ADAPTIVE CANONICAL CORRELATION ANALYSIS WITH CONSIDERATIONS
FOR HIGH DIMENSIONAL MATRICES:
A WEIGHTED RANK SELECTION CRITERION APPROACH
WITH
A HIV/NEUROCOGITIVE APPLICATION

By

JENNIFER ANN GEIS

A Dissertation submitted to the
Department of Statistics
in partial fulfillment of the
requirements for the degree of
Doctorate of Philosophy

Degree Awarded:
Spring Semester, 2012

The members of the supervisory committee were:

Dr. Yiyuan She  
Professor Directing Thesis

Dr. Anke Meyer-Baese  
University Representative

Dr. Florentina Bunea  
Committee Member, Cornell University

Dr. Xufeng Niu  
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with the university requirements.
To Mom and Dad:

For always reminding me...
   ... it’s not how fast you’re going but where you’re headed.
ACKNOWLEDGMENTS

I would like to acknowledge and give my most heartfelt thanks to my two advisors and mentors whom, with incredible patience and understanding, have seen me through this: Dr. Florentina Bunea and Dr. Yiyuan She. An additional thanks goes to the professors of the Department of Statistics at Florida State University, in particular Dr. Xufeng Niu, for professing their knowledge to me. A special thanks goes to Dr. Marten Wegkamp of Cornell University for all of his past support. I would also like to thank the department staff, especially Pamela McGhee, for always keeping me on track.
# TABLE OF CONTENTS

List of Tables ......................................................... vii
List of Symbols and Notations ........................................ ix
Abbreviations ......................................................... x
Abstract ........................................................... xi

1 Introduction .......................................................... 1

2 Background Methodology ............................................ 5
   2.1 Canonical Correlation Analysis ................................. 6
      2.1.1 Population Canonical Correlations and Variates .......... 6
      2.1.2 Sample Canonical Correlations and Variates ............ 10
      2.1.3 Regularized Canonical Correlation Analysis ............ 11
      2.1.4 Relationship to Reduced-Rank Regression ............... 12
      2.1.5 Additional Descriptive Measures ........................ 14
   2.2 Rank Selection Criterion (RSC) ................................ 15
      2.2.1 Tuning Parameter ........................................ 16
      2.2.2 Dimension Reduction .................................... 17
   2.3 Choice of Weight Matrix ...................................... 17

3 Adaptive Canonical Correlation Analysis and Weighted Rank Selection Criterion .............. 20
   3.1 Weighted Rank Selection Criterion ............................ 21
      3.1.1 Tuning Parameter Selection ............................... 22
   3.2 Adaptive Canonical Correlation Analysis ....................... 25
      3.2.1 Tuning Parameter Selection ............................... 26
   3.3 High Dimensional Considerations ............................. 27

4 Simulation Exploration .............................................. 32
   4.1 Experiment Setup ............................................. 32
   4.2 Results ...................................................... 33

5 Low Dimensional Neuroimaging Application of ACCA ............................................. 37
   5.1 Data Description ............................................. 37
   5.2 Methodology ................................................ 38
   5.3 Results ..................................................... 40
5.3.1 Original Predictors, All Responses .................. 40
5.3.2 Learning and Memory Domains with Original Predictors .... 42
5.3.3 Executive Functioning and Information Processing/Attention Domains with Original Predictors .................. 43
5.3.4 Verbal Fluency and Motor Domains with Original Predictors .... 43
5.4 Summary ............................................. 44

6 ACCA with Variable Selection 45
6.1 Equivalence of Group Selection in Univariate Models ............... 45
6.2 Group Lasso ........................................... 47
6.3 High Dimensional Neuroimaging Application ....................... 47
6.3.1 Results ............................................. 48
6.3.2 Summary ............................................. 50

7 Conclusion 51

A Theorems 53

B Large Sample Simulation Results 54

C High Dimensional Simulation Results 56

D Large Sample Neuroimaging Data Analysis Results 59

E High Dimensional Neuroimaging Data Analysis Results 67

Bibliography ................................................ 70
Biographical Sketch ....................................... 74
## LIST OF TABLES

4.1 Covariance Error Structures .......................................................... 32
5.1 Description of Neurocognitive Response Variables .......................... 38
5.2 Description of Predictor Variables .................................................. 39
5.3 Response Variables by Domain ....................................................... 40
B.1 **Experiment 1, Setup 1**- Rank recovery rates. \( \Gamma = \Sigma_e^{-1} \) with \( \mu_{\text{adap}} \). .......... 54
B.2 **Experiment 1, Setup 1**- Average Mean Squared Error. \( \Gamma = \Sigma_e^{-1} \) with \( \mu_{\text{adap}} \). 54
B.3 **Experiment 1, Setup 2**- Rank recovery rates. \( \Gamma = \hat{\Sigma}_Y^{-1} \) with \( \mu_1^{(y)} \) .... 55
B.4 **Experiment 1, Setup 2**- Average Mean Square Error. \( \Gamma = \hat{\Sigma}_Y^{-1} \) with \( \mu_1^{(y)} \). 55
C.1 **Experiment 2, Setup 1**- Rank recovery rates. \( \Gamma = \Sigma_e^{-1} \) with \( \mu_{\text{adap}1} \). .......... 56
C.2 **Experiment 2, Setup 1**- Average Mean Square Error. \( \Gamma = \Sigma_e^{-1} \) with \( \mu_{\text{adap}1} \). 56
C.3 **Experiment 2, Setup 2**- Rank recovery rates. \( \Gamma = \hat{\Sigma}_Y^{-1} \) with \( \mu_1^{(y)} \) .... 57
C.4 **Experiment 2, Setup 2**- Average Mean Squared Error. \( \Gamma = \hat{\Sigma}_Y^{-1} \) with \( \mu_1^{(y)} \). 57
C.5 **Experiment 2, Setup 3**- Rank recovery rates. \( \Gamma = (\hat{\Sigma}_Y + \delta I_n)^{-1} \) with \( \mu_1^{(y)} \). 58
C.6 **Experiment 2, Setup 3**- Average Mean Square Error. Unknown \( \Gamma = (\hat{\Sigma}_Y + \delta I_n)^{-1} \) with \( \mu_1^{(y)} \). . 58
D.1 Large Sample Neuroimaging ACCA with All Response Variables: Predictor Set 60
D.2 Large Sample Neuroimaging ACCA with All Response Variables: Response Set 61
D.3 Large Sample Neuroimaging ACCA with Learning and Memory Domain Variables: Predictor Set 62
D.4 Large Sample Neuroimaging ACCA with Learning and Memory Domain Variables: Response Set 62
LIST OF SYMBOLS AND NOTATIONS

Let $M$ and $N$ be some generic matrices.

- $(M)^{-}$: the Moore-Penrose pseudoinverse of $M$
- $P = X(X'X)^{-}X'$: the projection matrix
- $r(M)$: the rank of the matrix $M$
- $(M)_{ij}$: the entry of $M$ in the $i$-th row and $j$-th column
- $\|M\|_F = (\sum_i \sum_j (M)_{ij}^2)^{1/2}$: the Frobenius norm of the matrix $M$
- $\mathbb{I}_k$: the $k \times k$ identity matrix
- $\sigma_i^2(M)$: the $i$-th largest singular value of $M$
- $\lambda_i(M)$: the $i$-th largest eigenvalue of $M$
- $\mu$: the tuning parameter for RSC
- $\mu_{\text{adap}}$: the adaptive tuning parameter for RSC
- $\mu_1$: the tuning parameter for WRSC
- $\mu_{\text{adap1}}$: the adaptive tuning parameter for WRSC
- $S^2$: the unbiased estimator of the variance $\sigma^2$
- $\hat{\sigma}^2$: the biased ML estimator of the variance $\sigma^2$
- $M \otimes N$: Kronecker product between the matrices $M$ and $N$
- $\text{vec}(M)$: the vectorized version of $M$
- $\|m_j\|_2$: the Euclidean norm of the $j$-th row of $M$
- $\|M\|_{2,1} = \sum_{j=1}^p \|m_j\|_2$: the sum of the Euclidean norms $\|m_j\|_2$ of rows $m_j$ of $M$. 
ABBREVIATIONS

i.i.d. independently and identically distributed
p.d. positive-definite
p.s.d. positive-semidefinite
s.t. subject to
w.r.t. with respect to
ACCA Adaptive Canonical Correlation Analysis
AR(1) Autoregressive 1
CCA Canonical Correlation Analysis
CV Cross-Validation
DTI Diffusion Tensor Imaging
FA Fractional Anisotropy
GLASSO Group Lasso
H-AR(1) Heterogeneous Autoregressive 1
HIV Human Immunodeficiency Virus
MD Mean Diffusivity
MSE Mean Squared Error
RCCA Regularized Canonical Correlation Analysis
RRR Reduced-Rank Regression
RSC Rank Selection Criterion
SNR signal to noise ratio
SVD singular value decomposition
UN Unstructured
WRSC Weighted Rank Selection Criterion
ABSTRACT

Multivariate response regression models are being employed increasingly more in almost all fields. They are particularly complex by considering the relationships between predictor and response variables without neglecting the within set correlations. A typical inferential method for such models is Canonical Correlation Analysis. This is a fairly novel technique when approached from a multivariate point of view, rather than the traditional correlation maximization technique. However, the difficulty lies in the estimation of the number of significant canonical relationships, equivalently so, determining the rank of the coefficient estimator in Reduced-Rank Regression.

In the past, determining the appropriate rank included making distributional assumptions and restricting the number of predictor and response variables or by subjective judgment. Recent work in this area by Bunea, She, and Wegkamp has shown that this is unnecessary: consistent rank estimation may be achieved without this restrictions. This Rank Selection Criterion is a data-adaptive technique that is applicable even in the high dimensional setting.

There is, however, still an assumption that the error terms are independently and identically distributed (i.i.d.). While this is necessary to show the strong theoretical results proven by Bunea et al., some practical considerations and flexibility is required in many applications. That is, the i.i.d. assumption cannot always be safely made. This requires introduction of a weight matrix to the RSC. Theoretics are developed here for the large sample setting that parallel Bunea et al.'s work, providing support to use a decorrelator weight matrix, some good estimator of the error covariance. One such possibility is the sample residual matrix. However, a computationally more convenient weight matrix is the sample response covariance. When such a weight matrix is chosen, CCA is directly accessible by this weighted version of RSC, forming the Adaptive CCA. Now, the number of canonical relationships may be adaptively estimated while simultaneously estimating the canonical variates.

However, particular consideration is required for the high dimensional setting as the theoretics no longer hold. What will be offered instead is extensive simulations that will reveal that using the sample response covariance as the weight matrix still provides good rank recovery and estimation of the coefficient matrix, and hence, also good estimation of the number of canonical relationships and variates. It will be argued precisely why other versions of the sample response covariance are no longer valid in the high dimensional setting and how even a regularized version may be inferior to the sample response covariance.

Another approach, which avoids these issues in the high dimensional setting, is to use some type of group variable selection methodology to first reduce the dimension of the
predictor set before applying ACCA for inferential conclusions. Truly, any group selection methods may be employed prior to ACCA as variable selection in the multivariate response regression model is the same as group selection in the univariate response regression model.

To offer a practical application of these ideas, ACCA will be applied to a neuroimaging dataset. A high dimensional dataset will be generated from this large sample set to which Group Lasso will be first utilized before ACCA.
While multivariate modeling and analysis techniques are common, they are far from simple. Models with multiple predictor and response variables are ideal in many data analysis settings in nearly any field of interest. But extracting pertinent information from each variable while also considering their within set and opposite set relationships is hardly trivial.

Consider, for instance, a cohort of HIV-positive patients. HIV is known to infect nerve cells and the brain itself, not long after contraction, potentially causing systemic issues in attention and working memory, information processing speed, psychomotor abilities, executive functions, and in learning and memory [45]. These various neurocognitive “domains” may be measured by a variety of assessments and indices. However, how can the physical ramifications be assessed? Physical changes in the brain may be measured by in numerous diffusion tensor imaging (DTI) and brain volumetric (BV) measures. These quantify the restriction of water diffusion and the size of various parts of the brain, respectively. So, can these be linked to known neurocognitive problems while accounting for standard clinical variables?

It is desirable to withdraw the maximal amount of information from the correlation within the predictor set without forgetting the potential correlation in the response set. The multiple regression model has both multiple predictor and response variables, but standard least-squares estimators regress on the response variables separately and independently, failing to utilize the within set correlation of the responses. Estimators with reduced-ranks were developed to remedy this problem. Such estimation techniques were introduced by Anderson in the 1950s [1], and Reduced-Rank Regression (RRR) was later formally coined and developed by Izenman in 1975 [23]. Similar works were also produced by Robinson [34] [35] and Rao [31] around that time period. While innovative in their own right, the limitations and assumption requirements were less than ideal. These estimators were for fixed ranks only and asymptotic in nature, mostly derived in a likelihood framework with a Gaussian assumption on the errors. In 1999, Anderson was able to bypass the Gaussian assumption on the errors [2]. However, this still assumed that the true rank of the coefficient matrix was known and fixed. Later in 2002, he attempted to fix this problem by creating asymptotic rank selection tests, yet these were only valid when the number of predictors was small and fixed [3].

In addition to these issues, RRR, by itself, does provides limited inferential insight. With
such a multivariate model, where correlation exists within both sets, how can significant, multiple, uncorrelated relationships be identified? And after doing so, how can inferences be made? Hotelling’s Canonical Correlation Analysis (CCA) offers inferential methodology for identifying and interpreting such relationships [22]. Not surprising is that it is a unique version of RRR. While CCA provides a number of useful inference techniques, it has shown to work poorly in the high dimensional setting or even when the number of observations is close to the dimension size. Specifically, that is, as the number of variables increases, the canonical correlations become almost 1 and provides no meaningful interpretations. See Eaton and Perlman [11] for a full account. Also, the required sample covariances needed to perform CCA are sometimes ill-conditioned and often times, singular and non-invertible. While a regularized version of CCA has been offered as a solution by Vinod [41] and Leurgans et al. [27], Regularized Canonical Correlation Analysis (RCCA) is computationally expensive and time consuming as a 2-dimensional grid of regularization parameters needs to be cross-validated. In addition to this, RCCA only examines the largest canonical correlation corresponding to only the first relationship between sets while disregarding the rest. So even if the number of variables may be copious, such as the numerous available DTI and BV measures, if the cohort is relatively small, CCA is difficult to apply. Even disregarding this issue, as a first step, CCA requires the number of significant relationships to be known. This is equivalent to determining the rank of the coefficient matrix in RRR with the previously listed problems. Only after this is found, the canonical weights may be derived from the appropriate decomposition of the coefficient estimator to exercise the full inferential power of CCA.

It was only in 2011 that an innovative data-adaptive solution to the rank estimation problem was provided by Bunea, She, and Wegkamp in their paper “Optimal selection of reduced rank estimators of high-dimensional matrices” [8]. While others have been constrained by rank and distributional assumptions, Bunea et al. have introduced methodology to find a consistent estimator of the effective rank of the coefficient matrix while concurrently estimating the coefficient matrix itself. This is done using the Rank Selection Criterion (RSC) and opens a wide range of potential applications for dataset reduction as it is valid for any sample size and any number of response and predictor variables, even in the high dimensional setting. Bunea et al. show that the estimated rank is characterized by a tuning parameter. While this tuning parameter may be cross-validated, the addition of an adaptive tuning parameter in a closed form makes RSC extremely convenient and easy to use. However, there is a constraint: the error terms are assumed to be independently and identically distributed (i.i.d.), a not necessarily practical assumption in application.

The generalization of RSC without this assumption on the error terms will be carefully examined here in estimation of the rank of the coefficient matrix. This is also the first step in determining the number of significant canonical correlations which may be thought of as the number of uncorrelated relationships between the two sets. Examination of how to do so will proceed carefully as to understand the two important settings: large sample and high dimension.

The large sample setting with relevant theoretical backing will be the basis for the empirical results for the high dimensional setting. This will be done through a simple model transformation which scales the errors by a weight matrix so that RSC may be applied without issue. Recovery of the final coefficient estimator will be only one additional step, a
complete process referred to here as the Weighted Rank Selection Criterion (WRSC). WRSC will be supported in the large sample setting by the fundamental theorems of Bunea et al. with the addition of a weight matrix that acts as a “decorrelator” of the model. Inspection of this weight matrix will show that in fact, a particular selection of the weight matrix yields an adaptive version of CCA with simultaneous estimation of the number significant canonical correlations and of the canonical weights, here, the so-called Adaptive Canonical Correlation Analysis (ACCA). This relationship is only maintained through a convenient association between the residual covariance and the response covariance matrices. Without such, ACCA is unable to simultaneously estimate the canonical weights with the estimation of the number of significant canonical relationships. It also requires modification of the adaptive tuning parameter that is typically utilized in RSC. It will be shown in the large sample setting that appropriate scaling of the adaptive tuning parameter is possible. This is essential in ACCA so that the recovery of canonical weights occurs in addition to the estimation of the number of significant canonical relationships.

Treatment of the sample versions of these two weight matrices is not thoroughly examined in literature. This becomes particularly relevant in the computational aspects of the high dimensional setting. The connection between ACCA and WRSC only holds if the weight matrix can act as a true “decorrelator” of the model, a difficult requirement in the high dimensional setting. In addition to this, singularity issues arise as invertibility of the predictor and response covariance is required. A naive alternative would be to simply use the Moore-Penrose pseudoinverse whenever inversion of the sample residual covariance is required. However, this will show to be at a computational disadvantage to using the sample response covariance. While regularization is also an option for handling these singularity issues, theoretics so that the connection between ACCA and WRSC is maintained are nonexistent and will be established here. Yet through such investigations, it will be readily apparent that simple regularization on the sample residual covariance is at disadvantage to regularization on the sample response covariance, supporting the argument that the latter is a far more practical choice than the former. Extensive simulations will support the validity of the theoretical findings in the large sample setting and demonstrate in the high dimensional setting, good rank recovery and coefficient estimation is possible.

A slightly different methodology for handling high dimension reduction is possible through group variable selection techniques. Variable selection in multivariate response regression models is equivalent with group variable selection in the univariate response regression models. Hence, the multivariate model may be treated to consider both row sparsity and rank sparsity (or rank deficient) where the predictor set, if row sparse, may be reduced by group selection techniques, and then even furthermore so by ACCA. Bunea, She, and Wegkamp are amongst the first two employee such techniques that treat models at both sparsity levels in [7]. Their work is continued here with group selection performed by the population Group Lasso (GLASSO), followed by the implementation of ACCA.

To substantiate the application of ACCA, the inferential techniques of ACCA will be illustrated using a sample of HIV patients to establish unique relationships between the two sets of neurocognitive variables and with clinical and neuroimaging measures. ¹ The original dataset will be analyzed in Chapter 5 in the large sample setting. From the original

¹Provided by Dr. Hernando Ombao, Brown University, Research Center for Statistical Sciences.
predictor, a larger predictor set will be generated so that the high dimensional application may also be shown in Chapter 6 using a combination of GLASSO and ACCA.

This will be preceded by first, careful examination of CCA as a unique variation of RRR in Chapter 2, followed by RSC with the i.i.d. restriction, and a discussion about the choice of the weight matrix. This provides the framework for Chapter 3 where ACCA is developed, and WRSC is formed as consequence. Special considerations will be discussed for treatment of the high dimensional setting. Following directly, the supporting simulations for both the large sample and high dimensional setting. A detailed analysis using the large sample version of ACCA of the neuroimaging dataset of HIV patients will be provided in Chapter 5, and Chapter 6 will introduce the two step process in dimension reduction before application to a higher order neuroimaging dataset. This will be concluded with a summary of the new, relevant findings here, both of theoretical and empirical nature and a discussion of continuing work in this area.
CHAPTER 2

BACKGROUND METHODOLOGY

The basic methodology of Canonical Correlation Analysis (CCA) is introduced here as understanding its relationship with Reduced-Rank Regression (RRR) is crucial in the development of Adaptive Canonical Correlation Analysis (ACCA) and thus, also of Weighted Rank Selection Criterion (WRSC). In particular, treatment of the weight matrix will be examined very closely as it is the crux upon which the rest of the theory will be built. First, some basic notation is introduced:

Let \( \mathbf{x} \) and \( \mathbf{y} \) be two generic vectors of variables of sizes \( p \times 1 \) and \( n \times 1 \), respectively, with joint means

\[
E\left[\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}\right] = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}
\]

(2.1)

and covariances matrix

\[
E\left[\begin{pmatrix} \mathbf{x} - \mu_X \\ \mathbf{y} - \mu_Y \end{pmatrix}\left(\begin{pmatrix} \mathbf{x} - \mu_X \\ \mathbf{y} - \mu_Y \end{pmatrix}\right)\right] = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}
\]

(2.2)

Assume they are related via

\[
\mathbf{y}' = \mathbf{x}' \mathbf{A} + \mathbf{e}'
\]

(2.3)

under the following assumptions:

A.1. Without loss of generality that \( \mu_X = 0 \) and \( \mu_Y = 0 \).

A.2. The covariance matrices \( \Sigma_{XX} \) and \( \Sigma_{YY} \) are assumed to be non-singular unless otherwise noted.

A.2. The generic error vector \( \mathbf{e} \) has mean \( E(\mathbf{e}) = 0 \) and covariance \( Cov(\mathbf{e}) = \Sigma_{e} \).

Suppose that for \( i = 1, \ldots, m \) observations there are corresponding \( \mathbf{y}_i \) response and \( \mathbf{x}_i \) predictor vectors that are the \( i \)-th realization of the generic vectors \( \mathbf{x} \) and \( \mathbf{y} \). Let \( \mathbf{e}_i \) be the \( i \)-th realization of \( \mathbf{e} \) and assume that the \( \mathbf{e}_i \)'s are independently distributed for \( i = 1, \cdots, m \).

The model may be written in the standard matrix format for a multivariate response regression model using the data matrices formed from the \( i = 1, \cdots, m \) observations:

\[
\mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_m' \end{pmatrix} \quad \mathbf{Y} = \begin{pmatrix} \mathbf{y}_1' \\ \mathbf{y}_2' \\ \vdots \\ \mathbf{y}_m' \end{pmatrix}
\]
The model between \( X \) and \( Y \) may then be written as

\[
Y = XA + E
\]

where \( E \) is the \( m \times n \) matrix of error terms, \( Y \) is a \( m \times n \) response matrix, and \( X \) is a \( m \times p \) predictor matrix. Assume that \( X \) and \( Y \) have been column centered by their sample means. The corresponding maximum likelihood (ML) sample covariances are calculated as

\[
\hat{\Sigma}_{XX} = \frac{1}{m}X'X,
\]
\[
\hat{\Sigma}_{XY} = \frac{1}{m}X'Y = \hat{\Sigma}_{YX}',
\]
\[
\hat{\Sigma}_{YY} = \frac{1}{m}Y'Y
\]

where no assumptions are made on the invertibility of the sample covariance matrices \( \hat{\Sigma}_{XX} \) and \( \hat{\Sigma}_{YY} \) unless otherwise noted.

### 2.1 Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) is used to examine the relationship between two sets of variables. It is a particularly powerful inferential technique as it does not disregard the potential within set dependencies within either of the sets. Hence, the shared influences of the two sets upon the common sample may be examined carefully. While it is typically not used to create a direct model, as one set predicting the other, it maintains close ties to Reduced-Rank Regression (RRR). Many sources provide documentation of CCA including Hair et al. [18], Izenman [24], Johnson and Wichern [25], and Reinsel and Velu [32]. For the purposes of this section, the terms “predictor” and “response” will be slightly abused to label the two sets \( x \) and \( y \), respectively. In traditional CCA, there is no distinction between the two as treatment is symmetric.

While the previously given references provide the details of the population version of CCA, the sample version of CCA is less carefully documented. In particular, the details of the high dimensional setting are unclear. The population version is first introduced from two perspectives. The first of these is the Hotelling’s traditional, sequential way [22]. The second of these is a multivariate approach, particularly well-documented by Reinsel and Velu [32]. These two approaches are first given from the population point of view. The sample version of CCA follows along with a brief description of a particular type of CCA that is applicable in the high dimensional setting, Regularized Canonical Correlation Analysis (RCCA). CCA’s close relationship to Reduced-Rank Regression (RRR) as a weighted constrained problem will then be examined, along with some useful additional descriptive measures that will be applied later.

#### 2.1.1 Population Canonical Correlations and Variates

Following the notation from the beginning of Chapter 2, let \((\xi_k, \omega_k)\) denote a pair of new variables for \( k = 1, \ldots, t \leq \min(n, p) \) referred to as canonical variates where \( t \) is the
number of non-zero canonical correlations. For now, assume that \( t \) is given and that \( \Sigma_{XX} \) and \( \Sigma_{YY} \) are nonsingular and invertible. The canonical variates \( \xi_k \) and \( \omega_k \) are formed by

\[
\xi_k = x'f_k \\
\omega_k = y'h_k
\]

where \( f_k \) and \( h_k \) are the *canonical coefficients* or *canonical weights* found such that the \( k \)-th largest canonical correlation

\[
\rho_k = \text{Corr}(\xi_k, \omega_k) = \frac{f_k' \Sigma_{XY} h_k}{(f_k' \Sigma_{XX} f_k)^{1/2}(h_k' \Sigma_{YY} h_k)^{1/2}}
\]

is maximized with the properties

(i) \( \text{Cov}(\xi_k, \xi_l) = f_k' \Sigma_{XX} f_l = 0, \) for \( l \neq k \)

(ii) \( \text{Cov}(\omega_k, \omega_l) = h_k' \Sigma_{YY} h_l = 0, \) for \( l \neq k. \)

That is, \( \xi_l \) is uncorrelated with all the other \( \xi_k \)'s, and \( \omega_l \) is uncorrelated with all the other \( \omega_k \)'s. The new pairs of variables \((\xi_k, \omega_k)\) are ordered based upon their corresponding canonical correlations so that \( \rho_1 \geq \rho_2 \geq \cdots \geq \rho_t \). The appropriate canonical weights \( f_k \) and \( h_k \) may be derived in one of two ways:

**Way 1- Sequentially** This is Hotelling approach to CCA where CCA is a correlation-maximization technique [22]. Let \( \xi = x'f \) and \( \omega = y'h \) be the generic linear projections. First, continue to assume without loss of generality that \( \mathbb{E}(x) = 0 \) and \( \mathbb{E}(y) = 0 \) so that \( \mathbb{E}(\xi) = 0 \) and \( \mathbb{E}(\omega) = 0. \) Furthermore, assume that \( f' \Sigma_{XX} f = 1 \) and \( h' \Sigma_{YY} h = 1. \)

The problem reduces to finding the vectors of \( f \) and \( h \) such that

\[
\text{Corr}(\xi, \omega) = f' \Sigma_{XY} h
\]

has maximal correlation among all linear functions of \( x \) and \( y. \) Set

\[
f(f, h) = f' \Sigma_{XY} h - \frac{1}{2} \sqrt{\phi}(f' \Sigma_{XX} f - 1) - \frac{1}{2} \sqrt{\theta}(h' \Sigma_{YY} h - 1)
\]

where \( \sqrt{\phi} \) and \( \sqrt{\theta} \) are Lagrangian multipliers. Taking the partial derivatives w.r.t \( f \) and \( h \) gives

\[
\frac{\partial f(f, h)}{\partial f} = \Sigma_{XY} h - \sqrt{\phi} \Sigma_{XX} f \tag{2.6}
\]

\[
\frac{\partial f(f, h)}{\partial h} = \Sigma_{XY} f - \sqrt{\theta} \Sigma_{YY} h. \tag{2.7}
\]

Setting these partials to zero and multiplying (2.6) and (2.7) by \( f' \) and \( h' \), respectively gives now

\[
f' \Sigma_{XY} h - \sqrt{\phi} f' \Sigma_{XX} f = 0 \tag{2.8}
\]

\[
h' \Sigma_{XY} f - \sqrt{\theta} h' \Sigma_{YY} h = 0 \tag{2.9}
\]
so that
\[ f'\Sigma_{XY}h = \sqrt{\phi} = \sqrt{\theta}. \]

Using this to substitute \( \sqrt{\phi} \) for \( \sqrt{\theta} \) and with rearrangement of terms gives
\[ -\sqrt{\phi}\Sigma_{XX}f + \Sigma_{XY}h = 0 \quad (2.10) \]
\[ \Sigma_{YX}f - \sqrt{\phi}\Sigma_{YY}h = 0. \quad (2.11) \]

By multiplying the former of these by \( \Sigma_{YX}\Sigma_{XX}^{-1} \) and substituting it into the latter gives
\[ (\Sigma_{XY}\Sigma_{XX}^{-1}\Sigma_{XY} - \sqrt{\phi}\Sigma_{YY})h = 0 \]
or equivalently
\[ (\Sigma_{YY}^{-1/2}\Sigma_{XX}\Sigma_{XY}^{-1}\Sigma_{XY}^{-1/2} - \sqrt{\phi}\|n\|)h = 0. \]

This requires the determinant
\[ |\Sigma_{YY}^{-1/2}\Sigma_{XX}\Sigma_{XY}^{-1}\Sigma_{YY}^{-1/2} - \sqrt{\phi}\|n\|| = 0 \]
for there to be a nontrivial solution which is a polynomial of \( \sqrt{\phi} \) with \( n \) real roots. These are the ordered eigenvalues \( \phi_1 \geq \phi_2 \geq \cdots \geq \phi_n \geq 0 \) of
\[ \Sigma_{YY}^{-1/2}\Sigma_{XX}\Sigma_{XY}^{-1}\Sigma_{YY}^{-1/2} \]
with corresponding eigenvectors \( v_1, v_2, \cdots, v_n \). Hence, the maximal correlation between \( \xi \) and \( \omega \) is given by \( \sqrt{\phi} = \sqrt{\phi_1} \). This then provides that the choice of the canonical coefficients \( f \) and \( h \) are given by the vectors
\[ f_1 = \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}v_1 \]
\[ h_1 = \Sigma_{YY}^{-1/2}v_1. \]

Thus, the first pair of canonical variates is \( (\xi_1, \omega_1) \) where \( \xi_1 = x'f_1 \) and \( \omega_1 = y'h_1 \) with correlation \( \text{Corr}(\xi_1, \omega_1) = f_1'\Sigma_{XY}h_1 = \sqrt{\phi_1} \).

Now, given the first pair of canonical variate \( (\xi_1, \omega_1) \), let \( \xi = x'f \) and \( \omega = y'h \) denote a second arbitrary pair of variates with unit variances. The second pair of canonical variates have maximal correlation amongst all possible linear combinations of \( x \) and \( y \) which are uncorrelated with the first pair of canonical variates \( (\xi_1, \omega_1) \). That is,
\[ f'\Sigma_{XX}f_1 = 0 \]
\[ h'\Sigma_{YY}h_1 = 0. \]

Following from (2.10) and (2.11),
\[ \text{Corr}(\xi, \omega_1) = f_1'\Sigma_{XY}h_1 = \sqrt{\phi_1}f_1'\Sigma_{XX}f_1 = 0 \quad (2.12) \]
\[ \text{Corr}(\omega, \xi_1) = h_1'\Sigma_{YX}f_1 = \sqrt{\phi_1}h_1'\Sigma_{YY}h_1 = 0. \quad (2.13) \]
Again, \( f \) and \( h \) are chosen to maximize \( f(f, h) \) where

\[
f(f, h) = f'\Sigma_{XY}h - \frac{1}{2} \sqrt{\phi}(f'\Sigma_{XX}f - 1) - \frac{1}{2} \sqrt{\theta}(h'\Sigma_{YY}h - 1) + \sqrt{\eta}f'\Sigma_{XX}f_1 + \sqrt{\nu}h'\Sigma_{YY}h_1
\]

where \( \sqrt{\phi}, \sqrt{\theta}, \sqrt{\eta}, \sqrt{\nu} \) are Lagrangian multipliers. By taking the partial derivations of \( f(f, h) \) with respect to \( f \) and \( h \) and setting them equal to zero gives:

\[
\frac{\partial f}{\partial f} = \Sigma_{XY}h - \sqrt{\phi}\Sigma_{XX}f + \sqrt{\eta}\Sigma_{XX}f_1 = 0 \quad (2.15)
\]

\[
\frac{\partial f}{\partial h} = \Sigma_{YX}f - \sqrt{\theta}\Sigma_{YY}h + \sqrt{\nu}\Sigma_{YY}h_1 = 0. \quad (2.16)
\]

Similar to before, multiplying (2.15) and (2.16) by \( f' \) and \( h' \) respectively and then noting (2.12) and (2.13), these equations reduce to (2.10) and (2.11). Therefore, the second pair of canonical variates \((\xi_2, \omega_2)\) are calculated using the canonical coefficients

\[
f_2 = \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}v_2 \quad (2.17)
\]

\[
h_2 = \Sigma_{YY}^{-1/2}v_2 \quad (2.18)
\]

with correlation \( \text{Corr}(\xi_2, \omega_2) = f_2'\Sigma_{XY}h = \sqrt{\phi}_2 \).

The remaining canonical variates and correlations are derived in a similar sequential manner until no further solutions can be found.

**Way 2- Multivariately** CCA may be approached from a multivariate setting as in [32] and [24] by achieving a minimum in some least-squares sense. Let \( F_t = (f_1, f_2, \ldots, f_t) \) and \( H_t = (h_1, h_2, \ldots, h_t) \) so that \( \xi \) and \( \omega \) are now vector variates

\[
\xi = x'F_t \quad \omega = y'H_t
\]

again assuming that \( x \) and \( y \) have zero means. The canonical weight matrices may be found by minimizing

\[
\mathbb{E}
\left[
(YH_t - XF_t)(YH_t - XF_t)^\prime
\right]
\]

for a given \( t \) where \( \text{Cov}(\omega) = \Sigma_{\omega\omega} = H\Sigma_{YY}H' = I_t \).

To see this, first fix the matrix \( H \) and minimize the criterion (2.19):

\[
\min_{F} \mathbb{E}
\left[
(YH_t - XF_t)(YH_t - XF_t)^\prime
\right]
= tr
\left[
\Sigma_{\omega\omega} - \Sigma_{\omega\omega}X\Sigma_{XX}^{-1}X\Sigma_{\omega\omega}
\right]
\geq tr
\left[
\Sigma_{XX}^{-1/2}X\Sigma_{\omega\omega}X\Sigma_{XX}^{-1/2}X
\right]
\geq tr
\left[
H\Sigma_{YY}H' - H\Sigma_{YY}X\Sigma_{XX}^{-1}X\Sigma_{YY}H'
\right]
= t - \sum_{j=1}^{p_j}
\rho_j
\]
where $\tilde{\rho}_j$ is the $j$-th largest eigenvalue of $\mathbf{H} \Sigma_Y \Sigma_X^{-1} \Sigma_X \Sigma_Y \mathbf{H}' = \mathbf{H} \Sigma_Y^{1/2} \mathbf{R} \Sigma_Y^{1/2} \mathbf{H}$ where

$$
\mathbf{R} = \Sigma_Y^{-1/2} \Sigma_Y \Sigma_X^{-1} \Sigma_X \Sigma_Y \Sigma_Y^{-1/2}
$$

and $\mathbf{H} \Sigma_Y^{1/2} \Sigma_Y^{1/2} \mathbf{H}' = \mathbf{H} \Sigma_Y \mathbf{H}' = \mathbb{I}_t$. Thus, by Theorem A.0.2, the Poincaré Separation Theorem:

$$
t - \sum_{j=1}^t \tilde{\rho}_j \geq t - \sum_{j=1}^t \rho_j
$$

where $\rho_j$ is the $j$-th largest eigenvalue of $\mathbf{R}$.

Using Theorem A.0.1,

$$
\mathbf{F}_t = \Sigma_X^{-1} \Sigma_X \Sigma_Y \Sigma_Y^{-1/2} \hat{\mathbf{U}}_t \quad (2.20)
$$

$$
\mathbf{H}_t = \Sigma_Y^{-1/2} \hat{\mathbf{U}}_t \quad (2.21)
$$

where $\hat{\mathbf{U}}_t = (\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \ldots, \hat{\mathbf{u}}_t)$ and $\hat{\mathbf{u}}_k$ is the eigenvector corresponding to the $k$-th largest eigenvalue of

$$
\hat{\mathbf{R}} = \Sigma_Y^{-1/2} \Sigma_Y \Sigma_X^{-1} \Sigma_X \Sigma_Y \Sigma_Y^{-1/2}.
$$

Notice that both the traditional, sequential method and the multivariate method yield identical formulas for finding the canonical correlations and the canonical variates. While CCA was initially established in the sequential way, the multivariate way is a more novel approach. It is far more convenient and quicker for matrix-based programming. However, the sample canonical correlations and variates have computational issues, particularly in the high dimensional setting. To examine these, first the sample canonical correlations and variates are given for the large sample setting.

### 2.1.2 Sample Canonical Correlations and Variates

The sample canonical correlations and variates may be found using the matrices $\mathbf{X}$ and $\mathbf{Y}$ which are $m \times p$ and $m \times n$ matrices, respectively. The canonical coefficient matrices are easily estimated by $\mathbf{X}$ and $\mathbf{Y}$, using the ML sample covariances given at the beginning of the chapter in place of their theoretical versions in the previous two ways where $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ are assumed to be nonsingular and invertible.

The sample estimates of the canonical weight matrices $\hat{\mathbf{F}}_t$ and $\hat{\mathbf{H}}_t$ are then given by

$$
\hat{\mathbf{F}}_t = \hat{\Sigma}_X^{-1} \hat{\Sigma}_X \hat{\Sigma}_Y \hat{\Sigma}_Y^{1/2} \hat{\mathbf{U}}_t \quad (2.23)
$$

$$
\hat{\mathbf{H}}_t = \hat{\Sigma}_Y^{-1/2} \hat{\mathbf{U}}_t \quad (2.24)
$$

where $\hat{\mathbf{U}}_t = (\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \ldots, \hat{\mathbf{u}}_t)$ and $\hat{\mathbf{u}}_k$ is the eigenvector corresponding to the $k$-th largest eigenvalue of

$$
\hat{\mathbf{R}} = \hat{\Sigma}_Y^{-1/2} \hat{\Sigma}_Y \hat{\Sigma}_X^{-1} \hat{\Sigma}_X \hat{\Sigma}_Y \hat{\Sigma}_Y^{-1/2}.
$$

The corresponding sample canonical variates are formed by

$$
\hat{\Xi} = \mathbf{X} \hat{\mathbf{F}} \quad (2.26)
$$

$$
\hat{\Omega} = \mathbf{Y} \hat{\mathbf{H}} \quad (2.27)
$$
where $\hat{\mathbf{F}} = \left( \hat{f}_1, \hat{f}_2, \ldots, \hat{f}_t \right)$ and $\hat{\mathbf{H}} = \left( \hat{h}_1, \hat{h}_2, \ldots, \hat{h}_t \right)$. Now, $\hat{\Xi} = \left( \hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_t \right)$ and $\hat{\Omega} = \left( \hat{\omega}_1, \hat{\omega}_2, \ldots, \hat{\omega}_t \right)$ are matrices both of size $m \times t$ where the columns make up the canonical variates. Note that each estimated canonical variate pair $(\hat{\xi}_k, \hat{\omega}_k)$ has $m$ rows corresponding to the $m$ observations.

The corresponding canonical correlations may be calculated for each canonical variate pair as

$$\hat{\rho}_k = \frac{\hat{f}_k^\prime \hat{\Sigma}_{XY} \hat{h}_k}{(\hat{f}_k^\prime \hat{\Sigma}_{XX} \hat{f}_k)^{1/2}(\hat{h}_k^\prime \hat{\Sigma}_{YY} \hat{h}_k)^{1/2}}$$

for $k = 1, \ldots, t$.

Notice that the non-singularity requirement on $\hat{\Sigma}_{XX}$ and $\hat{\Sigma}_{YY}$ is needed so that these matrices are invertible. While Regularized Canonical Correlation Analysis (RCCA) was introduced to help deal with these invertibility issues, there exists certain computational limitations making it less than ideal. It is detailed next.

### 2.1.3 Regularized Canonical Correlation Analysis

The theory in CCA holds if $t$ is given and when the sample is large compared to the number of predictors and responses. However, when the sample size is less than the maximum of the number of predictors and responses or is only a bit larger, the inverses of $\hat{\Sigma}_{XX}$ and $\hat{\Sigma}_{YY}$ do not exist or are ill conditioned. In addition to this, the canonical correlations will be nearly 1 as the number of variables increases. Due to this, Eaton and Perlman advised that the CCA should only be performed if $m \geq n + p + 1$ [11].

Regularized Canonical Correlation Analysis (RCCA) offers a solution to these problems. It introduces regularization parameters that essentially add just enough to the estimators to make them invertible. This technique was proposed first by Vinod [41] and developed by Leurgans et al. [27] and is similar to the application of regularization applied in ridge regression.

Estimators of $\hat{\Sigma}_{XX}$ and $\hat{\Sigma}_{YY}$ are required and are typically the sample covariance matrices $\hat{\Sigma}_{XX}$ and $\hat{\Sigma}_{YY}$. These in RCCA are replaced by

$$\tilde{\Sigma}_{XX} = \hat{\Sigma}_{XX} + \delta_x \mathbb{I}_p$$
$$\tilde{\Sigma}_{YY} = \hat{\Sigma}_{YY} + \delta_y \mathbb{I}_n$$

where $\delta_x$ and $\delta_y$ are regularization parameters.

These regularization parameters may be determined using cross-validation as was proposed by González et al. [16] by the following steps.

1. Randomly divide the $m$ observations into $\nu$ folds. Let $\delta = (\delta_x \ \delta_y)$.

2. Compute the first canonical correlation $\rho_{\delta}^{(-i)}$ for a particular value $\delta$ without the observations from the $i$-th fold along with the corresponding first canonical coefficients $\hat{f}_{\delta}^{(-i)}$ and $\hat{h}_{\delta}^{(-i)}$.

3. Find the cross-validation score (CV) for $\delta$ as

$$CV(\delta_x, \delta_y) = \text{Corr} \left( \left\{ x_i^\prime \hat{f}_{\delta}^{(-i)} \right\}_{i=1}^{\nu}, \left\{ y_i^\prime \hat{h}_{\delta}^{(-i)} \right\}_{i=1}^{\nu} \right).$$
4. The final estimates of $\delta = (\delta_x, \delta_y)$ are chosen from a grid of values that maximize (2.31) so that

$$\hat{\delta} = \left( \hat{\delta}_1, \hat{\delta}_2 \right) = \arg\max_{\delta_x, \delta_y} CV(\delta_x, \delta_y).$$

While this does provide one option to handle the large dimensional setting, there are some serious setbacks. First of all, cross-validating a 2-dimensional grid of regularization parameters can be computation expensive and time consuming. Second of all, the cross-validation score only considers the first canonical variate pair and completely ignores the others. What will be established in the coming chapters is that first of all, it may be sufficient to cross-validate only a 1-dimensional grid, specifically to regularize $\hat{\Sigma}_{YY}$. Second of all, the given cross-validation score, which considers maximizing a canonical correlation for only the first canonical relationship, is not needed and that minimizing the typical prediction error is all that is required.

2.1.4 Relationship to Reduced-Rank Regression

Reduced-Rank Regression (RRR) has a direct relationship with CCA. This comes in handy when connecting CCA to WRSC later. To establish this, consider the model in (2.4) with the same given assumptions. RRR produces estimators of $A$ of reduced-rank to account for the correlation within the response set. That is, suppose $r(A) \leq \min(n, p)$. Then, $A$ may be expressed as

$$A = CD$$

and the model in (2.4) may be written as the reduced-rank regression model

$$Y = XCD + E.$$  \hspace{1cm} (2.32)

In their most basic form, estimators of the coefficient matrix $A$ minimize the weighted sum of squares with given rank, say $k$, with a positive-definite matrix of weights $\Gamma$. For clarity, these reduced-rank estimators of given rank $k$ are denoted here as $\hat{B}_k$. For a positive-definite weight matrix $\Gamma$, the estimator of $A$ of given rank $k$ denoted as $\hat{B}_k$ may be found by

$$\min_B \left\| (Y - XB)\Gamma^{1/2} \right\|^2_F \quad \text{s.t.} \quad r(B) \leq k.$$  \hspace{1cm} (2.33)

This is a weighted constrained problem.

To compute $\hat{B}_k$ for any $\Gamma$, a computational efficient procedure suggested by Reinsel and Velu [32] may be used with the ML sample covariances $\hat{\Sigma}_{XX}, \hat{\Sigma}_{XY} = \hat{\Sigma}'_{YX},$ and $\hat{\Sigma}_{YY}$. For now, assume that $\hat{\Sigma}_{XX}$ is nonsingular. The steps for computing $\hat{B}_k$ of given rank $k$ are then:

1. Find the (normalized) eigenvectors $\hat{V}_k = (\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_k)$, where $\hat{v}_j$ is the eigenvector corresponding to the $j$-th largest eigenvalue of the symmetric matrix

$$\hat{R} = \Gamma^{1/2}\hat{\Sigma}_{YX}\hat{\Sigma}_{XX}^{-1}\hat{\Sigma}_{XY}\Gamma^{1/2}.$$
2. Calculate the (full-rank) least-squares estimator $\hat{B} = \hat{\Sigma}^{-1}_{XX} \hat{\Sigma}_{XY}$. Then, form $\hat{C} = \hat{B} \Gamma^{1/2} \hat{V}_k$ and $\hat{D} = \hat{V}_k' \Gamma^{-1/2}$.

3. Compute the estimator $\hat{B}_k = \hat{C}_k \hat{D}_k$ where $\hat{C}_k = \hat{C}[1: k]$ denotes the matrix $\hat{C}$ retaining only its first $k$ columns and $\hat{D}_k = \hat{D}[1: k]$ denotes the matrix $\hat{D}$ retaining only its first $k$ rows.

These steps give the estimated reduced-rank regression coefficient matrix $\hat{B}_k$ of rank $k$ as

$$\hat{B}_k = \hat{C}_k \hat{D}_k = \hat{\Sigma}^{-1}_{XX} \hat{\Sigma}_{XY} \Gamma^{1/2} \left( \sum_{j=1}^{k} \hat{v}_j \hat{v}_j' \right) \Gamma^{-1/2} \tag{2.35}$$

where

$$\hat{C}_k = \hat{\Sigma}^{-1}_{XX} \hat{\Sigma}_{XY} \Gamma^{1/2} \hat{V}_k \tag{2.36}$$
$$\hat{D}_k = \hat{V}_k \Gamma^{-1/2}. \tag{2.37}$$

The two matrices $\hat{C}_k$ and $\hat{D}_k$ constructed above yield the unique decomposition of $\hat{B}_k$ with the following properties:

(i) $\hat{D}_k \Gamma \hat{D}_k' = \mathbb{I}_k$,

(ii) $\hat{C}_k' (XX) \hat{C}_k$ is a diagonal matrix

provided that $\Gamma$ is positive-definite.

Then, following, e.g. Izenman [24], finding the first $t = k$ canonical variates reduces to first minimizing the criterion in (2.34) with $\Gamma = \hat{\Sigma}^{-1}_{YY}$ (recall, $\hat{\Sigma}_{YY}$ is nonsingular). The two canonical weight matrices needed to construct the canonical variates are given by the decomposition of $\hat{B}_k$. By Theorem 2.3 in Reinsel and Velu [32], the matrices of canonical variates $\hat{\Xi}$ and $\hat{\Omega}$ may be constructed via the matrices $\hat{C}_k$ and $\hat{D}_k$ as

$$\hat{\Xi} = X \hat{C}_k \tag{2.38}$$
$$\hat{\Omega} = Y \left( \hat{D}_k \right)^{-}. \tag{2.39}$$

That is,

$$\hat{F}_t = \hat{C}_k \tag{2.40}$$

and

$$\hat{H}_t \hat{D}_k \hat{H}_t = \hat{H}_t \tag{2.41}$$
$$\hat{D}_k \hat{H}_t \hat{D}_k = \hat{D}_k$$

so that,

$$\hat{H}_t = \left( \hat{D}_k \right)^{-} \tag{2.41}$$

where the number of canonical relationships $t = k$, rank of the coefficient matrix.
2.1.5 Additional Descriptive Measures

While the previous has addressed the computational aspects of CCA, it is left to be said how inferences may be derived. Interpretation of canonical correlations and canonical variates is highly subjective. However, many additional descriptive measures aid in identifying relationships. Basic interpretation includes examination of the magnitude and sign of the estimated canonical correlations and the estimated canonical coefficients. However, evaluation may also include various additional calculations. These calculations include percentages of shared variance, percentages of explained variance, redundancy indexes, and percentages of contributed variance. An extensive discussion of these is found in [17] and are briefly reviewed directly.

The amount of shared variance is calculated from the canonical loadings. Let \( L_x(k, j) \) and \( L_y(k, j) \) denote the \( k \)-th canonical loadings with the \( j \)-th variable of the predictor and response set, respectively. Then, \( L_x(k, j) \) is the simple linear correlation between the \( j \)-th predictor variable \( x_j \) and \( k \)-th predictor canonical variate \( \hat{\xi}_k \) and \( L_y(k, j) \) is the simple linear correlation between the \( j \)-th response variable \( y_j \) and \( k \)-th response canonical variate \( \hat{\omega}_k \).

The percentage of shared variance for each variable with its canonical variate is simply the square of the canonical loadings. This may be thought of as the amount of variation in each variable explained by its respective canonical variate. Here, the relationship of each variable with its own canonical variate is isolated.

On the other hand, the amount of explained variance is the percentage of variance in one canonical variate that can be explained by the other canonical variate. The two canonical variates are each treated as a whole, and the percentage of explained variance is the canonical correlations squared. That is, let \( E_k \) denote the percentage of explained variance of the \( k \)-th canonical variate. Then,

\[
E_k = \hat{\rho}^2_k. \tag{2.42}
\]

An overall measure between each pair of canonical variates was developed by Stewart and Love [39] as the product of the average amount of shared variance and the amount of explained variance. This gives what is referred to as the redundancy index and is calculated as

\[
\begin{align*}
\text{k-th Redundancy Index of the Predictor Set} & = \frac{1}{p} \sum_{j=1}^{p} L_x^2(k, j) \times E_k \tag{2.43} \\
\text{k-th Redundancy Index of the Response Set} & = \frac{1}{n} \sum_{j=1}^{n} L_y^2(k, j) \times E_k. \tag{2.44}
\end{align*}
\]

In order to have a high redundancy, there must exist a high canonical correlation and a high degree of shared variance. Typically, a researcher is only interested in the variance extracted from the response set as a measure of the predictive ability of the estimated canonical function, but the variance extracted from the predictor set may also provide useful insight.

One caution in using these type of descriptive measures is that they are subject to a great deal of variability from one sample to another, the canonical coefficients even more
so than these measures. Problems such as suppression and multicollinearity may influence these type of statistics and make the interpretations invalid. Thus, the canonical cross-loadings are also useful in interpretation as they are the most resistant to these problems. Let \( C_x(k, j) \) and \( C_y(k, j) \) denote the canonical cross-loading of a predictor variable and a response variable, respectively. Then, they are simply the simple linear correlation between the \( j \)-th predictor variable \( x_j \) and the \( k \)-th response canonical variate \( \hat{\omega}_k \) and the simple linear correlation between the \( j \)-th response variable \( y_j \) and the \( k \)-th predictor canonical variate \( \hat{\xi}_k \). The square of these give percentages of contributed variance a variable gives to its opposite canonical variate. This provides a better measure of the relationship between the predictor and response sets, as the percentage of contributed variance isolates the influence of one variable across to the other canonical variate and helps in individually examining the predictor-response relationship.

### 2.2 Rank Selection Criterion (RSC)

The Rank Selection Criterion (RSC) is a novel approach by Bunea et al. for determining reduced-rank estimators through a data-adaptive method. It has been shown to be a computationally elegant solution to the long-standing rank selection problem and treats it in a penalized fashion. Because it is valid for any number of \( p \) predictors, \( n \) responses, and \( m \) observations, there is a wide range of applications. RSC is outlined here, but see [8] for the theoretical details.

Following from model (2.4), make the additional assumption that the covariance of the error terms is \( \Sigma_e = \sigma^2 I_n \). The \( p \times n \) estimator of \( A \) is found by minimizing the Frobenius norm of the model fit with a penalty term. That is, the penalized problem is

\[
\hat{A} = \arg\min_B \left\{ \| Y - XB \|^2_F + \mu r(B) \right\}
\]  

(2.45)

where \( \mu > 0 \) is the tuning parameter. Note that the i.i.d. assumption on the error matrix is required in order to obtain precise numerical constants for the penalty term in the criterion. The following two-step procedure may be used to compute \( \hat{A} \):

**Step 1:** Let \( \hat{B}_k \) denote the minimizer of \( \| Y - XB \|^2_F \) over all matrices \( B \) of fixed rank \( k \). This may be done by the following steps:

1. Find the (normalized) eigenvectors \( \hat{V}_k = (\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_k) \), where \( \hat{v}_j \) is the eigenvector corresponding to the \( j \)-th largest eigenvalue of the symmetric matrix

\[
\hat{R} = \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY}.
\]

2. Calculate the (full-rank) least-squares estimator \( \hat{B} = \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY} \). Then, form \( \hat{C} = \hat{B} \Gamma^{1/2} \hat{V}_k \) and \( \hat{D} = \hat{V}_k \Gamma^{-1/2} \).

3. Compute the estimator \( \hat{B}_k = \hat{C}_k \hat{D}_k \) where \( \hat{C}_k = \hat{C}[1 : k] \) denotes the matrix \( \hat{C} \) retaining only its first \( k \) columns and \( \hat{D}_k = \hat{D}[1 : k,] \) denotes the matrix \( \hat{D} \) retaining only its first \( k \) rows.
**Step 2:** The final estimator is $\hat{A} = \hat{B}_k$ where $\hat{k}$ is given by the following proposition:

**Proposition 2.2.1.** Let $\lambda_1 \geq \lambda_2 \geq \ldots$ denote the ordered eigenvalues of $\hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY}$.

Then,

$$\hat{k} = \max\{k : \lambda_k \geq \frac{\mu}{m}\}$$

(2.46)

where $\mu$ is the tuning parameter.

**Proof.** See Proposition 1 of [8].

Thus by definition, $\hat{k}$ is the number of eigenvalues of $\hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY}$ that exceeds $\frac{\mu}{m}$.

□

Note that Step 1 is essentially the same as in Section 2.1.4 with $\Gamma = \mathbb{I}_n$ for the large sample setting. In the high dimensional setting, Bunea et al. demonstrate, both empirically and theoretically, that even if $\hat{\Sigma}_{XX}$ is singular that if the Moore-Penrose generalized inverse is used, consistent rank estimation is still possible. That is, $\hat{B}_k$ as found above is the globally optimal solution to this penalized constrained problem regardless of the sample size.

### 2.2.1 Tuning Parameter

If an additional assumption of Gaussian errors is employed so that $E$ has independent $N(0, \sigma^2)$ entries, the penalty term $\text{pen}(B) = \mu r(B)$ may be further characterized by the following corollary.

**Corollary 2.2.2.** Assume that $E$ has independent $N(0, \sigma^2)$ entries. For any $\theta > 0$, the penalty term may be defined up to constants as

$$\text{pen}(B) = (1 + \theta)(1 + \xi)^2(\sqrt{n} + \sqrt{q})^2\sigma^2 r(B)$$

(2.47)

with $\theta, \xi > 0$ arbitrary and $q = r(X)$.

**Proof.** See Corollary 4 of [8].

This ensures good rank selection and prediction performance provided that the tuning parameter $\mu$ is just a bit larger than $(\sqrt{n} + \sqrt{q})^2$.

However, the practical case is when the variance $\sigma^2$ is unknown. Then, the penalty term may be defined using an estimated variance so that for any $\theta, \xi > 0$, and $0 < \delta < 1$,

$$\text{pen}(B) = \frac{(1 + \theta)}{1 - \delta}(1 + \xi)^2(\sqrt{n} + \sqrt{q})^2 S^2 r(B)$$

(2.48)

where $S^2$ is the unbiased estimator

$$S^2 = \frac{1}{(m - q)n} \left\| Y - PY \right\|_F^2$$

(2.49)

under the assumption that $q < m$. 

16
The choice of the tuning parameter $\mu$ may be determined in one of two ways: (1) utilizing cross-validation (typically 5-fold cross-validation is sufficient) or (2) using the data-dependent closed form as suggested by Bunea et al. The final closed form of the so-called adaptive tuning parameter is

$$\mu_{\text{adap}} = 2 \sigma^2 (n + q).$$  \hfill (2.50)

Bunea et al. have shown through extensive simulation experiments that constants slightly larger or smaller than 2 give similar results and that $\mu_{\text{adap}}$ has demonstrated excellent performance in both cases of large sample and large dimension. In practice, choice between cross-validation and the adaptive tuning parameter may be based upon the one that yields the best performance through validation. That is, the one that gives the smallest mean squared error (MSE) when calculated on an independent part of the data that is set aside prior to estimation ought to be used.

2.2.2 Dimension Reduction

RSC may also be applied as a dimension reduction technique of the predictor space. The two properties given in Section 2.1.4 provide that new, uncorrelated predictors may be found. Recall, the final estimator of the coefficient matrix is $\hat{A} = \hat{B}_k = \hat{C}_k \hat{D}_k$ with $\Gamma = \mathbb{I}_n$. To see this, consider the predicted response matrix rewritten as

$$\hat{Y} = X \hat{A} = X \hat{C}_k \hat{D}_k = Q \hat{D}_k.$$ 

The $\hat{k}$ columns of $Q = X \hat{C}_k$ are orthogonal by construction and may be regarded as new predictors. Since typically $\hat{k}$ is much smaller than $p$, this procedure offers significant dimension reduction.

The previous considerations and optimality results of the RSC estimator are well established under the assumption that $\Sigma = \sigma^2 \mathbb{I}_n$ with the weight matrix set to $\Gamma = \mathbb{I}_n$. In Chapter 3, this will be generalized for any known $\Sigma$, forming the Weighted Rank Selection Criterion (WRSC). This will be developed to connect WRSC to RSC to form ACCA.

2.3 Choice of Weight Matrix

The choice of the weight matrix $\Gamma$ is very important in Reduced-Rank Regression (RRR). In the Rank Selection Criterion, $\Gamma = \mathbb{I}_n$. However, in Canonical Correlation Analysis (CCA), CCA is recover through RRR by setting $\Gamma = \Sigma^{-1}_Y$. Here, two very popular choices of the weight matrix will be discussed: $\Sigma^{-1} = (\Sigma_{YY} - \Sigma_{YX} \Sigma^{-1}_X \Sigma_{XY})^{-1}$ and $\Sigma^{-1}_Y$. These two particular weight matrices are closely related in the Reduced-Rank Regression (RRR) setting. Reinsel and Velu is one place where the relationship between the two weight matrices is clearly established, but only from the population point of view [32].
While it is noted, however, that the sample estimators $\Gamma = \hat{\Sigma}^{-1}_{YY}$ and $\Gamma = \hat{\Sigma}^{-1}_r = \hat{\Sigma}_{YY} - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY}$ (the sample residual covariance) yield the same maximum likelihood (ML) estimates for $\hat{\Lambda} = \hat{\Theta} \hat{D}$ in the large sample setting, details for the high dimensional setting are lacking. These are examined carefully later in the next chapter as they become extremely important in the establishment of ACCA and WRSC. For convenience of notation, $\Sigma_c := \Sigma_r = \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}$, the population version of the residual covariance.

For the population version given in [32] to be valid, assume that $\Sigma = \Sigma_{YY}$ is positive-definite, i.e. $\Gamma = \Sigma^{-1}_Y$ is positive-definite and $\Gamma = \Sigma^{-1}_r$ is positive-definite. Then the relationship between the two weight matrices is given by the following:

**Lemma 2.3.1.** Let $\rho_j$ be the $j$-th largest eigenvalue of $\Sigma_{YY}^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1/2}$ and let $\lambda_j$ be the $j$-th largest eigenvalue of $\Sigma_r^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_r^{-1/2}$. Then, $\rho_j$ and $\lambda_j$ are related by

$$\lambda_j = \frac{\rho_j}{1 - \rho_j}$$

and

$$\rho_j = \frac{\lambda_j}{1 + \lambda_j}.$$

**Proof.** See [32]. ∎

**Lemma 2.3.2.** Let $\rho_j$ be the $j$-th largest eigenvalue of $\Sigma_{YY}^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1/2}$ with $v_j$ being the corresponding eigenvector. Let $\lambda_j$ be the $j$-th largest eigenvalue of $\Sigma_r^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_r^{-1/2}$ with the corresponding eigenvector $v^*_j$. Denote the eigenvectors as $\mathbf{V} = (v_1, v_2, \ldots, v_k)$ and $\mathbf{V}^* = (v_1^*, v_2^*, \ldots, v_k^*)$. Then, the eigenvectors are related via

$$\mathbf{V} = \Sigma_{YY}^{1/2} \Sigma_r^{-1/2} \mathbf{V}^* (1 - \Lambda)^{1/2}$$

where $\Lambda = \text{diag}(\rho_1, \rho_2, \ldots, \rho_k)$.

**Proof.** See [32]. ∎

These lemmas are used to establish the following relationship between the two coefficient estimators from $\Gamma = \Sigma^{-1}_Y$ and $\Gamma = \Sigma^{-1}_r$.

**Lemma 2.3.3.** For $\Gamma = \Sigma^{-1}_r$, let the estimator of the coefficient matrix be denoted as $\mathbf{B}_k^* = C_k^* \mathbf{D}_k^*$ of rank $k$. For $\Gamma = \Sigma^{-1}_Y$, let the estimator of the coefficient matrix be denoted as $\mathbf{B}_k = C_k \mathbf{D}_k$. Then,

$$\mathbf{B}_k^* = C_k^* \mathbf{D}_k^* \equiv C_k \mathbf{D}_k = \mathbf{B}_k$$

even though

$$C_k^* \neq C_k \text{ and } \mathbf{D}_k^* \neq \mathbf{D}_k.$$

**Proof.** This proof simply requires the use of the relationship established in Lemma 2.3.2 so that

$$C_k^* = \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_r^{-1/2} \Sigma_{YX}^{-1/2} \mathbf{V} (\mathbf{I} - \Lambda^2)^{-1/2}$$

$$= \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_r^{-1/2} \mathbf{V} (\mathbf{I} - \Lambda^2)^{-1/2}$$

$$= C_k (\mathbf{I} - \Lambda^2)^{-1/2}$$
\[ D_k^* = \left( \Sigma_r^{1/2} \Sigma_{YY}^{-1/2} V (I - \Lambda^2)^{-1/2} \right)' \Sigma_r^{1/2} \]

\[ = (I - \Lambda^2)^{-1/2} V' \Sigma_{YY}^{-1/2} \Sigma_r \]

\[ = (I - \Lambda^2)^{-1/2} V' \Sigma_{YY}^{-1/2} (\Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}) \]

\[ = (I - \Lambda^2)^{-1/2} V' \left( \Sigma_{YY}^{1/2} - \Sigma_{YY}^{-1/2} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1/2} \right) \Sigma_{YY}^{1/2} \]

\[ = (I - \Lambda^2)^{1/2} V' \Sigma_{YY}^{1/2} \]

So even though

\[ C_k \neq C_k^* \]

\[ D_k \neq D_k^* , \]

the following still holds.

\[ C_k^* D_k^* = C_k (I - \Lambda^2)^{-1/2} (I - \Lambda^2)^{1/2} D_k \]

\[ = C_k D_k . \]
CHAPTER 3

ADAPTIVE CANONICAL CORRELATION ANALYSIS AND WEIGHTED RANK SELECTION CRITERION

Adaptive Canonical Correlation Analysis (ACCA) is developed here from the weighted version of the Rank Selection Criterion (RSC), Weight Rank Selection Criterion (WRSC). ACCA adds the additional step to standard CCA in which the number of significant canonical correlations (relationships) $t$ is data-adaptively estimated. This lies directly in the addition of a weight matrix to RSC.

To develop these ideas, first recall that CCA is simply RRR with $\Gamma = \Sigma_{YY}^{-1}$. What was noted in Section 2.3 is that the eigenvalues from

$$\Gamma^{1/2} \Sigma_{XX} \Sigma_{XY}^{-1} \Sigma_{XY} \Sigma_{YY}^{1/2}$$

for $\Gamma = \Sigma_{YY}^{-1}$ and $\Gamma = \Sigma_{e}^{-1} = \Sigma_{r}^{-1}$ are related through Lemma 2.3.1. On the other hand, RSC is simply RRR working under the assumption that $\Sigma_{e} = \sigma^2 I_n$ and that $\Gamma = I_n$ and that the estimated rank is characterized by the eigenvalues of the above with the weight matrix set to the identity matrix. However, if $\Sigma_{e}$ is not i.i.d., then a new criterion needs to be developed with a weighting matrix that decorrelates the correlated model, ideally $\Gamma = \Sigma_{e}^{-1}$, the so-called WRSC here. If $\Sigma_{e}$ were known, the model may be decorrelated, the number of significant canonical correlations may be estimated by the estimated rank in the weighted version of RSC $\hat{k}$ and the canonical weights recovered from the decomposition of the coefficient estimator.

This, however, invites a variety of issues, including whether the continued use of an adaptive tuning parameter may be justified and the invertibility issues that arise from the high dimensional setting. Since the relationship between RRR and CCA is direct, WRSC is first developed under no assumptions of the error covariance for $\Gamma = \Sigma_{e}^{-1}$ where $\Sigma_{e}$ is assumed to be positive definite and nonsingular. Making the assumption there exists a good estimator of $\Sigma_{e}$ that is nonsingular, extensions of WRSC hold, for the most part, to the sample version. The typical estimator of $\Sigma_{e}$ is the sample residual covariance $\hat{\Sigma}_{e}$, but as noted before, $\hat{\Sigma}_{e}$ is only a good estimator in the large sample setting. If only considering the large sample setting, $\Gamma = \hat{\Sigma}_{YY}^{-1}$ may be used in place of $\hat{\Sigma}_{r}^{-1}$ as the relationship established in Section 2.3 holds. Noted here though is that careful treatment of the adaptive tuning parameter is required. This is addressed in the coming chapter. With the choice of $\Gamma = \hat{\Sigma}_{YY}^{-1}$,
now, the canonical weights may be found with $\hat{t}$ the estimated number of significant canonical correlations from the estimated rank in WRSC.

The large dimensional setting will require a more delicate handling. While theoretical properties are still needed, extensive empirical simulations will show that good rank recovery and prediction error are still possible. What is argued here is that there are a number of good reasons to use either $\hat{\Sigma}YY$ or a regularized version thereof, rather than any alternative version of $\hat{\Sigma}_r$. A more direct way to bypass these issues will be addressed in Chapter 6 where the additional step of variable selection will be advocated.

While an adaptive version of CCA is the final result, the methodology is developed from WRSC and hence is examined first. After the connection is drawn to ACCA for the population and the large sample versions, high dimensional considerations will be returned to. These are followed-up with simulations of a variety of settings, demonstrating that even without complete theoretics, empirical results are still promising.

### 3.1 Weighted Rank Selection Criterion

Recall that the general multivariate response model is

$$Y = XA + E$$  \hspace{1cm} (3.1)

where $Y$ is $m \times n$, $X$ is $m \times p$, and $A$ is $p \times n$ for $m$ observations, $p$ predictor variables, and $n$ response variables. Assume for now that

A.1. The matrices $X$ and $Y$ have been centered by their column means.

A.2. The error terms are distributed with mean zero and covariance $\Sigma e$.

A.3. The weight matrix $\Gamma = \Sigma e^{-1}$ is positive-definite (p.d.) unless otherwise noted.

If $\Sigma e$ were known, then the original model (3.1) could be immediately transformed into a truly decorrelated model

$$Z = XA_1 + T$$  \hspace{1cm} (3.2)

where, $Z = Y\Sigma e^{-1/2}$, $A_1 = A\Sigma e^{-1/2}$, and $T = E\Sigma e^{-1/2}$. The new error matrix $T$ is a matrix of $m$ independent realizations of the generic vector $t$ where $Cov(t) = \sigma^2I_n$ has uncorrelated errors. So, as in RSC, the penalized criterion may be written as

$$\min_M \left\| Z - XM \right\|_F^2 + \mu r(M).$$  \hspace{1cm} (3.3)

Denote the estimator of $M$ of rank $\hat{k}$ recovered from RSC as $\hat{M}_k$. Then,

$$\left\| (Y - XB)\Sigma e^{1/2} \right\|_F^2 + \mu r(B) \geq \left\| Z - XM \right\|_F^2 + \mu r(M) \geq \left\| Z - X\hat{M}_k \right\|_F^2 + \mu r(\hat{M}_k)$$

where $r(M) \leq r(B)$. The final global minimizer is recovered from as $\hat{M}_k$,

$$\hat{B}_k = \hat{M}_k \Sigma e^{1/2}.$$
Remark 3.1.1. The estimator $\hat{B}_k$ is also the globally optimal solution to the weighted constrained problem:

$$\min_{r(B) \leq k} \left\| (Y - XB)\Sigma_e^{1/2} \right\|^2_F.$$

In fact, if there exists a $\tilde{B}$ such that $r(\tilde{B}) \leq k$ and

$$\left\| (Y - \tilde{XB})\Sigma_e^{1/2} \right\|^2_F < \left\| (Y - X\hat{B}_k)\Sigma_e^{1/2} \right\|^2_F,$$

then

$$\left\| (Y - \tilde{XB})\Sigma_e^{-1/2} \right\|^2_F + \mu r(\tilde{B}) < \left\| (Y - X\hat{B}_k)\Sigma_e^{-1/2} \right\|^2_F + \mu r(\hat{B}_k),$$

which is a contradiction.

Now, consider the possibility that $\Gamma$ is positive-semidefinite. Then, there exists infinitely many $B$'s such that $r(B) \leq k$ and

$$\left\| (Y - XB)\Sigma_e^{1/2} \right\|^2_F \leq \left\| (Y - X\hat{B}_k)\Sigma_e^{1/2} \right\|^2_F,$$

However, of these $B$'s, the one with the smallest Frobenius norm may always be chosen so that

$$\hat{B}_k \triangleq \hat{M}_k \left( \Gamma^{1/2} \right)^{-}. $$

Hence, even if the weight matrix is positive-semidefinite, the global minimizer of the weighted penalized problem may always be found through a weighted version of RSC, referred here as Weighted Rank Selection Criterion (WRSC).

Let $\hat{k}$ be the estimated rank of $B$ recovered based upon the appropriate tuning parameter (to be discussed next). If $\hat{k}$ is given, then the estimator of $A$ is

$$\hat{B}_k = \hat{C}_k \hat{D}_k$$

where

$$\hat{C}_k = \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} \Gamma^{1/2} \hat{\nu}_k,$$

$$\hat{D}_k = \hat{\nu}_k \Gamma^{-1/2},$$

where $\hat{\nu}_k = (\hat{\nu}_1 \ldots \hat{\nu}_k)$ and $\hat{\nu}_j$ is the eigenvector corresponding to the $j$-th largest eigenvalue of

$$\hat{R} = \Gamma^{1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} \Gamma^{1/2}. $$

3.1.1 Tuning Parameter Selection

The choice of tuning parameter follows as in RSC. However, now it is defined for $\Gamma = \Sigma_e^{-1}$, instead of $\Gamma = I_n$. The rank $\hat{k}$ here follows from a similar proposition as to Proposition 2.2.1. A necessary lemma is first stated and proven.

Lemma 3.1.2. Let $\hat{M} = \hat{B} \Sigma_e^{1/2}$ be the full-rank coefficient estimator of $M$ where $\hat{B} = \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY}$ and let $\hat{M}_k = \hat{B}_k \Sigma_e^{1/2}$ be the reduced-rank estimator of $M_k$ of rank $k$ for the positive-definite matrix $\Sigma_e$ with $\hat{B}_k$ defined in (2.35). Then,

$$\frac{1}{m} \left\| XM - XM_k \right\|^2_F = \frac{1}{m} \left\| (XB - X\hat{B}_k)\Sigma_e^{1/2} \right\|^2_F = \sum_{j > k} \lambda_j,$$

where $\lambda_j$ denotes the $j$-th largest eigenvalue of $\Sigma_e^{1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} \Sigma_e^{1/2}$. 

22
Proof. By Izenman [24],
\[
\frac{1}{m} \| (Y - X\hat{B}_k) \Sigma_e^{1/2} \|_F^2 = \frac{1}{m} tr \left\{ (Y - X\hat{B}_k)' (Y - X\hat{B}_k) \Sigma_e \right\} \\
= \frac{1}{m} tr \left\{ (\hat{\Sigma}_{YY} - \Sigma_e^{-1/2} \left( \sum_{j=1}^{k} \lambda_j v_j v_j' \right) \Sigma_e^{-1/2} ) \Sigma_e \right\} \\
= \frac{1}{m} tr \left\{ (\hat{\Sigma}_{YY} - \hat{\Sigma}_{XY} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY} ) \Sigma_e + \sum_{j>k} \lambda_j v_j v_j' \right\} \\
= \frac{1}{m} tr \left\{ (\hat{\Sigma}_{YY} - \hat{\Sigma}_{XY} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY} ) \Sigma_e \right\} + \sum_{j>k} \lambda_j
\]
where \( \lambda_j \) is the \( j \)-th largest eigenvalue of \( \Sigma_e^{1/2} \hat{\Sigma}_{XY} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY} \Sigma_e^{1/2} \) and \( v_j \) is its corresponding normalized eigenvector. With a rearrangement of terms,
\[
\sum_{j>k} \lambda_j = \frac{1}{m} tr \left\{ (Y - X\hat{B}_k)' (Y - X\hat{B}_k) \Sigma_e \right\} - tr \left\{ (\hat{\Sigma}_{YY} - \hat{\Sigma}_{XY} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY} ) \Sigma_e \right\} \\
= \frac{1}{m} tr \left\{ (Y - X\hat{B}_k)' (Y - X\hat{B}_k) - (Y'Y - Y'PY) \right\} \Sigma_e \]
\[
= \frac{1}{m} tr \left\{ (X\hat{B} - X\hat{B})' (X\hat{B} - X\hat{B}) \Sigma_e \right\} \\
= \frac{1}{m} \| (X\hat{B} - X\hat{B}) \Sigma_e^{1/2} \|_F^2.
\]

This lemma is now used in the following proposition that defines the estimated rank \( \hat{k} \).

**Proposition 3.1.3.** Let \( \lambda_1 \geq \lambda_2 \geq \ldots \) be the ordered eigenvalues of \( \Sigma_e^{1/2} \hat{\Sigma}_{XY} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY} \Sigma_e^{1/2} \) with \( \Sigma_e = \Sigma_e^{-1} \). Then,
\[
\hat{k} = \max \left\{ k : \lambda_k \geq \frac{\mu_1}{m} \right\}
\]
where \( \mu_1 \) is the appropriate tuning parameter and \( \hat{A} = \hat{M}_k \Sigma_e^{1/2} \).

**Proof.** Assume first that the rank \( k \) is fixed. For \( \hat{B}_k \),
\[
\| (Y - X\hat{B}_k) \Sigma_e^{-1/2} \|_F^2 = \| (Y - PY) \Sigma_e^{-1/2} \|_F^2 + \| (PY - X\hat{B}_k) \Sigma_e^{-1/2} \|_F^2
\]
by the Pythagorean theorem. Note that \( X\hat{B} = PY \). By Lemma 3.1.2 with \( \Gamma = \Sigma_e^{-1} \), the second term may be written as
\[
\| (X\hat{B} - X\hat{B}) \Sigma_e^{-1/2} \|_F^2 = m \sum_{j>k} \lambda_j.
\]

Then, the weighted penalized least squares criterion reduces to
\[
\| (Y - PY) \Sigma_e^{-1/2} \|_F^2 + \left\{ m \sum_{j>k} \lambda_j + \mu_1 k \right\}.
\]
Now it is given that
\[
\min_B \left\{ \| (Y - XB)\Sigma_e^{-1/2}\|_F^2 + \mu_1 r(B) \right\} =
\|(Y - PY)\Sigma_e^{-1/2}\|_F^2 - \mu_1 n + \min_k \sum_{j>k} \{m \lambda_j - \mu_1\}.
\]

By taking \(k\) as the largest index \(j\) for which \(m \lambda_j - \mu_1 \geq 0\) or \(\lambda_j \geq \frac{\mu_1}{m}\), \(\sum_{j>k} \{m \lambda_j - \mu_1\}\) is minimized.

Therefore, \(\hat{k}\) is the the number of eigenvalues of \(\Sigma_e^{-1/2} \hat{\Sigma}_{XX} \hat{\Sigma}_{XY} \Sigma_e^{1/2}\) that is greater than \(\frac{\mu_1}{m}\).

**In computation**, a successful, practical application of this will require employing an estimator of \(\Sigma_e\) in the above. If the number of observations \(m\) is large enough, the estimator of the covariance of residual \(\hat{\Sigma}_r\) is such an estimator where
\[
\hat{\Sigma}_r = \hat{\Sigma}_{YY} - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY}.
\]

If \(\Sigma_e\) were known, then ideally the adaptive tuning parameter \(\mu_{\text{adap}}\) may be used where
\[
\mu_{\text{adap}} = 2S_1^2(n + q)
\]
and
\[
S_1^2 = \frac{\|(Y - PY)\Sigma_e^{-1/2}\|_F^2}{mn - qn}
\]
where \(S_1^2\) is just the unbiased estimator of \(\sigma^2\). However in this honest composition, it is only hypothesized that when
\[
\mu_1 = \frac{(1 + \theta)}{1 - \delta} (1 + \xi)^2 S_1^2 (\sqrt{n} + \sqrt{q})
\]
for any \(\theta, \xi > 0\) and \(0 < \delta < 1\), that the rank consistency as from RSC still holds as in (2.48).

However, \(\Sigma_e\) is rarely known. If the sample size is large enough, then the residual covariance \(\hat{\Sigma}_r\) is a consistent estimator of \(\Sigma_e\). If the covariance of the residuals is employed, then the ML biased estimator of the variance is
\[
\hat{\sigma}_1^2 = \frac{\|(Y - PY)\hat{\Sigma}_r^{-1/2}\|_F^2}{mn} = \frac{1}{mn} \text{tr} \left\{ (Y - PY)(\hat{\Sigma}_{YY} - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY})^{-1} (Y - PY)' \right\}
\]
\[
= \frac{1}{mn} \text{tr} \left\{ (Y - PY)'(Y - PY)(\hat{\Sigma}_{YY} - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY})^{-1} \right\}
\]
\[
= \frac{1}{mn} \text{tr} \left\{ (Y'Y - Y'PY)(\hat{\Sigma}_{YY} - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY})^{-1} \right\}
\]
\[
= \frac{1}{mn} \text{tr} \left\{ m(\hat{\Sigma}_{YY} - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY}) (\hat{\Sigma}_{YY} - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XY})^{-1} \right\}
\]
\[
= \frac{1}{mn} \text{tr} \left\{ mI_n \right\} = 1.
\]
Therefore, denote the adaptive tuning parameter with the weight matrix set to the residual covariance as \( \mu_{\text{adap}} = \mu_{\text{adap}}^{(r)} \). Then, it reduces to

\[
\mu_{\text{adap}}^{(r)} = 2(n + q).
\]

If only a model is desired and the sample size is large enough, then the problem stops by using \( \Gamma = \hat{\Sigma}_r^{-1} \) and the corresponding adaptive tuning parameter \( \mu_{\text{adap}}^{(r)} \) in WRSC. However, it may be desirable to make inferences based upon the model, using CCA. Since estimating the rank of the coefficient matrix \( \hat{k} \) is the same as estimating the number of significant canonical correlations \( \hat{t} \), WRSC may also be employed to recover \( \hat{t} \). This is established through the choice of the weight matrix and is closely reviewed next before examining the high dimensional setting.

3.2 Adaptive Canonical Correlation Analysis

ACCA complements the construction of WRSC by adding a step to CCA in which the number of significant canonical correlations \( t \) is chosen adaptively from the data in an optimal fashion. Since estimating the rank of the coefficient matrix \( \hat{k} \) is equivalent to estimating the number of significant canonical correlations \( \hat{t} \), an estimate of \( \hat{k} \) may be found using WRSC, with some requirements. WRSC requires an estimator of \( \Sigma \). One such estimator is the covariance of the residuals \( \hat{\Sigma}_r \). However, it was established in Section 2.3 that using \( \Gamma = \hat{\Sigma}_r^{-1} \) and \( \Gamma = \hat{\Sigma}_{YY}^{-1} \) results in the same estimated coefficient matrix in the population version. The same properties carry over to the large sample setting only. Thus, either choice of weight matrix results in the same estimate of the rank \( \hat{k} \) and thus \( \hat{t} \).

The difficulty lies in the proper scaling of the tuning parameter. Even the relationship in Lemma 2.3.3 exists and assuming that \( \hat{\Sigma}_r \) and \( \hat{\Sigma}_{YY} \) are positive definite and nonsingular, it is still true that

\[
\min_B \left\| (Y - XB)\hat{\Sigma}_r^{-1/2} \right\|_F^2 \neq \min_B \left\| (Y - XB)\hat{\Sigma}_{YY}^{-1/2} \right\|_F^2.
\]

That is, the minimum value that is attained by the criterion is not the same for \( \Gamma = \hat{\Sigma}_r^{-1} \) and \( \Gamma = \hat{\Sigma}_{YY}^{-1} \). Hence, scaling of the tuning parameter is required in the WRSC criterion.

First to observe how the canonical variates may be recovered, set \( \Gamma = \hat{\Sigma}_{YY}^{-1} \) in place of \( \Gamma = \Sigma_e^{-1} \). Assume \( \mu_1 = \mu_1^{(y)} \) denotes the tuning parameter. Then the criterion is written as

\[
\hat{A} = \hat{C}_k\hat{D}_k = \arg\min_{B \in \mathbb{C}^{D \times P}} \left\{ \left\| (Y - XB)\hat{\Sigma}_{YY}^{-1} \right\|_F + \mu_1^{(y)} r(B) \right\}.
\]

The canonical variates may then be recovered by

\[
\hat{\Xi} = X\hat{C}_k \tag{3.6}
\]
\[
\hat{\Omega} = Y(\hat{D}_k)^{-} \tag{3.7}
\]

where \( \hat{k} = \hat{t} \) the number of significant canonical correlations.

Now for a closer examination of the tuning parameter.
3.2.1 Tuning Parameter Selection

Good rank recovery is dependent upon the tuning parameter $\mu_1$ selection in WRSC. If $\mu_1^{(r)}$ is a known tuning parameter (adaptive or otherwise) when $\Gamma = \Sigma_r^{-1}$, then a tuning parameter $\mu_1^{(y)}$ for $\Gamma = \Sigma_{YY}^{-1}$ may be easily derived though the relationship established in Lemma 2.3.1 (providing, again of course, that $\hat{\Sigma}_r$ is a good estimator of $\Sigma_e$. This is shown in the following theorem (continuing to assume that both are positive-definite and nonsingular).

**Theorem 3.2.1.** For any given tuning parameter $\mu_1^{(r)}$ in

$$\min_B \left\{ \left\| (Y - XB)\hat{\Sigma}_r^{-1/2}\right\|_F^2 + \mu_1^{(r)}(B) \right\},$$

if

$$\frac{\mu_1^{(y)}}{m} = \frac{\mu_1^{(r)}}{m} \frac{1}{1 + \frac{\mu_1^{(r)}}{m}}$$

in (3.5), then the rank recovered from setting the weight matrix to $\Gamma = \hat{\Sigma}_r^{-1}$ using the tuning parameter $\mu_1^{(r)}$ is identical to the rank recovered from setting the weight matrix to $\Gamma = \hat{\Sigma}_{YY}^{-1}$ using the tuning parameter $\mu_1^{(y)}$.

**Proof.** Let $\Gamma = \Sigma_e^{-1}$ be replaced with $\Gamma = \hat{\Sigma}_r^{-1}$ in Proposition 3.1.3 with a corresponding tuning parameter denoted as $\mu_1^{(r)}$. Then by definition,

$$\lambda_k \geq \frac{\mu_1^{(r)}}{m} > \lambda_{k+1}$$

where $\lambda_k$ is the $k$-th largest eigenvalue of $\hat{\Sigma}_r^{-1/2} \Sigma_{XY} \hat{\Sigma}_{XX} \Sigma_{XY} \Sigma_r^{-1/2}$. Let $\rho_k$ denote the $k$-th largest eigenvalue of $\hat{\Sigma}_{YY}^{-1/2} \Sigma_{YY} \hat{\Sigma}_{XX} \Sigma_{XY} \Sigma_{YY}^{-1/2}$. By Lemma 2.3.1, $\lambda_k$ and $\rho_k$ are related by

$$\rho_k = \frac{\lambda_k}{1 + \lambda_k}$$

$$\lambda_k = \frac{\rho_k}{1 - \rho_k}.$$ 

So,

$$\frac{\rho_k}{1 - \rho_k} \geq \frac{\mu_1^{(r)}}{m} > \frac{\rho_{k+1}}{1 - \rho_{k+1}}$$

implying

$$\rho_k \geq \frac{\mu_1^{(r)}}{m + \mu_1^{(r)}} > \rho_{k+1}.$$ 

Therefore, the tuning parameter for $\mu_1^{(y)}$ for $\Gamma = \hat{\Sigma}_{YY}^{-1}$ that yields the same estimated rank $\hat{k}$ as from using $\Gamma = \hat{\Sigma}_r^{-1}$ is

$$\mu_1^{(y)} = \frac{\mu_1^{(r)}}{1 + \frac{\mu_1^{(r)}}{m}}.$$ 

(3.8)
Hence, given a tuning parameter \( \mu^{(r)}_1 \) and that \( \Gamma = \hat{\Sigma}_r^{-1} \) in WRSC, the tuning parameter \( \mu^{(y)}_1 \) as defined in (3.8) with \( \Gamma = \hat{\Sigma}_{YY}^{-1} \) may be found using this theorem that will provide rank recovery.

Since the adaptive tuning parameter when \( \Gamma = \hat{\Sigma}_r^{-1} \) is \( \mu^{(r)}_{\text{adap}} = 2(n + q) \), using Theorem 3.2.1 provides a closed-form of the adaptive tuning parameter for \( \Gamma = \hat{\Sigma}_{YY}^{-1} \) as

\[
\mu^{(y)}_{\text{adap}} = \frac{2(n + q)}{1 + \frac{2(n+q)}{m}}. \tag{3.9}
\]

Note that up until now, in this chapter, only the population versions and large sample estimators of the weight matrices have been discussed. A careful discussion of their sample versions is in the high dimensional setting is now required.

### 3.3 High Dimensional Considerations

WRSC has established that a globally optimal solution to this weighted constrained problem exists. The properties and theorems of RSC hold only if the weight matrix \( \Gamma \) can act as a true model decorrelator, such as \( \Gamma = \Sigma_e^{-1} \). If \( \Sigma_e \) were known, no further issues need to be addressed as Bunea et al. have established properties for high dimensional rank recovery. Through the population version established by Reinsel and Velu, the relationship to \( \Gamma = \Sigma_{YY}^{-1} \) is direct, provided that \( \Sigma_{YY} \) is positive definite. This is then directly CCA with the additional step of adaptive estimation of the number of significant canonical correlations. In the large sample setting, these concepts carry over to the sample version.

However, in the high dimensional setting, the sample version introduces a plethora of problems. It is desirable to achieve two goals of interest: (1) estimating the rank of the coefficient matrix \( \hat{k} \) and (2) recovering a good estimator of the unknown coefficient matrix \( A \). The idea here is to gently, and empirically, explore options for high dimensional application. It is noted that when \( p > m \), \( \hat{\Sigma}_r \) is singular and not invertible. An alternative is to use Moore-Penrose generalized inverse \( \Gamma = \hat{\Sigma}_r^{-} \). However, this will show to be unstable as finding this reduces to taking the inverse of the singular values. If the singular values are small, these approach \( \infty \). On the other hand, \( \hat{\Sigma}_{YY} \) is still nonsingular even in \( p > m \) as long as \( n < m \). More recently, estimators with regularization parameters have been utilized to bypass these issues. That is, the sample residual covariance could be regularized so that \( \Gamma = (\hat{\Sigma}_r + \delta I_n)^{-1} \). But what can be directly established is that even for the regularized versions of the weight matrices \( \hat{\Sigma}_r \) and \( \hat{\Sigma}_{YY} \), the relationships in Lemma 2.3.1 still hold. But the latter has a small advantage over the former as its corresponding eigenvalues have a stable limit while those for the regularized version of \( \hat{\Sigma}_r \) may not.

To establish similar relationships as in Lemmas 2.3.1 for the two regularized weight matrices \( \hat{\Sigma}_r + \delta I_n \) and \( \hat{\Sigma}_{YY} + \delta I_n \), consider the following:

**Lemma 3.3.1.** Let \( \rho_j^{(\delta)} \) be the \( j \)-th largest eigenvalue of

\[
(\hat{\Sigma}_{YY} + \delta I_n)^{-1/2} \hat{\Sigma}_{YY} \hat{\Sigma}_X \hat{\Sigma}_{XX} \hat{\Sigma}_{XY} (\hat{\Sigma}_{YY} + \delta I_n)^{-1/2}
\]
and let $\lambda_j^{(\delta)}$ be the $j$-th largest eigenvalue of 

$$(\hat{\Sigma}_r + \delta \mathbb{I}_n)^{-1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{-1/2}.$$ 

Then, $\rho_j^{(\delta)}$ and $\lambda_j^{(\delta)}$ are related by 

$$\lambda_j^{(\delta)} = \frac{\rho_j^{(\delta)}}{1 - \rho_j^{(\delta)}} \quad (3.10)$$

and 

$$\rho_j^{(\delta)} = \frac{\lambda_j^{(\delta)}}{1 + \lambda_j^{(\delta)}}. \quad (3.11)$$

**Proof.** By definition, 

$$|\rho_j^{(\delta)} - (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{-1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{-1/2}| = 0 \quad (3.12)$$

$$|\lambda_j^{(\delta)} - (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{-1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{-1/2}| = 0 \quad (3.13)$$

where $\hat{\Sigma}_{YX} + \delta \mathbb{I}_n$ and $\hat{\Sigma}_r + \delta \mathbb{I}_n$ are nonsingular for $\delta > 0$. From (3.13),

$$|(\hat{\Sigma}_r + \delta \mathbb{I}_n)^{1/2} \rho_j^{(\delta)} \mathbb{I}_n (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{1/2}$$

$$- (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{1/2} (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{-1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{-1/2} (\hat{\Sigma}_r + \delta \mathbb{I}_n)^{1/2} = 0$$

$$\Rightarrow |\lambda_j^{(\delta)} (\hat{\Sigma}_r + \delta \mathbb{I}_n) - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY}| = 0.$$ 

From this, it follows that 

$$|\lambda_j^{(\delta)} (\hat{\Sigma}_r + \delta \mathbb{I}_n) - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY}| = |\lambda_j^{(\delta)} (\hat{\Sigma}_{YX} - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} + \delta \mathbb{I}_n) - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY}|$$

$$= |\lambda_j^{(\delta)} (\hat{\Sigma}_{YX} + \delta \mathbb{I}_n) - (1 + \lambda_j^{(\delta)}) \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY}|$$

$$= 0.$$

Thus, 

$$\left| \frac{\lambda_j^{(\delta)}}{1 + \lambda_j^{(\delta)}} - (\hat{\Sigma}_{YX} + \delta \mathbb{I}_n)^{-1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} (\hat{\Sigma}_{YX} + \delta \mathbb{I}_n)^{-1/2} \right| = 0.$$ 

Because a function of the form $h(t) = \frac{t}{1+t}$, $t > 0$ is strictly monotone, $\frac{\lambda_j^{(\delta)}}{1 + \lambda_j^{(\delta)}}$ yields all eigenvalues of 

$$(\hat{\Sigma}_{YX} + \delta \mathbb{I}_n)^{-1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX}^{-} \hat{\Sigma}_{XY} (\hat{\Sigma}_{YX} + \delta \mathbb{I}_n)^{-1/2}.$$ 

That is, 

$$\rho_j^{(\delta)} = \frac{\lambda_j^{(\delta)}}{1 + \lambda_j^{(\delta)}}.$$ 

28
and it is immediate that

\[ \lambda_j^{(\delta)} = \frac{\rho_j^{(\delta)}}{1 - \rho_j^{(\delta)}}. \]

**Note on Convergence** Lemma lem:4 assumes that \( \delta > 0 \) and that \( \rho_j^{(\delta)} \) and \( \lambda_j^{(\delta)} \) are dependent on \( \delta \). When \( \delta \to 0^+ \), \( \rho_j^{(\delta)} \) has a stable limit while \( \lambda_j^{(\delta)} \) may not. This is noted using the following lemma.

**Lemma 3.3.2.** For \( \delta > 0 \),

\[
\lim_{\delta \to 0^+} (\hat{\Sigma}_{YY} + \delta \mathbb{I}_n)^{-1/2} Y' = \left( \left( \frac{1}{m} Y'Y \right)^{1/2} \right)^{-} Y'
\]

**Proof.** Assume that \( Y \) has been centered by its column means. Let \( Y = U D V' \) be the reduced singular value decomposition of \( Y \) where \( r = r(Y) \). Hence, \((D)_{ii} > 0\) for all \( i \) and \( D \) is a \( r \times r \) matrix. Also, \( U \) is a \( m \times r \) matrix and \( V \) is a \( n \times r \) matrix with \( U'U = \mathbb{I}_m \) and \( V'V = \mathbb{I}_r \). Thus, \( \hat{\Sigma}_{YY} \) may be written as

\[
\hat{\Sigma}_{YY} = \frac{1}{m} Y'Y = \frac{1}{m} V D^2 V'.
\]

Let \( \tilde{V} = (V \ V_\perp) \in \mathbb{R}^{n \times n} \) be an orthonormal matrix so that \( \tilde{V} \tilde{V}' = \tilde{V}'\tilde{V} = \mathbb{I}_n \). Thus,

\[
\tilde{V} (\hat{\Sigma}_{YY} + \delta \mathbb{I}_n) \tilde{V}' = \tilde{V} (\hat{\Sigma}_{YY} - \delta \mathbb{I}_n) \tilde{V}'
\]

\[
\Rightarrow (\hat{\Sigma}_{YY} + \delta \mathbb{I}_n)^{-1/2} Y' = \tilde{V} \left( \left( \frac{1}{m} D^2 + \delta \mathbb{I}_r \right)^{-1/2} \circ \mathbb{I}_r \circ \mathbb{I}_n \right) \tilde{V}' \left( \tilde{V}' \tilde{V} \mathbb{I}_r \mathbb{I}_\perp \right) V D U' \]

\[
= (V \ V_\perp) \left( \left( \frac{1}{m} D^2 + \delta \mathbb{I}_r \right)^{-1/2} \circ \mathbb{I}_r \circ \mathbb{I}_n \right) \left( \tilde{V}' \tilde{V} \mathbb{I}_r \mathbb{I}_\perp \right) V D U'
\]

\[
= (V \ V_\perp) \left( \left( \frac{1}{m} D^2 + \delta \mathbb{I}_r \right)^{-1/2} \circ \mathbb{I}_r \circ \mathbb{I}_n \right) \left( \tilde{V}' \tilde{V} \mathbb{I}_r \mathbb{I}_\perp \right) V D U'
\]

\[
= (V \ V_\perp) \left( \left( \frac{1}{m} D^2 + \delta \mathbb{I}_r \right)^{-1/2} \circ \mathbb{I}_r \circ \mathbb{I}_n \right) \left( \mathbb{I}_r \mathbb{I}_\perp \mathbb{I}_r \mathbb{I}_\perp \right) V D U'
\]

\[
= (V \ V_\perp) \left( \left( \frac{1}{m} D^2 + \delta \mathbb{I}_r \right)^{-1/2} \circ \mathbb{I}_r \circ \mathbb{I}_n \right) \left( \mathbb{I}_r \mathbb{I}_\perp \mathbb{I}_r \mathbb{I}_\perp \right) V D U'
\]

\[
= (V \ V_\perp) \left( \left( \frac{1}{m} D^2 + \delta \mathbb{I}_r \right)^{-1/2} \circ \mathbb{I}_r \circ \mathbb{I}_n \right) \left( \mathbb{I}_r \mathbb{I}_\perp \mathbb{I}_r \mathbb{I}_\perp \right) V D U'
\]

From the reduced singular value decomposition, \( D \) has positive diagonal entries. Hence,

\[
\lim_{\delta \to 0^+} \left( \frac{1}{m} D^2 + \delta \mathbb{I}_r \right)^{-1/2} D = \sqrt{m} \mathbb{I}_r.
\]
So,
\[ \lim_{\delta \to 0^+} (\hat{\Sigma}_{YY} + \delta \mathbb{I}_n)^{-1/2} = \sqrt{m} V U' \]

where
\[
\sqrt{m} V U' = \sqrt{m} V D^{-1} D U' = \sqrt{m} V D^{-1} V' V D U' = \left( \frac{1}{m} V D^2 V' \right)^{1/2} V D U' = \left( \frac{1}{m} Y' Y \right)^{1/2} Y' \]

\[ \square \]

Lemma 3.3.2 implies that
\[
(\hat{\Sigma}_{YY} + \delta \mathbb{I}_n)^{-1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX} \hat{\Sigma}_{XY} (\hat{\Sigma}_{YY} + \delta \mathbb{I}_n)^{-1/2} \rightarrow (\hat{\Sigma}_{YY}^{1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX} \hat{\Sigma}_{XY} (\hat{\Sigma}_{YY}^{1/2})^{-} \text{ as } \delta \rightarrow 0^+.
\]

Thus \( \rho_j^{(\delta)} \) converges to the eigenvalues of \( (\hat{\Sigma}_{YY}^{1/2} \hat{\Sigma}_{YX} \hat{\Sigma}_{XX} \hat{\Sigma}_{XY} (\hat{\Sigma}_{YY}^{1/2})^{-} \).

The eigenvalues \( \rho_j^{(\delta)} \) have some advantage over the eigenvalues \( \lambda_j^{(\delta)} \) since \( \lambda_j^{(\delta)} \) may not be stable. To see this, consider a small example:

**Example 3.3.3.** Let
\[
Y = \begin{pmatrix} \mathbb{I} & \mathbb{I} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} \text{ and } P = X (X' X)^{-} = \begin{pmatrix} \mathbb{O} & \mathbb{I} \\ \mathbb{O} & \mathbb{O} \end{pmatrix}.
\]

Since
\[
(\mathbb{I} - P) Y = \begin{pmatrix} \mathbb{I} \\ \mathbb{O} \\ \mathbb{O} \end{pmatrix},
\]
then
\[
\hat{\Sigma}_r = \hat{\Sigma}_{YY} - \hat{\Sigma}_{YX} \hat{\Sigma}_{XX} \hat{\Sigma}_{XY}
\]
\[
= \frac{1}{m} (Y' Y - Y' X (X' X)^{-} X Y)
\]
\[
= \frac{1}{m} (Y' Y - Y' P Y)
\]
\[
= \frac{1}{m} Y' (\mathbb{I}_n - P) Y
\]
\[
= \frac{1}{m} Y' (\mathbb{I}_n - P) (\mathbb{I}_n - P) Y
\]
\[
= \frac{1}{m} \begin{pmatrix} \mathbb{I} \\ \mathbb{O} \\ \mathbb{O} \end{pmatrix}.
\]
Hence,

\[
(\hat{\Sigma}_r + \delta I_n)^{-1/2} = \left(\begin{array}{cc}
(1 + \delta) & \delta \\
\delta & \delta
\end{array}\right)^{-1/2} = \left(\begin{array}{cc}
\frac{1}{1+\delta} & \frac{1}{\delta} \\
\frac{1}{\delta} & \frac{1}{\delta}
\end{array}\right)
\]

\Rightarrow (\hat{\Sigma}_r + \delta_m I_n)^{-1/2} = \left(\begin{array}{cc}
\frac{1}{\sqrt{1+\delta}} & \frac{1}{\sqrt{\delta}} \\
\frac{1}{\sqrt{\delta}} & \frac{1}{\sqrt{\delta}}
\end{array}\right) \left(\begin{array}{c}
\odot \\
\odot
\end{array}\right)
\]

= \left(\begin{array}{c}
\odot \\
\frac{1}{\sqrt{\delta}} \odot
\end{array}\right) \Rightarrow \left(\begin{array}{c}
\odot \\
\infty \\
\odot
\end{array}\right) as \delta \rightarrow 0 + .

Therefore \(\lambda_j^{(\delta)}\) becomes more unstable as \(\delta\) gets smaller.

Hence, \(\rho_j^{(\delta)}\) has some advantage over \(\lambda_j^{(\delta)}\). Also noted that in this example that

\[
\left(\hat{\Sigma}_r^{1/2}\right)^{-Y'PY} = \left(\begin{array}{c}
\odot \\
\odot
\end{array}\right) \left(\begin{array}{c}
\odot \\
\odot
\end{array}\right)
\]

= \left(\begin{array}{c}
\odot \\
\odot \\
\odot
\end{array}\right),

so using \(\hat{\Sigma}_r^{-}\) does not give good results either.

These results advocate the use of \(\hat{\Sigma}_Y^{-1}\) when \(n < m\) or \((\hat{\Sigma}_Y + \delta I_n)^{-1}\) in the high dimensional setting over using \(\hat{\Sigma}_r^{-}\) or \((\hat{\Sigma}_Y + \delta I_n)^{-1}\). While the proofs of the theoretical properties are obscure, simulations in the following chapter offer strong support for these ideas.
CHAPTER 4

SIMULATION EXPLORATION

Data was simulated here to compare and contrast WRSC and RSC when the error covariance took on a variety of forms. Two types of settings or “experiments” were examined: large sample and high dimension. These were designed to offer perspective in application and practicality of WRSC in computation. Details of the simulation setups are first presented, followed by a discussion of the results.

4.1 Experiment Setup

The various covariance structures included the i.i.d. structure, autoregressive 1 (AR(1)) structure, heterogeneous autoregressive 1 (H-AR(1)) structure, and an unstructured (UN) one. Table 4.1 contains the calculation for each (i,j) entry of each covariance structure. For the i.i.d. structure, \( \sigma^2 = 1 \) so that it was simply the identity matrix. The AR(1) structure had \( \rho = 0.9 \) and \( \sigma^2 = 3 \). The H-AR(1) structure randomly generated integers between 1 and 5 for the variances on the diagonal, used their square roots to find \( \sigma_i^{1/2} \) and \( \sigma_j^{1/2} \), and set \( \rho = 0.9 \). The unstructured covariance matrix is randomly generated.

<table>
<thead>
<tr>
<th>Name</th>
<th>(i,j)-th element</th>
</tr>
</thead>
<tbody>
<tr>
<td>i.i.d.</td>
<td>( \sigma^2 I_n )</td>
</tr>
<tr>
<td>AR(1)</td>
<td>( \sigma^2 \rho^{\mid i-j \mid} )</td>
</tr>
<tr>
<td>H-AR(1)</td>
<td>( \sigma_i \sigma_j \rho^{\mid i-j \mid} )</td>
</tr>
<tr>
<td>UN</td>
<td>( \sigma_{ij} )</td>
</tr>
</tbody>
</table>

Table 4.1: Covariance Error Structures

**Experiment 1: Large Sample Size** Generate \( i = 1, ..., m \) observations of \( x_i \sim N_p(0, \Sigma_{XX}) \) i.i.d. with \( (\Sigma_{XX})_{jk} = \rho_x^{\mid j-k \mid} \) for \( \rho_x > 0 \) and \( j, k = 1, ..., p \). The true (unknown) coefficient matrix is set to \( A = b \cdot B_0 B_1 \) where \( b > 0 \), \( B_0 \) is a \( p \times r \) matrix, and \( B_1 \) is a \( r \times n \) matrix so that \( \text{rank}(A) = r \). Both \( B_0 \) and \( B_1 \) have i.i.d. \( N(0,1) \) entries. The \( m \times n \) noise matrix \( E \) is composed of \( e_i \) observations that are also generated as independent \( N(0,1) \). Then observation \( y_i' \) of the \( m \times n \) \( Y \) matrix is calculated as \( y_i' = x_i' A = e_i' \).
The control parameters are initialized as \( m = 100, \ p = 25, \ n = 25, \) and \( r = 10 \) for combinations of \( \rho_x = 0.1, 0.5, 0.9 \) and \( b = 0.2, 0.4, 0.6, 0.8 \). The following two setups are used:

**Setup 1:** Assume that \( \Sigma_e \) is known. Utilize WRSC with the tuning parameter \( \mu_{adap1} \) and RSC with the tuning parameter \( \mu_{adap} \).

**Setup 2:** Use \( \Gamma = \hat{\Sigma}_{YY}^{-1} \). Utilize WRSC with tuning parameter \( \mu_{adap1} = \mu_{adap}^{(y)} \) and RSC with the tuning parameter \( \mu_{adap} \).

**Experiment 2: Large Dimensionality** The matrix \( X \) is generated as \( X_0 \Sigma_{XX}^{1/2} \) where \( (\Sigma_{XX})_{jk} = \rho_x^{\text{\{j}} - k]) \) for \( j, k = 1, \ldots, p \) and \( X_0 = X_1 X_2 \). Both \( X_1 \) and \( X_2 \) are simulated separately with independent \( N(0, 1) \) entries of sizes \( m \times q \) and \( q \times p \), respectively so that \( \text{rank}(X) = q \). The coefficient matrix is generated the same way as in in Experiment 1, and \( q \) is chosen so that \( q < m \). Again, this is done for combinations of \( \rho_x = 0.1, 0.5, 0.9 \) and \( b = 0.2, 0.4, 0.6, 0.8 \). The setups used here are:

**Setup 1:** Assume that \( \Sigma_e \) is known. Utilize WRSC with the tuning parameter \( \mu_{adap1} \) and RSC with the tuning parameter \( \mu_{adap} \). Here, the control parameters are initialized as \( m = 20, \ p = 100, \ n = 25, \ q = 10, \) and \( r = 5 \).

**Setup 2:** Use \( \Gamma = (\hat{\Sigma}_{YY})^{-1} \). Utilize WRSC with tuning parameter \( \mu_{adap1} = \mu_{adap}^{(y)} \) and RSC with the tuning parameter \( \mu_{adap} \). Here, the control parameters are initialized as \( m = 75, \ p = 200, \ n = 15, \ q = 10, \) and \( r = 5 \). This is so that \( n < m \) but \( p > m \).

**Setup 3:** Use \( \Gamma = (\hat{\Sigma}_{YY} + \delta I_n)^{-1} \) where \( \delta \) is tuned using 5-fold cross-validation. Utilize WRSC with tuning parameter \( \mu_{adap1} = \mu_{adap}^{(y)} \) and RSC with the tuning parameter \( \mu_{adap} \). Here, the control parameters are initialized as \( m = 20, \ p = 100, \ n = 25, \ q = 10, \) and \( r = 5 \) so that \( m < n < p \).

The control parameter \( b \) is used here to control the signal-to-noise ratio

\[
\text{SNR} = d_r^2(XA) / (\sqrt{q} + \sqrt{n})
\]

where \( d_r^2(XA) \) denotes the \( r \)-th largest singular value of \( XA \). Bunea et al. proved that recovery of the correct rank is only possible if the SNR is large enough [8]. The parameters \( b = 0.2, 0.4, 0.6, 0.8 \) correspond to moderate to high SNR values for low to moderate correlation between the predictors \( (\rho_x = 0.1, 0.5, 0.9) \).

### 4.2 Results

The results were evaluated at two levels: (1) correct rank recovery and (2) accuracy of the estimator. Rank recovery is measured as the percentage of times the correct rank is recovered out of 100 iterations. Note that this simulation has a “fixed X design” as \( X \) and
A are simulated once per b and ρx combinations for all 100 iterations. Therefore for each b and ρx combination, only E is re-generated for each 100 iterations (and Y is re-calculated). Accuracy of the coefficient estimator is measured by the scaled average mean squared error (MSE) where

\[ MSE(XA) = 100 \cdot \frac{\|XA - \hat{X}A\|^2}{mn}. \]

The most complex setup here is for a moderate signal strength \( b = 0.2 \) with high correlation within the predictor set \( \rho_x = 0.9 \). Results are expected to improve as the predictor correlation decreases and as the signal strength increases. See Tables B.1 and B.2 for rank recovery rates and average MSEs for Experiment 1, Setup 1. Tables B.3 and B.4 contain similar results for Experiment 1, Setup 2. The high dimensional results from Experiment 2, Setup 1 are given in Tables C.1 and C.2, followed by Tables C.3 and C.4 for Setup 2, and by Tables C.5 and C.6.

**Experiment 1: Large Sample** For Setup 1, rank recovery rates and accuracy measures were identical for RSC and WRSC. This was expected since WRSC treats \( \Gamma = \Sigma_e^{-1} \) as a known matrix and WRSC reduces to RSC when \( \Gamma = I_n \). But, for the non-i.i.d. covariances, WRSC outperforms RSC, particularly in the high signal strengths and the low predictor correlations, particularly noted for signal strengths of greater than \( b = 0.4 \) and a corresponding high predictor correlation \( \rho_x = 0.9 \). However, even moderate signal strengths (\( b = 0.2 \)) yielded good results for WRSC for predictor correlations of \( \rho_x = 0.5, 0.1 \). Differences in recovery rates were sometimes 100% better for WRSC than RSC. These results offer support of WRSC provided that \( \Sigma_e \) is known indicating the validity of WRSC’s basic theory.

Similar results were seen in the average MSEs for this setup as the average MSEs of WRSC and RSC were identical for i.i.d. errors but were much smaller for WRSC for the non-i.i.d. errors when rank recovery was high. This indicates that good recovery of the coefficient matrix (and thus, the canonical weights) is possible. There is a noted increase in the overall average MSEs as the i.i.d. structure had the lowest overall average MSEs (in the 15-18 range), followed by heterogeneous autoregressive 1 (in the 27-30 range), then autoregressive 1 (in the 32-35 range), and finally the unstructured covariance (in the 67-73 range). Notice that the overall ranges of the MSEs increase with the complexity of the covariance structure. Aside from this, there seems to be no obvious pattern between average MSEs and signal strength or predictor correlation.

**Setup 2** treated as \( \Sigma_e \) as an unknown covariance and replaces it with \( \Gamma = \hat{\Sigma}_{YY}^{-1} \). The corresponding adaptive tuning parameter \( \mu_y \) is also used here in WRSC while RSC remains the same as in Setup 1. Rank recovery rates for i.i.d. errors were better for RSC than WRSC. However, WRSC did competitively well provided that the signal strength was moderate to high (\( b = 0.2 \) to \( b = 0.8 \)) and the predictor correlation was weak to moderate (\( \rho_x = 0.1 \) to \( \rho_x = 0.5 \)). The worse cases for WRSC were for signal strength \( b = 0.2 \) or for high predictor correlation \( \rho_x = 0.9 \). The rank recovery rates for the non-i.i.d. covariance structures are consistently better for WRSC than RSC for most signal strengths and predictor correlations, particularly after the most
challenging setup of $b = 0.2$ and $\rho_x = 0.9$. While rank recovery rates for WRSC were not as strong as in Setup 1 overall, they were still very good for strong signals and low predictor correlation.

The respective ranges of the average MSEs for the various structures remain very similar to those in Setup 1. This validates that using $\Gamma = \hat{\Sigma}_{YY}^{-1}$ as the weight matrix provides good rank recovery and estimation of the coefficient matrix.

**Experiment 2: High Dimension** In Setup 1, when the covariance of the errors is treated as known, rank recovery from RSC was, as expected in the i.i.d. case, near perfect. Identical results were given by WRSC, too. However, rank recovery by RSC in non-i.i.d. settings was less than ideal with rank recovery rates in the 38% to 65% range. But, WRSC performed excellently across all types of covariance errors. Rates were almost 100% for all signal strengths and predictor correlations.

Average MSEs were identical in the i.i.d. covariance. But in the non-i.i.d. covariances, WRSC had noticeably smaller average MSEs than RSC. Again, as in Experiment 1, the average MSEs overall increase with the complexity of the covariance structure. Thus, rank recovery and good coefficient estimation is possible for WRSC. This offers promising support for WRSC application to the high dimensional setting.

**Setup 2** is designed so that $m > n$ but $m < p$. Because $m > n$, $\hat{\Sigma}_{YY}$ should still be a reasonable estimator of $\Sigma_{YY}$ and nonsingular. RSC has perfect rank recovery when errors are i.i.d., but WRSC does competitively well. As seen previously, when errors are generated with non-i.i.d. errors, WRSC does much better than RSC with rank recovery rates in the high 90% range. Rates for RSC vary in the non-i.i.d. settings in the 60% to 80% range (with the exception in AR(1) when $b = 0.2$ and $\rho_x = 0.9$).

The average MSEs for RSC when errors are generated i.i.d. are slightly better than those of WRSC, corresponding also to its slightly better rank recovery rates. In the other situations when errors are non-i.i.d., WRSC has lower average MSEs than RSC. Those that are generated from the unstructured covariance matrix are the largest for WRSC, in the mid-40s range while RSC are in the 70s range. The average MSEs that are from AR(1) errors are the second largest, followed by H-AR(1), still with WRSC outperforming RSC.

The simulations for this high dimension version of WRSC were performed as the given in Setup 3 with the weighting matrix now set to $\Gamma = (\hat{\Sigma}_{YY} + \hat{\delta}I_n)^{-1}$. In the i.i.d. setting, rank recovery rates of WRSC were competitive with RSC provided that the signal strength was strong enough. Interestingly, WRSC did very well when $b = 0.2$ (moderate signal strength) with a moderate predictor correlation of $\rho_x = 0.5$, but then did comparatively poorer for a low predictor correlation $\rho_x = 0.1$. For non-i.i.d. structures, WRSC performed well for a signal strength $b = 0.4$ and a low predictor correlation $\rho_x = 0.5$, rank recovery rates did moderately well. However, in general, WRSC out-
performs RSC in the non-i.i.d. setting with near perfect recovery rates for a strong signal $b = 0.4$. While RSC does overall better than WRSC in the i.i.d. setting, WRSC does competitively well in the noted signal and correlation strength.

The average MSEs were similar in the i.i.d. setting between RSC and WRSC, provided that WRSC performed as well as RSC. However, in the non-i.i.d. setting, WRSC had much lower average MSEs than RSC where rank recovery rates were better. That is, typically for predictor correlations $\rho_x = 0.9$ and $\rho_x = 0.1$. Again, interestingly, the moderate predictor correlation $\rho_x = 0.5$ gave higher average MSEs for some of the lower signals examined here. In general, if RSC had better rank recovery rates, the corresponding average MSEs were also poor. The overall average MSE errors were larger in this high dimensional setting, but again had similar patterns in the ranges as in Experiment 1. This shows that with the regularization parameter, coefficient recovery for WRSC is also better than RSC in the non-i.i.d. setting given a strong signal strength. However, the accuracy of the coefficient estimator does seem to be dependent upon the correct rank recovery.
CHAPTER 5

LOW DIMENSIONAL NEUROIMAGING
APPLICATION OF ACCA

CCA has shown to be a useful inferential tool for examining multiple relationships between two multivariate sets in a variety of fields. As an extension of such, the practical application of ACCA will be demonstrated here using a dataset compiled of clinical, neurocognitive, and neuroimaging measures. This dataset, recently collected by the Neurobehavioral Laboratory at the Miriam Hospital, was provided by Dr. Hernando Ombao of Brown University. It is a unique compilation of data as neurocognitive assessments have often been examined with clinical variables but the collection of neuroimaging data is still a fairly new discipline. In addition to this, these measures are taken from a relatively small cohort of HIV-positive patients. The application examined here extends the work of [9] where the same dataset was utilized in a penalized least squares regression application. ACCA methodology will be implemented along with the standard inferential techniques as described in Chapter 3 to offer a new perspective to the relationships between the two sets. A brief description of the data will be provided followed by the analysis techniques applied. Results will then be summarized and significant connections will be drawn between the two sets.

5.1 Data Description

Measurements from sixty-two HIV-positive participants were used to examine the relationship between brain volumetric measures and diffusion tensor imaging (DTI) derived measures with those of HIV associated neurocognitive deficits. Variables of the predictor set included clinical descriptors, as well as brain volumetric and DTI-derived measures, totaling 31 original predictor variables.

Automated brain segmentation was performed as described in [9] to produce measures of cortical grey matter, white matter, caudate, putamen, pallidum, thalamus, hippocampus, amygdala, and corpus callosum. The DTI derived measures give fractional anisotropy (FA) and mean diffusivity (MD) measurements. These were taken on the internal capsule of the brain on 5 consecutive axial slices on the white matter segmentation images creating four regions of interest with the first being the anterior-most and the last being the posterior-most sections of the brain. This was also done for the corpus callosum which was divided into three subregions: genu (anterior subregion), body (middle subregion), and splenium.
Anisotropy is defined as the property of having different values when measured in different directions. Fractional anisotropy is simply a scalar value that characterizes the amount of anisotropy of the diffusion in areas of the brain, reflecting fiber density, axonal diameter, and myelination in white matter [4]. Mean diffusivity measures the water diffusion or flow in the brain. In addition to these were clinical measures of HIV stage (early or advanced stages), Hepatitis C status, age, education, alcohol use, and cocaine/opiate use. Table 5.2 contains a complete list of all predictor variables.

The response set includes 13 variables collected from the following tests: WAIS-III Symbol Search, Digit Span, and Letter-Number Sequencing, Grooved Pegboard (Gpeg), Trail Making A, Trail Making B, Controlled Oral Word Association Test (COWAT), Hopkins Verbal Learning Test- Revised (HVLT), and Brief Visuospatial Memory Test- Revised. See Table 5.1 for a complete list of all response variables. These are widely used neuropsychological tests and make up the following response domains of interest: 1) verbal fluency, 2) psychomotor speed, 3) information processing speed and attention, 4) executive function, 5) learning, and 6) memory. Table 5.3 gives the response variables by each domain. The raw data collected from these assessments has been transformed so that all values are positive with smaller values associated with debilitated neurocognitive functions.

<table>
<thead>
<tr>
<th>Response variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HVLT sum</td>
<td>Total free recall for verbal learning test</td>
</tr>
<tr>
<td>HVLT delay</td>
<td>Delayed recall for verbal memory test</td>
</tr>
<tr>
<td>BVMT sum</td>
<td>Total free recall for non-verbal learning test</td>
</tr>
<tr>
<td>BVMT delay</td>
<td>Delayed recall for non-verbal memory test</td>
</tr>
<tr>
<td>WAIS LNS</td>
<td>Measure of attention and working memory with letter number sequencing</td>
</tr>
<tr>
<td>WAIS DigSym</td>
<td>Measures processing speed with digit symbol</td>
</tr>
<tr>
<td>WAIS SymSmSrch</td>
<td>Measures processing speed with symbol search</td>
</tr>
<tr>
<td>Gpeg dom</td>
<td>Measures motor speed of putting pegs into holes with dominant hand</td>
</tr>
<tr>
<td>Gpeg nondom</td>
<td>Measures motor speed of putting pegs into holes with non-dominant hand</td>
</tr>
<tr>
<td>Trail A</td>
<td>Measures motor speed through simple test of visuomotor sequencing</td>
</tr>
<tr>
<td>Trail B</td>
<td>Measures complex visuomotor sequencing</td>
</tr>
<tr>
<td>COWAT</td>
<td>Measures verbal fluency through word generations within one minute</td>
</tr>
<tr>
<td>Animal</td>
<td>Measures of verbal fluency</td>
</tr>
</tbody>
</table>

Table 5.1: Description of Neurocognitive Response Variables

## 5.2 Methodology

The primary goal here is to estimate \( \hat{\mathbf{t}} \) the number of significant relationships between the response variables from each respective set and the predictor variables and identify which variables contribute the most to these relationships. While there are many ways to examine and interpret the results from canonical analysis, a unique approach will be taken here to better understand the independent-dependent relationship between the predictor and response sets. This is done despite the ability of canonical analysis to examine the relationships from a symmetric point of view and to better understand the data from a modeling aspect.
### Table 5.2: Description of Predictor Variables

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV stage</td>
<td>Clinical</td>
<td>Binary measure based upon severity of immunosuppression</td>
</tr>
<tr>
<td>hcv current</td>
<td>Clinical</td>
<td>Binary measure Hepatitis C status</td>
</tr>
<tr>
<td>age</td>
<td>Clinical</td>
<td>Number of years</td>
</tr>
<tr>
<td>Education</td>
<td>Clinical</td>
<td>Number of years</td>
</tr>
<tr>
<td>kmsk alc</td>
<td>Clinical</td>
<td>Binary measure of alcohol use</td>
</tr>
<tr>
<td>kmsk cocopi</td>
<td>Clinical</td>
<td>Binary measure of cocaine and opiate use</td>
</tr>
<tr>
<td>fa cc genu</td>
<td>DTI derived</td>
<td>Fractional anisotropy in anterior subregion of corpus callosum</td>
</tr>
<tr>
<td>fa cc body</td>
<td>DTI derived</td>
<td>Fractional anisotropy in body subregion of corpus callosum</td>
</tr>
<tr>
<td>fa cc splenium</td>
<td>DTI derived</td>
<td>Fractional anisotropy in posterior subregion of corpus callosum</td>
</tr>
<tr>
<td>md cc genu</td>
<td>DTI derived</td>
<td>Mean diffusivity in anterior subregion of corpus callosum</td>
</tr>
<tr>
<td>md cc body</td>
<td>DTI derived</td>
<td>Mean diffusivity in body subregion of corpus callosum</td>
</tr>
<tr>
<td>md cc splenium</td>
<td>DTI derived</td>
<td>Mean diffusivity in posterior subregion of corpus callosum</td>
</tr>
<tr>
<td>fa ic1</td>
<td>DTI derived</td>
<td>Fractional anisotropy in 1st (anterior) subregion of internal capsule</td>
</tr>
<tr>
<td>fa ic2</td>
<td>DTI derived</td>
<td>Fractional anisotropy in 2nd subregion of internal capsule</td>
</tr>
<tr>
<td>fa ic3</td>
<td>DTI derived</td>
<td>Fractional anisotropy in 3rd subregion of internal capsule</td>
</tr>
<tr>
<td>fa ic4</td>
<td>DTI derived</td>
<td>Fractional anisotropy in 4th (posterior) subregion of internal capsule</td>
</tr>
<tr>
<td>md ic1</td>
<td>DTI derived</td>
<td>Mean diffusivity in 1st (anterior) subregion of internal capsule</td>
</tr>
<tr>
<td>md ic2</td>
<td>DTI derived</td>
<td>Mean diffusivity in 2nd subregion of internal capsule</td>
</tr>
<tr>
<td>md ic3</td>
<td>DTI derived</td>
<td>Mean diffusivity in 3rd subregion of internal capsule</td>
</tr>
<tr>
<td>md ic4</td>
<td>DTI derived</td>
<td>Mean diffusivity in 4th (posterior) subregion of internal capsule</td>
</tr>
<tr>
<td>whitematter</td>
<td>Brain Volumetric</td>
<td>Volume measurement</td>
</tr>
<tr>
<td>cortex</td>
<td>Brain Volumetric</td>
<td>Volume measurement</td>
</tr>
<tr>
<td>thalamus</td>
<td>Brain Volumetric</td>
<td>Volume measurement</td>
</tr>
<tr>
<td>caudate</td>
<td>Brain Volumetric</td>
<td>Volume measurement</td>
</tr>
<tr>
<td>putamen</td>
<td>Brain Volumetric</td>
<td>Volume measurement</td>
</tr>
<tr>
<td>pallidum</td>
<td>Brain Volumetric</td>
<td>Volume measurement</td>
</tr>
<tr>
<td>hippocampus</td>
<td>Brain Volumetric</td>
<td>Volume measurement</td>
</tr>
<tr>
<td>amygdala</td>
<td>Brain Volumetric</td>
<td>Volume measurement</td>
</tr>
<tr>
<td>cc splenium</td>
<td>Brain Volumetric</td>
<td>Corpus callosum posterior subregion</td>
</tr>
<tr>
<td>cc body</td>
<td>Brain Volumetric</td>
<td>Corpus callosum middle 3 subregions</td>
</tr>
<tr>
<td>cc genu</td>
<td>Brain Volumetric</td>
<td>Corpus callosum anterior subregion</td>
</tr>
</tbody>
</table>

At the variable level, only the predictor cross-loadings will be examined to quantify the relationship between each predictor variable and the response canonical variate as a whole. That is, for \( k = 1, \ldots, \hat{t} \) and \( j = 1, \ldots, p \), \( C_x(k, j) \) as given in Section 2.1.3 will be examined for each \( k \) response canonical variates with each \( j \) predictor variable. Only the canonical loadings of the response variables will be looked at to determine which response variables contribute the most to its respective variate. So, for \( k = 1, \ldots, \hat{t} \) and \( j = 1, \ldots, n \), \( L_y(k, j) \) will be examined as the relationship between the \( k \)-th canonical response variate with the \( j \)-th response variable. The canonical coefficients, predictor loadings, and response cross-loadings are also provided, but not examined here for reasons previously stated.

All variables will be first centered and scaled by their columns. This is because the canonical variates are considered to be “artificial.” By scaling and centering, the canonical coefficients may be examined as standardized variables [25]. Following this, the number of significant relationships will be determined adaptively using WRSC/ACCA methodology. This leads to the recovery of the canonical coefficients and canonical variates. Generally speaking, only the largest of the canonical loadings/cross-loadings will be examined as the
largest contributors to the variable-to-variate correlation. A larger canonical loadings/cross-loadings indicates that the particular variable is a larger contributor to its variate/opposite variate. The canonical loadings and cross-loadings may be understood as stated in previous: values closer to ±1 indicate stronger relationships while values close to 0 indicating a weaker relationship, positive values indicate as the variable increases, the variate as a whole also increases giving a direct relationship while negative values indicate that as the variable increases the variate as a whole decreases giving an inverse relationship.

First, the variables will be examined from a large sample/low dimension point of view with all original predictors and all response variables. Results from this will lead to the examination of subsets, grouping by various domains. Specifically, 1.) learning and memory domains, 2.) executive functioning and information processing speed/attention domains, and 3.) verbal fluency and pyshomotor speed. Then, the data will be examined from a high dimensional setting using the artificial data created as given above.

## 5.3 Results

The results are all tabulated in Appendix D.

### 5.3.1 Original Predictors, All Responses

Utilizing the original 31 predictor variables and all 13 response variables, the estimated number of significant canonical correlations was \( t = 4 \). Therefore, ACCA estimated that there are four significant, uncorrelated relationships between the two sets. The canonical correlations are:

\[
\hat{\rho}_1 = 0.799 \\
\hat{\rho}_2 = 0.754 \\
\hat{\rho}_3 = 0.739 \\
\hat{\rho}_4 = 0.719
\]

The normalized canonical coefficients, loadings, and cross-loadings are given in Table D.1 and D.2. The canonical response variates can be thought of as a compilation of all neuropsychological assessments as an overall index with larger values in the response set reflecting better cognitive performance.
**First Relationship**  The first relationship yields strong response loadings for HVLT sum, HVLT delay, and WAIS DigSym ($\geq 0.50$), followed closely by BVMT sum and BVMT delay (0.43 and 0.42). The HVLT and BVMT variables correspond to learning and memory domains while WAIS DigSym corresponds to the information processing and attention domain, indicating a certain amount of information processing speed/attention are required for learning and memory function.

The predictor cross-loadings that are the strongest are hcv current, Education, kmsk alc, kmsk cocopi, and md cc splenium. The binary variables corresponding to Hepatitis C status, alcohol use, and cocaine/opiate use all have negative cross-loadings, giving an inverse relationship with the response variate. Hinkin et al. and Ryan et al. had similar findings as they found that the co-infection of HIV and Hepatitis C were almost three times more likely to have impairments in learning and memory than those infected by HIV alone [20] [36]. Alcohol and drug use have also been long established to affect learning and memory functions. The subject’s education is the only positive cross-loading here, as those with greater amounts of education tend to have greater learning and memory function.

**Second Relationship**  In the second relationship, the strongest loadings in the response variate are COWAT and Animal, both positive and $\geq 0.50$. These correspond to the verbal fluency domain and influence the response variate the most in a positive, direct direction. Loadings in the 0.40-0.50 range were associated primarily with executive functioning and psychomotor speed.

The predictor cross-loadings in this relationship are greatest for age (0.47), followed by HIV stage and Education (both $\geq 0.25$) and then by fa cc body and pallidum. All of these predictor cross-loadings are positive in direction. The age of the subject is the strongest influencing factor in the response variate, followed by the amount of education the subject has. While these are logical influencing factors in verbal fluency, the positive cross-loading of the HIV stage is counter-intuitive. The variable pallidum corresponds to the size of the pallidum in the brain, which is typically associated with motor control [42]. Its positive sign is not surprising as diminishing brain volume is associated with increase in poor neurocognitive performance [30]. A moderate, direct relationship also is shown in the fractional anisotropy of the middle regions of the corpus callosum.

**Third Relationship**  The third relationship had negative predictor cross-loadings in the brain volumetric measures of whitematter, cortex, thalamus, putamen, and pallidum. Also, significant negative cross-loadings were fa ic1, fa ic2, and fa cc genu, all fractional anisotropy measures in the internal capsule and the corpus capsule. Here, Education also had a negative cross-loading where age, hcv current, kmsk cocopi, and md cc genu had positive cross-loadings.

While this may seem initially surprising, examination of the response loadings will show that most are negative, notably BVMT sum, BVMT delay, WAIS DigSym, WAIS SymSrch, Gpeg dom, Gpeg nondom, Tail B, and Animal. While these variables cover
a variety of domains, they are primarily associated with the information processing/attention and motor domains.

**Fourth Relationship** Examining the canonical loadings of the response variables for the fourth relationship, the strongest loadings correspond to variables WAIS DigSym, Gpeg dom, Gpeg nondom, and Trail A, all with positive loadings $\geq 0.50$. WAIS SymSrch has the next largest loading of 0.35. All of these variables correspond to the information and attention and motor domains. Thus, the largest contributors to this response variate are associated with information processing/attention, and psychomotor domains.

The predictor cross-loadings for the fourth relationship are the strongest for age, md cc genu, md cc body, md cc splenium, fa ic2, and md ic1 to md ic4. Both the age of the subject and the fractional anisotropy of the second region of the internal capsule are positive cross-loadings indicating a direct relationship with the response variate. All of the mean diffusivity measures cover the complete areas of the corpus callosum and the internal capsule and are negative cross-loadings. Thus, they maintain an inverse relationship with the response variate.

This corresponds with the findings by Wu et al. [43] that high mean diffusivity and low fractional anisotropy values in portions of the corpus callosum are correlated with defects in motor speed in HIV positive patients.

### 5.3.2 Learning and Memory Domains with Original Predictors

Next, a subset of the domains was analyzed using the 31 original predictors and all of the predictor variables associated with the learning and memory domains. This returned $\hat{t} = 1$ significant relationship with $\hat{\rho} = 0.753$. Learning and memory are two very closely related domains. In the previous analysis, it was seen that variables relating to these domains from the first relationship loaded significantly together. Hence, it is unsurprising to see all four response variables from these two domains with very large, positive loadings. Thus, as any of these response variables increase, the overall response variate increases. The fact that only one significant relationship was estimated indicates that these two domains are correlated with each other.

The strongest cross-loadings of the predictors were hcv current, Education, and kmsk cocopi. Of these, only the measurement of the subjects education had a positive relationship with the response variate. Both Hepatitis C and cocaine/opiate use had strong, negative relationships with the response variate. Co-infection of Hepatitis C and HIV has shown greater learning and memory impairment than just HIV infection alone [36] [20]. Drug use is well-known to adversely affect both learning and memory abilities [28]. Thus the strong, inverse relationship is expected.

Following these in strength are fa cc genu, thalamus, putamen, pallidum, and hippocampus, all positive in the 0.20 to 0.22 range. The latter four are brain volumetric measures. Again, decrease in brain volume is associated with impaired neurocognitive function so the positive relationship with learning and memory functions is expected [30].
5.3.3 Executive Functioning and Information Processing/Attention Domains with Original Predictors

The 6 response variables corresponding to either the executive functioning or the information processing/attention domains and the 31 original predictor variables returned $t = 2$ significant relationships with canonical correlations $\hat{\rho}_1 = 0.783$ and $\hat{\rho}_2 = 0.735$.

**First Relationship** The three response variables that load into the first relationship are WAIS DigSym, WAIS SymSrch, and Trail A, all of which are related to the information processing/attention domain. They are much stronger than the other three response variates and are all positive so that as each one of them increases, the response variate as a whole also increases.

The predictor variables that cross-load with the greatest strength in a negative direction are hcv current, kmsk alc, kmsk cocopi, md cc genu, md cc body, and md cc splenium. The inverse relationship between the information/attention domain and Hepatitis C status, alcohol use, and cocaine/opiate use indicate that as they increase, the response variate decreases as a whole. The remaining three variables are all associated with the mean diffusivity of the corpus callosum, also indicating that as the diffusivity in the corpus callosum increases, the response variate decreases.

The predictor variables that cross-load positively are fa cc splenium, whitematter, putamen, and hippocampus. The later three are all brain volumetric measures, indicating an increase in brain volume in these areas can be associated with the increase in the information/attention domain. The first of these is the fractional anisotropy in the posterior region corpus callosum.

**Second Relationship** Interestingly, the variables that most strongly load into this response variate are those that are associated with the executive functioning domain. These are Trail B, COWAT, and WAIS LNS, with loadings $\geq 0.60$.

The predictor variable that cross-loads with the most strength is age. This cross-loading is positive and far stronger (0.59) than the rest of the predictor variables. The next three predictor variables in the 0.20-0.30 range are hcv current, kmsk cocopi, and cc splenium. These are all positive, but surprising as having a positive Hepatitis C status and cocaine/opiate use would not intuitively thought to related to executive function in a positive manner. However, the gap between age and the next largest cross-loading is vastly different indicating that this is not as strong contributor.

It is noted here that these two domains can be considered to be uncorrelated as two relationships were established, with loadings of response variables associated with each domain.

5.3.4 Verbal Fluency and Motor Domains with Original Predictors

With the verbal fluency and motor domains combined, there were a total of 5 predictor variables that were analyzed with the 31 original predictor variables. Only $t = 1$ significant
The response loadings that were the strongest were COWAT, Animal, Gpeg dom, and Gpeg nondom. The first two of these correspond with verbal fluency while the latter two correspond with psychomotor abilities. The single relationship indicates a close correlated relationship between the two domains.

Of the predictor cross-loadings, age, Education, fa cc body, fa ic1, pallidum, and amygdala were the strongest, with a positive relationship with the response variate. Increase in age and the number of years of education that subject has had corresponds with greater verbal fluency and psychomotor abilities. The two brain volumetric indicate that larger volumes in these areas correspond also with an overall greater response variate. Also, the increase in fractional anisotropy in the middle three sections of the corpus callosum and the anterior most subregion of the internal capsule also corresponds with an overall greater response variate.

5.4 Summary

The application of ACCA has been shown here in the large sample setting to show how unique relationships may be identified. By using all response variables with all original predictors, four uncorrelated relationships were returned. These corresponded to 1.) learning and memory domains 2.) verbal fluency 3.) a subset of the response variables across all of the domains and 4.) information processing/attention and motor domains. The corresponding strongest cross-loadings of the predictor variables were unsurprising and documented by others.

Further subsetting steps were taken to examine these relationships closer into the sets of 1.) learning and memory domains 2.) executive functioning and information processing/attention domains and 3.) verbal fluency and motor domains. While the first and last of these estimated only one significant relationship apiece, the executive functioning and information/processing/attention domains estimated two significant relationships. Hence, it is appropriate to group learning and memory together and verbal fluency and motor skills together.

Generally speaking, support evidence has been given here to associate increase mean diffusivity, alcohol use, and cocaine/opiate use with decrease neurocognitive performance. Also that education, age, increased fractional anisotropy and brain volumetrics correspond with increased neurocognitive functioning.
Often times, particularly when the data is of high dimension, it is desirable to first select variables before performing any type of additional analysis. A multivariate response regression model may be sparse in the sense that not all variables may be needed in the model, that the model is row sparse. Methods that handle row sparsity are variable selection techniques. So one way to address the issues from a high dimensional dataset is through a combination of both variable selection and some other multivariate technique to offer a more complete methodology. If Adaptive Canonical Correlation Analysis (ACCA) is implemented in the second step, then since the predictor set is first reduced by variable selection, it may not be necessary to have any other additional high dimensional considerations for ACCA. This two step type of methodology is similar to that of Bunea et al. in [7] as models that are both row sparse and rank sparse (i.e. rank deficient) are explored. They address the rank spareness aspect through techniques like the Rank Selection Criterion, that produces low rank estimators. See [7] for a complete discussion of these models. While ACCA is not immediately a rank reduction technique, it is very closely related as discussed previously.

Thus, a two-step process of first of all, variable selection followed by ACCA, will be offered to manage high dimensional issues. Variable selection in multivariate response models is equivalent to group selection in the univariate response models. That is, if the $j$-th predictor variable is not in (2.4), this is equivalent with assuming that the $j$-th row in the coefficient matrix $A$ is zero. Since the rows of $A$ may be treated in groups of coefficients, then any group selection method for univariate response models may be employed to variable select in multivariate response models. This will be made clear in the following section. What is then a nice consequence is that results such consistency in the univariate setting carry over to the multivariate setting. A common group selection method for high dimensional models is Group Lasso (GLASSO) [44]. This will be introduced and then applied to a high dimensional dataset derived from the neuroimaging data from Chapter 5. Then ACCA will be utilized to identify and infer the canonical relationships in the predictors only selected by GLASSO.

### 6.1 Equivalence of Group Selection in Univariate Models

Variable selection in multivariate response regression model as given in (2.4) may be shown to be equivalent with group selection of variables in the univariate regression model.
If variables may be clustered together in predefined groups, then group selection of variables is simply selecting whole groups of variables to be included in the model.

To note this, first denote the predictor matrix as $\tilde{X}$ as the predictor matrix prior to variable selection. Recall the multivariate response model

$$\mathbf{Y} = \tilde{X}\mathbf{A} + \mathbf{E}$$

where the rows of $\tilde{X}$ and $\mathbf{Y}$ make-up the $i = 1, \ldots, m$ observations, each row of size $1 \times p$ and $1 \times n$, respectively. Continue to assume that $\tilde{X}$ and $\mathbf{Y}$ have been centered by their column means and that $\mathbf{E}$ is a random zero mean matrix. This may then be written in a vectorized version of the transposed model as

$$\text{vec}(\mathbf{Y}') = (\tilde{X} \otimes \mathbb{I}_n) \text{vec}(\mathbf{A}') + \text{vec}(\mathbf{E}')$$  \hspace{1cm} (6.1)$$

where $\mathbf{M} \otimes \mathbf{N}$ denotes the Kronecker product of two generic matrices $\mathbf{M}$ and $\mathbf{N}$. Hence, since $\tilde{X}$ is a $m \times p$ matrix and $\mathbb{I}_n$ is a $n \times n$ matrix, $\tilde{X} \otimes \mathbb{I}_n$ is a $mn \times pn$ size matrix with block elements of $(\tilde{X})_{ij}\mathbb{I}_n$ where $(\tilde{X})_{ij}$ denotes the $(i, j)$-th element of $\tilde{X}$.

Let $\mathbf{A}' = (\beta_1 \ldots \beta_p)$ where $\beta_j \in \mathbb{R}^n$ is the $j$-th row of $\mathbf{A}$. Define

$$\gamma := \text{vec}(\mathbf{A}')$$  \hspace{1cm} (6.2)$$

$$\mathbf{y} := \text{vec}(\mathbf{Y}')$$  \hspace{1cm} (6.3)$$

$$\mathbf{Z} := (\tilde{X} \otimes \mathbb{I}_n)$$  \hspace{1cm} (6.4)$$

$$\mathbf{e} := \text{vec}(\mathbf{E}')$$  \hspace{1cm} (6.5)$$

where $\gamma$ is a $pn \times 1$ vector and $\mathbf{y}$ and $\mathbf{e}$ are a $mn \times 1$ vectors. Thus, (6.1) may be written as

$$\mathbf{y} = \mathbf{Z}\gamma + \mathbf{e}$$  \hspace{1cm} (6.6)$$

where

$$\gamma = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} a_{11} \\ \vdots \\ a_{1n} \\ a_{21} \\ \vdots \\ a_{p1} \\ \vdots \\ a_{pm} \end{pmatrix}.$$  

Denote the columns of $\tilde{X}$ as $\tilde{X}^{(1)}, \tilde{X}^{(2)}, \ldots, \tilde{X}^{(p)}$, each a vector of size $m \times 1$. Then (6.6) can be re-written as

$$\mathbf{y} = \sum_{k=1}^{p} (\tilde{X}^{(k)} \otimes \mathbb{I}_n) \beta_k + \mathbf{e} := \sum_{k=1}^{p} \mathbf{Z}_k \beta_k + \mathbf{e}$$

where

$$\mathbf{Z}_k := \begin{pmatrix} x_{1,k} \mathbb{I}_n \\ x_{2,k} \mathbb{I}_n \\ \vdots \\ x_{m,k} \mathbb{I}_n \end{pmatrix}.$$
Consistent group selection holds now on \( Z \), but using properties of the Kroenecker product, reduce to conditions holding on \( \tilde{X} \). If \( \tilde{X}^{(j)} \) is not in the model, then the \( j \)-th row of \( A \) is identically zero which is just \( \beta_j \) in the vectorized model. By treating the \( \beta_j \)'s as groups of coefficients, any group selection method for univariate response regression models may now be employed.

6.2 Group Lasso

Group LASSO is one group selection technique for univariate models (and thus also a variable selection technique in the multivariate setting) that is popular for high dimensional data [44]. Continuing in the above notation, suppose there are \( p \) groups, all of the same size \( m \). The GLASSO minimizes the following convex criterion:

\[
\min_B \left\{ \frac{1}{2} \left\| \text{vec}(Y') - (\tilde{X} \otimes I_n)\text{vec}(B') \right\|_F^2 + \lambda \left\| B \right\|_{2,1} \right\} = \min_B \left\{ \frac{1}{2} \left\| Y - \tilde{X}B \right\|_F^2 + \lambda \left\| B \right\|_{2,1} \right\}
\]

(6.7)

where \( \tilde{X} \) denotes the predictor set prior to variable selection, \( \left\| B \right\|_{2,1} = \sum_{j=1}^p \left\| b_j \right\|_2 \), and \( \lambda \) is the tuning parameter.

To apply this, the rank \( k \) is first fixed to \( k = m \wedge p \wedge n \) where \( m \) is the sample size, \( p \) is the number of predictors in \( \tilde{X} \), and \( n \) is the number of response variables. While this is a convex problem, it may be computationally expensive to find a global minimum, particularly for a large dataset. This may be remedy by using thresholding operations instead such as soft-thresholding.

Let \( \Theta \) be the soft-thresholding operator and \( \overrightarrow{\Theta} \) be the multivariate version. Then, define

\[
T \circ B = \overrightarrow{\Theta} \left( \frac{1}{K} \tilde{X}'Y + (I - \frac{1}{K} \tilde{X}'\tilde{X})B; \frac{\lambda}{K} \right)
\]

(6.8)

for all \( B \) matrices of size \( p \times k \), where \( K \) is a constant satisfying \( K > \| \tilde{X} \| / 2 \). Refer to [37] and [7] for further details.

6.3 High Dimensional Neuroimaging Application

Consider, once again, the neuroimaging dataset from Chapter 5. Recall that the predictor set contains 31 original predictors of clinical, DTI-imaging, and brain volumetric measures. The response set contains 13 variables from neurocognitive assessments that looked at the following domains: 1) Verbal, 2) Motor, 3) Information Processing and Attention, 4) Executive, 5) Learning, and 6) Memory. These measurements were for a total of 62 HIV-positive patients.

To create a predictor set of higher dimension, quadratic DTI and quadratic brain volumetric measures were added, along with interaction variables between HIV stage, the DTI measures, and the brain volumetric measures. This resulted in a total of 235 predictors of which 31 were the original predictors. Hence, this new dataset has \( p = 235 \) predictors, \( n = 13 \) responses, and \( m = 62 \) observations.
The methodology is as stated previously in this chapter, with inferences made using the standard additional measures of traditional CCA. Group Lasso was used with five-fold cross-validation. Both $\tilde{X}$ and $Y$ are standardized prior to analysis where $\tilde{X}$ denotes the original predictor set as it is prior to the GLASSO step. After predictor variables are selected in the first step by GLASSO, then ACCA is applied to identify significant relationships. The additional descriptive measures that are examined here are the same as those in Chapter 5, for the same listed reasons. These include the canonical loadings of the response variate, and the canonical cross-loadings of the predictor set.

### 6.3.1 Results

The analysis is performed here with the high dimensional predictor set along with first of all, all response variables, then response variable associated with the learning and memory domains, followed by the response variables associated with the executive functioning and information processing/attention domains, and finally, the verbal and motor domains. Results are tabulated in Appendix E.

**All Response Variables** When all response variables were retained in the response set, only four predictors of the 235 were selected by GLASSO. ACCA was then performed and returned two canonical, uncorrelated relationships with canonical correlations of $\hat{\rho}_1 = 0.706$ and $\hat{\rho}_2 = 0.560$. The four variables that are selected by GLASSO are the predictor variables age, Education, $\text{kmsk cocopi}$, and $\text{fa ic2 quad}$. That is, the subject’s age, number of years of education, their past cocaine/opiate usage, and the quadratic term of the fractional anisotropy in the second subregion of the internal capsule, respectively.

**First Relationship** When the loadings of the response variate are examined in the first relationship, HVLT sum and delay and BVMT sum and delay are all negative in direction while the remaining nine are positive in direction. They are mid-range in strength in -0.20’s for the HVLT assessments and -0.40’s for the BVMT assessments. These variables with negative loadings correspond to the learning and memory domains. The strongest of the positive loadings are $\text{GPeg nondom}$, $\text{Trail B}$, $\text{Trail A}$ followed by $\text{COWAT}$ and $\text{GPeg dom}$, all in excess of 0.50.

Of the cross-loadings in the predictor set, the subject’s age and past cocaine/opiate usage are the strongest, both positive in direction indicating that as they increase, the response variate as a whole increases. The only cross-loading that is negative is number of years of education but is fairly weak with a cross-loading of $-0.21$.

**Second Relationship** In the second relationship, all of the response loadings are positive. The strongest ones here are $\text{WAIS SymSrch}$, $\text{BVMT sum}$, and $\text{BVMT delay}$, followed by $\text{WAIS DigSym}$. These were roughly in the 0.70’s range. The predictor set has only one negative cross-loadings which corresponds to the subject’s past cocaine/opiate use. While only -0.34, it is still the second strongest cross-loading in the predictor set. This inverse relationship with the response variate is unsurprising as it is known to negatively affect cognitive function. The other three response variables have positive cross-loadings, of which education is the strongest, indicating a direct relationship with the response variate.
Learning and Memory Domains With the learning and memory domains combined, GLASSO selected 9 predictor variables. After ACCA was performed, only one significant canonical relationship was returned with a canonical correlation of \( \hat{\rho}_1 = 0.730 \). Of the response loadings, the largest contributors are those from the BVMT assessment. These are all positive and fairly strong, all being \( \geq 0.55 \).

There were three cross-loadings in the predictor set that were negative, of which only hepatitis C status and past cocaine/opiate use were of notable strength. The inverse relationship with the learning and memory domains as a whole is noted again. The positive cross-loading of the education variable is also expected, as it is of positive strength as much as past cocaine/opiate use is of negative strength (0.50). There are two interaction terms that are of mid, positive strength that involve the brain volumetric of the putamen. While there is one interaction term that involves HIV stage, this is positive of relatively low strength (0.19).

Executive Functioning and Information Processing/Attention Domains These domains combined returned only three predictor variables after GLASSO: age, past cocaine/opiate use, and the quadratic mean diffusivity in the body of the corpus callosum. Interestingly, these three predictors returned two significant canonical relationships with canonical correlations of \( \hat{\rho}_1 = 0.699 \) and \( \hat{\rho}_2 = 0.531 \).

First Relationship The strongest response loadings were for COWAT, Trail B, and Trail A, and the other being much smaller. The three variables from the WAIS assessments are the smallest contributors to this response variate. In the predictor set, both age and past cocaine/opiate use are fairly strong and positive while the one quadratic term is negative but relatively small.

Second Relationship In the second relationship, all response variables had positive loadings but now with WAIS DigSym and WAIS SymSrch being the strongest. They are larger contributors in the second relationship than in the first with loadings in the high 0.80’s. Both of these are, in particular, in the information processing/attention domain. The two predictor variables with the strongest cross-loadings are both negative and give an interesting relationship with this response variate and past cocaine/opiate use and the quadratic term.

Verbal and Motor Domains The verbal and motor domains combined returned 7 predictors using GLASSO. Then, only one significant relationship was returned by ACCA with a correlation of \( \hat{\rho}_1 = 0.651 \). The response loadings are all fairly large and positive with COWAT being the weakest at 0.47, which is associated with the verbal domain. The cross-loadings of the predictors are all positive, aside from the quadratic mean diffusivity in the body of the corpus callosum. However, this is also the weakest of the cross-loadings. The stronger cross-loadings were in the 0.30’s range with age and three interaction terms involving the brain volumetric of the amygdala. This is followed in strength by two interaction variables involving HIV stage with two brain volumetric measures of, again, amygdala and whitematter.
6.3.2 Summary

To summarize these results, GLASSO significantly reduced the predictor space, making ACCA possible to apply without any further high dimensional considerations. When all response variables are retained, it is difficult to see and understand the relationships with the domains as they are combined together into the response variate. However, breaking down the response variables into domains that are grouped in pairs (as was done in Chapter 5) helps identify the more important predictor variables. Also is seen where there may exist more than one relationship within a pair of domains. In general, some of the more important predictors returned were age, education, and past cocaine/opiate usage. Interestingly, none of the lower order DTI and brain volumetric measures were ever retained by GLASSO. They were only included in ACCA as quadratic terms or interaction terms.
CHAPTER 7

CONCLUSION

What has been offered here is a new, data-adaptive method for canonical correlation analysis that estimates the number of significant canonical relationships while at the same time estimating the canonical weights themselves. The novelty in this approach is from a multivariate point of view in the Reduced-Rank Regression setting, rather than the traditional sequential way. While plenty of literature documents CCA from a population version or in a large sample setting, only Regularized CCA (RCCA) has ever been offered as solution for dealing with high dimensional issues. RCCA has computation limits as it is expensive to cross-validate across a 2-dimensional grid and completely neglects to consider the other canonical relationship aside from the strongest one. ACCA offers empirical support for alternative options to RCCA that do not have these issues.

This new Adapative Canonical Correlation Analysis (ACCA) is built upon the theoretical findings of Bunea et al. from their formulation of the Rank Selection Criterion (RSC). RSC estimates the rank in a penalized fashion through a tuning parameter with the nice addition of the close-form adaptive tuning parameter. This eliminates any need to perform any tuning. In order to draw the connection between the two, a weighted version of RSC has been developed here, the Weighted Rank Selection Criterion (WRSC), which requires a weight matrix and careful treatment of the tuning parameter. This was done in Chapter 3.

The choice of the weight matrix has been carefully examined from a large sample and high dimensional point of view. The theoretics of RSC carry over to WRSC only through a true decorrelator weight matrix. The obvious of choice is the sample residual covariance matrix but is only practical in the large sample setting. An alternative weight matrix is the sample response matrix which while lacking some of the theoretical support, has demonstrated empirically to recover rank and estimate the coefficient matrix well, hence estimating the number of significant canonical relationship and the canonical weight well also. This is arguably a more suitable choice of a weight matrix in the high dimensional setting as alternatives such as the pseudoinverse of the sample residual covariance or the regularized version of the sample residual covariance have been shown here to be, in computation, less feasible.

Another approach to the high dimensional setting in by adding a variable selection technique to reduce the dimension of the predictors prior to performing ACCA. Variable selection in multivariate response models is the same as group selection in univariate response regression models. With the dimension reduced using a technique such as Group
Lasso, ACCA may be then applied without further issue.

These concepts have been supported through simulation findings that indicate the applicability of ACCA in both the large sample and high dimensional setting. Furthermore, true application has been shown through a neuroimaging dataset from HIV-positive patients. Treatment of this dataset has been both through a low dimensional view using only the original predictors, and through a high dimensional view, using generated predictors. It is further analyzed with additional techniques from ACCA, broken down by neurocognitive domains of interest. In the high dimensional treatment, GLASSO is first applied to reduce the dimension before ACCA is used. This offers a complete picture of the theoretics, concepts, and ideas developed in this composition.
APPENDIX A

THEOREMS

Theorem A.0.1. Let \( S \in \mathbb{R}^{m \times n} \) of rank \( r(S) = m \). Then the minimum of the Euclidean norm

\[
\text{tr} \left[ (S - P)(S - P)' \right]
\]

over all matrices \( P \) of size \( m \times n \) with rank \( r(P) = r \leq m \) is achieved when \( P = MM' \) where \( M \in \mathbb{R}^{m \times r} \) and has columns of \( r \) eigenvectors corresponding to the first \( r \) largest eigenvalues of \( SS' \).

Theorem A.0.2 (Poincaré Separation Theorem). Let \( A \) be a \( m \times m \) matrix and let \( U \) be a \( m \times k \) matrix with \( k \leq m \) such that \( U'U = I_k \). Then,

\[
\tilde{\lambda}_j \leq \lambda_j
\]

where \( \tilde{\lambda}_j \) is the \( j \)-th largest eigenvalue of \( U'AU \) and \( \lambda_j \) is the \( j \)-th largest eigenvalue of \( A \).
APPENDIX B

LARGE SAMPLE SIMULATION RESULTS

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\rho_x$</th>
<th>IID</th>
<th>AR(1)</th>
<th>H-AR(1)</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RSC</td>
<td>WRSC</td>
<td>RSC</td>
<td>WRSC</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.25</td>
</tr>
<tr>
<td>0.5</td>
<td>0.62</td>
<td>0.06</td>
<td>0.07</td>
<td>1</td>
<td>0.49</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>0.56</td>
<td>0.76</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>0.08</td>
<td>0.35</td>
<td>1</td>
<td>0.48</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.58</td>
<td>0.61</td>
<td>1</td>
<td>0.81</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>0.48</td>
<td>0.7</td>
<td>1</td>
<td>0.35</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>0.99</td>
<td>0.35</td>
<td>1</td>
<td>0.48</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.58</td>
<td>0.61</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>0.48</td>
<td>0.7</td>
<td>1</td>
<td>0.35</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>1</td>
<td>0.69</td>
<td>1</td>
<td>0.61</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.69</td>
<td>0.59</td>
<td>1</td>
<td>0.57</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>0.44</td>
<td>1</td>
<td>0.57</td>
<td>1</td>
</tr>
</tbody>
</table>

Table B.1: **Experiment 1, Setup 1** - Rank recovery rates. $\Gamma = \Sigma_e^{-1}$ with $\mu_{adap}$.

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\rho_x$</th>
<th>IID</th>
<th>AR(1)</th>
<th>H-AR(1)</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RSC</td>
<td>WRSC</td>
<td>RSC</td>
<td>WRSC</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>17.2</td>
<td>17.2</td>
<td>72.0</td>
<td>33.4</td>
</tr>
<tr>
<td>0.5</td>
<td>17.2</td>
<td>17.2</td>
<td>70.7</td>
<td>34.0</td>
<td>61.1</td>
</tr>
<tr>
<td>0.1</td>
<td>16.2</td>
<td>16.2</td>
<td>66.7</td>
<td>32.6</td>
<td>58.6</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>16.7</td>
<td>16.7</td>
<td>70.1</td>
<td>34.5</td>
</tr>
<tr>
<td>0.5</td>
<td>16.1</td>
<td>16.1</td>
<td>65.1</td>
<td>34.3</td>
<td>54.7</td>
</tr>
<tr>
<td>0.1</td>
<td>15.8</td>
<td>15.8</td>
<td>61.6</td>
<td>33.3</td>
<td>52.3</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>16.3</td>
<td>16.3</td>
<td>65.0</td>
<td>33.2</td>
</tr>
<tr>
<td>0.5</td>
<td>16.0</td>
<td>16.0</td>
<td>59.7</td>
<td>32.5</td>
<td>52.8</td>
</tr>
<tr>
<td>0.1</td>
<td>15.9</td>
<td>15.9</td>
<td>62.6</td>
<td>33.3</td>
<td>50.8</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>16.3</td>
<td>16.3</td>
<td>67.8</td>
<td>33.3</td>
</tr>
<tr>
<td>0.5</td>
<td>15.9</td>
<td>15.9</td>
<td>61.3</td>
<td>32.2</td>
<td>52.6</td>
</tr>
<tr>
<td>0.1</td>
<td>15.9</td>
<td>15.9</td>
<td>60.9</td>
<td>32.4</td>
<td>50.4</td>
</tr>
</tbody>
</table>

Table B.2: **Experiment 1, Setup 1** - Average Mean Squared Error. $\Gamma = \Sigma_e^{-1}$ with $\mu_{adap}$. 54
### Table B.3: Experiment 1, Setup 2- Rank recovery rates. $\Gamma = \Sigma_{YY}^{-1}$ with $\mu_1^{(y)}$.

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\rho_x$</th>
<th>IID</th>
<th>AR(1)</th>
<th>H-AR(1)</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RSC</td>
<td>WRSC</td>
<td>RSC</td>
<td>WRSC</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.62</td>
<td>0.15</td>
<td>0.06</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>1</td>
<td>0.68</td>
<td>0.56</td>
<td>0.99</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>0.08</td>
<td>0</td>
<td>0.16</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.81</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>0.58</td>
<td>1</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>0.99</td>
<td>0.61</td>
<td>0.35</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.58</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>0.48</td>
<td>1</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>1</td>
<td>0.98</td>
<td>0.69</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.48</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>0.44</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table B.4: Experiment 1, Setup 2- Average Mean Square Error. $\Gamma = \Sigma_{YY}^{-1}$ with $\mu_1^{(y)}$.

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\rho_x$</th>
<th>IID</th>
<th>AR(1)</th>
<th>H-AR(1)</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RSC</td>
<td>WRSC</td>
<td>RSC</td>
<td>WRSC</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>17.2</td>
<td>21.4</td>
<td>72.0</td>
<td>33.9</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>17.2</td>
<td>19.8</td>
<td>70.7</td>
<td>35.1</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>16.2</td>
<td>19.1</td>
<td>66.7</td>
<td>33.6</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>16.7</td>
<td>18.8</td>
<td>70.1</td>
<td>35.2</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>16.1</td>
<td>17.6</td>
<td>65.1</td>
<td>35.3</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>15.8</td>
<td>17.4</td>
<td>61.6</td>
<td>34.3</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>16.3</td>
<td>19.4</td>
<td>65.0</td>
<td>34.1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>16.0</td>
<td>17.5</td>
<td>59.7</td>
<td>33.4</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>15.9</td>
<td>17.4</td>
<td>62.6</td>
<td>34.1</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>16.3</td>
<td>18.1</td>
<td>67.8</td>
<td>34.1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>15.9</td>
<td>17.4</td>
<td>61.3</td>
<td>32.9</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>15.9</td>
<td>17.4</td>
<td>60.9</td>
<td>33.3</td>
</tr>
</tbody>
</table>
APPENDIX C

HIGH DIMENSIONAL SIMULATION RESULTS

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\rho_x$</th>
<th>IID</th>
<th>AR(1)</th>
<th>H-AR(1)</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RSC</td>
<td>WRSC</td>
<td>RSC</td>
<td>WRSC</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>0.63</td>
<td>0.99</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>0.57</td>
<td>1</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>0.49</td>
<td>0.99</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>0.53</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>0.99</td>
<td>0.99</td>
<td>0.65</td>
<td>1</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>0.99</td>
<td>0.99</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>0.99</td>
<td>0.99</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>0.99</td>
<td>0.99</td>
<td>0.65</td>
<td>1</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>0.99</td>
<td>0.99</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

Table C.1: Experiment 2, Setup 1- Rank recovery rates. $\Gamma = \Sigma^{-1}$ with $\mu_{adap}$.

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\rho_x$</th>
<th>IID</th>
<th>AR(1)</th>
<th>H-AR(1)</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RSC</td>
<td>WRSC</td>
<td>RSC</td>
<td>WRSC</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>29.7</td>
<td>29.7</td>
<td>110.7</td>
<td>74.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>29.8</td>
<td>29.8</td>
<td>117.1</td>
<td>75.8</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>30.4</td>
<td>30.4</td>
<td>127.3</td>
<td>82.5</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>29.7</td>
<td>29.7</td>
<td>118.7</td>
<td>79.1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>30.0</td>
<td>30.0</td>
<td>107.7</td>
<td>76.7</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>30.2</td>
<td>30.2</td>
<td>121.2</td>
<td>78.1</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>29.8</td>
<td>29.8</td>
<td>121.1</td>
<td>77.3</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>30.0</td>
<td>30.0</td>
<td>113.4</td>
<td>75.5</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>30.0</td>
<td>30.0</td>
<td>124.1</td>
<td>81.0</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>29.9</td>
<td>29.9</td>
<td>116.0</td>
<td>77.3</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>30.0</td>
<td>30.0</td>
<td>112.8</td>
<td>75.3</td>
</tr>
<tr>
<td>0.1</td>
<td>29.4</td>
<td>29.4</td>
<td>111.9</td>
<td>79.1</td>
<td>92.0</td>
</tr>
</tbody>
</table>

Table C.2: Experiment 2, Setup 1- Average Mean Square Error. $\Gamma = \Sigma^{-1}$ with $\mu_{adap}$. 

56
<table>
<thead>
<tr>
<th>$b$</th>
<th>$\rho$</th>
<th>IID</th>
<th>RSC</th>
<th>WRSC</th>
<th>AR(1)</th>
<th>RSC</th>
<th>WRSC</th>
<th>H-AR(1)</th>
<th>RSC</th>
<th>WRSC</th>
<th>UN</th>
<th>RSC</th>
<th>WRSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>1</td>
<td>0.98</td>
<td>0.84</td>
<td>0.97</td>
<td>0.7</td>
<td>0.99</td>
<td>0.69</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>1</td>
<td>0.99</td>
<td>0.77</td>
<td>0.97</td>
<td>0.61</td>
<td>0.98</td>
<td>0.67</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>1</td>
<td>0.98</td>
<td>0.78</td>
<td>0.97</td>
<td>0.63</td>
<td>0.99</td>
<td>0.72</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>1</td>
<td>0.97</td>
<td>0.73</td>
<td>0.98</td>
<td>0.69</td>
<td>0.98</td>
<td>0.71</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>1</td>
<td>0.99</td>
<td>0.64</td>
<td>0.96</td>
<td>0.65</td>
<td>1</td>
<td>0.65</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>1</td>
<td>0.97</td>
<td>0.75</td>
<td>0.95</td>
<td>0.73</td>
<td>1</td>
<td>0.67</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>0.78</td>
<td>1</td>
<td>0.69</td>
<td>0.98</td>
<td>0.67</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>1</td>
<td>0.95</td>
<td>0.71</td>
<td>0.99</td>
<td>0.59</td>
<td>0.98</td>
<td>0.74</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>1</td>
<td>0.98</td>
<td>0.71</td>
<td>0.98</td>
<td>0.73</td>
<td>0.99</td>
<td>0.69</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>1</td>
<td>0.97</td>
<td>0.69</td>
<td>0.95</td>
<td>0.75</td>
<td>0.98</td>
<td>0.73</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>1</td>
<td>0.98</td>
<td>0.73</td>
<td>0.98</td>
<td>0.7</td>
<td>0.97</td>
<td>0.79</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>1</td>
<td>0.94</td>
<td>0.72</td>
<td>0.99</td>
<td>0.63</td>
<td>0.97</td>
<td>0.7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table C.3: **Experiment 2, Setup 2** - Rank recovery rates. $\Gamma = \hat{\Sigma}^{-1}_{YY}$ with $\mu_1^{(g)}$

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\rho$</th>
<th>IID</th>
<th>RSC</th>
<th>WRSC</th>
<th>AR(1)</th>
<th>RSC</th>
<th>WRSC</th>
<th>H-AR(1)</th>
<th>RSC</th>
<th>WRSC</th>
<th>UN</th>
<th>RSC</th>
<th>WRSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>9.0</td>
<td>9.5</td>
<td>28.7</td>
<td>20.7</td>
<td>30.1</td>
<td>20.3</td>
<td>69.9</td>
<td>47.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>8.7</td>
<td>9.1</td>
<td>30.3</td>
<td>20.5</td>
<td>29.1</td>
<td>17.8</td>
<td>68.3</td>
<td>44.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>9.0</td>
<td>9.5</td>
<td>30.1</td>
<td>21.6</td>
<td>29.6</td>
<td>18.7</td>
<td>64.9</td>
<td>44.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>9.0</td>
<td>9.4</td>
<td>30.6</td>
<td>21.1</td>
<td>28.0</td>
<td>19.1</td>
<td>66.3</td>
<td>45.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>8.7</td>
<td>9.1</td>
<td>33.9</td>
<td>21.7</td>
<td>28.6</td>
<td>18.6</td>
<td>68.8</td>
<td>45.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>8.7</td>
<td>9.2</td>
<td>31.7</td>
<td>22.4</td>
<td>27.3</td>
<td>18.1</td>
<td>68.4</td>
<td>46.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>8.9</td>
<td>9.4</td>
<td>30.8</td>
<td>21.8</td>
<td>27.4</td>
<td>18.1</td>
<td>70.6</td>
<td>47.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>8.9</td>
<td>9.4</td>
<td>31.4</td>
<td>21.3</td>
<td>31.4</td>
<td>19.7</td>
<td>68.0</td>
<td>45.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>9.0</td>
<td>9.5</td>
<td>31.1</td>
<td>21.0</td>
<td>28.3</td>
<td>19.3</td>
<td>66.6</td>
<td>45.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>8.9</td>
<td>9.5</td>
<td>32.9</td>
<td>22.2</td>
<td>26.4</td>
<td>18.7</td>
<td>61.9</td>
<td>42.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>9.0</td>
<td>9.5</td>
<td>30.4</td>
<td>20.3</td>
<td>27.3</td>
<td>18.5</td>
<td>62.7</td>
<td>45.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>9.1</td>
<td>9.6</td>
<td>31.9</td>
<td>22.3</td>
<td>30.2</td>
<td>19.5</td>
<td>66.0</td>
<td>44.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table C.4: **Experiment 2, Setup 2** - Average Mean Squared Error. $\Gamma = \hat{\Sigma}^{-1}_{YY}$ with $\mu_1^{(g)}$
Table C.5: **Experiment 2, Setup 3** - Rank recovery rates. \( \Gamma = (\hat{\Sigma}_{YY} + \delta I_n)^{-1} \) with \( \mu^{(y)}_1 \).

<table>
<thead>
<tr>
<th>( b )</th>
<th>( \rho_\sigma )</th>
<th>IID</th>
<th>AR(1)</th>
<th>H-AR(1)</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>1</td>
<td>0.87</td>
<td>0.63</td>
<td>0.66</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.57</td>
<td>1</td>
<td>0.59</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>0.88</td>
<td>0.49</td>
<td>0.51</td>
<td>0.52</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>0.53</td>
<td>0.76</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.65</td>
<td>0.68</td>
<td>0.61</td>
</tr>
<tr>
<td>0.1</td>
<td>0.99</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0.63</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>0.99</td>
<td>1</td>
<td>0.47</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.59</td>
<td>0.79</td>
<td>0.5</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>0.52</td>
<td>1</td>
<td>0.57</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>0.56</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.56</td>
<td>0.98</td>
<td>0.61</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>0.65</td>
<td>1</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Table C.6: **Experiment 2, Setup 3** - Average Mean Square Error. Unknown \( \Gamma = (\hat{\Sigma}_{YY} + \delta I_n)^{-1} \) with \( \mu^{(y)}_1 \).

<table>
<thead>
<tr>
<th>( b )</th>
<th>( \rho_\sigma )</th>
<th>IID</th>
<th>AR(1)</th>
<th>H-AR(1)</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>29.7</td>
<td>57.9</td>
<td>110.7</td>
<td>152.7</td>
</tr>
<tr>
<td>0.5</td>
<td>29.8</td>
<td>29.8</td>
<td>117.1</td>
<td>84.8</td>
<td>100.1</td>
</tr>
<tr>
<td>0.1</td>
<td>30.4</td>
<td>60.4</td>
<td>127.3</td>
<td>210.3</td>
<td>104.9</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9</td>
<td>29.7</td>
<td>29.7</td>
<td>118.7</td>
<td>236.8</td>
</tr>
<tr>
<td>0.5</td>
<td>30.0</td>
<td>30.0</td>
<td>107.7</td>
<td>262.7</td>
<td>99.9</td>
</tr>
<tr>
<td>0.1</td>
<td>30.2</td>
<td>30.1</td>
<td>121.2</td>
<td>87.2</td>
<td>96.1</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td>29.8</td>
<td>29.7</td>
<td>121.1</td>
<td>87.7</td>
</tr>
<tr>
<td>0.5</td>
<td>30.0</td>
<td>30.1</td>
<td>113.4</td>
<td>345.6</td>
<td>103.6</td>
</tr>
<tr>
<td>0.1</td>
<td>30.0</td>
<td>30.0</td>
<td>124.1</td>
<td>91.6</td>
<td>98.4</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>29.9</td>
<td>29.9</td>
<td>116.0</td>
<td>89.6</td>
</tr>
<tr>
<td>0.5</td>
<td>30.0</td>
<td>30.0</td>
<td>112.8</td>
<td>211.2</td>
<td>95.5</td>
</tr>
<tr>
<td>0.1</td>
<td>29.4</td>
<td>29.4</td>
<td>111.9</td>
<td>92.3</td>
<td>92.0</td>
</tr>
</tbody>
</table>
APPENDIX D

LARGE SAMPLE NEUROIMAGING DATA ANALYSIS RESULTS
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f_1$</td>
<td>$f_2$</td>
<td>$f_3$</td>
</tr>
<tr>
<td>HIV stage</td>
<td>0.10</td>
<td>0.22</td>
<td>-0.04</td>
</tr>
<tr>
<td>hcv current</td>
<td>-0.04</td>
<td>0.57</td>
<td>0.10</td>
</tr>
<tr>
<td>age</td>
<td>-0.12</td>
<td>0.07</td>
<td>-0.25</td>
</tr>
<tr>
<td>Education</td>
<td>0.06</td>
<td>0.66</td>
<td>-0.40</td>
</tr>
<tr>
<td>kmsk alc</td>
<td>-0.05</td>
<td>0.00</td>
<td>-0.08</td>
</tr>
<tr>
<td>kmsk cocopi</td>
<td>-0.33</td>
<td>0.31</td>
<td>0.09</td>
</tr>
<tr>
<td>fa cc genu</td>
<td>-0.04</td>
<td>-0.10</td>
<td>-0.16</td>
</tr>
<tr>
<td>fa cc body</td>
<td>-0.10</td>
<td>0.50</td>
<td>0.21</td>
</tr>
<tr>
<td>fa cc splenium</td>
<td>-0.04</td>
<td>-0.31</td>
<td>-0.26</td>
</tr>
<tr>
<td>md cc genu</td>
<td>0.48</td>
<td>-0.04</td>
<td>0.76</td>
</tr>
<tr>
<td>md cc body</td>
<td>-0.50</td>
<td>-0.07</td>
<td>-0.60</td>
</tr>
<tr>
<td>md cc splenium</td>
<td>-0.64</td>
<td>-0.14</td>
<td>-0.39</td>
</tr>
<tr>
<td>fa ic1</td>
<td>-0.13</td>
<td>0.40</td>
<td>-0.01</td>
</tr>
<tr>
<td>fa ic2</td>
<td>-0.22</td>
<td>0.03</td>
<td>-0.22</td>
</tr>
<tr>
<td>fa ic3</td>
<td>-0.16</td>
<td>0.08</td>
<td>0.19</td>
</tr>
<tr>
<td>fa ic4</td>
<td>0.55</td>
<td>0.33</td>
<td>-0.32</td>
</tr>
<tr>
<td>md ic1</td>
<td>-0.58</td>
<td>0.39</td>
<td>0.09</td>
</tr>
<tr>
<td>md ic2</td>
<td>0.55</td>
<td>-0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>md ic3</td>
<td>0.88</td>
<td>1.18</td>
<td>0.57</td>
</tr>
<tr>
<td>md ic4</td>
<td>-0.33</td>
<td>-1.01</td>
<td>-0.73</td>
</tr>
<tr>
<td>whitematter</td>
<td>0.36</td>
<td>0.01</td>
<td>0.18</td>
</tr>
<tr>
<td>cortex</td>
<td>-0.17</td>
<td>0.43</td>
<td>-0.10</td>
</tr>
<tr>
<td>thalamus</td>
<td>-0.66</td>
<td>-0.64</td>
<td>-0.12</td>
</tr>
<tr>
<td>caudate</td>
<td>-0.40</td>
<td>-0.40</td>
<td>0.14</td>
</tr>
<tr>
<td>putamen</td>
<td>0.33</td>
<td>-0.79</td>
<td>-0.57</td>
</tr>
<tr>
<td>pallidum</td>
<td>0.11</td>
<td>0.54</td>
<td>0.08</td>
</tr>
<tr>
<td>hippocampus</td>
<td>0.22</td>
<td>0.51</td>
<td>0.38</td>
</tr>
<tr>
<td>amygdala</td>
<td>0.47</td>
<td>0.21</td>
<td>-0.04</td>
</tr>
<tr>
<td>cc splenium</td>
<td>-0.13</td>
<td>-0.27</td>
<td>0.71</td>
</tr>
<tr>
<td>cc body</td>
<td>-0.30</td>
<td>-0.14</td>
<td>-0.27</td>
</tr>
<tr>
<td>cc genu</td>
<td>-0.14</td>
<td>-0.04</td>
<td>-0.61</td>
</tr>
</tbody>
</table>

Table D.1: Large Sample Neuroimaging ACCA with All Response Variables: Predictor Set
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h_1$</td>
<td>$h_2$</td>
<td>$h_3$</td>
</tr>
<tr>
<td>HVLT sum</td>
<td>0.51</td>
<td>0.24</td>
<td>0.33</td>
</tr>
<tr>
<td>HVLT delay</td>
<td>0.20</td>
<td>0.22</td>
<td>0.14</td>
</tr>
<tr>
<td>BVMT sum</td>
<td>0.12</td>
<td>0.00</td>
<td>-0.31</td>
</tr>
<tr>
<td>BVMT delay</td>
<td>0.00</td>
<td>0.03</td>
<td>-0.52</td>
</tr>
<tr>
<td>WAIS LNS</td>
<td>-0.01</td>
<td>0.29</td>
<td>0.18</td>
</tr>
<tr>
<td>WAIS LNS</td>
<td>0.39</td>
<td>-0.22</td>
<td>0.01</td>
</tr>
<tr>
<td>WAIS SymSrch</td>
<td>-0.06</td>
<td>-0.20</td>
<td>-0.31</td>
</tr>
<tr>
<td>GPeg dom</td>
<td>-0.03</td>
<td>0.37</td>
<td>-0.10</td>
</tr>
<tr>
<td>GPeg nondom</td>
<td>-0.15</td>
<td>0.06</td>
<td>-0.14</td>
</tr>
<tr>
<td>Trail A</td>
<td>0.37</td>
<td>-0.04</td>
<td>0.34</td>
</tr>
<tr>
<td>Trail B</td>
<td>-0.33</td>
<td>0.05</td>
<td>-0.13</td>
</tr>
<tr>
<td>COWAT</td>
<td>-0.29</td>
<td>0.43</td>
<td>0.00</td>
</tr>
<tr>
<td>Animal</td>
<td>0.15</td>
<td>0.38</td>
<td>-0.19</td>
</tr>
</tbody>
</table>

Table D.2: Large Sample Neuroimaging ACCA with All Response Variables: Response Set
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV stage</td>
<td>-0.01</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>hcv current</td>
<td>0.07</td>
<td>-0.44</td>
<td>-0.42</td>
</tr>
<tr>
<td>age</td>
<td>0.15</td>
<td>-0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>Education</td>
<td>0.24</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td>kmsk alc</td>
<td>0.07</td>
<td>-0.08</td>
<td>-0.08</td>
</tr>
<tr>
<td>kmsk cocopi</td>
<td>-0.48</td>
<td>-0.57</td>
<td>-0.49</td>
</tr>
<tr>
<td>fa cc genu</td>
<td>0.03</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>fa cc body</td>
<td>0.27</td>
<td>0.23</td>
<td>0.15</td>
</tr>
<tr>
<td>fa cc splenium</td>
<td>0.07</td>
<td>0.18</td>
<td>0.14</td>
</tr>
<tr>
<td>md cc genu</td>
<td>0.16</td>
<td>-0.10</td>
<td>-0.11</td>
</tr>
<tr>
<td>md cc body</td>
<td>-0.16</td>
<td>-0.11</td>
<td>-0.04</td>
</tr>
<tr>
<td>md cc splenium</td>
<td>-0.31</td>
<td>-0.20</td>
<td>-0.11</td>
</tr>
<tr>
<td>fa ic1</td>
<td>0.12</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>fa ic2</td>
<td>-0.33</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>fa ic3</td>
<td>-0.12</td>
<td>-0.16</td>
<td>-0.10</td>
</tr>
<tr>
<td>fa ic4</td>
<td>-0.15</td>
<td>-0.14</td>
<td>-0.06</td>
</tr>
<tr>
<td>md ic1</td>
<td>-0.05</td>
<td>-0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>md ic2</td>
<td>-0.03</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>md ic3</td>
<td>1.04</td>
<td>0.24</td>
<td>0.17</td>
</tr>
<tr>
<td>md ic4</td>
<td>-1.11</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>whitematter</td>
<td>0.26</td>
<td>0.30</td>
<td>0.17</td>
</tr>
<tr>
<td>cortex</td>
<td>-0.10</td>
<td>0.23</td>
<td>0.16</td>
</tr>
<tr>
<td>thalamus</td>
<td>-0.08</td>
<td>0.29</td>
<td>0.21</td>
</tr>
<tr>
<td>caudate</td>
<td>-0.75</td>
<td>-0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td>putamen</td>
<td>0.11</td>
<td>0.31</td>
<td>0.22</td>
</tr>
<tr>
<td>pallidum</td>
<td>0.41</td>
<td>0.31</td>
<td>0.21</td>
</tr>
<tr>
<td>hippocampus</td>
<td>0.18</td>
<td>0.27</td>
<td>0.20</td>
</tr>
<tr>
<td>amygdala</td>
<td>0.09</td>
<td>0.17</td>
<td>0.12</td>
</tr>
<tr>
<td>cc splenium</td>
<td>-0.11</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td>cc body</td>
<td>0.30</td>
<td>0.29</td>
<td>0.15</td>
</tr>
<tr>
<td>cc genu</td>
<td>-0.24</td>
<td>0.06</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Table D.3: Large Sample Neuroimaging ACCA with Learning and Memory Domain Variables: Predictor Set

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HVLT sum</td>
<td>0.28</td>
<td>0.64</td>
<td>0.50</td>
</tr>
<tr>
<td>BVMT sum</td>
<td>0.42</td>
<td>0.87</td>
<td>0.75</td>
</tr>
<tr>
<td>HVLT delay</td>
<td>0.21</td>
<td>0.69</td>
<td>0.38</td>
</tr>
<tr>
<td>BVMT delay</td>
<td>0.35</td>
<td>0.88</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Table D.4: Large Sample Neuroimaging ACCA with Learning and Memory Domain Variables: Response Set
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f_1$</td>
<td>$f_2$</td>
<td>$f_1$</td>
</tr>
<tr>
<td>HIV stage</td>
<td>-0.02</td>
<td>-0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>hcv current</td>
<td>-0.16</td>
<td>0.44</td>
<td>-0.33</td>
</tr>
<tr>
<td>age</td>
<td>0.74</td>
<td>0.28</td>
<td>-0.05</td>
</tr>
<tr>
<td>Education</td>
<td>-0.36</td>
<td>0.14</td>
<td>0.13</td>
</tr>
<tr>
<td>kmsk alc</td>
<td>-0.17</td>
<td>-0.05</td>
<td>-0.29</td>
</tr>
<tr>
<td>kmsk cocopi</td>
<td>-0.53</td>
<td>0.29</td>
<td>-0.31</td>
</tr>
<tr>
<td>fa cc genu</td>
<td>0.32</td>
<td>0.00</td>
<td>0.21</td>
</tr>
<tr>
<td>fa cc body</td>
<td>-0.76</td>
<td>0.19</td>
<td>0.13</td>
</tr>
<tr>
<td>fa cc splenium</td>
<td>0.44</td>
<td>-0.13</td>
<td>0.41</td>
</tr>
<tr>
<td>md cc genu</td>
<td>-0.02</td>
<td>-0.12</td>
<td>-0.22</td>
</tr>
<tr>
<td>md cc body</td>
<td>-1.09</td>
<td>-0.35</td>
<td>-0.41</td>
</tr>
<tr>
<td>md cc splenium</td>
<td>-0.07</td>
<td>0.11</td>
<td>-0.38</td>
</tr>
<tr>
<td>fa ic1</td>
<td>-0.16</td>
<td>0.20</td>
<td>0.16</td>
</tr>
<tr>
<td>fa ic2</td>
<td>0.18</td>
<td>0.43</td>
<td>0.06</td>
</tr>
<tr>
<td>fa ic3</td>
<td>-0.33</td>
<td>-0.35</td>
<td>0.03</td>
</tr>
<tr>
<td>fa ic4</td>
<td>0.29</td>
<td>0.19</td>
<td>0.12</td>
</tr>
<tr>
<td>md ic1</td>
<td>-0.38</td>
<td>0.31</td>
<td>-0.14</td>
</tr>
<tr>
<td>md ic2</td>
<td>0.23</td>
<td>-0.18</td>
<td>-0.06</td>
</tr>
<tr>
<td>md ic3</td>
<td>-0.21</td>
<td>-0.25</td>
<td>-0.11</td>
</tr>
<tr>
<td>md ic4</td>
<td>1.12</td>
<td>0.17</td>
<td>-0.12</td>
</tr>
<tr>
<td>whitematter</td>
<td>0.39</td>
<td>0.63</td>
<td>0.18</td>
</tr>
<tr>
<td>cortex</td>
<td>-0.37</td>
<td>-0.35</td>
<td>0.11</td>
</tr>
<tr>
<td>thalamus</td>
<td>0.64</td>
<td>0.21</td>
<td>0.10</td>
</tr>
<tr>
<td>caudate</td>
<td>0.31</td>
<td>0.20</td>
<td>0.11</td>
</tr>
<tr>
<td>putamen</td>
<td>0.75</td>
<td>-0.52</td>
<td>0.21</td>
</tr>
<tr>
<td>pallidum</td>
<td>-0.99</td>
<td>-0.54</td>
<td>-0.01</td>
</tr>
<tr>
<td>hippocampus</td>
<td>-0.17</td>
<td>0.46</td>
<td>0.14</td>
</tr>
<tr>
<td>amygdala</td>
<td>-0.20</td>
<td>-0.34</td>
<td>0.05</td>
</tr>
<tr>
<td>cc splenium</td>
<td>-0.26</td>
<td>-0.05</td>
<td>-0.01</td>
</tr>
<tr>
<td>cc body</td>
<td>-0.10</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td>cc genu</td>
<td>0.01</td>
<td>-0.21</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table D.5: Large Sample Neuroimaging ACCA with Executive and Information/Attention Domain Variables: Predictor Set
Table D.6: Large Sample Neuroimaging ACCA with Executive and Information/Attention Domain Variables: Response Set

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h_1$</td>
<td>$h_2$</td>
<td>$h_1$</td>
</tr>
<tr>
<td>Trail B</td>
<td>0.04</td>
<td>0.58</td>
<td>0.37</td>
</tr>
<tr>
<td>COWAT</td>
<td>-0.35</td>
<td>0.35</td>
<td>-0.09</td>
</tr>
<tr>
<td>WAIS LNS</td>
<td>0.01</td>
<td>0.30</td>
<td>0.14</td>
</tr>
<tr>
<td>WAIS DigSym</td>
<td>0.57</td>
<td>-0.21</td>
<td>0.87</td>
</tr>
<tr>
<td>WAIS SymSrch</td>
<td>0.38</td>
<td>-0.08</td>
<td>0.70</td>
</tr>
<tr>
<td>Trail A</td>
<td>0.31</td>
<td>0.31</td>
<td>0.61</td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>HIV stage</td>
<td>0.33</td>
<td>0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>hcv current</td>
<td>0.30</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>age</td>
<td>0.15</td>
<td>0.11</td>
<td>0.31</td>
</tr>
<tr>
<td>Education</td>
<td>0.53</td>
<td>0.28</td>
<td>0.25</td>
</tr>
<tr>
<td>kmsk alc</td>
<td>0.02</td>
<td>-0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>kmsk cocopi</td>
<td>0.06</td>
<td>-0.17</td>
<td>-0.02</td>
</tr>
<tr>
<td>fa cc genu</td>
<td>-0.05</td>
<td>0.12</td>
<td>0.15</td>
</tr>
<tr>
<td>fa cc body</td>
<td>0.36</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>fa cc splenium</td>
<td>-0.06</td>
<td>-0.13</td>
<td>-0.12</td>
</tr>
<tr>
<td>md cc genu</td>
<td>-0.29</td>
<td>-0.03</td>
<td>-0.10</td>
</tr>
<tr>
<td>md cc body</td>
<td>0.28</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>md cc splenium</td>
<td>0.20</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>fa ic1</td>
<td>0.39</td>
<td>0.14</td>
<td>0.24</td>
</tr>
<tr>
<td>fa ic2</td>
<td>-0.35</td>
<td>0.23</td>
<td>0.36</td>
</tr>
<tr>
<td>fa ic3</td>
<td>0.54</td>
<td>0.07</td>
<td>-0.00</td>
</tr>
<tr>
<td>fa ic4</td>
<td>-0.48</td>
<td>-0.05</td>
<td>-0.00</td>
</tr>
<tr>
<td>md ic1</td>
<td>0.53</td>
<td>0.02</td>
<td>-0.10</td>
</tr>
<tr>
<td>md ic2</td>
<td>-0.66</td>
<td>-0.09</td>
<td>-0.10</td>
</tr>
<tr>
<td>md ic3</td>
<td>1.28</td>
<td>0.19</td>
<td>0.09</td>
</tr>
<tr>
<td>md ic4</td>
<td>-1.27</td>
<td>0.08</td>
<td>-0.01</td>
</tr>
<tr>
<td>whitematter</td>
<td>-0.99</td>
<td>-0.07</td>
<td>0.13</td>
</tr>
<tr>
<td>cortex</td>
<td>0.84</td>
<td>0.12</td>
<td>0.14</td>
</tr>
<tr>
<td>thalamus</td>
<td>-0.25</td>
<td>-0.03</td>
<td>0.16</td>
</tr>
<tr>
<td>caudate</td>
<td>-0.27</td>
<td>0.03</td>
<td>0.09</td>
</tr>
<tr>
<td>putamen</td>
<td>-0.50</td>
<td>0.03</td>
<td>0.11</td>
</tr>
<tr>
<td>pallidum</td>
<td>0.90</td>
<td>0.14</td>
<td>0.28</td>
</tr>
<tr>
<td>hippocampus</td>
<td>-0.14</td>
<td>-0.03</td>
<td>0.16</td>
</tr>
<tr>
<td>amygdala</td>
<td>0.24</td>
<td>0.11</td>
<td>0.22</td>
</tr>
<tr>
<td>cc splenium</td>
<td>-0.35</td>
<td>-0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>cc body</td>
<td>0.16</td>
<td>-0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>cc genu</td>
<td>0.46</td>
<td>0.10</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Table D.7: Large Sample ACCA Neuroimaging ACCA with Verbal Fluency and Motor Domain Variables: Predictor Set
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>COWAT</td>
<td>0.63</td>
<td>0.72</td>
<td>0.34</td>
</tr>
<tr>
<td>Animal</td>
<td>0.48</td>
<td>0.72</td>
<td>0.26</td>
</tr>
<tr>
<td>Gpeg dom</td>
<td>0.60</td>
<td>0.44</td>
<td>0.33</td>
</tr>
<tr>
<td>Gpeg nondom</td>
<td>-0.09</td>
<td>0.43</td>
<td>-0.05</td>
</tr>
<tr>
<td>Trail A</td>
<td>-0.46</td>
<td>0.05</td>
<td>-0.25</td>
</tr>
</tbody>
</table>

Table D.8: Large Sample ACCA Neuroimaging ACCA with Verbal Fluency and Motor Domain Variables: Response Set
### APPENDIX E

## HIGH DIMENSIONAL NEUROIMAGING DATA ANALYSIS RESULTS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f_1$</td>
<td>$f_2$</td>
<td>$f_1$</td>
</tr>
<tr>
<td>age</td>
<td>0.58</td>
<td>0.23</td>
<td>0.68</td>
</tr>
<tr>
<td>Education</td>
<td>-0.17</td>
<td>0.51</td>
<td>-0.24</td>
</tr>
<tr>
<td>kmsk cocopi</td>
<td>0.56</td>
<td>-0.36</td>
<td>0.68</td>
</tr>
<tr>
<td>fa ic2 quad</td>
<td>0.51</td>
<td>0.58</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table E.1: ACCA: High Dimensional Predictor Set with All Response Variables After GLASSO and ACCA

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h_1$</td>
<td>$h_2$</td>
<td>$h_1$</td>
</tr>
<tr>
<td>HVLT sum</td>
<td>-0.09</td>
<td>0.09</td>
<td>-0.20</td>
</tr>
<tr>
<td>HVLT delay</td>
<td>-0.16</td>
<td>0.07</td>
<td>-0.29</td>
</tr>
<tr>
<td>BVMT sum</td>
<td>-0.22</td>
<td>0.23</td>
<td>-0.41</td>
</tr>
<tr>
<td>BVMT delay</td>
<td>-0.18</td>
<td>0.25</td>
<td>-0.43</td>
</tr>
<tr>
<td>WAIS LNS</td>
<td>0.05</td>
<td>0.12</td>
<td>0.15</td>
</tr>
<tr>
<td>WAIS DigSym</td>
<td>-0.09</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>WAIS SymSrch</td>
<td>-0.02</td>
<td>0.18</td>
<td>0.09</td>
</tr>
<tr>
<td>GPeq dom</td>
<td>0.16</td>
<td>0.12</td>
<td>0.51</td>
</tr>
<tr>
<td>GPeq nondom</td>
<td>0.24</td>
<td>0.10</td>
<td>0.68</td>
</tr>
<tr>
<td>Trail A</td>
<td>0.22</td>
<td>-0.02</td>
<td>0.61</td>
</tr>
<tr>
<td>Trail B</td>
<td>0.32</td>
<td>0.12</td>
<td>0.63</td>
</tr>
<tr>
<td>COWAT</td>
<td>0.23</td>
<td>0.08</td>
<td>0.51</td>
</tr>
<tr>
<td>Animal</td>
<td>0.18</td>
<td>0.18</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Table E.2: ACCA: High Dimensional Response Set with All Response Variables After GLASSO and ACCA
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>hcv current</td>
<td>0.03</td>
<td>-0.58</td>
<td>-0.44</td>
</tr>
<tr>
<td>Education</td>
<td>0.49</td>
<td>0.64</td>
<td>0.50</td>
</tr>
<tr>
<td>kmsk cocopi</td>
<td>-0.45</td>
<td>-0.69</td>
<td>-0.50</td>
</tr>
<tr>
<td>fa cc genu quad</td>
<td>0.22</td>
<td>0.40</td>
<td>0.24</td>
</tr>
<tr>
<td>fa ic2 quad</td>
<td>0.02</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td>HIV stage*cc body</td>
<td>0.14</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>fa cc genu*putamen</td>
<td>0.16</td>
<td>0.48</td>
<td>0.30</td>
</tr>
<tr>
<td>md cc splenium*caudate</td>
<td>-0.48</td>
<td>-0.11</td>
<td>-0.09</td>
</tr>
<tr>
<td>md ic3*putamen</td>
<td>0.30</td>
<td>0.48</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Table E.3: ACCA: High Dimensional Predictor Set with Learning and Memory Response Variables After GLASSO and ACCA

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HVLT sum</td>
<td>0.18</td>
<td>0.56</td>
<td>0.31</td>
</tr>
<tr>
<td>HVLT delay</td>
<td>0.20</td>
<td>0.64</td>
<td>0.34</td>
</tr>
<tr>
<td>BVMT sum</td>
<td>0.42</td>
<td>0.91</td>
<td>0.72</td>
</tr>
<tr>
<td>BVMT delay</td>
<td>0.42</td>
<td>0.92</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Table E.4: ACCA: High Dimensional Response Set with Learning and Memory Response Variables After GLASSO and ACCA

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>0.72</td>
<td>0.59</td>
<td>0.81</td>
</tr>
<tr>
<td>kmsk cocopi</td>
<td>0.51</td>
<td>-0.78</td>
<td>0.72</td>
</tr>
<tr>
<td>md cc body quad</td>
<td>-0.31</td>
<td>-0.66</td>
<td>-0.13</td>
</tr>
</tbody>
</table>

Table E.5: ACCA: High Dimensional Predictor Set with Executive and Info/Attention Response Variables After GLASSO and ACCA

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Trail B</td>
<td>0.57</td>
<td>0.04</td>
<td>0.69</td>
</tr>
<tr>
<td>COWAT</td>
<td>0.46</td>
<td>-0.09</td>
<td>0.70</td>
</tr>
<tr>
<td>WAIS LNS</td>
<td>0.06</td>
<td>0.23</td>
<td>0.28</td>
</tr>
<tr>
<td>WAIS DigSym</td>
<td>-0.38</td>
<td>0.54</td>
<td>0.00</td>
</tr>
<tr>
<td>WAIS SymSrch</td>
<td>-0.30</td>
<td>0.49</td>
<td>-0.01</td>
</tr>
<tr>
<td>Trail A</td>
<td>0.49</td>
<td>0.00</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table E.6: ACCA: High Dimensional Response Set with Executive and Info/Attention Response Variables After GLASSO and ACCA
### Table E.7: ACCA: High Dimensional Predictor Set with Verbal and Motor Response Variables After GLASSO and ACCA

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>0.60</td>
<td>0.54</td>
<td>0.37</td>
</tr>
<tr>
<td>md cc body quad</td>
<td>-0.28</td>
<td>-0.22</td>
<td>-0.19</td>
</tr>
<tr>
<td>fa ic2 quad</td>
<td>0.70</td>
<td>0.47</td>
<td>0.35</td>
</tr>
<tr>
<td>HIV stage*whitematter</td>
<td>-0.26</td>
<td>0.54</td>
<td>0.27</td>
</tr>
<tr>
<td>HIV stage*amygdala</td>
<td>0.69</td>
<td>0.56</td>
<td>0.26</td>
</tr>
<tr>
<td>fa cc genu*amygdala</td>
<td>0.67</td>
<td>0.46</td>
<td>0.31</td>
</tr>
<tr>
<td>fa ic2*amygdala</td>
<td>-0.53</td>
<td>0.51</td>
<td>0.34</td>
</tr>
</tbody>
</table>

### Table E.8: ACCA: High Dimensional Response Set with Verbal and Motor Response Variables After GLASSO and ACCA

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>COWAT</td>
<td>0.24</td>
<td>0.47</td>
<td>0.38</td>
</tr>
<tr>
<td>Animal</td>
<td>0.35</td>
<td>0.71</td>
<td>0.54</td>
</tr>
<tr>
<td>GPeG dom</td>
<td>0.30</td>
<td>0.75</td>
<td>0.47</td>
</tr>
<tr>
<td>GPeG nondom</td>
<td>0.26</td>
<td>0.81</td>
<td>0.40</td>
</tr>
<tr>
<td>Trail A</td>
<td>0.27</td>
<td>0.71</td>
<td>0.43</td>
</tr>
</tbody>
</table>


BIBLIOGRAPHY


BIOGRAPHICAL SKETCH

The author, born in Kwangju, South Korea, adopted by Beth and Jerome Geis, was raised in the frozen tundra of the Twin Cities of Minnesota. After dropping out of high school, she returned to higher education, after receiving her GED, to Augsburg College in Minneapolis. After receiving her B.A. in Mathematics and her B.S. in Actuarial Science, she was accepted to Northern Illinois University where she completed her M.S. in Applied Probability and Statistics. Following this, she moved to Tallahassee, Florida and was enrolled at Florida State University where she received a second M.S. in Biostatistics. She currently works to complete her Ph.D. in Biostatistics.