB NV	-0.70	(-1.43, -0.28)	(-1.08, -0.31)	-0.38	(-1.05, 0.15)
Brief	-0.38	(-0.96, -0.05)	(-0.72, -0.04)	-0.52	(-1.19, -0.14)
SOC	0.09	(-0.51, 0.73)	(-0.39, 0.57)	0.27	(-0.22, 0.92)
SSP	-0.17	(-1.05, 0.74)	(-0.95, 0.61)	-0.21	(-0.90, 0.40)
Cor	0.55			0.47	
Jackknife Cor	0.34			0.31	
Pval	0.04			0.01	
N	214			214	

Table A17: Estimates for first canonical correlation and directions for white matter tract AD values and EF tests for transelliptical and standard CCA

DTI Vars	Transelliptical CCA Loadings	Boot CI	Asymp CI	Standard CCA Loadings	Boot CI
ARC FT Left	0.33	(-0.45, 1.34)	(-0.42, 1.08)	0.14	(-0.40, 0.77)
ARC FT Right	0.22	(-0.79, 1.56)	(-0.89, 1.33)	0.00	(-0.67, 0.68)
ARC FP Left	-0.42	(-1.76, 0.35)	(-1.50, 0.66)	-0.11	(-0.74, 0.42)
ARC FP Right	-0.29	(-1.29, 0.25)	(-0.88, 0.29)	-0.40	(-1.12, 0.02)
ARC TP Left	0.19	(-0.44, 1.09)	(-0.38, 0.75)	0.11	(-0.27, 0.59)
ARC TP Right	-0.16	(-1.04, 0.46)	(-0.60, 0.28)	-0.24	(-0.73, 0.02)
CGC Left	-0.50	(-1.54, 0.28)	(-1.26, 0.27)	-0.62	(-1.45, -0.33)
CGC Right	0.21	(-0.49, 1.06)	(-0.62, 1.03)	0.42	(0.02, 1.20)
CTPF Left	-0.01	(-0.72, 0.91)	(-0.65, 0.63)	0.22	(-0.13, 0.75)
CTPF Right	0.32	(-0.93, 1.57)	(-0.71, 1.35)	-0.15	(-0.71, 0.33)
Genu	0.45	(-0.43, 1.95)	(-0.77, 1.68)	0.72	(0.43, 1.59)
ILF Left	-0.47	(-2.13, 1.76)	(-1.26, 0.33)	0.05	(-0.51, 0.63)
ILF Right	0.75	(-1.19, 2.54)	(-0.46, 1.97)	0.33	(-0.07, 0.99)
IFOF Left	0.55	(-1.70, 2.29)	(-0.25, 1.36)	-0.08	(-0.69, 0.49)
IFOF Right	-1.23	(-3.44, 0.68)	(-2.42, -0.05)	-0.66	(-1.49, -0.39)
SLF Left	-0.18	(-1.12, 0.59)	(-0.95, 0.59)	-0.30	(-0.85, 0.00)
SLF Right	0.49	(0.12, 1.44)	(-0.06, 1.05)	0.59	(0.39, 1.27)
Splenium	0.03	(-0.69, 0.83)	(-0.74, 0.80)	0.00	(-0.42, 0.42)
UNC Left	-0.42	(-1.36, 0.67)	(-1.05, 0.22)	0.26	(-0.21, 0.94)
UNC Right	0.26	(-1.08, 1.39)	(-0.46, 0.98)	0.07	(-0.40, 0.58)
EF Vars					
SB V	0.42	(-0.30, 1.35)	(-0.45, 1.29)	0.40	(-0.13, 1.13)
SB NV	-0.34	(-1.14, 0.37)	(-1.00, 0.33)	-0.13	(-0.87, 0.54)
Brief	-0.38	(-1.27, 0.20)	(-1.49, 0.73)	-0.78	(-1.70, -0.39)
SOC	0.22	(-0.32, 1.05)	(-0.43, 0.86)	0.14	(-0.38, 0.76)
SSP	0.73	(0.08, 2.01)	(-0.08, 1.54)	0.30	(-0.41, 1.21)
Cor	0.53			0.44	
Jackknife Cor	0.00			0.22	
Pval	0.60			0.06	
N	214			214	

APPENDIX B: TECHNICAL DETAILS AND ADDITIONAL SIMULATION RESULTS FOR CHAPTER 3

B.1 Proof of Theorems

Proof of Theorem 4.3

Proof. Define $\theta_C = [S^{(1)}, \ldots, S^{(p)}, \Lambda^{(1)}, \ldots, \Lambda^{(p)}, S^{(1,2)}, \ldots, S^{(p-1,p)}]^T$ and $\hat{\theta}_C = [\hat{S}^{(1)}, \ldots, \hat{S}^{(p)}, \hat{\Lambda}^{(1)}, \ldots, \hat{\Lambda}^{(p)}, \hat{S}^{(1,2)}, \ldots, \hat{S}^{(p-1,p)}]^T$ where $\hat{S}^{(j)}$ is the Kaplan-Meier estimator, $\hat{\Lambda}^{(j)}$ is the Nelson-Aalen estimator, and $\hat{S}^{(j,j')}$ is the Dabrowska estimator. Based previously shown results for the Kaplan-Meier, Nelson-Aalen estimator and Dabrowska, (see Kosorok (2008) for further details on the Kaplan-Meier and Nelson-Aalen estimators and Cheng et al. (2007) for further details on the Dabrowska estimator)

$$\sqrt{n}[\hat{\theta}_C - \theta_C] \to Z_{\theta_C},$$

where $Z_{\theta_C} = [Z_{S_1}, \ldots, Z_{S_p}, Z_{\Lambda_1}, \ldots, Z_{\Lambda_p}, Z_{S_{1,2}}, \ldots, Z_{S_{p-1,p}}]^T$ is a $2p + (p^2 - p)/2$ dimensional mean 0 Gaussian process. Using this set up we can show the results for $\widehat{CN}$ and $\widehat{CM}$

1. $\widehat{CN}$: Consider the mapping

$$\phi_{CNjj'}(\theta_C) = S^{(j,j')} - S^{(j)}S^{(j')}.$$

Using this mapping

$$\sqrt{n}[\text{Vec}(\widehat{CN}) - \text{Vec}(CN)] = \sqrt{n} \begin{bmatrix} \phi_{CN11}(\hat{\theta}_C) - \phi_{CN11}(\theta_C) \\ \phi_{CN12}(\hat{\theta}_C) - \phi_{CN12}(\theta_C) \\ \vdots \\ \phi_{CNp-1p}(\hat{\theta}_C) - \phi_{CNp-1p}(\theta_C) \\ \phi_{CNpp}(\hat{\theta}_C) - \phi_{CNpp}(\theta_C) \end{bmatrix}.$$

The Hadamard derivative of $\phi_{CNjj'}(\theta_C)$ in the direction of Z_{θ_C} is

$$\phi'_{\theta_C CNjj'}(Z) = S^{(j)} Z_{S_j^-} + S^{(j')} Z_{S_j} - Z_{S_{j,j'}}.$$

Therefore by the functional delta method, Theorem 2.8 in Kosorok (2008), $\sqrt{n}[\text{Vec}(\widehat{CN}) - \text{Vec}(CN)] \rightsquigarrow Z_{CN}$ where Z_{CN} is a p^2 dimensional mean 0 Gaussian process.

2. $\widehat{CM}$: Consider the mapping

$$\begin{aligned}\phi_{CMjj'}(\theta_C)(t_j, t_{j'}) = S^{(jj')}(t_j, t_{j'}) - 1 + \int_0^{t_j} S^{(j,j')}(s_j^-, t_{j'})\lambda^{(j)}(s_j)ds_j + \\ \int_0^{t_{j'}} S^{(j,j')}(t_j, s_{j'}^-)\lambda^{(j')}(s_{j'})ds_{j'} + \\ \int_0^{t_j}\int_0^{t_{j'}} S^{(j,j')}(s_j^-, s_{j'}^-)\lambda^{(j)}(s_j)\lambda^{(j')}(s_{j'})ds_j ds_{j'}. \qquad \text{(B.55)}\end{aligned}$$

Using this mapping

$$\sqrt{n}[\text{Vec}(\widehat{CM}) - \text{Vec}(CM)] = \sqrt{n}\begin{bmatrix} \phi_{CM11}(\hat{\theta_C}) - \phi_{CM11}(\theta_C) \\ \phi_{CM12}(\hat{\theta_C}) - \phi_{CM12}(\theta_C) \\ \vdots \\ \phi_{CMp-1p}(\hat{\theta_C}) - \phi_{CMp-1p}(\theta_C) \\ \phi_{CMpp}(\hat{\theta_C}) - \phi_{CMpp}(\theta_C) \end{bmatrix}.$$

As with the results for $\widehat{CN}$, the desired result will follow if it can be shown that $\phi_{CMjj'}(\theta_C)(t_j, t_{j'})$ is Hadamard differentiable in the direction of Z_{θ_C}. In order to do this we can consider each of the five parts on the right hand side of Equation (B.55) separately. It is straightforward to show that $S^{(j,j')}(t_j, t_{j'})$ and -1 are Hadamard differentiable. $\int_0^{t_j} S^{(j,j')}(s_j^-, t_{j'})\lambda^{(j)}(s_j)ds_j$ and $\int_0^{t_{j'}} S^{(j,j')}(t_j, s_{j'}^-)\lambda^{(j')}(s_{j'})ds_{j'}$ can both be shown to be Hadamard differentiable through Lemma 12.3 from Kosorok (2008), and $\int_0^{t_j}\int_0^{t_{j'}} S^{(j,j')}(s_j^-, s_{j'}^-)\lambda^{(j)}(s_j)\lambda^{(j')}(s_{j'})ds_j ds_{j'}$ can be shown to be Hadamard differentiable

through Lemma A5 from Cheng et al. (2007). Therefore by the functional delta method $\sqrt{n}[\text{Vec}(\widehat{CM}) - \text{Vec}(CM)] \rightsquigarrow Z_{CM}$ where Z_{CM} is a p^2 dimensional mean 0 Gaussian process.

In order to show the results for $\widehat{RN}$ and $\widehat{RM}$ we define $\theta_R = [S^{(1)}, \ldots, S^{(p)}, CN^{(1,1)}, CN^{(1,2)}, \ldots, CN^{(p,p)}, CM^{(1,1)}, CM^{(1,2)}, \ldots, CM^{(p,p)}]^T$ and $\hat{\theta}_R = [\hat{S}^{(1)}, \ldots, \hat{S}^{(p)}, \widehat{CN}^{(1,1)}, \widehat{CN}^{(1,2)}, \ldots, \widehat{CN}^{(p,p)}, \widehat{CM}^{(1,1)}, \widehat{CM}^{(1,2)}, \ldots, \widehat{CM}^{(p,p)}]^T$. Then using results from above

$$\sqrt{n}[\hat{\theta}_R - \theta_R] \rightsquigarrow Z_{\theta_R},$$

where $Z_{\theta_R} = [Z_{S_1}, \ldots, Z_{S_p}, Z_{CN_{1,1}}, Z_{CN_{1,2}}, \ldots, Z_{CN_{p,p}}, Z_{CM_{1,1}}, Z_{CM_{1,2}}, \ldots, Z_{CM_{p,p}}]^T$ is a $p + 2p^2$ dimensional mean 0 Gaussian process. From here we can show the results for $\widehat{RN}$ and $\widehat{RM}$ using the functional delta method as above.

1. $\widehat{RN}$ Define

$$\phi_{RNjj'}(\theta_R) = \frac{CN^{(j,j')}}{\sqrt{S^{(j)}(1 - S^{(j)})}\sqrt{S^{(j')}(1 - S^{(j')})}}.$$

Using this mapping

$$\sqrt{n}[\text{Vec}(\widehat{RN}) - \text{Vec}(RN)] = \sqrt{n} \begin{bmatrix} \phi_{RN11}(\hat{\theta}_R) - \phi_{RN11}(\theta_R) \\ \phi_{RN12}(\hat{\theta}_R) - \phi_{RN12}(\theta_R) \\ \vdots \\ \phi_{RNp-1p}(\hat{\theta}_R) - \phi_{RNp-1p}(\theta_R) \\ \phi_{RNpp}(\hat{\theta}_R) - \phi_{RNpp}(\theta_R) \end{bmatrix}.$$

The Hadamard differentiability of $\phi_{RNjj'}(\theta_R)$ in the direction of Z_{θ_R} can be shown through repeated application of the chain rule and the following results,

- If $\phi_1(A) = A^2$ then $\phi'_{A1}(\alpha) = 2A\alpha$.
- If $\phi_2(A) = \sqrt{A}$ then $\phi'_{A2}(\alpha) = \frac{\alpha}{2\sqrt{A}}$.
- If $\phi_3(A, B) = \frac{A}{B}$ then $\phi'_{AB3}(\alpha, \beta) = \frac{B\alpha - A\beta}{B^2}$.

Therefore by the functional delta method $\sqrt{n}[\text{Vec}(\widehat{RN}) - \text{Vec}(RN)] \rightsquigarrow Z_{RN}$ where Z_{RN} is a p^2 dimensional mean 0 Gaussian process.

2. $\widehat{RM}$ Results are similar to those for $\widehat{RN}$ if we consider the mapping

$$\phi_{RMjj'}(\theta_R) = \frac{CM^{(j,j')}}{\sqrt{1 - S^{(j)}}\sqrt{1 - S^{j'}}}$$

■

Proof of Theorem 4.4

Proof. Similar to the proof of Theorem 4.3 we define $\ddot{\theta}_C = [F_1^{(1)}, \ldots, F_1^{(p)}, F_{11}^{(1,1)}, F_{11}^{(1,2)}, \ldots, F_{11}^{(p,p)}, F_{12}^{(1,1)}, F_{12}^{(1,2)}, \ldots, F_{12}^{(p,p)}, F_{21}^{(1,1)}, F_{21}^{(1,2)}, \ldots, F_{21}^{(p,p)}, F_{22}^{(1,1)}, F_{22}^{(1,2)}, \ldots, F_{22}^{(p,p)}, \ddot{\Lambda}_1^{(1)}, \ldots, \ddot{\Lambda}_1^{(j)}]^T$ and $\widehat{\ddot{\theta}}_C = [\hat{F}_1^{(1)}, \ldots, \hat{F}_1^{(p)}, \hat{F}_{11}^{(1,1)}, \hat{F}_{11}^{(1,2)}, \ldots, \hat{F}_{11}^{(p,p)}, \hat{F}_{12}^{(1,1)}, \hat{F}_{12}^{(1,2)}, \ldots, \hat{F}_{12}^{(p,p)}, \hat{F}_{21}^{(1,1)}, \hat{F}_{21}^{(1,2)}, \ldots, \hat{F}_{21}^{(p,p)}, \hat{F}_{22}^{(1,1)}, \hat{F}_{22}^{(1,2)}, \ldots, \hat{F}_{22}^{(p,p)}, \hat{\ddot{\Lambda}}_1^{(1)}, \ldots, \hat{\ddot{\Lambda}}_1^{(j)}]^T$. Using results from Cheng et al. (2007) it can be shown that

$$\sqrt{n}[\hat{\ddot{\theta}}_C - \ddot{\theta}_C] \rightsquigarrow Z_{\ddot{\theta}_C},$$

where $Z_{\ddot{\theta}_C} = [Z_{F_{11}}, \ldots, Z_{F_{1p}}, Z_{F_{1111}}, Z_{F_{1112}}, \ldots, Z_{F_{11pp}}, Z_{F_{1211}}, Z_{F_{1212}}, \ldots, Z_{F_{12pp}}, Z_{F_{2111}}, Z_{F_{2112}}, \ldots, Z_{F_{21pp}}, Z_{F_{2211}}, Z_{F_{2212}}, \ldots, Z_{F_{22pp}}, Z_{\Lambda_1}, \ldots, Z_{\Lambda_p}]^T$ is a $2p + 3p^2$ dimensional mean 0 Gaussian process. The proofs for $\widehat{\ddot{CN}}$ and $\widehat{\ddot{CM}}$ follow in a similar manner to $\widehat{CN}$ and $\widehat{CM}$.

1. $\widehat{\ddot{CN}}$: The relevant mapping is for $\widehat{\ddot{CN}}$

$$\phi_{\ddot{CN}jj'}(\ddot{\theta}_C) = F_{11}^{(j,j')} - F_1^{(j)} F_1^{(j')}$$

Using this mapping the desired results can be shown using similar methods to $\widehat{CN}$ in the proof of Theorem 4.3 above.

2. $\widehat{\ddot{CM}}$: The relevant mapping for $\widehat{\ddot{CM}}$ is

$$\begin{aligned}\phi_{\ddot{CM}jj'}(\ddot{\theta}_C)(t) = \ddot{\Lambda}_1^{(j)}(t)\ddot{\Lambda}_1^{(j')}(t)\ddot{S}^{(j,j')}(t,t) + \int_0^t \ddot{\Lambda}_1^{(j)}(s)\ddot{\Lambda}_1^{(j')}(s)F_{22}^{(j,j')}(ds) \\ + \int_0^t \int_0^t \{1-\ddot{\Lambda}_1^{(j)}(s_1)\}\{1-\ddot{\Lambda}_1^{(j')}(s_2)\}F_{11}^{(j,j')}(ds_1,ds_2) \\ + \int_0^t \{1-\ddot{\Lambda}_1^{(j)}(s)\}\{-\ddot{\Lambda}_1^{(j')}(t)\}\{F_1^{(j)}(ds) - F_{11}^{(j,j')}(ds,t) - F_{12}^{(j,j')}(ds,t)\} \\ + \int_0^t \{1-\ddot{\Lambda}_1^{(j')}(s)\}\{-\ddot{\Lambda}_1^{(j)}(t)\}\{F_1^{(j')}(ds) - F_{11}^{(j,j')}(t,ds) - F_{21}^{(j,j')}(t,ds)\} \\ + \int_0^t \int_{s_1}^t \{1-\ddot{\Lambda}_1^{(j)}(s_1)\}\{-\ddot{\Lambda}_1^{(j')}(s_2)\}F_{12}^{(j,j')}(ds_1,ds_2) \\ + \int_0^t \int_{s_1}^t \{1-\ddot{\Lambda}_1^{(j')}(s_1)\}\{-\ddot{\Lambda}_1^{(j)}(s_2)\}F_{12}^{(j',j)}(ds_1,ds_2),\end{aligned}$$

This mapping can be shown to be Hadamard differentiable in the direction of $Z_{\ddot{\theta}_C}$ through repeated application of Lemma 12.3 from Kosorok (2008) and Lemma A5 from Cheng et al. (2007). The results then follow using similar methods to the proof for $\widehat{CM}$ in Theorem 4.3 above.

The results for $\widehat{\ddot{RN}}$ and $\widehat{\ddot{RM}}$ can be obtained by defining $\ddot{\theta}_R = [F_1^{(1)},\ldots,F_1^{(p)},\ddot{CN}^{(1,1)},\ddot{CN}^{(1,2)},\ldots,\ddot{CN}^{(p,p)},\ddot{CM}^{(1,1)},\ddot{CM}^{(1,2)},\ldots,\ddot{CM}^{(p,p)}]^T$ and $\hat{\ddot{\theta}}_R = [\hat{F}_1^{(1)},\ldots,\hat{F}_1^{(p)},\widehat{\ddot{CN}}^{(1,1)},\widehat{\ddot{CN}}^{(1,2)},\ldots,\widehat{\ddot{CN}}^{(p,p)},\widehat{\ddot{CM}}^{(1,1)},\widehat{\ddot{CM}}^{(1,2)},\ldots,\widehat{\ddot{CM}}^{(p,p)}]^T$. Using the results from above

$$\sqrt{n}[\hat{\ddot{\theta}}_R - \ddot{\theta}_R] \rightsquigarrow Z_{\ddot{\theta}_R},$$

where $Z_{\ddot{\theta}_R} = [Z_{F11},\ldots,Z_{F1p},Z_{\ddot{CN}11},Z_{\ddot{CN}12},\ldots,Z_{\ddot{CN}pp},Z_{\ddot{CM}11},Z_{\ddot{CM}12},\ldots,Z_{\ddot{CM}pp}]^T$ is a $p+2p^2$ dimensional mean 0 Gaussian process. The results for $\widehat{\ddot{RN}}$ and and $\widehat{\ddot{RM}}$ can be shown using similar methods to $\widehat{RN}$ and $\widehat{RM}$ above using the two mappings below.

1. $\widehat{\ddot{RN}}$: The relevant mapping is

$$\phi_{\ddot{RN}jj'}(\ddot{\theta}_R) = \frac{\ddot{CN}^{(j,j')}}{\sqrt{F_1^{(j)}(1-F_1^{(j)}}\sqrt{F_1^{(j')}(1-F_1^{(j')}}}.$$

2. $\widehat{\ddot{RM}}$: The relevant mapping is

$$\phi_{\ddot{RM}jj'}(\ddot{\theta}_R) = \frac{\ddot{CM}^{(j,j')}}{\sqrt{F_1^{(j)}}\sqrt{F_1^{(j')}}}.$$

■

Proof of Theorem 4.5

Proof. The consistency of $\hat{V}$ can be established through a straightforward application of the Davis-Kahan Theorem, and the consistency of $\hat{\Xi}$ can be established through Weyl's inequality. From here asymptotic normality can be established using steps similar to Theorem 13.5.1 in Anderson (2003), which derives the asymptotic variances of the eigenvectors and eigenvalues for the sample covariance matrix when data come from a multivariate normal distribution and the sample covariance matrix has Wishart distribution. However the steps can be straightforwardly extended to any asymptotically normal positive definite estimate of the covariance or correlation matrix as shown below.

Consider the transformation $Q = V^T\widehat{\Sigma}V$. Then by the Delta method $\sqrt{n}(Q-\Xi) \to_d N(0, J_V\Psi_\Sigma J_V^T)$ where $J_V = V^T \otimes V^T$. It can be shown that $\widehat{\Sigma}$ and Q have the same singular values so Q can be represented as

$$Q = G\widehat{\Xi}G^T, \tag{B.56}$$

for orthogonal G. G can be uniquely defined with the constraint $g_{ii} \geq 0$. Note that $\widehat{\Xi} = \hat{V}^T\widehat{\Sigma}\hat{V}$, which together with the fact that $V^T\hat{V}$ is orthogonal implies that every column of G is equal to $\pm$ the corresponding column of $V^T\hat{V}$. Because $\hat{V} \to V$ the constraint that $g_{ii} \geq 0$ implies

that $G \to I$ where I the identity matrix. Define $U = \sqrt{n}(Q - \Xi)$, $D = \sqrt{n}(\hat{\Xi} - \Xi)$ and $W = \sqrt{n}(Y - I)$. When we combine Equation (B.56) with $GG^T = I$, and the conditions $g_{ii} > 0$ and $\hat{\xi}_1, > \ldots, > \hat{\xi}_p$, we get a set of one to one functions from Q to G and $\hat{\Xi}$ except on a set of measure zero which are continuously differentiable and have well defined inverses in the neighborhood of $\hat{\Xi} = \Xi$ and $G = I$. Therefore by the fact that U is asymptotically normal with mean zero, from the delta method W and D are also asymptotically normal with mean zero. Further because the column of $\hat{V}$ is equal to $\pm$ the corresponding column of VG and $G \to I$, $\sqrt{n}(\hat{V} - V)$ has the same limiting distribution as $\sqrt{n}(VG - V)$ which is asymptotically normal with mean zero by the delta method.

In order to show that the limiting variances are functions of Ψ_Σ, V, and Λ, using similar algebra from Theorem 13.5.1 in (Anderson, 2003) we get the following equalities

$$U = W\Lambda + D + \Lambda W^T + o_p(1) \tag{B.57}$$

$$0 = W + W^T + o_p(1) \tag{B.58}$$

By combining results from Equations (B.57) and (B.58) and ignoring the $o_p(1)$ terms we obtain

$$w_{ii} = 0 \tag{B.59}$$

$$d_{ii} = u_{ii} \tag{B.60}$$

$$w_{ij} = \frac{u_{ij}}{\lambda_j - \lambda_i} \text{ for } i \neq j \tag{B.61}$$

This allows the limiting distribution for W and D to be solved for using the limiting distribution of U. In turn the limiting distribution of W can be used to find the limiting distribution of $\sqrt{n}(\hat{V} - V)$. ■

B.2 Additional Simulation Results

B.2.1 True correlation matrices for martingales and counting processes

Below are the true values for $RM(\mathbf{t})$, $RN(\mathbf{t})$, $R\ddot{M}(\mathbf{t})$, and $R\ddot{N}(\mathbf{t})$, for $\mathbf{t} = [1, 1, \ldots]^T$ and $\mathbf{t} = [2, 2, \ldots]^T$ in the simulation settings described in Chapter 4 of the main text. These are all based off of a single simulation of 500,000 subjects with no censoring. We present the matrices for $p = 8$, because based on our simulation structure the the matrices for $p = 4$ are equal to the upper left or bottom right portion of the $p = 8$ matrices.

$$RM([1, 1, \ldots]^T) = \begin{bmatrix} 1 & 0.49 & 0.32 & 0.16 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.49 & 1 & 0.50 & 0.32 & 0.16 & 0.00 & 0.00 & 0.00 \\ 0.32 & 0.50 & 1 & 0.50 & 0.32 & 0.16 & 0.00 & 0.00 \\ 0.16 & 0.32 & 0.50 & 1 & 0.49 & 0.32 & 0.16 & 0.00 \\ 0.00 & 0.16 & 0.32 & 0.49 & 1 & 0.50 & 0.32 & 0.16 \\ 0.00 & 0.00 & 0.16 & 0.32 & 0.50 & 1 & 0.50 & 0.32 \\ 0.00 & 0.00 & 0.00 & 0.16 & 0.32 & 0.50 & 1 & 0.50 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.16 & 0.32 & 0.50 & 1 \end{bmatrix}$$

$$RM([2, 2, \ldots]^T) = \begin{bmatrix} 1 & 0.53 & 0.35 & 0.17 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.53 & 1 & 0.53 & 0.34 & 0.17 & 0.00 & 0.00 & 0.00 \\ 0.35 & 0.53 & 1 & 0.53 & 0.35 & 0.17 & 0.00 & 0.00 \\ 0.17 & 0.34 & 0.53 & 1 & 0.53 & 0.35 & 0.17 & 0.00 \\ 0.00 & 0.17 & 0.35 & 0.53 & 1 & 0.53 & 0.35 & 0.17 \\ 0.00 & 0.00 & 0.17 & 0.35 & 0.53 & 1 & 0.53 & 0.35 \\ 0.00 & 0.00 & 0.00 & 0.17 & 0.35 & 0.53 & 1 & 0.53 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.17 & 0.35 & 0.53 & 1 \end{bmatrix}$$

$$RN([1,1,\ldots]^T) = \begin{bmatrix} 1 & 0.40 & 0.26 & 0.12 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.40 & 1 & 0.40 & 0.26 & 0.12 & 0.00 & 0.00 & 0.00 \\ 0.26 & 0.40 & 1 & 0.40 & 0.26 & 0.12 & 0.00 & 0.00 \\ 0.12 & 0.26 & 0.40 & 1 & 0.40 & 0.25 & 0.12 & 0.00 \\ 0.00 & 0.12 & 0.26 & 0.40 & 1 & 0.40 & 0.26 & 0.12 \\ 0.00 & 0.00 & 0.12 & 0.25 & 0.40 & 1 & 0.40 & 0.26 \\ 0.00 & 0.00 & 0.00 & 0.12 & 0.26 & 0.40 & 1 & 0.41 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.12 & 0.26 & 0.41 & 1 \end{bmatrix}$$

$$RN([2,2,\ldots]^T) = \begin{bmatrix} 1 & 0.34 & 0.20 & 0.09 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.34 & 1 & 0.34 & 0.20 & 0.09 & 0.00 & 0.00 & 0.00 \\ 0.20 & 0.34 & 1 & 0.34 & 0.20 & 0.09 & 0.00 & 0.00 \\ 0.09 & 0.20 & 0.34 & 1 & 0.34 & 0.21 & 0.09 & 0.00 \\ 0.00 & 0.09 & 0.20 & 0.34 & 1 & 0.35 & 0.20 & 0.09 \\ 0.00 & 0.00 & 0.09 & 0.21 & 0.35 & 1 & 0.34 & 0.21 \\ 0.00 & 0.00 & 0.00 & 0.09 & 0.20 & 0.34 & 1 & 0.34 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.09 & 0.21 & 0.34 & 1 \end{bmatrix}$$

$$\ddot{RM}([1,1,\ldots]^T) = \begin{bmatrix} 1 & 0.47 & 0.30 & 0.15 & 0.00 & -0.01 & 0.00 & 0.00 \\ 0.47 & 1 & 0.47 & 0.30 & 0.14 & -0.01 & -0.01 & -0.01 \\ 0.30 & 0.47 & 1 & 0.47 & 0.30 & 0.15 & 0.00 & 0.00 \\ 0.15 & 0.30 & 0.47 & 1 & 0.47 & 0.30 & 0.15 & 0.00 \\ 0.00 & 0.14 & 0.30 & 0.47 & 1 & 0.47 & 0.30 & 0.15 \\ -0.01 & -0.01 & 0.15 & 0.30 & 0.47 & 1 & 0.47 & 0.30 \\ 0.00 & -0.01 & 0.00 & 0.15 & 0.30 & 0.47 & 1 & 0.47 \\ 0.00 & -0.01 & 0.00 & 0.00 & 0.15 & 0.30 & 0.47 & 1 \end{bmatrix}$$

$$\ddot{RM}([2,2,\ldots]^T) = \begin{bmatrix} 1 & 0.49 & 0.31 & 0.15 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.49 & 1 & 0.49 & 0.31 & 0.15 & -0.01 & -0.01 & -0.01 \\ 0.31 & 0.49 & 1 & 0.49 & 0.31 & 0.15 & 0.00 & 0.00 \\ 0.15 & 0.31 & 0.49 & 1 & 0.49 & 0.31 & 0.15 & 0.00 \\ 0.00 & 0.15 & 0.31 & 0.49 & 1 & 0.49 & 0.32 & 0.15 \\ 0.00 & -0.01 & 0.15 & 0.31 & 0.49 & 1 & 0.49 & 0.32 \\ 0.00 & -0.01 & 0.00 & 0.15 & 0.32 & 0.49 & 1 & 0.49 \\ 0.00 & -0.01 & 0.00 & 0.00 & 0.15 & 0.32 & 0.49 & 1 \end{bmatrix}$$

$$\ddot{RN}([1,1,\ldots]^T) = \begin{bmatrix} 1 & 0.48 & 0.36 & 0.25 & 0.14 & 0.14 & 0.14 & 0.14 \\ 0.48 & 1 & 0.48 & 0.36 & 0.25 & 0.14 & 0.14 & 0.14 \\ 0.36 & 0.48 & 1 & 0.48 & 0.36 & 0.25 & 0.14 & 0.14 \\ 0.25 & 0.36 & 0.48 & 1 & 0.49 & 0.36 & 0.25 & 0.14 \\ 0.14 & 0.25 & 0.36 & 0.49 & 1 & 0.49 & 0.36 & 0.25 \\ 0.14 & 0.14 & 0.25 & 0.36 & 0.49 & 1 & 0.48 & 0.36 \\ 0.14 & 0.14 & 0.14 & 0.25 & 0.36 & 0.48 & 1 & 0.48 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0.25 & 0.36 & 0.48 & 1 \end{bmatrix}$$

$$\ddot{RN}([2,2,\ldots]^T) = \begin{bmatrix} 1 & 0.55 & 0.44 & 0.34 & 0.26 & 0.26 & 0.26 & 0.26 \\ 0.55 & 1 & 0.55 & 0.44 & 0.34 & 0.26 & 0.26 & 0.26 \\ 0.44 & 0.55 & 1 & 0.54 & 0.44 & 0.34 & 0.26 & 0.26 \\ 0.34 & 0.44 & 0.54 & 1 & 0.55 & 0.44 & 0.34 & 0.26 \\ 0.26 & 0.34 & 0.44 & 0.55 & 1 & 0.55 & 0.44 & 0.35 \\ 0.26 & 0.26 & 0.34 & 0.44 & 0.55 & 1 & 0.55 & 0.44 \\ 0.26 & 0.26 & 0.26 & 0.34 & 0.44 & 0.55 & 1 & 0.55 \\ 0.26 & 0.26 & 0.26 & 0.26 & 0.35 & 0.44 & 0.55 & 1 \end{bmatrix}$$

Note that all entries for $\ddot{RN}(\mathbf{t})$ are greater than zero for $\mathbf{t} = [1,1\ldots]^T$ and $\mathbf{t} = [2,2\ldots]^T$, even for the entries corresponding to zero correlation for T. As noted in Section 4 of the text this is

due to the competing event creating correlation between $\ddot{N}^{(j)}(t)$ and $\ddot{N}^{(j')}(t)$ even when $T^{(j)}$ and $T^{(j')}$ are uncorrelated. Alternatively $R\ddot{M}$ does not have the same issue and is more similar to RM. Also note that in the non-competing risk setting the correlations for RN are much lower than for RM.

Because $R\ddot{N}$ has a higher overall level of correlation than the other correlation matrices it has a higher first eigenvalue and larger separation between the first and second eigenvalues. Because of this the first average angle between the estimated and true first principal direction for $R\ddot{N}$ is smaller than for $R\ddot{M}$. However, because the first direction is partially driven by correlation induced by $\tilde{T}$ and not correlation between elements of T, using $R\ddot{M}$ instead of $R\ddot{N}$ may still be preferred.

B.2.2 Additional simulation results

Tables B18 and B19 give additional simulation results for PCA using the counting process and martingale correlation matrices. Simulation set ups are the same as described in Section 4.4 of the main text. The principal components are estimated at $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$. All reported values are the average (SD) angle between the true and estimated principal directions over all 1,000 simulated data sets, except for the counting process in the non-competing risk set up using C_2 as the censoring distribution. In this case there are simulations $\hat{S}^{(j)}(2) = 0$ for some of the variables. This results in the counting process correlation not being well defined. The estimated counting process correlation not being defined at $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$ happens in 31 of the simulated data sets when $p = 4$ and $n = 200$, one of the simulated data sets when $p = 4$ and $n = 1000$, 86 of the data sets when $p = 8$ and $n = 200$, and 15 of the data sets when $p = 8$ and $n = 1000$.

When $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$ impact of sample size, dimension and censoring rate are similar to $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$. For $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$ the average angles between the estimated and true principal directions are higher than they are at $\mathbf{t} = [1, 1, 1, 1, \ldots]^T$ for both martingale and counting process PCA. This is likely due to unstable estimation as there are fewer subjects still in the risk set near $t = 2$. For both the Kaplan-Meier and Nelson-Aalen estimator this can lead to

large jumps based on just one failure. This is particularly true for counting process PCA when C_2 is the censoring distribution. This is likely due to the fact that even when the estimated survival curve doesn't go all the way to zero by time point two, the estimate still has a high variance.

Table B18: Average (SD) of the angle in radians between the true and estimated PCA directions based on counting process and martingale correlation matrices with no competing risks

$\mathbf{t} = [2, 2, 2, 2, \ldots]^T$									
		C_1				C_2			
		Counting Process		Martingale		Counting Process		Martingale	
		n=200	n=1000	n=200	n=1000	n=200	n=1000	n=200	n=1000
4 Dim	PC 1	0.17 (0.13)	0.07 (0.03)	0.06 (0.03)	0.03 (0.01)	0.65 (0.41)	0.33 (0.30)	0.13 (0.10)	0.07 (0.04)
	PC 2	0.41 (0.29)	0.16 (0.09)	0.16 (0.09)	0.07 (0.04)	0.90 (0.37)	0.67 (0.39)	0.36 (0.25)	0.18 (0.11)
	PC 3	0.74 (0.40)	0.50 (0.37)	0.54 (0.37)	0.28 (0.24)	0.99 (0.32)	0.88 (0.40)	0.73 (0.41)	0.58 (0.40)
	PC 4	0.68 (0.40)	0.49 (0.37)	0.53 (0.38)	0.28 (0.24)	0.88 (0.38)	0.79 (0.39)	0.68 (0.41)	0.57 (0.40)
8 Dim	PC 1	0.42 (0.29)	0.17 (0.09)	0.19 (0.13)	0.08 (0.04)	0.89 (0.33)	0.62 (0.33)	0.48 (0.34)	0.21 (0.14)
	PC 2	0.54 (0.31)	0.20 (0.09)	0.23 (0.13)	0.09 (0.04)	1.07 (0.31)	0.79 (0.36)	0.55 (0.33)	0.24 (0.14)
	PC 3	0.59 (0.28)	0.23 (0.08)	0.24 (0.08)	0.10 (0.03)	1.17 (0.27)	0.89 (0.34)	0.55 (0.25)	0.26 (0.10)
	PC 4	0.95 (0.33)	0.51 (0.24)	0.58 (0.29)	0.23 (0.09)	1.24 (0.23)	1.10 (0.30)	0.95 (0.33)	0.66 (0.32)
	PC 5	1.15 (0.29)	1.08 (0.32)	1.03 (0.34)	0.77 (0.37)	1.22 (0.21)	1.18 (0.27)	1.15 (0.28)	1.06 (0.32)
	PC 6	1.18 (0.26)	1.10 (0.32)	1.08 (0.31)	0.90 (0.38)	1.22 (0.25)	1.19 (0.27)	1.16 (0.29)	1.11 (0.31)
	PC 7	1.16 (0.29)	1.10 (0.31)	1.03 (0.34)	0.76 (0.35)	1.20 (0.25)	1.15 (0.28)	1.13 (0.30)	1.09 (0.33)
	PC 8	1.03 (0.32)	0.82 (0.34)	0.82 (0.36)	0.46 (0.26)	1.16 (0.27)	1.11 (0.30)	1.04 (0.33)	0.90 (0.35)

Table B19: Average (SD) of the angle in radians between the true and estimated PCA directions based on counting process and martingale correlation matrices in the presence of a competing risk

$\mathbf{t} = [2, 2, 2, 2, \ldots]^T$									
		C_1				C_2			
		Counting Process		Martingale		Counting Process		Martingale	
		n=200	n=1000	n=200	n=1000	n=200	n=1000	n=200	n=1000
4 Dim	PC 1	0.07 (0.04)	0.03 (0.01)	0.08 (0.05)	0.03 (0.02)	0.16 (0.14)	0.07 (0.04)	0.16 (0.12)	0.08 (0.04)
	PC 2	0.40 (0.24)	0.19 (0.10)	0.22 (0.15)	0.09 (0.05)	0.60 (0.32)	0.39 (0.25)	0.44 (0.29)	0.22 (0.15)
	PC 3	0.71 (0.38)	0.45 (0.32)	0.63 (0.39)	0.36 (0.30)	0.84 (0.39)	0.68 (0.38)	0.79 (0.40)	0.62 (0.40)
	PC 4	0.68 (0.37)	0.45 (0.32)	0.61 (0.40)	0.36 (0.30)	0.83 (0.38)	0.65 (0.37)	0.75 (0.40)	0.60 (0.40)
8 Dim	PC 1	0.12 (0.07)	0.05 (0.02)	0.31 (0.27)	0.11 (0.06)	0.26 (0.17)	0.11 (0.04)	0.53 (0.37)	0.26 (0.20)
	PC 2	0.43 (0.23)	0.21 (0.10)	0.37 (0.26)	0.13 (0.06)	0.62 (0.29)	0.38 (0.21)	0.63 (0.36)	0.30 (0.19)
	PC 3	0.56 (0.26)	0.25 (0.10)	0.36 (0.16)	0.14 (0.04)	0.83 (0.33)	0.52 (0.24)	0.68 (0.30)	0.32 (0.13)
	PC 4	0.90 (0.31)	0.51 (0.24)	0.75 (0.33)	0.34 (0.15)	1.08 (0.29)	0.88 (0.33)	1.00 (0.33)	0.74 (0.33)
	PC 5	1.14 (0.29)	1.04 (0.32)	1.10 (0.31)	0.87 (0.36)	1.17 (0.28)	1.15 (0.29)	1.16 (0.29)	1.10 (0.31)
	PC 6	1.15 (0.27)	1.08 (0.32)	1.11 (0.31)	1.00 (0.35)	1.18 (0.28)	1.15 (0.29)	1.16 (0.27)	1.12 (0.31)
	PC 7	1.17 (0.29)	1.06 (0.33)	1.09 (0.32)	0.90 (0.36)	1.18 (0.27)	1.14 (0.29)	1.15 (0.29)	1.09 (0.31)
	PC 8	1.07 (0.30)	0.79 (0.34)	0.94 (0.34)	0.60 (0.31)	1.16 (0.26)	1.03 (0.31)	1.09 (0.30)	0.93 (0.33)

B.3 Principal component Loadings for MPACT data

Figure B.3.5 gives the principal component loadings from day 30 to 360 for the third through ninth principal components. The loadings for each specific adverse event at day 360 can be found in Table 3 of the main text. The 3rd through 9th principal component loadings do not have as clear of a scientific interpretation as the first two principal component loading and are not as stable over time.

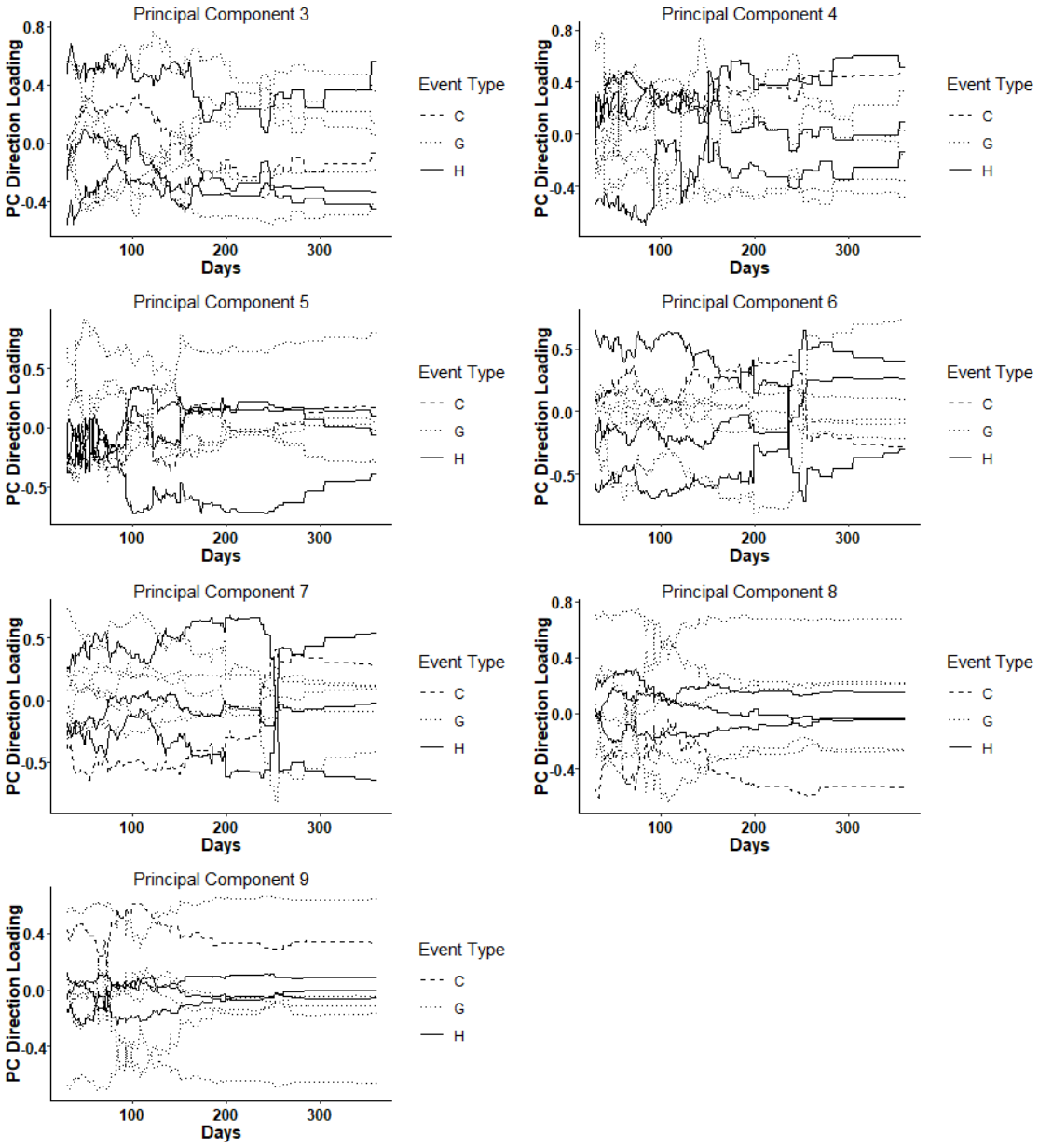

Figure B.3.5: Principal component direction loadings from day 30 to 360 for first two principal components using martingale correlation matrix estimates. Line types indicate constitutional, gastrointestinal and hematologic event types.

APPENDIX C: ADDITIONAL RESULTS FOR CHAPTER 4

C.1 Multi-set Canonical Directions for Executive Function and Brain Structure Data

Tables C20 through C24 give the first three latent mCCA directions using the MAXVAR/-VAR formulation for the four EF test scores, 20 DTI white matter tracts, and 88 GM brain regions. As noted in the main text the latent mCCA directions were estimated using the transformed Kendall's scatter matrix and three principal components for the EF variables, 16 principal components for the DTI variables, and seven for the GM brain regions.

Table C20: MAXVAR/VAR multi-set canonical correlation analysis direction loadings for executive function variables

	Dir 1	Dir 2	Dir 3
BASC	-0.16	-0.05	0.44
BRIEF	-0.06	-0.10	0.36
SB NV FR	-0.44	0.74	0.70
SB V FR	1.03	0.42	0.06

Table C21: MAXVAR/VAR multi-set canonical correlation analysis direction loadings for diffusion tensor imaging white matter tracts

	Dir 1	Dir 2	Dir 3
ARC FP Left	0.09	-0.16	0.42
ARC FP Right	0.23	-0.29	-0.09
ARC FT Left	0.18	0.22	-0.36
ARC FT Right	-0.03	-0.08	-0.14
ARC TP Left	0.07	-0.22	-0.01
ARC TP Right	-0.23	0.17	0.11
CGC Left	-0.17	0.46	0.21
CGC Right	0.22	0.04	-0.13
CTPF Left	-0.15	0.34	-0.07
CTPF Right	0.51	0.01	0.14
IFOF Left	0.19	-0.07	-0.36
IFOF Right	-0.23	-0.13	0.18
ILF Left	0.20	0.16	0.13
ILF Right	-0.22	0.21	0.19
SLF Left	-0.22	0.10	-0.17
SLF Right	-0.17	-0.18	-0.28
UNC Left	-0.11	-0.11	-0.63
UNC Right	0.16	-0.39	0.28
Genu	0.22	-0.23	0.26
Splenium	-0.45	-0.33	0.15

Table C22: MAXVAR/VAR multi-set canonical correlation analysis direction loadings for grey matter volume regions

	Dir 1	Dir 2	Dir 3
Precentral Gyrus Left	0.01	0.06	0.05
Precentral Gyrus Right	-0.01	0.02	0.08
Superial Frontal Gyrus Left	0.00	0.04	0.03
Superial Frontal Gyrus Right	0.02	0.07	0.01
Superial Frontal Gyrus Orbital Left	-0.01	0.04	0.02
Superial Frontal Gyrus Orbita Right	-0.03	0.03	-0.02
Middle Front Gyrus Left	-0.01	0.05	0.04
Middle Front Gyrus Right	-0.01	0.04	0.03
Middle Front Gyrus Orbital Left	0.00	0.06	0.01
Middle Front Gyrus Orbital Right	-0.02	0.03	0.01
Inferior Frontal Gyrus Opercular Left	0.02	0.03	-0.02
Inferior Frontal Gyrus Opercular Right	0.00	0.02	0.00
Inferior Frontal Gyrus Triangular Left	0.04	0.02	-0.04
Inferior Frontal Gyrus Triangular Right	-0.01	0.01	-0.05
Inferior Frontal Gyrus Orbigal Left	0.02	0.06	-0.04
Inferior Frontal Gyrus Orbital Right	0.04	0.06	-0.01
Rolandic Operulum Left	0.07	0.01	-0.06
Rolandic Operulum Right	0.04	0.00	-0.04
Supplementary Motor Area Left	0.03	0.02	0.04
Supplementary Motor Area Right	0.05	0.03	0.04
Olfactory Cortex Left	-0.02	0.05	-0.11
Olfactory Cortex Right	-0.02	0.05	-0.11
Superior Frontal Gyrus Medial Left	-0.04	0.08	-0.04
Superior Frontal Gyrus Medial Right	-0.04	0.08	-0.05
Superior Frontal Gyrus Medial Orbital Left	0.00	-0.01	-0.02
Superior Frontal Gyrus Medial Orbital Right	-0.03	-0.02	-0.09
Gyrus Rectus Left	-0.03	0.01	0.02
Gyrus Rectus Right	-0.02	0.00	-0.02
Insula Left	0.06	0.05	-0.09
Insula Right	0.10	0.05	-0.07
Anterior Cingulate and Paracingulate Gyrus Left	-0.05	0.05	-0.08
Anterior Cingulate and Paracingulate Gyrus Right	-0.02	0.04	-0.02
Median Cingulate and Paracingulate Gyrus Left	0.03	0.07	-0.08
Median Cingulate and Paracingulate Gyrus Right	0.05	0.05	-0.07
Posterior Cingulum Gyrus Left	0.01	0.05	-0.08
Posterior Cingulum Gyrus Right	0.02	0.03	-0.08
Hippocampus Left	0.05	0.03	0.00
Hippocampus Right	0.05	0.02	0.01
Parahippocampal Gyrus Left	0.05	0.01	0.02
Parahippocampal Gyrus Right	0.05	0.01	0.01

Table C23: MAXVAR/VAR multi-set canonical correlation analysis direction loadings for grey matter volume regions

	Dir 1	Dir 2	Dir 3
Amygdala Left	-0.07	0.01	-0.04
Amygdala Right	-0.04	-0.01	-0.01
Calcarine Fissure and Surrounding Cortex Left	0.02	-0.10	-0.06
Calcarine Fissure and Surrounding Cortex Right	0.02	-0.09	-0.05
Cuneus Left	0.04	-0.11	-0.04
Cuneus Right	0.05	-0.13	-0.03
Lingual Gyrus Left	0.00	-0.09	-0.02
Lingual Gyrus Right	0.03	-0.06	-0.01
Superior Occipital Gyrus Left	0.01	-0.13	0.00
Superior Occipital Gyrus Right	0.01	-0.07	0.02
Middle Occipital Gyrus Left	0.00	-0.10	-0.06
Middle Occipital Gyrus Right	-0.01	-0.06	0.05
Inferior Occipital Gyrus Left	-0.03	-0.03	-0.09
Inferior Occipital Gyrus Right	-0.03	-0.03	-0.05
Fusiform Gyrus Left	0.08	0.03	0.03
Fusiform Gyrus Right	0.09	0.02	0.04
Postcentral Gyrus Left	-0.09	0.05	-0.02
Postcentral Gyrus Right	-0.11	0.04	-0.05
Superior Parietal Gyrus Left	-0.09	0.04	-0.03
Superior Parietal Gyrus Right	-0.10	0.05	0.00
Inferior Parietal, but Supramarginal and Angular Gyrus Left	-0.03	0.04	0.01
Inferior Parietal, but Supramarginal and Angular Gyrus Right	-0.02	0.05	-0.07
Supramarginal Gyrus Left	0.01	0.01	-0.04
Supramarginal Gyrus Right	-0.05	0.02	-0.06
Angular Gyrus Left	0.05	0.08	0.01
Angular Gyrus Right	0.04	0.08	-0.03
Precuneus Left	-0.07	0.01	-0.09
Precuneus Right	-0.04	0.01	-0.09
Paracentral Lobule Left	-0.04	0.04	0.05
Paracentral Lobule Right	-0.01	0.05	0.02
Caudate Nucleus Left	-0.12	0.04	-0.09
Caudate Nucleus Right	-0.10	0.04	-0.09
Lenticular Nucleus Putamen Left	-0.09	-0.02	-0.04
Lenticular Nucleus Putamen Right	0.03	-0.02	-0.03
Thalamus Left	0.03	0.08	-0.03
Thalamus Right	0.03	0.07	-0.03
Heschl Gyrus Left	0.08	0.02	-0.09
Heschl Gyrus Right	0.05	0.00	-0.07
Superior Temporal Gyrus Left	0.07	-0.01	0.01
Superior Temporal Gyrus Right	0.01	0.03	-0.05

Table C24: MAXVAR/VAR multi-set canonical correlation analysis direction loadings for grey matter volume regions

	Dir 1	Dir 2	Dir 3
Temporal Pole: Superior Temporal Gyrus Left	-0.07	0.03	0.04
Temporal Pole: Superior Temporal Gyrus Right	-0.09	-0.04	0.02
Middle Temporal Gyrus Left	0.03	0.05	0.04
Middle Temporal Gyrus Right	0.05	0.03	0.05
Temporal Pole: Middle Temporal Gyrus Left	-0.04	0.00	0.08
Temporal Pole: Middle Temporal Gyrus Right	-0.05	-0.02	0.07
Inferior Temporal Gyrus Left	0.03	0.01	0.11
Inferior Temporal Gyrus Right	0.05	-0.03	0.08

BIBLIOGRAPHY

Alexander, A. L., Lee, J. E., Lazar, M., and Field, A. S. (2007). Diffusion tensor imaging of the brain. *Neurotherapeutics*, 4(3):316–329.

Alfons, A., Croux, C., and Filzmoser, P. (2017). Robust maximum association estimators. *Journal of the American Statistical Association*, 112(517):436–445.

Anderson, T. W. (1999). Asymptotic theory for canonical correlation analysis. *Journal of Multivariate Analysis*, 70(1):1–29.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley, 3rd edition.

Asendorf, N. A. (2015). *Informative data fusion: Beyond canonical correlation analysis*. PhD thesis, University of Michigan.

Branco, J. A., Croux, C., Filzmoser, P., and Oliveira, M. R. (2005). Robust canonical correlations: A comparative study. *Computational Statistics*, 20(2):203–229.

Bray, S., Krongold, M., Cooper, C., and Lebel, C. (2015). Synergistic effects of age on patterns of white and gray matter volume across childhood and adolescence. *eNeuro*.

CANTAB (2017). Cantab® cognitive assessment software. www.cantab.com.

Cheng, Y., Fine, J. P., and Kosorok, M. R. (2007). Nonparametric association analysis of bivariate competing-risks data. *Journal of the American Statistical Association*, 102(480):1407–1415.

Choi, Y., Taylor, J., Tibshirani, R., et al. (2017). Selecting the number of principal components: Estimation of the true rank of a noisy matrix. *The Annals of Statistics*, 45(6):2590–2617.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368):829–836.

Dabrowska, D. M. et al. (1988). Kaplan-meier estimate on the plane. *The Annals of Statistics*, 16(4):1475–1489.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 176223.

Feng, Q., Jiang, M., Hannig, J., and Marron, J. (2018). Angle-based joint and individual variation explained. *Journal of multivariate analysis*, 166:241–265.

Fu, X., Huang, K., Hong, M., Sidiropoulos, N. D., and So, A. M.-C. (2017). Scalable and flexible multiview max-var canonical correlation analysis. *IEEE Transactions on Signal Processing*, 65(16):4150–4165.

Gaynanova, I. and Li, G. (2019). Structural learning and integrative decomposition of multi-view data. *Biometrics*, 75(4):1121–1132.

Gill, R. D., Laan, M. J., and Wellner, J. A. (1995). Inefficient estimators of the bivariate survival function for three models. *Annales de l'IHP Probabilités et statistiques*, 31(3):545–597.

Gilmore, J. H., Schmitt, J. E., Knickmeyer, R. C., Smith, J. K., Lin, W., Styner, M., Gerig, G., and Neale, M. C. (2010). Genetic and environmental contributions to neonatal brain structure: a twin study. *Human brain mapping*, 31(8):1174–1182.

Gioia, G. A., Isquith, P. K., Guy, S. C., and Kenworthy, L. (2000). Test review behavior rating inventory of executive function. *Child Neuropsychology*, 6(3):235–238.

Girault, J. B., Cornea, E., Goldman, B. D., Knickmeyer, R. C., Styner, M., and Gilmore, J. H. (2019). White matter microstructural development and cognitive ability in the first 2 years of life. *Human Brain Mapping*, 40(4):1195–1210.

Han, F. and Liu, H. (2012). Tca: Transelliptical principal component analysis for high dimensional non-gaussian data. Technical report, Technical report.

Han, F. and Liu, H. (2013). Optimal rates of convergence for latent generalized correlation matrix estimation in transelliptical distribution. *arXiv preprint arXiv:1305.6916 v3*.

Han, F. and Liu, H. (2017). Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability*, 23(1):23.

Hannan, K. L., Wood, S. J., Yung, A. R., Velakoulis, D., Phillips, L. J., Soulsby, B., Berger, G., McGorry, P. D., and Pantelis, C. (2010). Caudate nucleus volume in individuals at ultra-high risk of psychosis: a cross-sectional magnetic resonance imaging study. *Psychiatry Research: Neuroimaging*, 182(3):223–230.

Hoeffding, W. (1961). The strong law of large numbers for u-statistics. Technical report, North Carolina State University. Dept. of Statistics.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28(3/4):321–377.

Jaser, M., Haug, S., and Min, A. (2017). A simple non-parametric goodness-of-fit test for elliptical copulas. *Dependence Modeling*, 5(1):330–353.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Karnofsky, D. A., Abelmann, W. H., Craver, L. F., and Burchenal, J. H. (1948). The use of the nitrogen mustards in the palliative treatment of carcinoma. with particular reference to bronchogenic carcinoma. *Cancer*, 1(4):634–656.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2):81–93.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451.

Kikinis, R., Pieper, S. D., and Vosburgh, K. G. (2014). 3d slicer: a platform for subject-specific image analysis, visualization, and clinical support. In *Intraoperative imaging and image-guided therapy*, pages 277–289. Springer.

Klüppelberg, C. and Kuhn, G. (2009). Copula structure analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):737–753.

Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J. K., Hamer, R. M., Lin, W., Gerig, G., and Gilmore, J. H. (2008). A structural mri study of human brain development from birth to 2 years. *Journal of Neuroscience*, 28(47):12176–12182.

Knickmeyer, R. C., Xia, K., Lu, Z., Ahn, M., Jha, S. C., Zou, F., Zhu, H., Styner, M., and Gilmore, J. H. (2016). Impact of demographic and obstetric factors on infant brain volumes: a population neuroscience study. *Cerebral Cortex*, 27(12):5616–5625.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.

Levitt, J. J., McCarley, R. W., Dickey, C. C., Voglmaier, M. M., Niznikiewicz, M. A., Seidman, L. J., Hirayasu, Y., Ciszewski, A. A., Kikinis, R., Jolesz, F. A., et al. (2002). Mri study of caudate nucleus volume and its cognitive correlates in neuroleptic-naive patients with schizotypal personality disorder. *American Journal of Psychiatry*, 159(7):1190–1197.

Lin, D. and Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, 80(3):573–581.

Lindskog, F., Mcneil, A., and Schmock, U. (2003). Kendall's tau for elliptical distributions. In *Credit Risk*, pages 149–156. Springer.

Liu, H., Han, F., and Zhang, C.-h. (2012). Transelliptical graphical models. In *Advances in neural information processing systems*, pages 800–808.

Lock, E. F., Hoadley, K. A., Marron, J. S., and Nobel, A. B. (2013). Joint and individual variation explained (jive) for integrated analysis of multiple data types. *The annals of applied statistics*, 7(1):523.

Lock, E. F., Park, J. Y., and Hoadley, K. A. (2020). Bidimensional linked matrix factorization for pan-omics pan-cancer analysis. *arXiv preprint arXiv:2002.02601*.

Menzel, U. (2009). Ccp r package. https://CRAN.R-project.org/package=CCP.

Muirhead, R. and Waternaux, C. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, 67(1):31–43.

Nielsen, A. A. (2002). Multiset canonical correlations analysis and multispectral, truly multi-temporal remote sensing data. *IEEE Transactions on Image Processing*, 11(3):293–305.

Nielsen, A. A. (2002). Multiset canonical correlations analysis and multispectral, truly multi-temporal remote sensing data. *IEEE transactions on image processing*, 11(3):293–305.

Niogi, S. and McCandliss, B. (2006). Left lateralized white matter microstructure accounts for individual differences in reading ability and disability. *Neuropsychologia*, 44(11):2178–2188.

Parkhomenko, E., Tritchler, D., and Beyene, J. (2007). Genome-wide sparse canonical correlation of gene expression with genotypes. In *BMC proceedings*, volume 1, page S119. BioMed Central.

Parkhomenko, E., Tritchler, D., and Beyene, J. (2009). Sparse canonical correlation analysis with application to genomic data integration. *Statistical applications in genetics and molecular biology*, 8(1):1–34.

Parra, L. C. (2018). Multi-set canonical correlation analysis simply explained. *arXiv preprint arXiv:1802.03759*.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, 79(3):495–512.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, pages 541–554.

Reynolds, C. and Kamphaus, R. (2010). Basc-2: Behavior assessment system for children , manual supplement for the clinical report.

Roid, G. H. (2003). *Stanford-Binet intelligence scales*. Riverside Pub.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388):871–880.

Rousseeuw, P. J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics–Theory and Methods*, 22(4):965–984.

Rublík, F. (2016). Estimates of the covariance matrix of vectors of u-statistics and confidence regions for vectors of kendall's tau. *Kybernetika*, 52(2):280–293.

Suo, X., Minden, V., Nelson, B., Tibshirani, R., and Saunders, M. (2017). Sparse canonical correlation analysis. *arXiv preprint arXiv:1705.10865*.

Taskinen, S., Croux, C., Kankainen, A., Ollila, E., and Oja, H. (2006). Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *Journal of Multivariate Analysis*, 97(2):359–384.

Todorov, V. and Filzmoser, P. (2009a). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32(3):1–47.

Todorov, V. and Filzmoser, P. (2009b). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32(3):1–47.

Tyler, D. E. (1987). A distribution-free m-estimator of multivariate scatter. *The Annals of Statistics*, pages 234–251.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289.

van der Laan, M. J. (1993). *Modified EM-estimator of the bivariate survival function*. Rijksuniversiteit Utrecht. Mathematisch Instituut.

Vinod, H. D. (1976). Canonical ridge and econometrics of joint production. *Journal of econometrics*, 4(2):147–166.

Visuri, S., Ollila, E., Koivunen, V., Möttönen, J., and Oja, H. (2003). Affine equivariant multivariate rank methods. *Journal of Statistical Planning and Inference*, 114(1-2):161–185.

Von Hoff, D. D., Ervin, T., Arena, F. P., Chiorean, E. G., Infante, J., Moore, M., Seay, T., Tjulandin, S. A., Ma, W. W., Saleh, M. N., et al. (2013). Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *New England Journal of Medicine*, 369(18):1691–1703.

Waaijenborg, S., de Witt Hamer, P. C. V., and Zwinderman, A. H. (2008). Quantifying the association between gene expressions and dna-markers by penalized canonical correlation analysis. *Statistical applications in genetics and molecular biology*, 7(1).

Wilms, I. and Croux, C. (2015). Sparse canonical correlation analysis from a predictive point of view. *Biometrical Journal*, 57(5):834–851.

Wilms, I. and Croux, C. (2016). Robust sparse canonical correlation analysis. *BMC systems biology*, 10(1):72.

Winkler, A. M., Renaud, O., Smith, S. M., and Nichols, T. E. (2020). Permutation inference for canonical correlation analysis. *arXiv preprint arXiv:2002.10046*.

Witten, D. M., Tibshirani, R., and Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10(3):515–534.

Witten, D. M. and Tibshirani, R. J. (2009). Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical applications in genetics and molecular biology*, 8(1):1–27.

Xia, M., Wang, J., and He, Y. (2013). Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one*, 8(7).

Xu, H., Lorbert, A., Ramadge, P. J., Guntupalli, J. S., and Haxby, J. V. (2012). Regularized hyperalignment of multi-set fmri data. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pages 229–232. IEEE.

Yoon, G., Carroll, R. J., and Gaynanova, I. (2018). Sparse semiparametric canonical correlation analysis for data of mixed types. *arXiv preprint arXiv:1807.05274*.