# New Estimation and Inferential Methods for Functional Connectivity Analysis

**A THESIS**
**SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL**
**OF THE UNIVERSITY OF MINNESOTA**
**BY**

**Andrew S. DiLernia**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS**
**FOR THE DEGREE OF**
**Doctor of Philosophy**

**Dr. Lin Zhang and Dr. Mark Fiecas, Advisors**

**June, 2021**

# Acknowledgements

It has been my pleasure to work with my advisors Dr. Lin Zhang and Dr. Mark Fiecas these past few years. I am truly grateful for their guidance and insight throughout my time at Minnesota. It was a privilege to have worked with Dr. Fiecas on multiple neuroimaging projects as he provided invaluable knowledge and expertise in our research endeavors. Having worked with Dr. Zhang every semester in graduate school, I have learned that she is an incredibly kind and knowledgeable person, and I could not have asked for a better advisor. I am also thankful to Dr. Julian Wolfson for the opportunity to work on several research projects as well, including my first project in graduate school. Not only was he a great research advisor, but he provided invaluable perspective in my job search as well.

Graduate school would not have been the same experience had it not been for my cohort of now great friends. The game nights throughout the years were some of the best times in graduate school, and I will cherish the memories and friendships made for the years to come.

I would also like to thank my cross country coach Dan Brunk who taught me to never give up in the face of adversity, and to appreciate life for the many special moments it provides. I am very thankful to my family, especially my parents for always being supportive of my endeavors and believing in me every step of the way. Their encouragement has been invaluable, and I am thankful to have such caring and loving parents.

I am also appreciative of my cat Lei. She kept me company when working late into the night on many occasions and taught me the importance of saving often. Most importantly, I want to thank my amazing significant other Lauren for her unconditional love and support. I am forever thankful to have such a wonderful partner in life.

# Dedication

To my significant other Lauren, who is my world.

## Abstract

Functional magnetic resonance imaging (fMRI) data is increasingly available and provides insight into the physiological mechanisms of the brain. As psychiatric disorders and many neurodegenerative diseases are intrinsically related to the brain, the availability of fMRI presents tremendous opportunities for improving understanding of these disorders and diseases. One approach for analyzing fMRI data is to describe functional connectivity (FC), the dependence of neuronal activity in regions of the brain. FC disruptions have been found in many mental disorders and diseases, so improving understanding of alterations in FC potentially underpinning mechanisms of these diseases is of clinical importance.

Several metrics are used to describe FC connections such as marginal correlations, partial correlations, mutual information, and coherence among others. In this dissertation, we propose novel methods for inference and estimation of partial correlations for FC analysis of multi-subject fMRI data. In our first project, we consider heterogeneity of FC patterns and aim to cluster multi-subject fMRI data based on each individual's FC patterns. We propose a novel penalized model-based clustering method which simultaneously estimates FC and clusters subjects into groups with similar FC patterns. The method estimates the precision matrix, the elements of which give partial correlations of all pairs of variables, at both the subject and cluster level for bi-level FC inference. We apply the method to a multi-subject fMRI data set collected on participants diagnosed with schizophrenia and healthy controls finding that participants with schizophrenia were more likely to be clustered into a group with reduced FC connections.

In our second project, we consider the issue of autocorrelation in fMRI data which is not accounted for in many existing methods when estimating and conducting inference of partial correlations. We derive an asymptotic joint distribution and novel covariance estimator for the partial correlations of a multivariate Gaussian process given mild regularity conditions. Based on the asymptotic distribution, we develop Wald confidence intervals and testing procedures for inference of individual partial correlations for inference of FC connections in single-subject fMRI data analysis. In our third project, we also use our theoretical result to propose a hierarchical model that directly accounts

for the autocorrelation in fMRI data and within group heterogeneity. We then develop a novel testing procedure for two-group comparisons of group-level FC in terms of the partial correlations which is robust to various levels of autocorrelation present in fMRI data.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Functional Neuroimaging and Mental Disorders

Over 50 million American adults were living with a mental illness such as an anxiety disorder, bipolar disorder, major depression, or other mental illness in 2019. Moreover, the percent of American adults diagnosed with some form of mental illness has increased from 17.7% of American adults in 2008 to 20.6% in 2019 (Substance Abuse and Mental Health Services Administration, 2020). Mental illness not only reduces an individuals' quality of life but presents additional costs and burdens for society as a whole (Kessler et al., 2009). Thus, improved understanding of the physiological mechanisms underlying these disorders is important for developing better treatments for individuals with these mental illnesses. As these disorders are intrinsically related to the brain, neuroimaging techniques for effectively describing the structure and neuronal activity patterns of the brain are vital for improving clinical practice moving forward.

Magnetic resonance imaging (MRI) consists of a broad area of non-invasive techniques implemented to describe brain function and structure, and it is one of the most widely used neuroimaging methods (Wager and Lindquist, 2015). MRI uses magnetic fields to produce three-dimensional images of organs such as the brain without using radiation (Logothetis, 2008). Structural MRI data involves capturing static images and is useful for examining gray and white matter of the brain, while functional magnetic resonance imaging (fMRI) commonly relies on observed blood oxygen level-dependent (BOLD) signals across time which serve as a surrogate measure for neuronal activity. As

a result, fMRI does not measure electrical currents as a surrogate measure for neuronal activity as other modalities do such as electroencephalography (EEG) and magnetoencephalography (MEG). However, fMRI can provide more granular measurements in terms of spatial resolution than EEG and MEG, and thus can facilitate more detailed descriptions of activity levels in various parts of the brain across time.

The BOLD signals underlying fMRI data are often collected at thousands of three-dimensional cubes called voxels across time as depicted in Figure 1.1, yielding high-dimensional data. Consequently, many statistical models cannot be easily used for analysis on this granular form of the data. However, dimension reduction prior to analysis can be done by averaging activation levels within sets of voxels called regions of interest (ROIs). Analyses can then be conducted on the ROI signals rather than the voxel-level signals, permitting use of a wider variety of statistical methods.

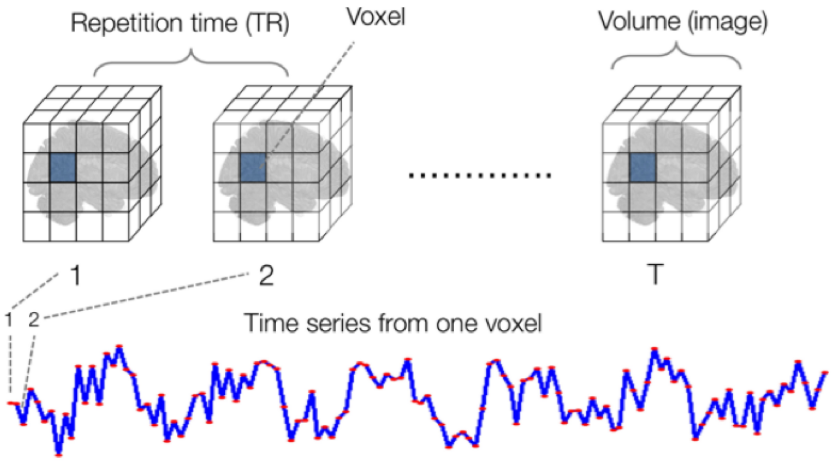


Figure 1.1: Depiction of fMRI data consisting of three-dimensional images, also called volumes, with a given repetition time or time resolution (TR) between each of the $T$ volumes. Image from Wager and Lindquist (2015).

Broadly, fMRI data collected on individuals in a research setting fall into two categories: resting-state fMRI and task-based fMRI. Resting-state fMRI data consist of fMRI data collected on individuals who are simply at rest and are not engaging in an assigned task or activity, whereas for task-based fMRI participants are engaged in an experimental condition. For resting-state fMRI, participants are commonly directed to

focus on a symbol like a cross in the scanner or to close their eyes for a duration of approximately 5 to 12 minutes. The resulting resting-state fMRI data that are collected are useful for exploring characteristics of cognitive performance, aging, psychopathology, and neurodegenerative diseases (Geuter et al., 2016).

## 1.2 Functional Connectivity Analysis of Multi-Subject FMRI Data

Examining the relatedness in activation levels in different parts of the brain across time, one approach for analyzing resting-state fMRI is to describe functional connectivity (FC), defined as the temporal dependence of neuronal activity in regions of the brain (Friston et al., 1993*b*). Alterations in FC have been associated with psychiatric disorders such as major depressive disorder (Zhu et al., 2012; Zeng et al., 2014) and schizophrenia (Yoon et al., 2008; Zhou et al., 2008; Pettersson-Yeo et al., 2011; Camchong et al., 2011; Fornito et al., 2012) as well as neurodegenerative diseases such as Alzheimer's (Dennis and Thompson, 2014). Thus, learning the disrupted patterns of FC networks for patients as compared to that of healthy controls is of great interest to underpin potential neurological mechanisms underlying these mental disorders and diseases.

Many methods have been developed for FC analysis of single and multi-subject fMRI data. For single-subject analyses, marginal correlations are commonly used to estimate FC connections between pairs of voxels or brain regions. However, other metrics can be used to describe FC connections as well such as the partial correlation, coherence, and mutual information among others (Bastos and Schoffelen, 2016). For resting-state fMRI data, Smith et al. (2011) found under various settings that the partial correlation tends to outperform other measures of association for detecting true FC connections. However, in higher dimensions partial correlations require significantly more computation than marginal correlations and implicitly require estimating more parameters than marginal correlations. For a thorough comparison of the utility of different metrics for FC analysis under various settings see Wang et al. (2014).

Joint FC network estimation can be conducted using independent component analysis (ICA), seed-based analyses, or graphical modeling approaches (Lee et al., 2013;

Smitha et al., 2017). For an ICA approach, the main idea is to estimate the independent source signals underlying a set of observed BOLD signals (Calhoun et al., 2009). Seed-based analyses commonly set certain ROIs as reference regions or seeds, and then estimate the relatedness of other ROIs to the seeds to produce a connectivity map of the brain (Li et al., 2009). In graphical modeling approaches, an undirected graph can be constructed with nodes commonly representing ROIs and edges connecting pairs of dependent nodes. Topological characteristics of the graphs can be used to characterize the FC networks, or one can analyze the strength of the connections between nodes as well. Yang et al. (2020) provides a broad overview of FC analysis methods for resting-state fMRI data.

With the availability of fMRI data from larger numbers of participants, multi-subject fMRI data analyses have become more popular as they facilitate the generalization of results to a broader population of individuals that eases clinically meaningful interpretations. The group ICA is one approach for estimating FC networks for multi-subject data without using brain regions that are defined a priori (Calhoun et al., 2001; Salaman et al., 2019). Another data-driven approach, independent vector analysis (IVA) is a generalization of ICA which jointly decomposes multi-subject data and better preserves individual-level heterogeneity than group ICA (Michael et al., 2014). Dictionary learning aims to find a parsimonious representation of the fMRI data using spatial information and has also been proposed for estimating both group and individual-level FC networks (Gong et al., 2018).

For multi-subject data sets in which brain regions are defined a priori, many have proposed methods for jointly estimating FC networks using the precision matrix (Guo et al., 2011; Zhu et al., 2014; Danaher et al., 2014; Cai et al., 2016; Qiu et al., 2016; Fan et al., 2018; Zhang et al., 2020). Although these methods provide sparse precision matrix estimates useful for FC inference of multi-subject fMRI data, they do not account for between-subject heterogeneity in FC structures that is commonly present (Price et al., 2017). Hierarchical models are a means of accounting for this within-group heterogeneity when estimating FC networks (Wager and Lindquist, 2015; Fiecas et al., 2017). Moreover, a common obstacle when analyzing resting-state fMRI data is the low signal-to-noise ratio. There are several sources of noise for fMRI data including heterogeneity in the magnetic field of the scanner itself and noise due to head motion or physiological

mechanisms of each individual. Some of these sources of noise can be removed through preprocessing prior to conducting an analysis, but it is difficult to remove their effects entirely (Lazar, 2008). In general, there are multiple sources of variability that should be accounted for when conducting FC analysis of multi-subject fMRI data: variability due to measurement error or sampling variability, and heterogeneity due to subject-specific differences in connectivity. It is important that one accounts for both sources of variation in the modeling process to be able to generalize the results to a larger population as is commonly the goal. For multi-subject fMRI data, hierarchical models can directly model these different sources of variability.

## 1.3 Dissertation Objectives and Overview

In this dissertation, we aim to provide methods for FC analysis of multi-subject resting-state fMRI data to improve understanding of the physiological mechanisms underlying psychiatric disorders, neurodegenerative diseases, and other aspects of the brain. Specifically, we provide the following:

1. A model-based clustering method for unsupervised learning of individuals in terms of FC based on resting-state fMRI data. This hierarchical model provides both individual and group-level estimates of FC in terms of the inverse covariance matrices, called precision matrices, while simultaneously grouping subjects into clusters with homogeneous FC patterns.

2. A derivation of the asymptotic distribution and covariance estimator for the empirical partial correlations of a multivariate Gaussian process. Using our derived asymptotic distribution, we propose methods of inference for the partial correlations of a multivariate Gaussian process which are useful for FC analysis of single-subject fMRI data.

3. A hierarchical model for group-level inference of FC as described by partial correlations. The model directly accounts for the autocorrelation present in fMRI data while allowing for within-group heterogeneity in terms of the true partial correlations for each subject. Based on the hierarchical model, we derive a novel

testing procedure for two-group comparisons of FC in terms of the group-level partial correlations.

The listed models and methods are provided in Chapters 2 through 4 in order. Next, we provide a summary of each of the three Chapters.

In Chapter 2, we present a hierarchical model for simultaneous clustering of subjects and estimation of FC using a multi-subject resting-state fMRI data set. This modeling framework clusters participants into homogeneous groups in terms of FC based on the precision matrix using a model-based clustering approach. Existing methods for clustering fMRI data at the subject-level mostly cluster subjects based on similar FC patterns and usually rely on prior subject-level FC estimates obtained from marginal correlations or covariance matrix estimates (Zeng et al., 2014). To the best of our knowledge, there are no existing methods that conduct clustering and FC estimation concurrently. Penalized Gaussian mixture models (GMM) have been proposed for simultaneous subject clustering and multiple network estimation for gene network analysis of microarray data (Zhou et al., 2009; Hill and Mukherjee, 2013; Gao et al., 2016; Hao et al., 2018). However, these models are only applicable when there is a single observation for each subject, and they often assume an identical covariance matrix for observations within a cluster. Hence, they cannot be directly used to cluster multi-subject fMRI data at the subject level as there are multiple observations for each subject. The novelty in our approach lies in allowing the data to drive the grouping of subjects based on FC, while pooling information across subjects within each group for improved estimates of FC. Through simulations, we demonstrate the utility of implementing both clustering and estimation concurrently when correlations between ROIs are smaller in magnitude. Additionally, we propose a modified stability approach to regularization selection which uses subsampling to select tuning parameters for our model as well as competing methods (Liu et al., 2010). Then, we propose use of a gap statistic for selecting the number of clusters or groups when the number of underlying groups is unknown (Tibshirani et al., 2001). Lastly, we demonstrate the utility of our proposed unsupervised method on a multi-subject fMRI data set consisting of participants diagnosed with chronic-episode schizophrenia, first-episode schizophrenia, and healthy controls, finding a tendency for those diagnosed with any degree of schizophrenia to be grouped together more so than healthy controls.

In Chapter 3, we derive the asymptotic distribution for the empirical partial correlations of a multivariate Gaussian process and derive an explicit form for the corresponding covariance matrix. Based on this derived asymptotic distribution, we propose inferential methods for partial correlations that directly account for autocorrelation present in the data. Partial correlations are one choice for describing FC, although there are alternatives such as the marginal correlation. However, some assert that partial correlations are more closely related to effective connectivity which describes the influence that ROIs exert on one another (Friston et al., 1993*a*; Marrelec et al., 2006). Although the asymptotic distribution of marginal correlations of a general multivariate time series have been derived (Roy, 1989), we addressed a similar issue for partial correlations to permit more valid inferential procedures when autocorrelation is present. We show the improvement in sensitivity and specificity for testing for non-null partial correlations compared to inference procedures which ignore autocorrelation in the data, and we show similar performance to a competitive inferential procedure based on the block-bootstrap which accounts for the temporal structure of the data in a nonparametric manner.

When groups of subjects are known a priori, Chapter 2 provides a straight-forward means of testing for group-level differences in FC for subjects with known group memberships. For example, a group-comparison could be conducted using a likelihood ratio test and the Wishart distribution. However, this testing procedure would not account for autocorrelation that is commonly exhibited in fMRI data. Addressing this need, in Chapter 4 we propose a hierarchical model for group-level inference of FC as described by partial correlations utilizing our derived asymptotic distribution from Chapter 3 to directly account for autocorrelation present in the data.

Our proposed framework in Chapter 4 bridges the ideas from both Chapters 2 and 3 since we are using a hierarchical model to account for both within-group heterogeneity and sampling variability in a similar manner as Chapter 2, while we also use the derived asymptotic distribution from Chapter 3 to explicitly account for autocorrelation in our modeling framework. Not only does our model allow for heterogeneity within each group in terms of the true partial correlations for each subject, but it also estimates the autocorrelation structure for each individual in a subject-specific manner. We display how inferential procedures for comparing subjects in terms of group-level FC based on our hierarchical model outperform competing methods in settings modeled after a real

fMRI data set. We then apply our proposed Wald testing procedure to a multi-subject resting-state fMRI data set collected on participants diagnosed with chronic episode schizophrenia and healthy controls.

# Chapter 2

# Penalized model-based clustering of fMRI data

## 2.1 Background

Describing FC among ROIs of the brain using fMRI data is of clinical importance, as it can be informative of neurological diseases and psychiatric disorders (Yoon et al., 2008; Zhou et al., 2008; Pettersson-Yeo et al., 2011; Camchong et al., 2011; Fornito et al., 2012; Dennis and Thompson, 2014). Methods for estimating sparse precision matrices have commonly been used to describe the FC structure among ROIs. For resting-state fMRI data from a single subject, graphical modeling approaches have commonly been used for joint analysis of FC connections among a set of ROIs (Lee et al., 2013; Smitha et al., 2017). Gaussian graphical models (GGM) assume a multivariate Gaussian distribution for the BOLD signals, under which the inverse covariance matrix, called the precision matrix, conveys conditional dependencies between ROIs (Lauritzen, 2004), and is used to describe FC through a graphical representation of the ROI dependence structure. Specifically, an undirected graph is constructed with nodes representing ROIs, and edges connecting pairs of conditionally dependent nodes. In this setting, methods for estimating sparse precision matrices have been used to obtain the FC structure among ROIs for a single subject. One commonly used method is the graphical lasso or GLasso (Friedman et al., 2007). For multi-subject fMRI data, there are often shared features of FC across subjects that can be accounted for to improve estimation.

Extending the GLasso, many have proposed methods for jointly estimating multiple sparse precision matrices (Guo et al., 2011; Zhu et al., 2014; Danaher et al., 2014; Cai et al., 2016; Qiu et al., 2016; Fan et al., 2018; Zhang et al., 2020), which can be used for inference of multiple subject-level FC networks for multi-subject fMRI data. However, these methods do not account for between-subject heterogeneity in FC structures that is commonly present (Fiecas et al., 2017; Price et al., 2017). This heterogeneity among subjects could be due to the presence of multiple subgroups or clusters of subjects with varying FC patterns (Mueller et al., 2013).

Although extensive work has been done on methods for classifying subjects based on resting-state and task-based fMRI signals (Davatzikos et al., 2005; Calhoun et al., 2008; Shen et al., 2010; Arribas et al., 2010; Castro et al., 2014), fewer studies have focused on unsupervised clustering of subjects based on FC patterns. Using an unsupervised maximum margin clustering method, Zeng et al. (2014) distinguished depressed patients from healthy controls based on FC estimates obtained from pairwise correlations. However, the maximum margin clustering approach does not yield interpretable estimates of FC network structures for each class or subject which are of clinical importance. Motivated by gene network analysis of microarray data, penalized Gaussian mixture models (GMM) have been proposed for simultaneous subject clustering and multiple network estimation (Zhou et al., 2009; Hill and Mukherjee, 2013; Gao et al., 2016; Hao et al., 2018). However, these methods apply to subject clustering with only one observation per subject as often seen in gene expression data and assume a common covariance matrix for observations within a cluster. Thus, they are not applicable to multi-subject fMRI data which have multiple observations per subject, and they do not allow for within-cluster heterogeneity.

In this chapter, we propose a penalized model-based clustering method for resting-state fMRI data which simultaneously estimates FC and groups patients based on FC features. Specifically, we propose a random covariance clustering method (RCCM) to simultaneously cluster subjects and obtain sparse precision matrix estimates for each cluster as well as each subject, producing interpretable estimates of subject- and group-level FC networks. A major contribution of the RCCM over existing penalized model-based clustering methods is to allow for clustering of entire subsets of fMRI data observed for each subject, rather than only individual observations, and that it allows subjects within

each cluster to have similar but not identical FC networks. This is implemented via a hierarchical structure, in which the subject-level precision matrices follow a mixture of Wishart distributions, each with mean matrix equal to a corresponding cluster-level precision matrix. Using a mixture of Wishart distributions for each subject's precision matrix rather than a Gaussian mixture for individual observations retains observations together for each participant. This is necessary for clustering participants based on fMRI data since it would not be interpretable to have observations from a single participant separated into multiple clusters. The degrees of freedom of each Wishart component controls the level of similarity between each subject-level matrix and its corresponding group-level matrix and is treated as a tuning parameter which is selected via an extended stability approach to regularization selection (stARS) method described in Section 2.2.3. Our simulations provide evidence that by conducting concurrent clustering and network estimation our proposed RCCM has improved performance in the estimation of subject-level networks due to sharing information across similar subjects, while better clustering is achieved due to improved subject-level estimates. We applied the RCCM to a resting-state fMRI data set collected on 61 participants diagnosed with schizophrenia and 43 healthy controls, finding a slight tendency for participants diagnosed with schizophrenia to be clustered together.

For the rest of the chapter, we present the proposed RCCM, the computational algorithm, and the selection of tuning parameters and the number of clusters in Section 2.2. Then, we describe simulations conducted to explore the relative performance of the RCCM to competitive two-step methods in Section 2.3 and illustrate the utility of the RCCM on a resting-state fMRI data set in Section 2.4. Lastly, in Section 2.5 we conclude the chapter with a discussion of our findings.

## 2.2 Method

### 2.2.1 Random Covariance Clustering Model

We consider the setting in which we have collected fMRI data on $K$ subjects for $p$ different ROIs with $n_k$ observations or time points on the $k^{th}$ subject. We let $y_{kjt}$ denote the $t^{th}$ observation or time point of the $j^{th}$ ROI for the $k^{th}$ subject for $k = 1, \ldots, K$, $j = 1, \ldots, p$, and $t = 1, \ldots, n_k$.

We assume that $\mathbf{y_{kt}} = (y_{k1t}, ..., y_{kpt})^T \sim \mathcal{N}_p(\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})$ are independent $p$-dimensional Gaussian random variables with mean vector $\boldsymbol{\mu_k}$ and covariance matrix $\boldsymbol{\Sigma_k}$. Moreover, we assume that the precision matrix of $\mathbf{y_{kt}}$, $\boldsymbol{\Omega_k} = \boldsymbol{\Sigma_k}^{-1}$, follows a mixture Wishart distribution with $G$ components:

$$\boldsymbol{\Omega_k} \sim p(\boldsymbol{\Omega_k}; \{\boldsymbol{\Omega_{0g}}, \pi_g\}_{g=1}^{G}) = \sum_{g=1}^{G} \pi_g p_g(\boldsymbol{\Omega_k}; \lambda_2, \boldsymbol{\Omega_{0g}}),$$

where $p_g(\boldsymbol{\Omega_k}; \lambda_2, \boldsymbol{\Omega_{0g}})$ is the probability density function (PDF) of the $g^{th}$ component corresponding to a Wishart random matrix with degrees of freedom $\lambda_2$ and mean $\boldsymbol{\Omega_{0g}}$. We note that $\boldsymbol{\Omega_k}$ describes the FC of the $k^{th}$ subject, $\boldsymbol{\Omega_{0g}}$ describes the cluster-level FC for the $g^{th}$ group, and that $\pi_g$ can be interpreted as the proportion of subjects belonging to cluster $g$ where $\sum_{g=1}^{G} \pi_g = 1$ for $g = 1, \ldots, G$. An essential element for the novelty of our model, this mixture Wishart distribution of the subject-level $\boldsymbol{\Omega}_k$ facilitates an interpretation of each subject's FC being similar to their corresponding cluster-level FC, but not necessarily identical. The degrees of freedom $\lambda_2$ is a tuning parameter controlling the degree of similarity between subject and cluster-level precision matrices, with higher values of $\lambda_2$ inducing more similarity between each subject-level matrix and its corresponding group-level matrix. The hierarchy of our proposed RCCM is illustrated in Figure 2.1, which illustrates the three-level structure of our model: the cluster level, within-cluster subject level, and within-subject observation level. Heterogeneity is assumed to be present in both the cluster- and subject-levels, while the observations for each subject are assumed to be homogeneous.

| Level | Model | Interpretation |
|---|---|---|
| Cluster | $\{\mathbf{\Omega_{0g}}\}_{g=1}^{G}$ | Set of $G$ cluster-level precision matrices describing group-level FC |
| | $\downarrow$ | |
| Subject | $\{\mathbf{\Omega_k} \overset{iid}{\sim} \sum_{g=1}^{G} \pi_g p_g(\mathbf{\Omega_k}; \lambda_2, \mathbf{\Omega_{0g}})\}_{k=1}^{K}$ | Set of $K$ subject-level precision matrices describing unique individual-level FC |
| | $\downarrow$ | |
| Observation | $\{\{\mathbf{y_{kt}}\|\mathbf{\Omega_k} \overset{ind}{\sim} \mathcal{N}_p(\boldsymbol{\mu_k}, \mathbf{\Omega_k^{-1}})\}_{t=1}^{n_k}\}_{k=1}^{K}$ | Set of fMRI signals consisting of $n_k$ observations of $p$ variables for subject $k$ where $p$ is the number of ROIs |

Figure 2.1: Hierarchy of Random Covariance Clustering Model (RCCM).

Assuming without loss of generality that our observed data is centered so that $\boldsymbol{\mu_k} = \mathbf{0}$ for $k = 1, \ldots, K$, the model likelihood for our observed data is

$$L = \prod_{k=1}^{K} \prod_{t=1}^{n_k} \left(f_k(\mathbf{y_{kt}}; \mathbf{\Omega_k})\right) p(\mathbf{\Omega_k}; \{\pi_g, \mathbf{\Omega_{0g}}\}_{g=1}^{G}),$$

where $f_k(\mathbf{y_{kt}}; \mathbf{\Omega_k}) = \frac{|\mathbf{\Omega_k}|^{1/2}}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}\mathbf{y_{kt}^T \Omega_k y_{kt}}\right)$ is the PDF of a mean $\mathbf{0}$ multivariate normal, $|\mathbf{\Omega}|$ denotes the determinant of a matrix $\mathbf{\Omega}$, and

$$p(\mathbf{\Omega_k}; \{\pi_g, \mathbf{\Omega_{0g}}\}_{g=1}^{G}) = \sum_{g=1}^{G} \pi_g p_g(\mathbf{\Omega_k}; \lambda_2, \mathbf{\Omega_{0g}}) = \sum_{g=1}^{G} \pi_g \left( \frac{|\mathbf{\Omega_k}|^{\frac{\lambda_2 - p - 1}{2}} \exp(-\mathrm{tr}(\lambda_2 \mathbf{\Omega_{0g}}^{-1} \mathbf{\Omega_k})/2)}{2^{\lambda_2 p/2} |\frac{1}{\lambda_2}\mathbf{\Omega_{0g}}|^{\lambda_2/2} \Gamma_p(\lambda_2/2)} \right)$$

is a Wishart mixture distribution with $G$ components where $\Gamma_p(\cdot)$ denotes the multivariate Gamma function. Hence, the corresponding model log-likelihood is

$$\ell = \log(L) = \sum_{k=1}^{K} \sum_{t=1}^{n_k} \log\left(f_k(\mathbf{y_{kt}}; \mathbf{\Omega_k})\right) + \sum_{k=1}^{K} \log p(\mathbf{\Omega_k}; \{\pi_g, \mathbf{\Omega_{0g}}\}_{g=1}^{G}),$$

and thus

$$-2\ell = \sum_{k=1}^{K}\sum_{t=1}^{n_k}\left(\mathbf{y_{kt}^T}\mathbf{\Omega_k}\mathbf{y_{kt}} - \log|\mathbf{\Omega_k}|\right) - 2\sum_{k=1}^{K}\log\left(\sum_{g=1}^{G}\pi_g p_g(\mathbf{\Omega_k};\lambda_2,\mathbf{\Omega_{0g}})\right).$$

To induce sparsity in our precision matrix estimates, we include lasso penalties on the subject- and cluster-level precision matrices with different regularization parameters. Thus, letting $\Theta = \{(\pi_g, \mathbf{\Omega}_k, \mathbf{\Omega}_{0g})\}$ for $g = 1,\ldots,G$ and $k = 1,\ldots,K$ denote the set of unknown parameters given the number of clusters or groups, $G$, we aim to estimate $\Theta$ by minimizing the following penalized objective function:

$$\begin{aligned}\sum_{k=1}^{K}\sum_{t=1}^{n_k}\left(\mathbf{y_{kt}^T}\mathbf{\Omega_k}\mathbf{y_{kt}} - \log|\mathbf{\Omega_k}|\right) - 2\sum_{k=1}^{K}\log\left(\sum_{g=1}^{G}\pi_g p_g(\mathbf{\Omega_k};\lambda_2,\mathbf{\Omega_{0g}})\right) + \lambda_1\sum_{k=1}^{K}||\mathbf{\Omega_k}||_1 \\ + \lambda_3\sum_{g=1}^{G}||\mathbf{\Omega_{0g}}||_1,\end{aligned} \tag{2.1}$$

where $||\mathbf{\Omega}||_1 = \sum_{i\neq j}|\omega_{i,j}|$ gives the sum of the absolute value of off-diagonal entries of the matrix $\mathbf{\Omega}$, and $\lambda_1$ and $\lambda_3$ are non-negative tuning parameters for the lasso penalties.

### 2.2.2 Computational Algorithm

#### E-Step

We seek to minimize the objective function in Equation (2.1) using the EM algorithm (Dempster et al., 1977). For the E-step, we first introduce latent variables which are indicators of cluster membership. Specifically, we let $z_{gk} = \mathbf{1}\{\mathbf{\Omega_k} \sim p_g(\mathbf{\Omega_k};\lambda_2,\mathbf{\Omega_{0g}})\} = \mathbf{1}\{\text{subject } k \text{ is from cluster } g\}$, and define

$$w_{gk} = \Pr(z_{gk}=1|\Theta) = \mathbb{E}[z_{gk}|\Theta] = \frac{\pi_g p_g(\mathbf{\Omega_k};\lambda_2,\mathbf{\Omega_{0g}})}{\sum_{c=1}^{G}\pi_c p_c(\mathbf{\Omega_k};\lambda_2,\mathbf{\Omega_{0c}})},$$

where $\Theta = \{\mathbf{\Omega_k}, \pi_g, \mathbf{\Omega_{0g}}\}$ for $g = 1,\ldots,G$ and $k = 1,\ldots,K$, and $\mathbf{1}\{\cdot\}$ is the indicator function. Thus, our complete objective function is

$$\sum_{k=1}^{K}\sum_{t=1}^{n_k}\left(\mathbf{y_{kt}^T\Omega_k y_{kt}} - \log|\mathbf{\Omega_k}|\right) - 2\sum_{k=1}^{K}\sum_{g=1}^{G} z_{gk}\left(\log(\pi_g) + \log(p_g(\mathbf{\Omega_k}; \lambda_2, \mathbf{\Omega_{0g}}))\right) + \lambda_1\sum_{k=1}^{K}||\mathbf{\Omega_k}||_1 + \lambda_3\sum_{g=1}^{G}||\mathbf{\Omega_{0g}}||_1,$$

since $\sum_{g=1}^{G} z_{gk} = 1$ for $k = 1, \ldots, K$ as each subject belongs to one and only one cluster. Therefore, since $\mathbb{E}_z[z_{gk}|\Theta^{(r)}] = w_{gk}^{(r)}$ it follows that the conditional expectation of our complete objective function is

$$Q(\Theta; \Theta^{(r)}) = \sum_{k=1}^{K}\sum_{t=1}^{n_k}\left(\mathbf{y_{kt}^T\Omega_k y_{kt}} - \log|\mathbf{\Omega_k}|\right) - 2\sum_{k=1}^{K}\sum_{g=1}^{G} w_{gk}^{(r)}\left(\log(\pi_g) + \log(p_g(\mathbf{\Omega_k}; \lambda_2, \mathbf{\Omega_{0g}}))\right) + \lambda_1\sum_{k=1}^{K}||\mathbf{\Omega_k}||_1 + \lambda_3\sum_{g=1}^{G}||\mathbf{\Omega_{0g}}||_1. \tag{2.2}$$

**M-Step**

We now seek to minimize the $Q(\Theta; \Theta^{(r)})$ function in Equation (2.2) by simultaneously estimating $G$ cluster-level matrices and $K$ subject-level precision matrices. We use an expectation/conditional maximization approach for optimization (Meng and Rubin, 1993). Specifically, we iteratively update two blocks: the set of subject-level precision matrices given by $\mathbf{\Omega} = \{\mathbf{\Omega_k}\}_{k=1}^{K}$ and the set of cluster-level precision matrices given by $\mathbf{\Omega_0} = \{\mathbf{\Omega_{0g}}\}_{g=1}^{G}$. We note that this objective function is non-convex, and due to the lasso penalties, we are no longer guaranteed to decrease our objective function at each iteration which could slow the method's convergence (Green, 1990). We proceed with the algorithm as follows:

1. Initialize $\mathbf{\Omega_k}^{(0)} = \hat{\mathbf{\Omega}}_{\mathbf{k_{gl}}}$ for $k = 1, \ldots, K$ where $\hat{\mathbf{\Omega}}_{\mathbf{k_{gl}}}$ is the individual GLasso precision matrix estimate for subject $k$ using a regularization parameter of 0.001.

2. Initialize $w_{gk}^{(0)}$ to be 0 or 1 based on a hard assignment of $\{\mathbf{\Omega_k}^{(0)}\}_{k=1}^{K}$ into $G$ disjoint clusters. Cluster assignments could be random, but we determined them using a

Ward hierarchical clustering method implemented via the *hclust* function from the *stats* package in R, using a matrix of the Frobenius norm of pair-wise differences as a distance matrix (Murtagh and Legendre, 2014; R Core Team, 2021).

3. Update each $\pi_g$ by calculating $\pi_g^{(r+1)} = \frac{1}{K}\sum_{k=1}^{K} w_{gk}^{(r)}$ for $g = 1, \ldots, G$.

4. Update each $\mathbf{\Omega_{0g}}$ by calculating

$$\mathbf{\Omega_{0g}}^{(r+1)} = \min_{\mathbf{\Omega_{0g}}}\{\text{tr}\left(\frac{\sum_{k=1}^{K} w_{gk}^{(r)}\mathbf{\Omega_k}^{(r)}}{\sum_{k=1}^{K} w_{gk}^{(r)}}\mathbf{\Omega_{0g}}^{-1}\right) + \log|\mathbf{\Omega_{0g}}| + \frac{\lambda_3}{\lambda_2\sum_{k=1}^{K} w_{gk}^{(r)}}||\mathbf{\Omega_{0g}}||_1\}$$

using Bien and Tibshirani's (2011) majorize-minimize algorithm for solving the covariance graphical lasso.

5. Update each $w_{gk}$ by calculating

$$w_{gk}^{(r+1)} = \frac{\pi_g^{(r+1)}\exp\left(-\frac{\lambda_2}{2}\text{tr}(\mathbf{\Omega_{0g}}^{(r+1)^{-1}}\mathbf{\Omega_k}^{(r)})\right)|\mathbf{\Omega_{0g}}^{(r+1)}|^{-\frac{\lambda_2}{2}}}{\sum_{c=1}^{G}\pi_c^{(r+1)}\exp\left(-\frac{\lambda_2}{2}\text{tr}(\mathbf{\Omega_{0c}}^{(r+1)^{-1}}\mathbf{\Omega_k}^{(r)})\right)|\mathbf{\Omega_{0c}}^{(r+1)}|^{-\frac{\lambda_2}{2}}},$$

for $k = 1, \ldots, K$ and $g = 1, \ldots, G$.

6. Update each $\mathbf{\Omega_k}$ by calculating

$$\mathbf{\Omega_k}^{(r+1)} = \min_{\mathbf{\Omega_k}}\{\text{tr}\left(\frac{n_k\mathbf{S_k} + \lambda_2\sum_{g=1}^{G} w_{gk}^{(r)}\mathbf{\Omega_{0g}}^{(r+1)^{-1}}}{n_k + \lambda_2 - p - 1}\mathbf{\Omega_k}\right) - \log|\mathbf{\Omega_k}| + \frac{\lambda_1}{n_k + \lambda_2 - p - 1}||\mathbf{\Omega_k}||_1\}$$

using the GLasso algorithm of Friedman et al. (2007), where $\mathbf{S_k} = \frac{1}{n_k}\sum_{t=1}^{n_k}\mathbf{y_{kt}}\mathbf{y_{kt}^T}$ for $k = 1, \ldots, K$.

7. Update each $w_{gk}$ by calculating

$$w_{gk}^{(r+1)} = \frac{\pi_g^{(r+1)} \exp\left(-\frac{\lambda_2}{2}\text{tr}(\mathbf{\Omega_{0g}}^{(r+1)^{-1}}\mathbf{\Omega_k}^{(r+1)})\right) |\frac{1}{\lambda_2}\mathbf{\Omega_{0g}}^{(r+1)}|^{-\frac{\lambda_2}{2}}}{\sum_{c=1}^{G} \pi_c^{(r+1)} \exp\left(-\frac{\lambda_2}{2}\text{tr}(\mathbf{\Omega_{0c}}^{(r+1)^{-1}}\mathbf{\Omega_k}^{(r+1)})\right) |\frac{1}{\lambda_2}\mathbf{\Omega_{0c}}^{(r+1)}|^{-\frac{\lambda_2}{2}}},$$

for $k = 1, \ldots, K$ and $g = 1, \ldots, G$.

8. Repeat steps 3 through 7 until convergence as determined by

$$\max\{\{|\mathbf{\Omega}_{k_{(i,j)}}^{(r+1)} - \mathbf{\Omega}_{k_{(i,j)}}^{(r)}|\} \cup \{|\mathbf{\Omega}_{g_{(i,j)}}^{(r+1)} - \mathbf{\Omega}_{g_{(i,j)}}^{(r)}|\}\} < \varepsilon,$$

where $\varepsilon > 0$ is a small number. That is, convergence is achieved when the largest change in magnitude of all individual entries of both the subject- and group-level estimates is smaller than $\varepsilon$, where we used $\varepsilon = 0.001$. Derivations of updates are included in Appendix A.

### 2.2.3 Selection of Tuning Parameters and Number of Clusters

There are three tuning parameters for the proposed RCCM denoted $\lambda_1$, $\lambda_2$, and $\lambda_3$ which control the sparsity of subject-level precision matrices, the within-cluster variability, and the sparsity of cluster-level precision matrices, respectively. Methods such as cross-validation (CV) and information criterion such as AIC and BIC can be used to determine the values of these tuning parameters, but we found for our real data analysis that CV yielded network estimates that were too dense with nearly all nodes connected to one another, while AIC yielded network estimates that were too sparse. Instead, we propose a modified stability approach for regularization selection (stARS) method (Liu et al., 2010). The stARS method uses a sub-sampling approach to select the tuning parameter that yields the least amount of regularization while still obtaining an estimate that is sparse and stable across subsamples. Although Liu et al. (2010) implemented stARS for a single-subject graphical lasso problem, we extend the approach to our RCCM method in the context of analyzing data from multiple subjects. Specifically, given a sample of $n_k$ observations for $k = 1, \ldots, K$, our extended stARS method consists of the following steps:

1. For $k = 1, \ldots, K$, draw $N$ distinct number of subsamples without replacement from the $n_k$ total observations denoted $S_{k1}, \ldots, S_{kN}$, each of size $b(n_k) = \lfloor 10\sqrt{n_k} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function.

2. For each candidate tuning parameter value, implement the desired method to obtain subject-level precision matrix estimates for each subsample yielding $N$ estimated edge sets for each subject denoted by $\hat{E}_{k1}(\Lambda), \ldots, \hat{E}_{kN}(\Lambda)$, where $\Lambda = (\lambda_1, \lambda_2, \lambda_3)$.

3. Calculate
$$\hat{\theta}_{k;st}(\Lambda) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}\{\hat{E}_{kj;st}\}$$
where $\mathbf{1}\{\hat{E}_{kj;st}\} = 1$ if the $j^{th}$ subsample implies an edge between nodes $s$ and $t$ for the $k^{th}$ subject and 0 otherwise. Thus, $\hat{\theta}_{k;st}(\Lambda)$ is the proportion of subsamples with an edge between $s$ and $t$ for the $k^{th}$ subject.

4. Calculate $\hat{\xi}_{k;st}(\Lambda) = 2\hat{\theta}_{k;st}(\Lambda)\left(1 - \hat{\theta}_{k;st}(\Lambda)\right)$, which can be interpreted as the proportion of times that each pair of subsamples disagrees on the presence of an edge between nodes $s$ and $t$ for the $k^{th}$ subject.

5. Calculate $\hat{D}(\Lambda) = \frac{1}{K} \sum_{k=1}^{K} \sum_{s<t} \hat{\xi}_{k;st}(\Lambda) / \binom{p}{2}$ which is a measure of instability, averaged across all subjects.

6. Calculate $\bar{D}(\Lambda) = \min_{\Lambda} \{\Lambda : \hat{D}(\Lambda) \leq \beta\}$ where we used $\beta = 0.05$.

7. Select $\Lambda$ that yields the least amount of sparsity among all candidate regularization parameter sets with instability limited by $\beta$.

For selecting the number of clusters, we used a gap statistic as proposed by Tibshirani et al. (2001). Generally, the gap statistic measures the within-cluster dispersion of a clustering of subjects given a certain number of clusters. The optimal number of clusters is then chosen as the smallest number that does not result in a significant increase in the gap statistic. A detailed description of calculating the gap statistic for our proposed RCCM is included in Appendix A.

## 2.3 Simulations

### 2.3.1 Simulation Settings

We conducted extensive simulations to examine the performance of our proposed RCCM, considering two different numbers of clusters, two different levels of magnitude of the precision matrix entries, and three different levels of similarity between the clusters. Specifically, data were generated with either $G = 2$ clusters containing 67 and 37 subjects in each group, or $G = 3$ clusters with 61, 24, and 19 subjects in each group which reflect the real data clustering results to be presented in Section 4. These data were generated for either a high or low magnitude setting, where the high magnitude setting corresponded to the off-diagonal entries of the true precision matrix having roughly three times the magnitude of the low magnitude setting entries on average. This introduced more distinction in precision matrices between clusters compared to the low magnitude setting. For all settings, 100 data sets were generated using R consisting of $n = 177$ observations of $p = 10$ variables on each of the $K = 104$ subjects (R Core Team, 2021). Dimensions in terms of the number of subjects, variables, and observations were chosen to match our data analysis described in Section 2.4. True networks and precision matrices were generated in a hierarchical manner beginning with group-level networks and precision matrices and then subject-level networks and matrices. That is, we first randomly generated either $G = 2$ or $G = 3$ hub-type networks, given by $\{N_g\}_{g=1}^{G}$, each with $\lfloor\sqrt{p}\rfloor$ hubs and thus $E = p - \lfloor\sqrt{p}\rfloor$ edges, while simultaneously forcing the networks to share $s = \lfloor\rho \times E\rfloor$ edges. We considered $\rho \in \{0.20, 0.50, 0.80\}$ for three different levels of overlap across the $G$ groups. Note that $\rho$ represents the approximate proportion of edges that are common across the cluster-level networks, and that these group-level networks were constant across the 100 simulations for a given $\rho$. Then, for each simulation we generated the cluster-level precision matrices, $\{\mathbf{\Omega_{0g}}\}_{g=1}^{G}$, as follows:

For $g = 1$ to $G$;

1. Begin with a $p \times p$ adjacency matrix with sparsity structure corresponding to $N_g$.
2. Randomly draw $E$ values from a uniform distribution with support on the interval $[-1, -0.50] \cup [0.50, 1]$ for elements corresponding to edges in the network to obtain $\mathbf{\Omega_{0g}}$.

3. Set the diagonal entries of $\mathbf{\Omega_{0g}}$ to 1.

4. If $\mathbf{\Omega_{0g}}$ is not positive definite, then divide each row by its number of non-zero elements.

The $s$ number of edges that were common across the groups were forced to have the same values in each $\mathbf{\Omega_{0g}}$.

For the $K = 104$ subjects, we first randomly assigned them to the $G$ clusters. Subject-level networks were then generated by randomly selecting $\lfloor 0.20 \times E \rfloor$ node pairs to add or remove an edge from their corresponding group-level network. For subject-level precision matrices, denoted $\{\mathbf{\Omega_k}\}_{k=1}^{104}$, common entries for subject- and corresponding group-level matrices were set the same as the group-level matrix with random noise generated from a $\mathcal{N}(0, 0.05^2)$ distribution added to non-zero entries. Entries for added subject-specific edges were generated from a uniform distribution with support on the interval $[-1, -0.50] \cup [0.50, 1]$. As before, if the generated $\mathbf{\Omega_k}$ was not positive definite, then we divided each row by its number of non-zero elements. Lastly, $n = 177$ observations were randomly generated for each subject from a $\mathcal{N}_p(\mathbf{0},\ \mathbf{\Omega_k}^{-1})$ distribution and were then centered and scaled prior to analysis. For tuning parameter selection, we implemented our modified stARS algorithm and included results for 5-fold CV in Appendix A.

### 2.3.2 Simulation Results

We assessed the performance of RCCM in two aspects: clustering and network estimation. Since existing methods only conduct clustering and FC network estimation separately, we considered competitive methods using a 2-step approach. For one approach, we started by obtaining GLasso estimates for each subject, and then clustering subjects based on vectorizing these estimates using K-means clustering, calling this GLasso & K-means (Friedman et al., 2007). For the second approach, we first used the Ward clustering method described in Section 2.2.2, and then implemented the group graphical lasso (GGL) of Danaher et al. (2014) for network estimation within each cluster, referred to as Ward & GGL (Murtagh and Legendre, 2014). GGL uses an $\ell_1$ penalty to encourage a shared sparsity pattern across subjects, but not necessarily entries of the same magnitude. For all approaches considered, we implemented our modified stARS

approach described in Section 2.2.3 for regularization selection. We note that the scale of what worked well for tuning parameters varied across the methods, so we tailored them accordingly.

To assess clustering performance, we calculated the rand index (RI) and adjusted rand index ($\text{RI}_{\text{adj}}$) with values of 1 indicating perfect clustering of subjects for both metrics (Rand, 1971; Hubert and Arabie, 1985). For RCCM, cluster memberships were determined by assigning each subject to the cluster with the highest posterior probability as described by the $\hat{w}_{gk}$ estimates. Clustering performances of RCCM, Ward & GGL, and GLasso & K-means averaged across 100 simulations are displayed in Tables 2.1 and 2.2. Overall, clustering accuracy across all methods tended to be better for $G = 2$ rather than $G = 3$ clusters, and better for the high magnitude setting compared to the low magnitude setting as expected. We believe the exceptionally poor performance of the competing methods in the low magnitude settings is likely due to the high-level of similarity between clusters not favoring two-step approaches which fail to extract useful information from subjects for clustering. For the low magnitude setting RCCM achieved at best an average RI = 0.997 and $\text{RI}_{\text{adj}} = 0.995$ in the setting with $G = 2$ clusters and only 20% of group edges overlapping, and at worst an average RI = 0.901 and $\text{RI}_{\text{adj}} = 0.805$ in the setting with $G = 3$ clusters and 50% of group edges overlapping. Generally, for the low magnitude settings RCCM performed the best in terms of clustering among the three methods considered. However, the Ward & GGL and GLasso & K-means performed more competitively in the high magnitude setting with both methods outperforming RCCM in the $G = 3$ setting with 80% of overlapping edges across the clusters.

For edge detection pertaining to network estimation, we calculated the true positive rate (TPR) or recall, the false positive rate (FPR), and the precision or positive predictive value (PPV) for both the subject and cluster-level networks. Performance in terms of edge-detection averaged across 100 simulations for $G = 2$ and 3 clusters are displayed in Tables 2.3, 2.4, 2.5, and 2.6 using our modified stARS method with RCCM, Ward & GGL and GLasso & K-means. In all tables, subject-level metrics are denoted with a subscript $k$, and group-level metrics subscript $g$.

Overall, we observe from Tables 2.3, 2.4, 2.5, and 2.6 that all three methods achieved higher power for the high magnitude settings compared to the low magnitude settings

as expected. Moreover, the three methods performed better in terms of power as the proportion of overlapping edges across the clusters decreased. For RCCM and Ward & GGL, this is likely due to improved clustering since across-cluster variability becomes comparable to within-cluster variability. For GLasso & K-means, we believe the improved power as overlap decreased is due to the true precision matrices having entries slightly higher in magnitude on average for lower overlap settings compared to higher ones. This is since less adjustments needed to be made to ensure the true precision matrices shared enough edges while still being positive definite.

For the low magnitude settings with $G = 2$ clusters displayed in Table 2.4, all three methods attained FPR values close to 0, displaying how stARS yields very sparse estimates. However, RCCM was the only method to maintain a reasonable power, displaying the advantage of concurrent estimation and clustering as opposed to conducting them sequentially. By pooling information across subjects in each group, RCCM was able to detect non-zero entries that were low in magnitude while the competing approaches with stARS yielded too much penalization. As expected, GLasso & K-means generally had the lowest power of the three methods, as it does not pool information across subjects in network estimation. Simulation results for 5-fold CV are included in Appendix A.

## 2.4 Data Analysis

We applied the proposed RCCM to a resting-state fMRI data set collected on 40 participants diagnosed with chronic schizophrenia, 21 participants diagnosed with first-episode schizophrenia, and 43 healthy controls using our modified stARS algorithm for tuning parameter selection. Resting-state fMRI data were collected during an approximately 6-minute period using a Siemens Trio 3T scanner (Erlangen, Germany) with the following acquisition parameters: gradient-echo echo-planar imaging 180 volumes, repetition time = 2 s, echo time = 30 ms, flip angle = 90°, 34 contiguous ACPC aligned axial slices, voxel size = $3.4 \times 3.4 \times 4.0$ mm, matrix = $64 \times 64 \times 34$ (Camchong et al., 2011). Additionally, a field map acquisition was collected and used to correct images for geometric distortion due to magnetic field inhomogeneities (repetition time = 300 ms, echo time = 1.91 ms/4.37 ms, flip angle = 55°, voxel size = $3.4 \times 3.4 \times 4.0$ mm).

Imaging data were preprocessed using the software tool FEAT (Woolrich et al., 2001). The first 3 volumes were removed prior to analysis, leaving 177 volumes in total. After preprocessing, data from 10 ROIs within the superior and inferior parietal lobules were extracted using an atlas developed through a parcellation study by Mars et al. (2011).

Resulting clusters of subjects obtained using our RCCM are referred to as groups A and B for $G = 2$ clusters and groups C, D, and E for $G = 3$. As displayed in Table 2.7, specifying $G = 2$ clusters yielded 67 and 37 subjects in groups A and B respectively. Most participants diagnosed with schizophrenia were classified into group A with 66.7% of participants diagnosed with first-episode schizophrenia and 75.0% of participants diagnosed with chronic-episode schizophrenia, while only about half of healthy controls were in the same group. Specifying $G = 3$ clusters yielded clusters with 61, 24, and 19 subjects in groups C, D, and E respectively. Most participants diagnosed with schizophrenia were clustered into group C, with 57.1% of participants diagnosed with first-episode schizophrenia and 70.0% of participants diagnosed with chronic-episode schizophrenia respectively, and about half of the healthy controls in the same cluster. Applying the competing methods to our data, we found somewhat similar results as that of RCCM. In terms of clustering, the estimated cluster memberships for RCCM and Ward clustering had rand indexes of 0.873 and 0.891 when specifying $G = 2$ and 3 clusters respectively, indicating that RCCM and Ward clustering agreed on roughly 90% of subject pairings. The estimated cluster memberships for GLasso & K-means were less similar to that of RCCM, with rand indexes of 0.675 and 0.673 when specifying $G = 2$ and 3 clusters respectively, indicating that RCCM and the GLasso & K-means approach agreed on only about 70% of subject pairings.

Using a gap statistic as described in Appendix A, we selected $G = 3$ as our final number of clusters (Tibshirani et al., 2001). Generally, the gap statistic compares the observed change in within-cluster dispersion of a clustering of subjects when specifying different numbers of clusters to what is expected under a corresponding null setting. The performance of using a gap statistic for the proposed RCCM was investigated via simulations with the results summarized in Table A.1 of Appendix A.

Specifying $G = 3$ clusters based on the results of the gap statistic, we observe that group C remained mostly the same as group A, but group B was further divided into groups D and E, suggesting more heterogeneity among subject-level estimates in group

B compared to group A. This within-cluster heterogeneity in terms of edge presence is displayed in Figure 2.2. Overall, subjects were somewhat similar within each cluster, always having precision matrix entries with the same sign and similar in magnitude. However, not all implied networks were identical. This individual heterogeneity is displayed in Figure 2.2 showing the variance in edge presence calculated as $V_{ij} = p_{ij} \times (1 - p_{ij})$ where $p_{ij}$ is the proportion of subjects within each cluster with an edge present between variables $i$ and $j$.

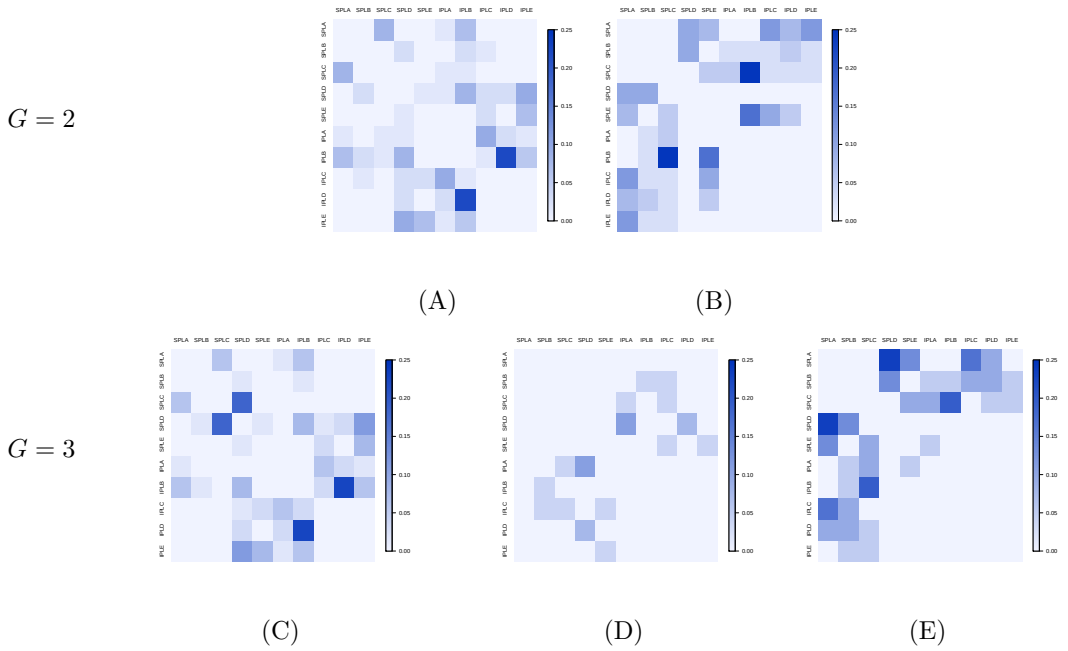


Figure 2.2: Variability in edge presence for groups A through E. A value of 0 indicates perfect agreement in edge presence within a cluster, while a value closer to 0.25 indicates a larger amount of heterogeneity in the presence of edges. Regions within the inferior parietal lobule are prefixed with IPL, while regions within the superior parietal lobule are prefixed with SPL.

Specifying either number of clusters, there was a tendency for participants diagnosed with schizophrenia to be clustered together more than healthy controls, but not significantly so. It is possible that subject-level heterogeneity in hemodynamic responses prevented better discrimination between those with schizophrenia and healthy controls.

Moreover, it is unreasonable to expect perfect discrimination of participants since psychiatric disorders such as schizophrenia are complex and may not always yield consistent patterns in fMRI data across individuals (Wager and Lindquist, 2015).

Plots of estimated networks made using the *igraph* R package (Gabor and Tamas, 2006) are displayed in Figure 2.3. Regions within the inferior parietal lobule are prefixed with IPL, while regions within the superior parietal lobule are prefixed with SPL. Specifying $G = 3$ clusters, we observe that the estimated group-level networks were somewhat similar, with all edges for group C, the majority schizophrenia group, also being present for groups D and E which had significantly fewer participants diagnosed with schizophrenia. Moreover, estimated precision matrices were also similar in that common non-zero entries had the same sign. For example, the SPLA and SPLC nodes were connected in all three networks, and all had negative entries in their corresponding precision matrix estimates. However, it is notable that group C, the majority schizophrenia group, had fewer estimated connections than groups D and E. This is consistent with others who have found evidence supporting a general disconnection hypothesis in schizophrenia based on disrupted or decreased FC (Honey et al., 2005; Zhou et al., 2008; Yoon et al., 2008; Chen et al., 2020*b*; Lottman et al., 2019).

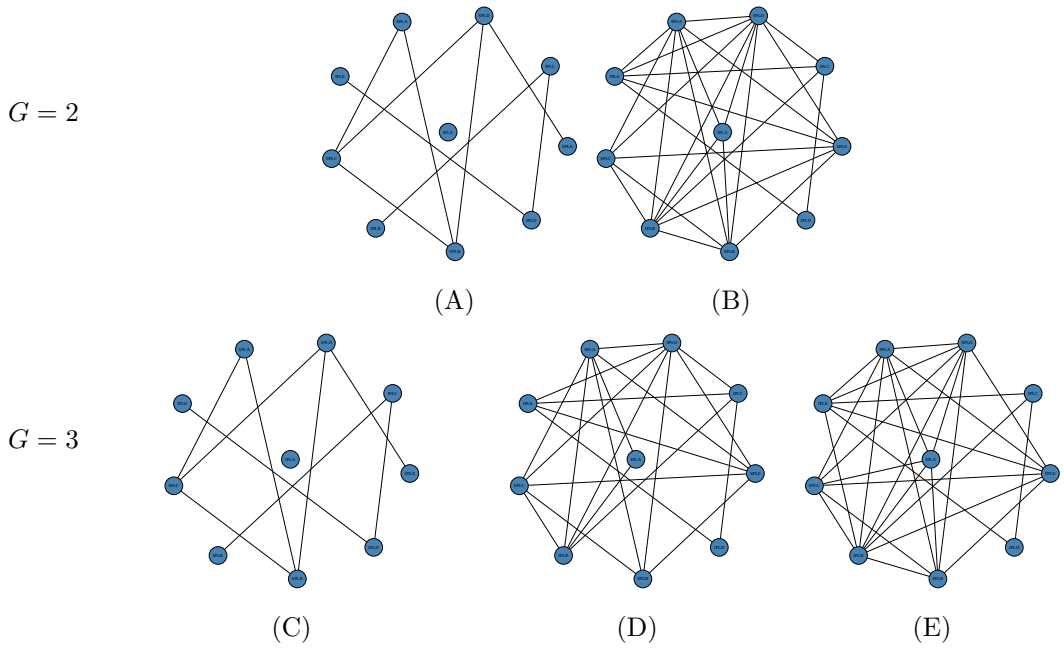


Figure 2.3: Estimated FC networks of the inferior and superior parietal lobules obtained using RCCM specifying $G = 2$ or 3 clusters for the top and bottom rows respectively. For $G = 2$ clusters, Group A contained the highest proportion of participants diagnosed with schizophrenia, and its estimated network had far fewer connections than Group B (top right) suggesting decreased FC among participants diagnosed with schizophrenia. For $G = 3$ groups, Group C contained the majority of participants diagnosed with schizophrenia, and similarly had decreased FC relative to Groups D and E. Notably, Groups D and E were mostly made up of subjects belonging to Group B in the $G = 2$ setting.

## 2.5 Discussion

Estimating FC for multi-subject fMRI data to better understand psychiatric disorders and neurodegenerative diseases is of clinical importance. We proposed a penalized model-based clustering method, the RCCM, to use fMRI data to simultaneously cluster subjects and provide interpretable estimates of both subject- and cluster-level FC networks. Additionally, we also proposed a modified stARS method for regularization

selection that obtains more interpretable estimates of FC networks than current approaches. We showed the competitive performance of RCCM compared to conducting clustering and estimation sequentially rather than concurrently through simulations under varying settings. Lastly, we displayed the utility of RCCM through application to a resting-state fMRI data set collected on participants diagnosed with schizophrenia and healthy controls finding evidence to support the disconnection hypothesis among those diagnosed with schizophrenia. A corresponding R package for the proposed RCCM and related functions is available for download at https://github.com/dilernia/rccm.

Although the proposed RCCM addresses short comings in current methods for analyzing multi-subject fMRI data sets, the method could be improved to better facilitate analyses in higher dimensions. Additionally, the proposed RCCM relies on the assumption of independent observations within subjects which is likely violated for fMRI data in which observations are known to exhibit autocorrelation. In Chapter 3, we present an asymptotic distribution for the empirical partial correlations of a multivariate time series to facilitate inference of FC that accounts for autocorrelation in the data.

Table 2.1: Clustering performance of RCCM, Ward, and K-means clustering using stARS for tuning parameter selection. Results are for $G = 2$ unbalanced groups, true precision matrices with entries large or small in magnitude, and group overlap being 0.20, 0.50, or 0.80 for each of 104 subjects averaged across 100 simulations. Ward clustering was based on a difference matrix of the Frobenius-norm differences between matrix estimates, while K-means clustering was based on vectorized GLasso matrix estimates.

| Magnitude | Overlap | Method | RI | $RI_{adj}$ |
|---|---|---|---|---|
| High | 0.2 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 1.000 (0.000) | 1.000 (0.000) |
| | 0.5 | RCCM | 0.995 (0.046) | 0.990 (0.100) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 0.987 (0.072) | 0.975 (0.145) |
| | 0.8 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 1.000 (0.000) | 1.000 (0.000) |
| Low | 0.2 | RCCM | 0.997 (0.008) | 0.995 (0.015) |
| | | Ward & GGL | 0.971 (0.031) | 0.942 (0.062) |
| | | GLasso & K-means | 0.529 (0.013) | -0.007 (0.023) |
| | 0.5 | RCCM | 0.993 (0.012) | 0.986 (0.023) |
| | | Ward & GGL | 0.962 (0.038) | 0.923 (0.076) |
| | | GLasso & K-means | 0.531 (0.031) | -0.002 (0.066) |
| | 0.8 | RCCM | 0.982 (0.030) | 0.963 (0.060) |
| | | Ward & GGL | 0.921 (0.067) | 0.842 (0.135) |
| | | GLasso & K-means | 0.527 (0.012) | -0.008 (0.024) |

Table 2.2: Clustering performance of RCCM, Ward, and K-means clustering using stARS for tuning parameter selection. Results are for $G = 3$ unbalanced groups, true precision matrices with entries large or small in magnitude, and group overlap being 0.20, 0.50, or 0.80 for each of 104 subjects averaged across 100 simulations. Ward clustering was based on a difference matrix of the Frobenius-norm differences between matrix estimates, while K-means clustering was based on vectorized GLasso matrix estimates.

| Magnitude | Overlap | Method | RI | $RI_{adj}$ |
|---|---|---|---|---|
| High | 0.2 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 0.865 (0.125) | 0.717 (0.263) |
| | 0.5 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 0.888 (0.127) | 0.765 (0.266) |
| | 0.8 | RCCM | 0.878 (0.116) | 0.767 (0.219) |
| | | Ward & GGL | 0.999 (0.003) | 0.998 (0.007) |
| | | GLasso & K-means | 0.975 (0.077) | 0.948 (0.156) |
| Low | 0.2 | RCCM | 0.910 (0.058) | 0.822 (0.112) |
| | | Ward & GGL | 0.892 (0.060) | 0.786 (0.117) |
| | | GLasso & K-means | 0.526 (0.031) | -0.002 (0.066) |
| | 0.5 | RCCM | 0.901 (0.062) | 0.805 (0.120) |
| | | Ward & GGL | 0.872 (0.067) | 0.748 (0.132) |
| | | GLasso & K-means | 0.513 (0.011) | -0.030 (0.019) |
| | 0.8 | RCCM | 0.917 (0.058) | 0.835 (0.119) |
| | | Ward & GGL | 0.885 (0.061) | 0.771 (0.121) |
| | | GLasso & K-means | 0.520 (0.015) | -0.013 (0.027) |

Table 2.3: Comparison of method performances for $G = 2$ unbalanced groups containing 67 and 37 subjects. Results are for the high magnitude setting with observations of $p = 10$ variables and $n = 177$ observations for each subject using stARs for tuning parameter selection, averaged across 100 simulations. Group-level performance measures for GLasso are missing since the method does not yield cluster-level estimates.

| Overlap | Method | $TPR_g$ | $FPR_g$ | $PPV_g$ | $TPR_k$ | $FPR_k$ | $PPV_k$ |
|---|---|---|---|---|---|---|---|
| 0.2 | RCCM | 0.913 | 0.091 | 0.673 | 1.000 | 0.134 | 0.580 |
| | | (0.000) | (0.006) | (0.014) | (0.000) | (0.004) | (0.008) |
| | Ward & GGL | 0.913 | 0.088 | 0.682 | 1.000 | 0.092 | 0.666 |
| | | (0.000) | (0.006) | (0.015) | (0.000) | (0.003) | (0.007) |
| | GLasso & K-means | | | | 1.000 | 0.093 | 0.664 |
| | | | | | (0.000) | (0.003) | (0.007) |
| 0.5 | RCCM | 1.000 | 0.066 | 0.724 | 0.929 | 0.120 | 0.587 |
| | | (0.000) | (0.005) | (0.017) | (0.004) | (0.006) | (0.012) |
| | Ward & GGL | 1.000 | 0.084 | 0.675 | 0.921 | 0.076 | 0.689 |
| | | (0.000) | (0.005) | (0.013) | (0.002) | (0.003) | (0.008) |
| | GLasso & K-means | | | | 0.921 | 0.077 | 0.687 |
| | | | | | (0.002) | (0.003) | (0.009) |
| 0.8 | RCCM | 0.705 | 0.078 | 0.710 | 0.910 | 0.148 | 0.641 |
| | | (0.155) | (0.025) | (0.094) | (0.033) | (0.009) | (0.028) |
| | Ward & GGL | 0.900 | 0.045 | 0.847 | 0.848 | 0.093 | 0.726 |
| | | (0.051) | (0.024) | (0.075) | (0.066) | (0.018) | (0.050) |
| | GLasso & K-means | | | | 0.856 | 0.091 | 0.731 |
| | | | | | (0.054) | (0.014) | (0.038) |

Table 2.4: Comparison of method performances for $G = 2$ unbalanced groups containing 67 and 37 subjects. Results are for the low magnitude setting with observations of $p = 10$ variables and $n = 177$ observations for each subject using stARs for tuning parameter selection, averaged across 100 simulations. Group-level performance measures for GLasso are missing since the method does not yield cluster-level estimates.

| Overlap | Method | $TPR_g$ | $FPR_g$ | $PPV_g$ | $TPR_k$ | $FPR_k$ | $PPV_k$ |
|---|---|---|---|---|---|---|---|
| 0.2 | RCCM | 0.588 | 0.120 | 0.507 | 0.887 | 0.011 | 0.947 |
| | | (0.099) | (0.036) | (0.093) | (0.037) | (0.005) | (0.021) |
| | Ward & GGL | 0.091 | 0.018 | 0.522 | 0.238 | 0.002 | 0.965 |
| | | (0.059) | (0.013) | (0.282) | (0.087) | (0.002) | (0.030) |
| | GLasso & K-means | | | | 0.124 | 0.000 | 0.994 |
| | | | | | (0.044) | (0.001) | (0.015) |
| 0.5 | RCCM | 0.592 | 0.148 | 0.524 | 0.806 | 0.020 | 0.924 |
| | | (0.095) | (0.050) | (0.092) | (0.063) | (0.007) | (0.024) |
| | Ward & GGL | 0.096 | 0.020 | 0.644 | 0.155 | 0.002 | 0.961 |
| | | (0.072) | (0.028) | (0.279) | (0.067) | (0.002) | (0.037) |
| | GLasso & K-means | | | | 0.159 | 0.000 | 0.993 |
| | | | | | (0.019) | (0.000) | (0.007) |
| 0.8 | RCCM | 0.574 | 0.115 | 0.650 | 0.639 | 0.017 | 0.936 |
| | | (0.125) | (0.076) | (0.081) | (0.136) | (0.006) | (0.017) |
| | Ward & GGL | 0.097 | 0.015 | 0.778 | 0.130 | 0.001 | 0.977 |
| | | (0.059) | (0.025) | (0.235) | (0.048) | (0.001) | (0.023) |
| | GLasso & K-means | | | | 0.118 | 0.000 | 0.993 |
| | | | | | (0.017) | (0.000) | (0.006) |

Table 2.5: Comparison of method performances for $G = 3$ unbalanced groups containing 61, 24, and 19 subjects. are for the high magnitude setting with observations of $p = 10$ variables and $n = 177$ observations for each subject using stARs for tuning parameter selection, averaged across 100 simulations. Group-level performance measures for GLasso are missing since the method does not yield cluster-level estimates.

| Overlap | Method | $TPR_g$ | $FPR_g$ | $PPV_g$ | $TPR_k$ | $FPR_k$ | $PPV_k$ |
|---|---|---|---|---|---|---|---|
| 0.2 | RCCM | 0.999 | 0.143 | 0.598 | 0.997 | 0.159 | 0.570 |
| | | (0.007) | (0.021) | (0.029) | (0.008) | (0.011) | (0.009) |
| | Ward & GGL | 0.998 | 0.108 | 0.663 | 0.988 | 0.120 | 0.636 |
| | | (0.011) | (0.016) | (0.030) | (0.010) | (0.013) | (0.017) |
| | GLasso & K-means | | | | 0.988 | 0.120 | 0.636 |
| | | | | | (0.010) | (0.013) | (0.016) |
| 0.5 | RCCM | 0.994 | 0.133 | 0.653 | 0.999 | 0.170 | 0.596 |
| | | (0.056) | (0.022) | (0.039) | (0.009) | (0.014) | (0.028) |
| | Ward & GGL | 1.000 | 0.086 | 0.745 | 0.997 | 0.124 | 0.667 |
| | | (0.000) | (0.017) | (0.039) | (0.002) | (0.007) | (0.012) |
| | GLasso & K-means | | | | 0.977 | 0.120 | 0.678 |
| | | | | | (0.116) | (0.022) | (0.056) |
| 0.8 | RCCM | 0.869 | 0.113 | 0.673 | 0.969 | 0.152 | 0.615 |
| | | (0.027) | (0.002) | (0.007) | (0.006) | (0.007) | (0.010) |
| | Ward & GGL | 0.858 | 0.093 | 0.714 | 0.913 | 0.094 | 0.709 |
| | | (0.033) | (0.015) | (0.032) | (0.012) | (0.008) | (0.017) |
| | GLasso & K-means | | | | 0.914 | 0.095 | 0.708 |
| | | | | | (0.011) | (0.008) | (0.016) |

Table 2.6: Comparison of method performances for $G = 3$ unbalanced groups containing 61, 24, and 19 subjects. are for the low magnitude setting with observations of $p = 10$ variables and $n = 177$ observations for each subject using stARs for tuning parameter selection, averaged across 100 simulations. Group-level performance measures for GLasso are missing since the method does not yield cluster-level estimates.

| Overlap | Method | $TPR_g$ | $FPR_g$ | $PPV_g$ | $TPR_k$ | $FPR_k$ | $PPV_k$ |
|---|---|---|---|---|---|---|---|
| 0.2 | RCCM | 0.984 | 0.020 | 0.912 | 0.935 | 0.017 | 0.926 |
| | | (0.065) | (0.019) | (0.083) | (0.040) | (0.010) | (0.040) |
| | Ward & GGL | 0.298 | 0.000 | 0.994 | 0.314 | 0.002 | 0.976 |
| | | (0.091) | (0.004) | (0.067) | (0.073) | (0.002) | (0.021) |
| | GLasso & K-means | | | | 0.158 | 0.000 | 0.996 |
| | | | | | (0.020) | (0.000) | (0.005) |
| 0.5 | RCCM | 0.801 | 0.005 | 0.974 | 0.796 | 0.006 | 0.975 |
| | | (0.124) | (0.024) | (0.125) | (0.041) | (0.002) | (0.008) |
| | Ward & GGL | 0.192 | 0.002 | 0.966 | 0.222 | 0.002 | 0.966 |
| | | (0.080) | (0.009) | (0.170) | (0.062) | (0.002) | (0.032) |
| | GLasso & K-means | | | | 0.107 | 0.001 | 0.990 |
| | | | | | (0.041) | (0.002) | (0.024) |
| 0.8 | RCCM | 0.746 | 0.028 | 0.897 | 0.702 | 0.028 | 0.899 |
| | | (0.066) | (0.025) | (0.088) | (0.051) | (0.017) | (0.052) |
| | Ward & GGL | 0.232 | 0.001 | 0.992 | 0.274 | 0.006 | 0.946 |
| | | (0.100) | (0.005) | (0.056) | (0.077) | (0.004) | (0.027) |
| | GLasso & K-means | | | | 0.162 | 0.001 | 0.991 |
| | | | | | (0.022) | (0.000) | (0.007) |

Table 2.7: Summary of clustering results for specifying $G = 2$ or 3 clusters. For $G = 2$, Group A had the highest proportion of participants diagnosed with schizophrenia with 66.7% of first-episode participants, and 75.0% of participants diagnosed with chronic-episode schizophrenia. Healthy controls, however, were more evenly distributed between the two groups, with 53.5% and 46.5% in Groups A and B respectively. Results for $G = 3$ groups were not as clear, with most subjects being clustered into Group C.

| $G$ | Group | Control | 1st Episode | Chronic |
|---|---|---|---|---|
| 2 | A | 23 (53.5%) | 14 (66.7%) | 30 (75.0%) |
| | B | 20 (46.5%) | 7 (33.3%) | 10 (25.0%) |
| 3 | C | 21 (48.8%) | 12 (57.1%) | 28 (70.0%) |
| | D | 10 (23.3%) | 6 (28.6%) | 8 (20.0%) |
| | E | 12 (27.9%) | 3 (14.3%) | 4 (10.0%) |

# Chapter 3

# Inference of Partial Correlations for a Stationary Multivariate Time Series

## 3.1 Introduction

The partial correlation coefficient measures the strength of the linear relationship between two variables of interest after removing the effect of other variables, and has been commonly used to describe conditional dependencies in areas such as ecology (Damos, 2016), geoscience (Erb, 2020), genomics (de la Fuente et al., 2004), and neuroimaging (Marrelec et al., 2006). The sampling distribution of a single partial correlation coefficient for independent and normally distributed data was first derived geometrically by Fisher (1924). Others have derived the distribution facilitating various procedures for conducting inference of individual partial correlations (Wilks, 1932; Cramer, 1974; Gupta, 1977; Williams, 1978; Bergsma, 2020). However, these methods of inference rely on an assumption of independent observations which is violated in time series data where observations are correlated over time, possibly resulting in spurious correlations (Student, 1914; Granger and Newbold, 1974). Addressing this issue, some have proposed use of adjusted degrees of freedom measures (Afyouni et al., 2019), pre-whitening data to remove autocorrelation present in the data prior to analysis (Haugh, 1976), and

modified standard error estimators (Bartlett, 1946). A limitation of these approaches though is the reliance on correctly specifying a model for autocorrelation present in the data (Box and Newbold, 1971). Moreover, these approaches provide inference procedures for correlations when autocorrelation is present, but they do not fully characterize the large-sample sampling distribution of the partial correlations.

Assuming independent observations, the asymptotic joint distribution of partial correlations has been derived (Hedges and Olkin, 1983). However, to the best of our knowledge an asymptotic distribution for the partial correlations of a general multivariate time series has not been formally derived. Partial correlations are distinctly important compared to marginal correlations since they adjust for the possible delayed effects of other variables for multivariate time series data. Moreover, inference of partial correlations is important for applications such as functional connectivity analysis in neuroimaging studies in which use of marginal or partial correlations can yield varying results (Marrelec et al., 2006; Kim et al., 2015).

We derive an asymptotic distribution for the partial correlations of a weakly stationary multivariate time series using a second-order Taylor series approximation and properties of quadratic forms of multivariate normal random vectors. Based on this derived distribution, we propose Wald based confidence intervals and testing procedures for inference of individual partial correlations when autocorrelation is present. We show through simulations the advantage of our proposed inferential procedures compared to others which assume independent observations, achieving closer to nominal false positive rates when autocorrelation is present.

In Section 3.2.1, we first review the definition of the partial correlation and then provide the large-sample distribution of the partial correlation coefficients. In Section 3.2.2, we derive a corresponding asymptotic covariance matrix for the empirical partial correlations. Then, in Section 3.3 we propose inferential methods based on our derived asymptotic distribution. In Section 3.4, we demonstrate the utility of our inference procedures for finite samples via simulation. Lastly, in Section 3.5 we conclude with a discussion of our findings and future directions.

## 3.2 Asymptotics of Sample Partial Correlations

### 3.2.1 Asymptotic Distribution

Let $\mathbf{x}(t) = \{\mathbf{x}_k\}_{k=1}^p$ be an $N$-length realization of a $p$-variate Gaussian process that is second-order stationary and ergodic with $\mathbf{x}_k \in \mathbb{R}^N$ for $k = 1, 2, \ldots, p$. Generally, a Gaussian process is stationary provided that it has a constant mean and that its covariance function $\mathrm{c}(x_i(t-\ell), x_j(t)) = \mathrm{c}_{ij}(\ell)$ only depends on the lag between points given by $\ell$. The sample partial correlation between any pair of variables can be expressed using the ordinary least squares (OLS) residuals. Specifically, letting $\mathbf{e}_{i\cdot(ij)}$ and $\mathbf{e}_{j\cdot(ij)}$ be the $N$-length vectors of OLS residuals from regressing $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively on the other $p-2$ variables $\{\mathbf{x}_k\}_{k\neq i,j}$, it follows that the empirical partial correlation between $\mathbf{x}_i$ and $\mathbf{x}_j$ is

$$r_{ij\cdot(ij)} = f(\begin{bmatrix} \boldsymbol{e}_{i\cdot(ij)} \\ \boldsymbol{e}_{j\cdot(ij)} \end{bmatrix}) = f(\boldsymbol{e}_{ij}) = \frac{\boldsymbol{e}_{i\cdot(ij)}^T \boldsymbol{e}_{j\cdot(ij)}}{\sqrt{\boldsymbol{e}_{i\cdot(ij)}^T \boldsymbol{e}_{i\cdot(ij)} \boldsymbol{e}_{j\cdot(ij)}^T \boldsymbol{e}_{j\cdot(ij)}}}, \tag{3.1}$$

so that $r_{ij\cdot(ij)}$ is equivalent to the sample marginal correlation between $\mathbf{e}_{i\cdot(ij)}$ and $\mathbf{e}_{j\cdot(ij)}$. Further assuming that each of the $p$ variables have square integrable spectral density functions, we will show that the empirical partial correlations $\{r_{ij\cdot(ij)}\}_{i\neq j}$ are jointly asymptotically normal. First, we present the necessary conditions to derive the asymptotic joint distribution of the partial correlations of a multivariate time series.

1. All fourth-order cumulants of $\mathbf{x}(t)$ are 0.
2. The spectral density functions of each component of $\mathbf{x}(t)$ are square-integrable.
3. Each component of $\mathbf{x}(t)$ is a stochastic linear process with finite variance meaning:
   (a) $\mathbf{x}(t) = \sum_{\ell=0}^{\infty} \mathbf{A}(\ell)\boldsymbol{\varepsilon}(t-\ell)$, where $\boldsymbol{\varepsilon}(t-\ell)$ is the vector of one-step linear prediction residuals at lag $\ell$.
   (b) $\mathbb{E}[\boldsymbol{\varepsilon}(t_1)\boldsymbol{\varepsilon}(t_2)^T] = \begin{cases} \mathbf{0}, & t_1 \neq t_2 \\ \boldsymbol{\Sigma}, & t_1 = t_2 \end{cases}$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{q\times q}$ is nonsingular.
   (c) $\mathbb{E}[\boldsymbol{\varepsilon}(t)] = \mathbf{0}$
   (d) $\sum_{\ell=0}^{\infty} ||\mathbf{A}(\ell)||^2 < \infty$, where $||\cdot||$ is the Euclidean norm.

4. The first through fourth moments of $\boldsymbol{\varepsilon}(t|\mathcal{F}_{t-1})$ are all finite constants, where $\mathcal{F}_{t-1}$ is the sub $\sigma$-algebra generated by $\{\mathbf{x}(t') : t' < t\}$.

We note that conditions 1 and 2 correspond to a result regarding the asymptotic normality of marginal correlations of a multivariate time series obtained using the multivariate delta method (Roy, 1989; Serfling, 1980), and conditions 3 and 4 correspond to a theorem establishing the joint asymptotic normality of empirical covariances for a general multivariate time series (Hannan, 1976).

**Theorem 1.** *Let* $\mathbf{x}(t) = \{\mathbf{x}_k(t)\}_{k=1}^{p}$ *be an ergodic, second-order stationary p-variate Gaussian process satisfying conditions 1 through 4 above. Then if* $r_{ij\cdot(ij)}$ *is the empirical partial correlation between* $\mathbf{x}_i$ *and* $\mathbf{x}_j$ *based on an* $N$*-length realization of* $\mathbf{x}(t)$*, it follows that* $\sqrt{N}(r_{ij\cdot(ij)} - \rho_{ij\cdot(ij)})$ *converges in distribution to a normal with mean the population partial correlation* $\rho_{ij\cdot(ij)}$ *for all* $i \neq j$*.*

A proof of Theorem 1 is provided in Appendix B. We next present a theorem providing an explicit form for the asymptotic covariance structure of the partial correlations in Section 3.2.2.

### 3.2.2 Asymptotic Covariance Estimator

Assuming the conditions of Theorem 1, it follows that $\{\boldsymbol{e}_{i\cdot(ij)}\}$ are $N$-length Gaussian vectors. In this setting, we present the asymptotic covariance structure for any two empirical partial correlations.

**Theorem 2.** *Let* $\mathbf{e}_{i\cdot(ij)}$*,* $\mathbf{e}_{j\cdot(ij)}$*,* $\mathbf{e}_{k\cdot(km)}$*, and* $\mathbf{e}_{m\cdot(km)}$ *be* $N$*-length Gaussian random vectors. Then the asymptotic variance of* $r_{ij\cdot(ij)}$ *is* $\frac{1}{2}\left[tr(\boldsymbol{H}[f(\boldsymbol{\theta}_{ij})]\boldsymbol{\Sigma}_{ij}\boldsymbol{H}[f(\boldsymbol{\theta}_{ij})]\boldsymbol{\Sigma}_{ij})\right]$*, where* $tr(\cdot)$ *denotes the trace function,* $\boldsymbol{\Sigma}_{ij} = Cov(\begin{bmatrix}\boldsymbol{e}_{i\cdot(ij)}^T & \boldsymbol{e}_{j\cdot(ij)}^T\end{bmatrix}^T)$*, and* $\boldsymbol{H}[f(\boldsymbol{\theta}_{ij})]$ *is a Hessian matrix derived in Appendix B. Moreover, the asymptotic covariance between* $r_{ij\cdot(ij)}$ *and* $r_{km\cdot(km)}$ *is* $\frac{1}{2}tr\left(\mathbf{H}[f(\boldsymbol{\theta}_{ij})]\boldsymbol{\Sigma}_{e_{ijkm}12}\mathbf{H}[f(\boldsymbol{\theta}_{km})]\boldsymbol{\Sigma}_{e_{ijkm}12}^T\right)$*, where* $\boldsymbol{\Sigma}_{e_{ijkm}12} = Cov(\begin{bmatrix}\boldsymbol{e}_{i\cdot(ij)}^T & \boldsymbol{e}_{j\cdot(ij)}^T\end{bmatrix}^T, \begin{bmatrix}\boldsymbol{e}_{k\cdot(km)}^T & \boldsymbol{e}_{m\cdot(km)}^T\end{bmatrix}^T)$.

A proof of Theorem 2 is also provided in Appendix B. In the proof, we approximate $r_{ij\cdot(ij)}$ and $r_{km\cdot(km)}$ as quadratic forms of multivariate normal random vectors which means they are asymptotically distributed as a generalized $\chi^2$ distribution (Imhof, 1961).

However, to facilitate more general inference we apply the conditions and result of Theorem 1 to obtain that the empirical partial correlations are asymptotically multivariate normal. The established asymptotic normality allows us to derive novel Wald-based confidence intervals and testing procedures for partial correlations of a multivariate time series.

## 3.3 Inferential Methods

We demonstrate the utility of our derived distribution and asymptotic covariance matrix in finite samples by implementing Wald confidence intervals for individual partial correlations and a Wald testing procedure for testing if each partial correlation is equal to 0. Specifically, we calculate $100 \times (1-\alpha)\%$ Wald confidence intervals for each $\rho_{ij\cdot(ij)}$, the population-level partial correlation, as $r_{ij\cdot(ij)} \pm Z_{\alpha/2} \times SE(r_{ij\cdot(ij)})$, where $r_{ij\cdot(ij)}$ is the empirical partial correlation as defined in Equation (3.1), $Z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution, $\alpha$ is the significance level where we use $\alpha = 0.05$ corresponding to 95% confidence intervals, and $SE(r_{ij\cdot(ij)})$ is the standard error of $r_{ij\cdot(ij)}$ for $i, j = 1, 2, \ldots, p$. The approximate standard error based on our second-order Taylor series approximation for $r_{ij\cdot(ij)}$ is the square root of $\frac{1}{2}\left[\text{tr}\left(\boldsymbol{H}[f(\boldsymbol{\theta}_{ij})]\boldsymbol{\Sigma}_{ij}\boldsymbol{H}[f(\boldsymbol{\theta}_{ij})]\boldsymbol{\Sigma}_{ij}\right)\right]$ as derived in Appendix B. There are several options for estimating the covariance matrix $\boldsymbol{\Sigma}_{ij}$, but we use tapered covariance estimators for each of the four blocks of $\boldsymbol{\Sigma}_{ij}$ and method of moments estimators for the Hessian matrices $\boldsymbol{H}[f(\boldsymbol{\theta}_{ij})]$ (Westgate, 2014).

We also propose a Wald test for testing whether individual partial correlations were 0 or not i.e., $H_0$: $\rho_{ij\cdot(ij)} = 0$ vs. $H_A$: $\rho_{ij\cdot(ij)} \neq 0$. The corresponding Wald test statistic is $T_{w;ij} = r^2_{ij\cdot(ij)}/SE(r_{ij\cdot(ij)})^2$. We note that this testing procedure is equivalent to using a likelihood ratio test, and thus under the null hypothesis $T_{w;ij}$ asymptotically follows a $\chi^2$ distribution with 1 degree of freedom (Wilks, 1938). Hence, we reject the null hypothesis if $T_{w;ij} \geq \chi^2_{(1;1-\alpha)}$, where $\chi^2_{(1;1-\alpha)} \approx 3.84$ for $\alpha = 0.05$.

## 3.4 Simulations

### 3.4.1 Setup

We implemented simulations to assess the performance of confidence intervals and inference procedures using our derived asymptotic distribution. We considered three different numbers of variables ($p \in \{5, 10, 15\}$), two different sample sizes ($N \in \{100, 500\}$), three different levels of autocorrelation ($\phi \in \{0, 0.40, 0.80\}$), and generated partial correlations either uniformly from the set $\{-0.30, 0, 0.30\}$ or as all being zero, yielding 36 unique settings in total. A total of 1,000 data sets were generated for each setting with $N$ observations of a $p$-variate first-order autoregressive model with correlation parameter $\phi$. Simulations were conducted using R (R Core Team, 2021).

For comparison approaches to construct confidence intervals, we considered three other approaches: a naïve confidence interval assuming normally distributed and independent observations, a Fisher-transformed interval also assuming normally distributed and independent observations, and a block-bootstrap interval for each partial correlation. For the naïve approach, the estimated standard error for $r_{ij\cdot(ij)}$ is $(1 - r^2_{ij\cdot(ij)})/(\sqrt{N-p})$ (Cramer, 1974; Ferguson, 2002). Thus, a naïve 95% confidence interval for $\rho_{ij\cdot(ij)}$ is $r_{ij\cdot(ij)} \pm t^*_{(N-p)}\sqrt{(1 - r^2_{ij\cdot(ij)})/(N-p)}$ where $t^*_{(N-p);\alpha/2}$ is the $\alpha/2$ quantile of a $t$-distribution with $N-p$ degrees of freedom (Levy and Narula, 1978). For the Fisher-transformed intervals, we constructed Wald intervals centered around the inverse hyperbolic tangent of the partial correlations given by

$$\tanh^{-1}(r_{ij\cdot(ij)}) = \frac{1}{2}\log\left((1 + r_{ij\cdot(ij)})/(1 - r_{ij\cdot(ij)})\right),$$

and under an assumption of independent and normally distributed observations $\tanh^{-1}(r_{ij\cdot(ij)})$ converges to a normal distribution with mean $\tanh^{-1}(\rho_{ij\cdot(ij)})$ and variance $1/(N-p-1)$ (Cramer, 1974; Fisher, 1915). Thus, we constructed Wald confidence intervals for $\tanh^{-1}(\rho_{ij\cdot(ij)})$ centered around $\tanh^{-1}(r_{ij\cdot(ij)})$ using an estimated standard error of $1/\sqrt{N-p-1}$, and then transformed the endpoints using the hyperbolic tangent function to obtain a corresponding interval for $\rho_{ij\cdot(ij)}$. For the block-bootstrap, we used the $\alpha/2$ and $1-(\alpha/2)$ quantiles calculated from 1,000 bootstrap samples, and we selected the block-length using an automatic selection algorithm for stationary multivariate time series data (Politis and White, 2004).

For comparison approaches for hypothesis testing of partial correlations, we implemented a naïve $t$-test, a Wald test based on a Fisher transformation, and a block-bootstrap testing procedure. The naïve $t$-test statistic is $t_{ij} = r_{ij\cdot(ij)}\sqrt{(N-p)/(1-r^2_{ij\cdot(ij)})}$ which we compared to the quantiles of a student's $t$ distribution with $N-p$ degrees of freedom (Levy and Narula, 1978). The Fisher-transformed test-statistic is $Z_{ij} = \tanh^{-1}(r_{ij\cdot(ij)})/\sqrt{(N-p-1)}$ which we compared to the quantiles of a standard normal distribution. For the block-bootstrap, we rejected the null hypothesis that the true partial correlation was 0 provided that 0 was outside the $\alpha/2$ and $1-\alpha/2$ quantiles of the 1,000 generated bootstrap samples.

### 3.4.2 Results

Simulation results are summarized in Tables 3.1 and 3.2. Table 3.1 displays the coverage rates from constructing 95% confidence intervals for individual partial correlations. Overall, all four approaches considered achieved close to the nominal coverage rate of 95% for the independence setting ($\phi = 0$). As expected, when autocorrelation was present the block-bootstrap and Taylor series approximation approach based on our derived asymptotic distribution attained coverage rates closer to the nominal rate than the naïve and Fisher transformation intervals which assume independent observations. We also observe that as the sample size ($N$) increased, the intervals based on our Taylor series approximation and the block-bootstrap approached the nominal coverage rate of 95% for all numbers of variables and all levels of autocorrelation. The Fisher transformation and naïve approaches, however, attained lower coverage rates as the sample size increased when a moderate to high amount of autocorrelation was present ($\phi \in \{0.40, 0.80\}$).

Table 3.2 displays the results of testing for a zero partial correlation for each of the $q$ unique partial correlations comparing our approach to the competing methods. Overall, from Table 3.2 we observe that in general the Taylor series Wald test and the block-bootstrap perform similarly in terms of Matthews correlation coefficient (MCC), both outperforming the Fisher transformation and naïve approaches when autocorrelation is present, while achieving similar MCC values in the independence setting. The Taylor approximation and the block-bootstrap also outperformed the Fisher and naïve testing approaches in terms of both the false positive rate (FPR) and false discovery rate

(FDR) when autocorrelation was present, while still achieving similar FPR and FDR values in the independence case. The only setting in which there was a meaningful difference between the block-bootstrap and our Wald test based on our Taylor series approximation was the high autocorrelation ($\phi = 0.80$) low sample size ($N = 100$) case, in which the block-bootstrap achieved somewhat lower FPR and FDR values. For the large sample size setting though ($N = 500$) the methods again performed very similarly in terms of FPR and FDR values. When implemented properly, the block-bootstrap provides a flexible approach for inference across a variety of settings. However, it can be computationally expensive especially in high-dimensional settings. An advantage of our approach compared to the block-bootstrap is potential gains in computational efficiency. Moreover, our derived covariance estimator can more easily be integrated into other modeling frameworks such as hierarchical models for multi-subject analysis.

## 3.5 Discussion

In this chapter, we provided the asymptotic distribution and covariance matrix for the empirical partial correlations of a second-order stationary multivariate time series. Although current methods exist for inference of partial correlations assuming independent and normally distributed observations, we showed how ignoring autocorrelation present in the data can result in significantly lower than nominal coverage rates for confidence intervals and inflated false positive rates in hypothesis testing, consistent with previous work (Student, 1914; Granger and Newbold, 1974). Using our derived asymptotic distribution for inference procedures performed similarly to the block-bootstrap which is a competitive method of inference for multivariate time series data. Our derived covariance estimator and asymptotic distribution provide tools for FC analysis of single-subject fMRI data sets based on the partial correlations. Moreover, our derived asymptotic distribution can be incorporated into other statistical models such as hierarchical models for analyzing data sets with multiple subjects. An R package implementing our derived asymptotic covariance estimator is available at https://github.com/dilernia/pcCov.

Although our derived asymptotic distribution and covariance matrix address short comings in current approaches for conducting inference of partial correlations for time

series data, our inferential procedures could be improved for smaller sample sizes. In particular, we observed that our derived asymptotic covariance between each pair of empirical partial correlations performed somewhat poorly for small sample sizes. In Chapter 4, we demonstrate the utility of our derived asymptotic distribution for analyzing multi-subject data sets in which information can be pooled across subjects allowing for improved estimation of the covariances between partial correlations.

Table 3.1: Individual coverage rates from 1,000 simulations for 95% confidence intervals of individual partial correlations based on $N$ observations. Note that $p$ is the total number of variables and $\phi$ is the autocorrelation parameter.

| $N$ | $p$ | $\phi$ | Taylor Approximation | Block-Bootstrap | Fisher Transform | Naïve |
|---|---|---|---|---|---|---|
| 100 | 5 | 0 | 0.936 | 0.934 | 0.939 | 0.956 |
| | | 0.4 | 0.924 | 0.918 | 0.899 | 0.922 |
| | | 0.8 | 0.826 | 0.874 | 0.653 | 0.692 |
| | 10 | 0 | 0.935 | 0.935 | 0.939 | 0.950 |
| | | 0.4 | 0.920 | 0.919 | 0.899 | 0.918 |
| | | 0.8 | 0.798 | 0.887 | 0.674 | 0.712 |
| | 15 | 0 | 0.910 | 0.912 | 0.914 | 0.919 |
| | | 0.4 | 0.890 | 0.899 | 0.874 | 0.884 |
| | | 0.8 | 0.775 | 0.897 | 0.692 | 0.719 |
| 500 | 5 | 0 | 0.947 | 0.946 | 0.948 | 0.960 |
| | | 0.4 | 0.941 | 0.933 | 0.902 | 0.918 |
| | | 0.8 | 0.919 | 0.906 | 0.640 | 0.664 |
| | 10 | 0 | 0.914 | 0.907 | 0.915 | 0.927 |
| | | 0.4 | 0.920 | 0.906 | 0.873 | 0.887 |
| | | 0.8 | 0.906 | 0.899 | 0.632 | 0.659 |
| | 15 | 0 | 0.726 | 0.714 | 0.728 | 0.748 |
| | | 0.4 | 0.772 | 0.751 | 0.703 | 0.722 |
| | | 0.8 | 0.843 | 0.841 | 0.562 | 0.580 |

Table 3.2: True positive rate (TPR), false positive rate (FPR), Matthews correlation coefficient (MCC), and false discovery rate (FDR) based on 1,000 simulations for testing if each individual partial correlation is nonzero.

| | | | $N = 100$ | | | | $N = 500$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Metric | $p$ | $\phi$ | Taylor | Bootstrap | Fisher | Naïve | Taylor | Bootstrap | Fisher | Naïve |
| TPR | 5 | 0 | 0.873 | 0.860 | 0.855 | 0.868 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.4 | 0.798 | 0.791 | 0.818 | 0.830 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.8 | 0.577 | 0.527 | 0.715 | 0.727 | 0.935 | 0.940 | 0.991 | 0.991 |
| | 10 | 0 | 0.774 | 0.748 | 0.749 | 0.768 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.4 | 0.704 | 0.685 | 0.720 | 0.737 | 0.999 | 0.999 | 1.000 | 1.000 |
| | | 0.8 | 0.549 | 0.464 | 0.650 | 0.663 | 0.887 | 0.894 | 0.977 | 0.977 |
| | 15 | 0 | 0.579 | 0.534 | 0.545 | 0.570 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | 0.4 | 0.539 | 0.491 | 0.543 | 0.566 | 0.990 | 0.991 | 0.995 | 0.995 |
| | | 0.8 | 0.480 | 0.344 | 0.544 | 0.560 | 0.769 | 0.771 | 0.927 | 0.928 |
| FPR | 5 | 0 | 0.063 | 0.064 | 0.054 | 0.060 | 0.051 | 0.053 | 0.050 | 0.051 |
| | | 0.4 | 0.074 | 0.082 | 0.089 | 0.099 | 0.058 | 0.064 | 0.095 | 0.097 |
| | | 0.8 | 0.171 | 0.122 | 0.336 | 0.351 | 0.084 | 0.095 | 0.360 | 0.363 |
| | 10 | 0 | 0.061 | 0.059 | 0.050 | 0.058 | 0.053 | 0.054 | 0.051 | 0.053 |
| | | 0.4 | 0.079 | 0.077 | 0.088 | 0.099 | 0.057 | 0.064 | 0.094 | 0.096 |
| | | 0.8 | 0.203 | 0.112 | 0.311 | 0.326 | 0.088 | 0.088 | 0.350 | 0.353 |
| | 15 | 0 | 0.058 | 0.051 | 0.048 | 0.055 | 0.052 | 0.055 | 0.050 | 0.052 |
| | | 0.4 | 0.083 | 0.068 | 0.086 | 0.097 | 0.059 | 0.066 | 0.094 | 0.096 |
| | | 0.8 | 0.213 | 0.086 | 0.276 | 0.293 | 0.100 | 0.091 | 0.349 | 0.352 |
| MCC | 5 | 0 | 0.822 | 0.810 | 0.815 | 0.821 | 0.954 | 0.952 | 0.955 | 0.954 |
| | | 0.4 | 0.744 | 0.728 | 0.747 | 0.748 | 0.948 | 0.942 | 0.915 | 0.914 |
| | | 0.8 | 0.437 | 0.446 | 0.397 | 0.392 | 0.861 | 0.855 | 0.678 | 0.676 |
| | 10 | 0 | 0.695 | 0.672 | 0.681 | 0.692 | 0.958 | 0.957 | 0.960 | 0.959 |
| | | 0.4 | 0.611 | 0.595 | 0.616 | 0.622 | 0.954 | 0.949 | 0.926 | 0.925 |
| | | 0.8 | 0.345 | 0.367 | 0.332 | 0.331 | 0.791 | 0.799 | 0.693 | 0.691 |
| | 15 | 0 | 0.539 | 0.511 | 0.524 | 0.536 | 0.954 | 0.951 | 0.956 | 0.954 |
| | | 0.4 | 0.476 | 0.454 | 0.477 | 0.484 | 0.938 | 0.933 | 0.914 | 0.912 |
| | | 0.8 | 0.277 | 0.304 | 0.269 | 0.267 | 0.664 | 0.675 | 0.614 | 0.612 |
| FDR | 5 | 0 | 0.059 | 0.060 | 0.052 | 0.057 | 0.041 | 0.043 | 0.040 | 0.041 |
| | | 0.4 | 0.073 | 0.084 | 0.085 | 0.093 | 0.046 | 0.051 | 0.074 | 0.075 |
| | | 0.8 | 0.200 | 0.157 | 0.298 | 0.304 | 0.071 | 0.080 | 0.246 | 0.248 |
| | 10 | 0 | 0.045 | 0.044 | 0.038 | 0.043 | 0.030 | 0.031 | 0.029 | 0.030 |
| | | 0.4 | 0.062 | 0.062 | 0.067 | 0.074 | 0.032 | 0.036 | 0.052 | 0.053 |
| | | 0.8 | 0.178 | 0.121 | 0.222 | 0.226 | 0.055 | 0.054 | 0.176 | 0.177 |
| | 15 | 0 | 0.069 | 0.065 | 0.060 | 0.066 | 0.037 | 0.039 | 0.036 | 0.037 |
| | | 0.4 | 0.101 | 0.091 | 0.103 | 0.111 | 0.042 | 0.047 | 0.065 | 0.067 |
| | | 0.8 | 0.245 | 0.152 | 0.273 | 0.279 | 0.087 | 0.080 | 0.219 | 0.220 |

# Chapter 4

# A Group-Comparison Procedure for Partial Correlations of a Stationary Multivariate Time Series

## 4.1 Background

The analysis of multi-subject fMRI data is increasingly important to facilitate understanding of neurodegenerative and psychiatric diseases (Craddock et al., 2013). One approach for examining differences in the fMRI data collected between clinically relevant groups is to describe alterations in FC of the brain. In this setting, many graphical modeling approaches have been proposed for detecting group-level differences in FC. Zalesky et al. (2010) proposed a network based statistic for group-level inference of FC that reduces the family-wise error rate while maintaining power by accounting for dependencies between differential edges in group-level networks. Also using a network based approach, Chen et al. (2020*a*) proposed a method for testing for topological differences in FC between clinically relevant groups via subnetwork identification. Chen et al. (2020*b*) proposed a novel Bayesian nonparametric model which uses the network topological structure based on the empirical covariances between connectivity edges for

improved inference. Using a mixed-effects model, Simpson and Laurienti (2015) modeled both the strength and probability of edges between ROIs in a graphical modeling approach. Fiecas et al. (2017) used a mixed effects model for comparing FC between groups of subjects using the marginal correlations between ROIs. Their model also accounted for within group heterogeneity of the autocorrelation structure and the true marginal correlations for each subject. For a more thorough review of group comparison methods for FC analysis of neuroimaging data see Kim et al. (2014) and Ginestet et al. (2014).

An important limitation of many of these approaches is an implicit assumption of independent observations even though fMRI data are known to exhibit autocorrelation (Friston et al., 1995). While the variance components model of Fiecas et al. (2017) does account for autocorrelation present in the data when conducting group-level inference, it uses an asymptotic covariance estimator for the marginal correlations, but not partial correlations. Smith et al. (2011) found under various settings for resting-state fMRI data that for detecting true differences in FC connections the partial correlation tends to outperform other measures of association such as the marginal correlation coefficient and coherence. However, others have found that it depends on the true sparsity structure of the group-level FC networks (Kim et al., 2015).

In this chapter, we propose a hierarchical model for modeling group-level partial correlations for multi-subject fMRI data based on the theoretical results presented in Chapter 3. Based on this hierarchical model, we propose a novel testing procedure for two-group comparisons of group-level FC. Similar to the variance components model of Fiecas et al. (2017), our model allows for within-group heterogeneity in terms of both the autocorrelation structure as well as the true correlations for each individual. The novelty in our proposed framework is that we are conducting inference based on the partial correlations rather than the marginal correlations by using our asymptotic covariance estimator derived in Chapter 3. We demonstrate the competitive performance of our modeling framework compared to other approaches in terms of power while maintaining nominal false positive rates through simulations.

For the rest of the chapter, we start with a presentation of our hierarchical model in Section 4.2. Then, we provide an algorithm for optimizing our corresponding model likelihood in Section 4.2.2. In Section 4.3, we propose an asymptotic Wald test for testing

an overall difference in group-level FC between two groups, and in Section 4.4.2 we describe alternative testing approaches. In Section 4.4 we describe simulations conducted to explore the relative performance of our proposed Wald test compared to competing methods, finding improved power across most settings while maintaining nominal false positive rates. In Section 4.5 we analyze a resting-state fMRI data set collected on participants diagnosed with schizophrenia and healthy controls. Lastly, in Section 4.6 we conclude with a discussion of our conclusions and future directions.

## 4.2 Method

### 4.2.1 A Hierarchical Model for the Partial Correlations of a Multivariate Gaussian Process

Suppose we have fMRI data collected from $G = 2$ groups each with $K_g$ subjects. For each subject, we have $n_{gk}$ observations or time points measured at $p$ ROIs. We let $y_{gkjt}$ denote the $t^{th}$ observation or time point of the $j^{th}$ ROI for the $k^{th}$ subject from the $g^{th}$ group for $g = 1, 2$, $k = 1, \ldots, K_g$, $j = 1, \ldots, p$, and $t = 1, \ldots, n_{gk}$. Using partial correlations to describe FC for each group, we assume a hierarchical structure for the data to allow for similar but different subject-level FC within each group which is illustrated in Figure 4.1. As depicted in Figure 4.1, we let $\{\boldsymbol{\rho_{0g}}\}_{g=1}^{G}$ be the vectors of true group-level partial correlations of length $q = p(p-1)/2$ which describe the average or group-level FC for each group. To account for inter-subject heterogeneity within each group, we assume that $\{\{\boldsymbol{\rho_{gk}} \overset{ind}{\sim} \mathcal{N}_q(\boldsymbol{\rho_{0g}}, \boldsymbol{\Sigma_{0g}})\}_{k=1}^{K_g}\}_{g=1}^{G}$. That is, we assume that each individual's vector of true partial correlations is normally distributed with mean being the group-level partial correlations. Lastly, to describe sampling variability in the empirical partial correlations for each individual, we assume that $\{\{\mathbf{r_{gk}}|\boldsymbol{\rho_{gk}} \overset{ind}{\sim} \mathcal{N}_q(\boldsymbol{\rho_{gk}}, \tilde{\boldsymbol{\Sigma}}_{\mathbf{gk}})\}_{k=1}^{K_g}\}_{g=1}^{G}$. That is, we assume that each individual's partial correlations are normally distributed with their own subject-specific mean and covariance matrix. The subject-specific mean allows for within-group heterogeneity in terms of FC, while the subject-specific covariance matrix allows for unique autocorrelation structures for each individual which in turn affect the uncertainty in the empirical partial correlations.

| Level | Model | Interpretation |
|---|---|---|
| Group | $\{\boldsymbol{\rho_{0g}}\}_{g=1}^{G}$ | Set of $G$ group-level partial correlations describing group-level FC |
| | $\downarrow$ | |
| Individual | $\{\{\boldsymbol{\rho_{gk}} \overset{ind}{\sim} \mathcal{N}_q(\boldsymbol{\rho_{0g}}, \mathbf{\Sigma_{0g}})\}_{k=1}^{K_g}\}_{g=1}^{G}$ | Set of $K$ individual-level partial correlations describing unique individual-level FC |
| | $\downarrow$ | |
| Observation | $\{\{\mathbf{r_{gk}}\|\boldsymbol{\rho_{gk}} \overset{ind}{\sim} \mathcal{N}_q(\boldsymbol{\rho_{gk}}, \tilde{\mathbf{\Sigma}}_{\mathbf{gk}})\}_{k=1}^{K_g}\}_{g=1}^{G}$ | Set of empirical partial correlations from $n_k$ observations of $p$ variables for subject $k$ of group $g$ where $p$ is the number of ROIs |

Figure 4.1: A normal-normal hierarchical model for the partial correlations of a multivariate Gaussian process.

Based on our proposed framework, our model likelihood for the subject-level empirical partial correlations is

$$L(\mathbf{\Theta}|\{y_{gkjt}\}) = \prod_{g=1}^{G}\prod_{k=1}^{K_g} f(\mathbf{r_{gk}}|\boldsymbol{\rho_{gk}}, \tilde{\mathbf{\Sigma}}_{\mathbf{gk}}) f(\boldsymbol{\rho_{gk}}|\boldsymbol{\rho_{0g}}, \mathbf{\Sigma_{0g}}),$$

where $\mathbf{\Theta} = \{\boldsymbol{\rho_{gk}}, \boldsymbol{\rho_{0g}}, \mathbf{\Sigma_{0g}}\}$ for $g = 1, 2, \ldots, G$ and $k = 1, 2, \ldots, K_g$, and $f(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Sigma})$ is the probability density function of a multivariate normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. Thus, our model log-likelihood is

$$\ell(\mathbf{\Theta}|\{y_{gkjt}\}) = \sum_{g=1}^{G}\sum_{k=1}^{K_g} \log\left(f(\mathbf{r_{gk}}|\boldsymbol{\rho_{gk}}, \tilde{\mathbf{\Sigma}}_{\mathbf{gk}})\right) + \log\left(f(\boldsymbol{\rho_{gk}}|\boldsymbol{\rho_{0g}}, \mathbf{\Sigma_{0g}})\right). \tag{4.1}$$

### 4.2.2 Computational Algorithm

Given an observed data set, we can calculate the empirical partial correlations for each subject and the estimated $\tilde{\mathbf{\Sigma}}_{\mathbf{gk}}$ using the covariance estimator derived in Chapter 3. We

now seek to maximize the model log-likelihood in Equation (4.1) with respect to $\boldsymbol{\Theta}$. In particular, we obtain estimates for $\boldsymbol{\Theta} = \{\boldsymbol{\rho_{gk}}, \boldsymbol{\rho_{0g}}, \mathbf{\Sigma_{0g}}\}$ using block-coordinate descent (Wright, 2015). Our optimization algorithm is as follows:

1. Initialize the $\boldsymbol{\rho_{0g}}$ and $\mathbf{\Sigma_{0g}}$. We suggest initializing $\boldsymbol{\rho_{0g}}$ as $\boldsymbol{\rho}_{\mathbf{0g}}^{(0)} = \frac{1}{K_g}\sum_{k=1}^{K_g} \mathbf{r_{gk}}$, each $\boldsymbol{\rho_{gk}}$ as $\boldsymbol{\rho}_{\mathbf{gk}}^{(0)} = \frac{1}{2}(\mathbf{r_{gk}} + \boldsymbol{\rho}_{\mathbf{0g}}^{(0)})$ and each $\mathbf{\Sigma_{0g}}$ as $\mathbf{\Sigma_{0g}}^{(0)} = \frac{1}{K_g}\sum_{k=1}^{K_g}(\boldsymbol{\rho}_{\mathbf{gk}}^{(0)} - \boldsymbol{\rho}_{\mathbf{0g}}^{(0)})(\boldsymbol{\rho}_{\mathbf{gk}}^{(0)} - \boldsymbol{\rho}_{\mathbf{0g}}^{(0)})^T$ for $g = 1, 2, \ldots, G$ and $k = 1, \ldots, K_g$.

2. Update each $\boldsymbol{\rho_{gk}}$ by calculating

$$\boldsymbol{\rho}_{\mathbf{gk}}^{(m+1)} = \left(\tilde{\mathbf{\Sigma}}_{\mathbf{gk}}^{-1} + \mathbf{\Sigma_{0g}}^{-1(m)}\right)^{-1}\left(\tilde{\mathbf{\Sigma}}_{\mathbf{gk}}^{-1}\mathbf{r_{gk}} + \mathbf{\Sigma_{0g}}^{-1(m)}\boldsymbol{\rho}_{\mathbf{0g}}^{(m)}\right)$$

   for $g = 1, 2, \ldots, G$ and $k = 1, \ldots, K_g$.

3. Update each $\boldsymbol{\rho_{0g}}$ by calculating

$$\boldsymbol{\rho}_{\mathbf{0g}}^{(m+1)} = \frac{1}{K_g}\sum_{k=1}^{K_g}\boldsymbol{\rho}_{\mathbf{gk}}^{(m+1)}$$

   for $g = 1, \ldots, G$.

4. Update each $\mathbf{\Sigma_{0g}}$ by calculating

$$\mathbf{\Sigma_{0g}}^{(m+1)} = \frac{1}{K_g}\sum_{k=1}^{K_g}(\boldsymbol{\rho}_{\mathbf{gk}}^{(m+1)} - \boldsymbol{\rho}_{\mathbf{0g}}^{(m+1)})(\boldsymbol{\rho}_{\mathbf{gk}}^{(m+1)} - \boldsymbol{\rho}_{\mathbf{0g}}^{(m+1)})^T$$

   for $g = 1, \ldots, G$.

5. Repeat steps 2 through 4 until convergence defined as

$$\left(\ell(\boldsymbol{\Theta}^{(m+1)}|\{y_{gkjt}\}) - \ell(\boldsymbol{\Theta}^{(m)}|\{y_{gkjt}\})\right)/\ell(\boldsymbol{\Theta}^{(m)}|\{y_{gkjt}\}) < \varepsilon,$$

   where $\varepsilon$ is a small positive number. Thus, convergence is attained when the percent increase in the model log-likelihood is less than $100 \times \varepsilon$, where we used $\varepsilon = 0.01$ corresponding to a 1% increase.

We note that the updates for the $\boldsymbol{\rho_{gk}}$ in step 2 are a weighted-average of the group-level partial correlations $\boldsymbol{\rho_{0g}}$ and the empirical individual-level partial correlations $\mathbf{r_{gk}}$. We also note for the update of the group-level covariance matrices $\mathbf{\Sigma_{0g}}$ in step 4 that the estimate of $\mathbf{\Sigma_{0g}}$ is positive definite if and only if the number of subjects in the group, $K_g$, is greater than the number of unique partial correlations, $q$. However, we suggest use of a regularized covariance estimator such as the graphical lasso or other shrinkage estimators to ensure that the resulting covariance matrices are positive definite (Friedman et al., 2007; Schäfer and Strimmer, 2005).

## 4.3 Wald Test

We consider the setting in which we have $G = 2$ groups each with $K_g$ subjects. Based on the calculated empirical partial correlations $\mathbf{r_{gk}}$, the asymptotic covariance estimators $\tilde{\mathbf{\Sigma}}_{\mathbf{gk}}$, and the estimated $\mathbf{\Theta}$ parameters obtained above, we propose a novel Wald hypothesis testing procedure to test the equality of the set of $q$ group-level partial correlations i.e., $H_0$: $\boldsymbol{\rho_{01}} = \boldsymbol{\rho_{02}}$ vs. $H_A$: $\boldsymbol{\rho_{01}} \neq \boldsymbol{\rho_{02}}$.

Based on our model above and letting $\bar{\mathbf{r}}_{0g} = \frac{1}{K_g} \sum_{k=1}^{K_g} \mathbf{r_{gk}}$, it follows that $\bar{\mathbf{r}}_{0g} \overset{ind}{\sim} \mathcal{N}_q(\boldsymbol{\rho_{0g}}, \tilde{\mathbf{\Sigma}}_{0g} + \frac{1}{K_g}\mathbf{\Sigma_{0g}})$, where $\boldsymbol{\rho_{0g}} \in \mathbb{R}^q$ is the vector of unique group-level partial correlations, and $\tilde{\mathbf{\Sigma}}_{0g} = \frac{1}{K_g^2} \sum_{k=1}^{K_g} \tilde{\mathbf{\Sigma}}_{\mathbf{gk}}$ for $g = 1, 2, \ldots, G$. Hence, the test statistic for the Wald test testing the equality of all $q$ partial correlations for $G = 2$ groups is then calculated as

$$T_w = (\bar{\mathbf{r}}_{01} - \bar{\mathbf{r}}_{02})^T \left(\tilde{\mathbf{\Sigma}}_{01} + \mathbf{\Sigma_{01}}/K_1 + \tilde{\mathbf{\Sigma}}_{02} + \mathbf{\Sigma_{02}}/K_2\right)^{-1} (\bar{\mathbf{r}}_{01} - \bar{\mathbf{r}}_{02}),$$

which asymptotically follows a $\chi^2$ distribution with $q$ degrees of freedom under the null of the group-level partial correlations being equal across the two groups. Thus, for an $\alpha$-level significance test we reject the null hypothesis of $\boldsymbol{\rho}_{01} = \boldsymbol{\rho}_{02}$ if $T_w \geq \chi^2_{(q;1-\alpha)}$ where $\chi^2_{(q;1-\alpha)}$ is the $100 \times (1-\alpha)$th percentile of a Chi-squared distribution with $q$ degrees of freedom. However, we found empirically that the test statistic was inflated when the number of subjects in each group was not very large. This was likely due to the asymptotic covariance estimates for the group-level covariance matrices $\mathbf{\Sigma_{0g}}$ being poorly estimated based on a smaller amount of data. Using a regularized

covariance estimator when the number of subjects in the group is small could potentially alleviate this issue (Schäfer and Strimmer, 2005; Friedman et al., 2007). To maintain a false positive rate closer to the nominal level, we use a permutation testing procedure using our proposed Wald test statistic. For the permutation test, we permute the group memberships of each of the $K_1 + K_2$ number of individuals $P$ times to obtain a distribution of $P$ permutation test statistics. Then, we reject the null hypothesis if $T_w \geq \chi^{2*}_{(1-\alpha)}$, where $\chi^{2*}_{(1-\alpha)}$ is the 95th percentile of the $P$ permutation test statistics for $\alpha = 0.05$.

## 4.4 Simulations

### 4.4.1 Setup

We implemented simulations to assess the performance of our proposed inference procedures for group comparisons of partial correlations. We considered two different numbers of variables ($p \in \{5, 10\}$), one sample size ($N = 255$), three different levels of autocorrelation ($\phi \in \{0, 0.40, 0.80\}$), and setting $q - p$ of the group-level partial correlations to be 0 and $p$ of the partial correlations to be 0.20 identically across the two groups in the null setting. For the non-null setting of the groups having different partial correlations, we reduced the magnitude of all nonzero partial correlations for one group by 30%, so nonzero partial correlations for the second group were set to be 0.14 instead of 0.20. Data were generated with either 16 and 19 individuals or 59 and 54 individuals to explore settings with a smaller and larger number of subjects in each group. A total of 500 data sets were generated for each setting with $N$ observations of a $p$-variate first-order autoregressive (AR) model with correlation parameter $\phi$.

There were two forms of heterogeneity that we introduced within each group: heterogeneity in the amount of autocorrelation in the data and heterogeneity in the true partial correlations for each subject. To induce heterogeneity for the autocorrelation present for each individual, we generated the autocorrelation parameter for each subject from a $\mathcal{N}(\phi, \sigma^2 = 0.05^2)$ distribution. Thus, it is possible that some subjects' data were generated from an AR model with an autocorrelation parameter larger than 1, violating the assumption of an ergodic time series. However, this was unlikely to happen often, even in the $\phi = 0.80$ setting, and would only attest to the robustness of our model when

misspecified in this manner. In addition to autocorrelation, we also introduced heterogeneity in each individual's true partial correlation $\boldsymbol{\rho_{gk}}$ by setting them equal to the true group-level partial correlations $\boldsymbol{\rho_{0g}}$ plus some noise generated from a $\mathcal{N}(0, \sigma^2 = 0.10^2)$ distribution. All of the parameter settings for our simulations were selected to match estimates from our motivating fMRI data set. Simulations were conducted using R (R Core Team, 2021).

### 4.4.2 Comparison Methods

For comparison approaches for testing the equality of partial correlations across two groups, we considered four other permutation testing approaches: a Wald test based on a variance components model (Fiecas et al., 2017), a Wald test using a block-bootstrap covariance estimator (Carlstein, 1986), a multivariate Hotelling $t$-test (Hotelling, 1931), and a permutation testing procedure based on the mean squared differences (MSD) of the partial correlations between the two groups. For each of the approaches considered, a permutation testing procedure based on $P = 500$ permutations of each individual's group membership was used to alleviate inflated false positive rates. The 95th percentile of each test's respective permutation distribution was used to determine statistical significance at the 5% significance level.

In general, the two-sample Hotelling $t$-test is not defined when the number of components of the vector of interest, $q$, exceeds $K_1 + K_2 - 1$. However, using the *Hotelling* package in R we implemented a shrinkage estimator for the covariance matrix ensuring the Hotelling test statistic is always defined (Curran, 2018). For the block-bootstrap approach, we obtained a block-bootstrap covariance estimator for the difference $\bar{\mathbf{r}}_{01} - \bar{\mathbf{r}}_{02}$ based on $B = 500$ bootstrap samples. The block length for the procedure was selected for each individual via an automatic selection algorithm for time series data (Politis and White, 2004). This covariance estimator was constant across the 500 permutations for the permutation testing procedure, but the difference between the estimated group-level correlation vectors varied across the permutations.

### 4.4.3 Results

Results for 500 simulations comparing the performance of our proposed Wald testing procedure to the other competing methods are summarized in Tables 4.1 and 4.2 and Figures 4.2 and 4.3. Tables 4.1 and 4.2 display the relative performances of the methods for testing for any difference in the group-level partial correlations evaluated by power and false positive rates for detecting true group-level differences in terms of partial correlations. Figures 4.2 and 4.3 display receiver operating characteristic (ROC) curves conveying the tradeoff between true and false positive rates for the methods considered. The ROC curves were created using the *ggplot2* package in R (Wickham, 2016).

From Figures 4.2 and 4.3, we observe that across all settings the methods tend to perform worse as the amount of autocorrelation increases as expected. We note that since the number of differences between each group's set of partial correlations increases linearly with the number of variables the methods are better able to detect differences between the groups for $p = 10$ variables as compared to $p = 5$ variables. We also note that all methods perform quite similarly in the independence case ($\phi = 0$) and when a small amount of autocorrelation is present ($\phi = 0.40$). However, when a high amount of autocorrelation is present ($\phi = 0.80$) with a smaller number of subjects in each group, our Normal-Normal Wald test and the VCM approaches outperform the other competing methods.

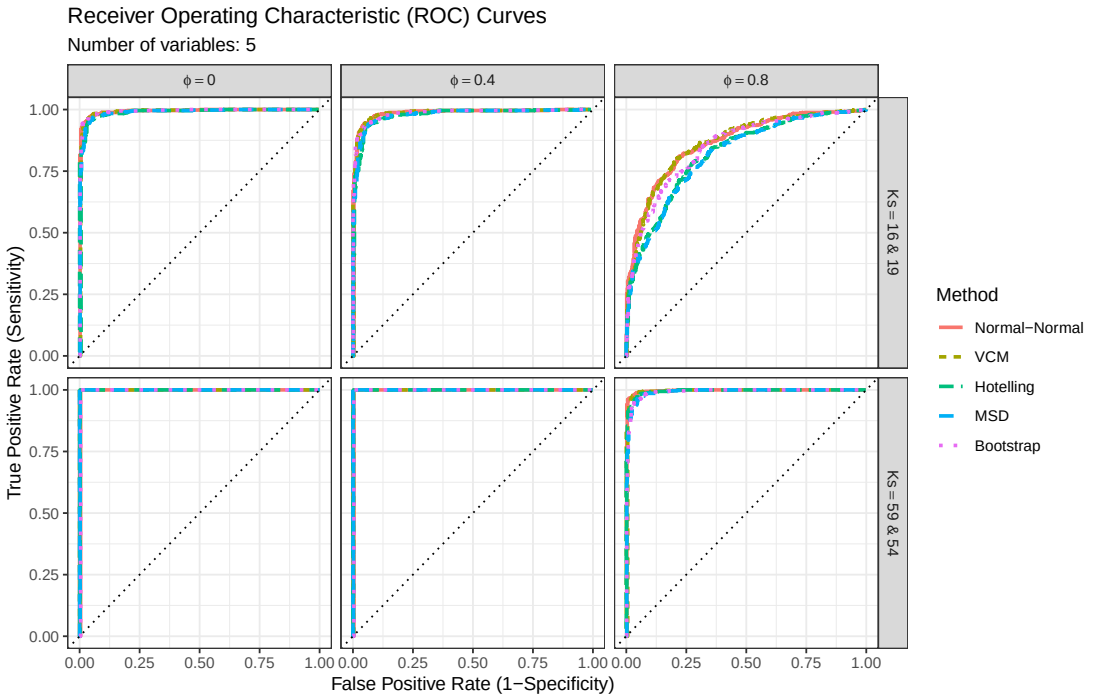


Figure 4.2: Receiver operating characteristic (ROC) curves displaying the trade-off between the true and false positive rates for detecting any difference in the set of group-level partial correlations for $p = 5$ variables.

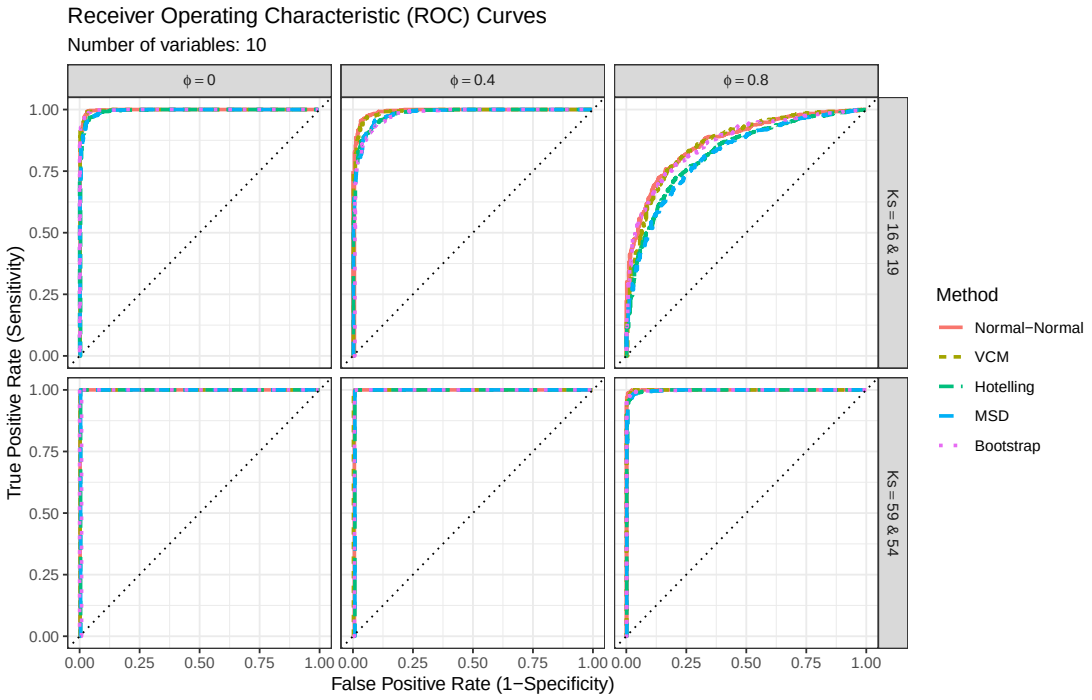


Figure 4.3: Receiver operating characteristic (ROC) curves displaying the trade-off between the true and false positive rates for detecting any difference in the set of group-level partial correlations for $p = 10$ variables.

Table 4.1 displays results for simulations with $p = 5$ variables for each subject and also shows the better performance overall by the Normal-Normal Wald test and the VCM compared to the other competing methods. The Normal-Normal Wald test and the VCM achieve the same or higher power than the other competing methods across all settings, most noticeably when a high amount of autocorrelation is present. Although all methods attained similar false positive rates due to using permutation testing procedures, the Normal-Normal model and the VCM approaches achieved higher power and thus better performance overall. This is likely since the Normal-Normal and VCM approaches account for the heterogeneity within each group in terms of autocorrelation and the true partial correlations for each subject, whereas the Hotelling and MSD methods do not. The block-bootstrap does account for autocorrelation, which is why

it has a slight improvement in performance compared to the Hotelling and MSD approaches when autocorrelation is present in the small sample size setting, but it is still outperformed by the Normal-Normal and VCM models. The only potentially meaningful difference between the Normal-Normal model and the VCM in terms of power is in the high autocorrelation ($\phi = 0.80$) setting with 16 and 19 subjects in each group, in which the Normal-Normal model achieved slightly higher power. We observed in simulations that this was likely due to the Normal-Normal model obtaining a better estimate of the difference between the group-level partial correlation vectors than the VCM model, which uses a weighted average of the empirical partial correlations within each group. Thus, if more heterogeneity were present in each subjects' autocorrelation or in the number of observations, we would expect the VCM to potentially outperform the Normal-Normal model.

Table 4.2 displays results for simulations with $p = 10$ variables for each subject. As with $p = 5$ variables, all methods perform similarly in terms of false positive rates, achieving values close to 0.05 across all settings due to all methods using a permutation testing procedure. Although the methods all perform somewhat similarly overall as displayed in Figures 4.2 and 4.3, a general trend across the settings again is that the Normal-Normal, VCM, and block-bootstrap approaches tend to outperform the Hotelling and MSD approaches in terms of power, indicating better overall performance especially in the high autocorrelation settings when each group has a smaller number of subjects. Hence, when a high amount of autocorrelation is present, we suggest use of the Normal-Normal, VCM, and block-bootstrap approaches for group comparisons of partial correlations. As with $p = 5$ variables, again the only noticeable difference in the performance of the Normal-Normal model and the VCM is in the high-autocorrelation setting ($\phi = 0.80$) when a smaller number of subjects are in each group. In this case the Normal-Normal model achieved roughly 10% higher power than the VCM, again likely due to better estimates of the difference between the group-level correlation vectors.

## 4.5 Data Analysis

Altered FC patterns among those with schizophrenia and related disorders have been found in a number of previous studies (Giraldo-Chica and Woodward, 2016). In a study

by Woodward and Heckers (2015), participants with chronic psychosis were found to have reduced FC of the thalamus with cerebellum and inferior parietal lobule regions. Giraldo-Chica et al. (2018) found increased thalamic anatomical connectivity with occipital regions among those with schizophrenia, and Li et al. (2019) found increased connectivity among the right medial prefrontal cortex and right superior temporal cortex for those diagnosed with schizophrenia compared to healthy controls. Woodward et al. (2012) found reduced FC between the thalamus and the prefrontal cortex among those with schizophrenia compared to healthy controls.

Our motivating data set consists of resting-state fMRI data collected on 64 total participants: 29 healthy controls and 35 participants diagnosed with chronic or first-episode schizophrenia. The mean age of participants diagnosed with schizophrenia was 29.4 years (range 20 to 53 years) and 35.6 years (range 18 to 59 years) for the healthy controls. Resting-state fMRI data were acquired using a Siemens Trio 3T scanner with the following acquisition parameters: gradient-echo echo-planar imaging 260 volumes, repetition time = 2 s, echo time = 30 ms, flip angle = $90^\circ$, 34 contiguous ACPC aligned axial slices, voxel size = $3.4 \times 3.4 \times 4.0$ mm, matrix = $64 \times 64 \times 34$. Imaging data were preprocessed using ICA-AROMA with the software tool fMRIPrep (Pruim et al., 2015; Esteban et al., 2019). The initial 5 volumes were removed before analysis, leaving 255 total volumes. After preprocessing, ROIs were extracted using an atlas created by Desikan et al. (2006). In an attempt to maintain statistical power, we only used a subset of these ROIs in our analysis. Based on previous studies, we selected 14 of the ROIs to compare participants diagnosed with schizophrenia and healthy controls in terms of group-level FC (Anticevic et al., 2014; Woodward and Heckers, 2015; Giraldo-Chica et al., 2018; Li et al., 2019; Woodward et al., 2012). The selected ROIs and their corresponding abbreviations are provided in Table 4.3.

We applied our proposed Wald testing procedure to our motivating data set to test for an overall difference in group-level FC between those with schizophrenia and healthy controls. Based on our permutation testing procedure, we did not find a statistically significant difference (permutation p-value = 0.360). This was consistent with the results of the other competing methods since all other methods obtained a p-value greater than or equal to 0.320. As displayed in Figures 4.4 and 4.5, the estimated group-level partial correlations do indeed look quite similar, consistent with the finding of

the methods. However, not finding a statistically significant difference in the overall group-level partial correlations is inconsistent with previous studies. This could be due to an insufficient number of subjects for simultaneously testing 91 correlations or the high level of autocorrelation in the data. Either of these aspects could potentially limit our statistical power as illustrated by our simulation results.

Although there was not a statistically significant overall difference in group-level FC, Figure 4.6 displays differences in FC connections for the control group compared to the schizophrenia group. We observe that the largest difference in strength was for the connection between the right hemisphere thalamus (TSR) and the left hemisphere lateral occipital cortex (LOL). This was the largest pairwise difference between the groups and is consistent with the results of Giraldo-Chica et al. (2018) who also found differential connectivity between thalamic and occipital regions among participants diagnosed with schizophrenia.

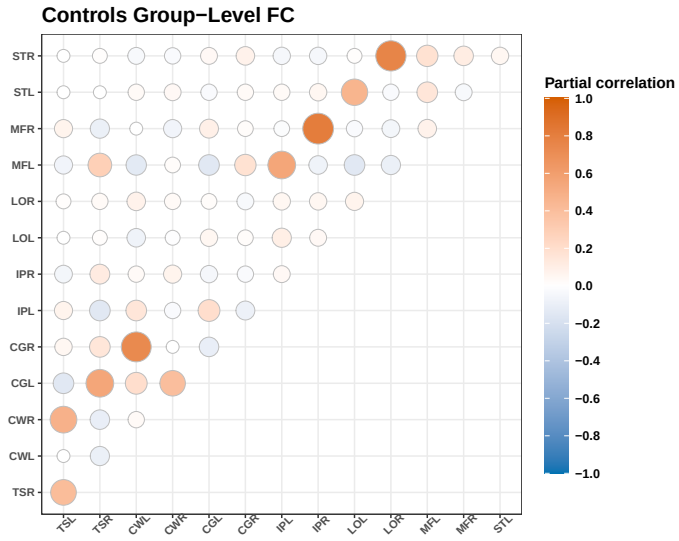


Figure 4.4: Estimated group-level partial correlations for the healthy controls. Full descriptions for the 14 selected ROIs are provided in Table 4.3.

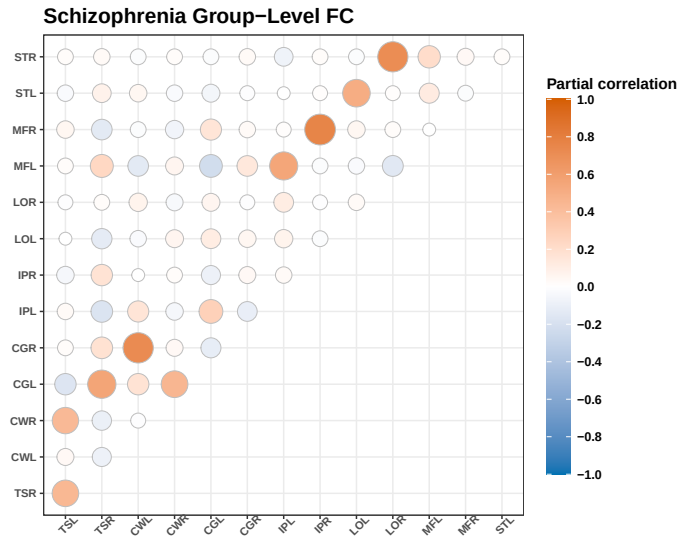


Figure 4.5: Estimated group-level partial correlations for the participants diagnosed with schizophrenia. Full descriptions for the 14 selected ROIs are provided in Table 4.3.

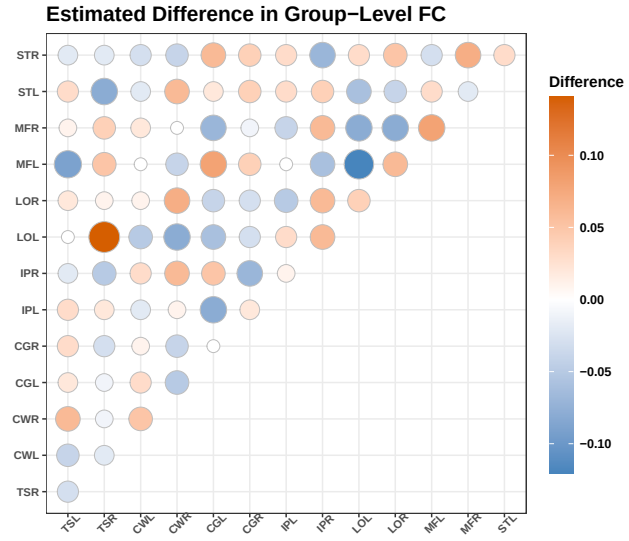


Figure 4.6: Difference in estimated group-level partial correlations between healthy controls and the schizophrenia group. Full descriptions for the 14 selected ROIs are provided in Table 4.3.

## 4.6 Discussion

In this chapter, we proposed a hierarchical model for comparing multiple subjects in terms of group-level FC that directly accounts for autocorrelation in the data and heterogeneity within each group in terms of FC. We showed how accounting for autocorrelation can lead to increased power when using permutation testing procedures compared to competing approaches, while still maintaining a nominal false positive rate. We did not find a statistically significant difference between healthy controls and participants diagnosed with schizophrenia in terms of group-level FC for our real data set. However, we saw from pairwise differences of group-level partial correlations that potential differences in FC did exist. Thus, our inference procedures may have been under-powered due to a high amount of autocorrelation in the data or testing too many correlations simultaneously. A larger number of subjects or using a smaller number of ROIs could increase our power to detect any difference in group-level FC.

Although we provided a flexible model for conducting group-level inference of partial correlations, our modeling framework is only feasible for a smaller number of variables ($p < 30$) due to computational constraints. This is since the number of estimated parameters for the asymptotic covariance estimator from Chapter 3 is quartic ($p^4$) in the number of variables. A more computationally efficient estimator for the covariance matrix would allow analyzing fMRI data with hundreds of ROIs which may provide a more detailed picture of group-level differences in FC. Another aspect that could be explored is the use of a regularized precision matrix estimator for obtaining estimated partial correlations (Kim et al., 2015). We did not use regularization in our simulations or analysis, but this could be potentially useful in higher-dimensional settings. To extend our modeling framework, one could allow for a dynamic correlation structure across time. A model that allowed for time varying correlations would facilitate FC analysis for task-based fMRI studies as well.

Table 4.1: Performances across 500 simulations for normal-normal hierarchical model (Normal-Normal), variance components model (VCM), a block-bootstrap based permutation test (Bootstrap), Hotelling $t$-test (Hotelling), and mean squared difference permutation test (MSD). Results are for $p = 5$ variables and 16 and 19 or 59 and 54 subjects in each group, with 255 observations for each of $p$ variables for each subject and autocorrelation parameter $\phi$. True positive rate (TPR) and false positive rate (FPR) are for a two group comparison testing a true difference in any of the group-level partial correlations.

| $K$s | $\phi$ | Metric | Normal-Normal | VCM | Bootstrap | Hotelling | MSD |
|---|---|---|---|---|---|---|---|
| 16 & 19 | 0 | TPR | 0.976 | 0.982 | 0.976 | 0.958 | 0.954 |
| | | FPR | 0.054 | 0.052 | 0.050 | 0.048 | 0.044 |
| | 0.4 | TPR | 0.940 | 0.938 | 0.924 | 0.882 | 0.886 |
| | | FPR | 0.046 | 0.050 | 0.044 | 0.044 | 0.038 |
| | 0.8 | TPR | 0.486 | 0.462 | 0.428 | 0.372 | 0.354 |
| | | FPR | 0.038 | 0.048 | 0.042 | 0.042 | 0.034 |
| 59 & 54 | 0 | TPR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | FPR | 0.048 | 0.058 | 0.054 | 0.042 | 0.052 |
| | 0.4 | TPR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | FPR | 0.042 | 0.052 | 0.046 | 0.040 | 0.046 |
| | 0.8 | TPR | 0.988 | 0.992 | 0.958 | 0.976 | 0.966 |
| | | FPR | 0.054 | 0.054 | 0.046 | 0.048 | 0.050 |

Table 4.2: Performances across 500 simulations for normal-normal hierarchical model (Normal-Normal), variance components model (VCM), a block-bootstrap based permutation test (Bootstrap), Hotelling $t$-test (Hotelling), and mean squared difference permutation test (MSD). Results are for $p = 10$ variables and 16 and 19 or 59 and 54 subjects in each group, with 255 observations for each of $p$ variables for each subject and autocorrelation parameter $\phi$. True positive rate (TPR) and false positive rate (FPR) are for a two group comparison testing a true difference in any of the group-level partial correlations.

| $K$s | $\phi$ | Metric | Normal-Normal | VCM | Bootstrap | Hotelling | MSD |
|---|---|---|---|---|---|---|---|
| 16 & 19 | 0 | TPR | 0.996 | 0.992 | 0.994 | 0.970 | 0.972 |
| | | FPR | 0.050 | 0.060 | 0.054 | 0.054 | 0.054 |
| | 0.4 | TPR | 0.972 | 0.970 | 0.906 | 0.916 | 0.908 |
| | | FPR | 0.058 | 0.058 | 0.070 | 0.072 | 0.062 |
| | 0.8 | TPR | 0.532 | 0.424 | 0.380 | 0.420 | 0.394 |
| | | FPR | 0.050 | 0.044 | 0.020 | 0.054 | 0.054 |
| 59 & 54 | 0 | TPR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | FPR | 0.044 | 0.050 | 0.056 | 0.052 | 0.048 |
| | 0.4 | TPR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | FPR | 0.052 | 0.056 | 0.048 | 0.058 | 0.054 |
| | 0.8 | TPR | 0.998 | 0.996 | 0.970 | 0.988 | 0.986 |
| | | FPR | 0.038 | 0.034 | 0.010 | 0.042 | 0.044 |

Table 4.3: Selected ROIs and their abbreviations.

| Region | Abbreviation |
|---|---|
| Thalamus (left hemisphere) | TSL |
| Thalamus (right hemisphere) | TSR |
| Cerebellum white matter (left hemisphere) | CWL |
| Cerebellum white matter (right hemisphere) | CWR |
| Cerebellum grey matter (left hemisphere) | CGL |
| Cerebellum grey matter (right hemisphere) | CGR |
| Inferior parietal cortex (left hemisphere) | IPL |
| Inferior parietal cortex (right hemisphere) | IPR |
| Lateral occipital cortex (left hemisphere) | LOL |
| Lateral occipital cortex (right hemisphere) | LOR |
| Medial orbital frontal cortex (left hemisphere) | MFL |
| Medial orbital frontal cortex (right hemisphere) | MFR |
| Superior temporal gyrus (left hemisphere) | STL |
| Superior temporal gyrus (right hemisphere) | STR |

# Chapter 5

# Discussion

In this thesis, we proposed multiple approaches for modeling and analyzing FC for multi-subject resting-state fMRI data. In Chapter 2 we proposed a penalized model-based clustering method, the RCCM, for fMRI data collected on multiple participants which concurrently estimates FC and groups subjects based on FC of the brain. We showed the competitive performance of RCCM compared to current methods which implement estimation and clustering sequentially. Then, in Chapter 3 we considered the issue of autocorrelation in fMRI data when testing partial correlation coefficients. Specifically, we presented an asymptotic distribution and covariance structure for the empirical partial correlations of a multivariate Gaussian process, facilitating inference of partial correlations for FC analysis. We also showed the competitive performance of our proposed inference procedures compared to current testing approaches such as the block-bootstrap for time series data and naïve approaches assuming independent observations. Lastly, in Chapter 4 we proposed a hierarchical model for comparing group-level FC for multi-subject fMRI data sets based on the partial correlations. Our hierarchical model accounts for both within and between group heterogeneity in terms of FC, and thus outperforms methods that assume identical FC structures within each group.

In future work, we envision developing an unsupervised or supervised method for grouping patients based on FC that directly accounts for temporal correlation commonly exhibited in fMRI data. This could be an extension of the RCCM proposed in

Chapter 2 and our hierarchical model proposed in Chapter 4. Overall, the methods proposed in this work contribute to a vast and continuously growing collection of methods for neuroimaging analysis and will hopefully prove useful for clinical practice moving forward.

# References


Afyouni, S., Smith, S. M. and Nichols, T. E. (2019). Effective degrees of freedom of the pearson's correlation coefficient under autocorrelation. *NeuroImage* **199**, 609–625.

Anticevic, A., Cole, M. W., Repovs, G., Murray, J. D., Brumbaugh, M. S., Winkler, A. M., Savic, A., Krystal, J. H., Pearlson, G. D. and Glahn, D. C. (2014). Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. *Cerebral Cortex* **24**(12), 3116–3130.

Arribas, J. I., Calhoun, V. D. and Adali, T. (2010). Automatic bayesian classification of healthy controls, bipolar disorder, and schizophrenia using intrinsic connectivity maps from fMRI data. *IEEE Transactions on Biomedical Engineering* **57**(12), 2850–2860.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society* **8**(1), 27–41.

Bastos, A. M. and Schoffelen, J. M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience* **9**, 175–175.

Bergsma, W. (2020). A note on the distribution of the partial correlation coefficient with nonparametrically estimated marginal regressions. *arXiv*.

Bien, J. and Tibshirani, R. J. (2011). Sparse estimation of a covariance matrix. *Biometrika* **98**(4), 807–820.

Box, G. E. P. and Newbold, P. (1971). Some comments on a paper of coen, gomme and kendall. *Journal of the Royal Statistical Society. Series A. General* **134**(2), 229–240.

Cai, T., Li, H., Liu, W. and Xie, J. (2016). Joint estimation of multiple high-dimensional precision matrices. *Statistica Sinica* **26**(2), 445–464.

Calhoun, V. D., Adali, T., Pearlson, G. D. and Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human brain mapping* **14**(3), 140–151.

Calhoun, V. D., Liu, J. and Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* **45**(1), S163–S172.

Calhoun, V. D., Maciejewski, P. K., Pearlson, G. D. and Kiehl, K. A. (2008). Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. *Human Brain Mapping* **29**(11), 1265.

Camchong, J., MacDonald, A. W., Bell, C., Mueller, B. A. and Lim, K. O. (2011). Altered functional and anatomical connectivity in schizophrenia. *Schizophrenia Bulletin* **37**(3), 640–650.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics* **14**(3), 1171 – 1179.

Castro, E., Gómez-Verdejo, V., Martínez-Ramón, M., Kiehl, K. A. and Calhoun, V. D. (2014). A multiple kernel learning approach to perform classification of groups from complex-valued fMRI data analysis: Application to schizophrenia. *NeuroImage* **87**, 1–17.

Chen, S., Bowman, F. D. and Xing, Y. (2020*a*). Detecting and testing altered brain connectivity networks with k-partite network topology. *Computational Statistics & Data Analysis* **141**, 109–122.

Chen, S., Xing, Y., Kang, J., Kochunov, P. and Hong, L. E. (2020*b*). Bayesian modeling of dependence in brain connectivity data. *Biostatistics* **21**(2), 269–286.

Craddock, R. C., Jbabdi, S., Yan, C., Vogelstein, J. T., Castellanos, F. X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S. and Milham, M. P. (2013). Imaging human connectomes at the macroscale. *Nature methods* **10**(6), 524–539.

Cramer, E. M. (1974). Brief report: The distribution of partial correlations and generalizations. *Multivariate Behavioral Research* **9**(1), 119–122. PMID: 26828735.

Curran, J. (2018). *Hotelling: Hotelling's $T^2$ Test and Variants*. R package version 1.0-5.

Damos, P. (2016). Using multivariate cross correlations, granger causality and graphical models to quantify spatiotemporal synchronization and causality between pest populations. *BMC ecology* **16**(1), 33–33.

Danaher, P., Wang, P. and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society* **76**(2), 373–397.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W., Gur, R. C. and Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *Neuroimage* **28**(3), 663–668.

de la Fuente, A., Bing, N., Hoeschele, I. and Mendes, P. (2004). Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics (Oxford, England)* **20**(18), 3565–3574.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society* **39**(1), 1–22.

Dennis, E. and Thompson, P. (2014). Functional brain connectivity using fMRI in aging and alzheimer's disease. *Neuropsychology Review* **24**(1), 49–62.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S. et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**(3), 968–980.

Erb, I. (2020). Partial correlations in compositional data analysis. *Applied Computing and Geosciences* **6**, 100026.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A. et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature methods* **16**(1), 111–116.

Fan, J., Liu, H., Wang, W. and Zhu, Z. (2018). Heterogeneity adjustment with applications to graphical model inference. *Electronic Journal of Statistics* **12**(2), 3908–3952.

Ferguson, T. S. (2002). *A course in large sample theory*, Texts in statistical science. Boca Raton, Florida: CRC Press.

Fiecas, M., Cribben, I., Bahktiari, R. and Cummine, J. (2017). A variance components model for statistical inference on functional connectivity networks. *NeuroImage* **149**, 256–266.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10**(4), 507–521.

Fisher, R. A. (1924). The distribution of the partial correlation coefficient. *Metron* **3**, 329–332.

Fornito, A., Zalesky, A., Pantelis, C. and Bullmore, E. T. (2012). Schizophrenia, neuroimaging and connectomics. *Neuroimage* **62**(4), 2296–2314.

Friedman, J., Hastie, T. and Tibshirani, R. (2007, dec). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**(3), 432–441.

Friston, K. J., Frith, C. D. and Frackowiak, R. S. J. (1993*a*). Time-dependent changes in effective connectivity measured with PET. *Human brain mapping* **1**(1), 69–79.

Friston, K. J., Frith, C. D., Liddle, P. F. and Frackowiak, R. S. J. (1993*b*). Functional connectivity: The principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow & Metabolism* **13**(1), 5–14.

Friston, K. J., Holmes, A. P., Poline, J-B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J. and Turner, R. (1995). Analysis of fMRI time-series revisited. *NeuroImage* **2**(1), 45–53.

Gabor, C. and Tamas, N. (2006). The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695.

Gao, C., Zhu, Y., Shen, X. and Pan, W. (2016). Estimation of multiple networks in gaussian mixture models. *Electronic Journal of Statistics* **10**(1), 1133–1154.

Geuter, S., Lindquist, M. A. and Wager, T. D. (2016). *Fundamentals of Functional Neuroimaging*, 4 edition., Cambridge Handbooks in Psychology. Cambridge University Press, p. 41–73.

Ginestet, C. E., Fournel, A. P. and Simmons, A. (2014). Statistical network analysis for functional MRI: summary networks and group comparisons. *Frontiers in computational neuroscience* **8**, 51–51.

Giraldo-Chica, M., Rogers, B. P., Damon, S. M., Landman, B. A. and Woodward, N. D. (2018). Prefrontal-thalamic anatomical connectivity and executive cognitive function in schizophrenia. *Biological Psychiatry* **83**(6), 509–517.

Giraldo-Chica, M. and Woodward, N. D. (2016). Review of thalamocortical resting-state fMRI studies in schizophrenia. *Schizophrenia research* **180**, 58–63.

Gong, J., Liu, X., Liu, T., Zhou, J., Sun, G. and Tian, J. (2018). Dual temporal and spatial sparse representation for inferring group-wise brain networks from resting-state fMRI dataset. *IEEE transactions on biomedical engineering* **65**(5), 1035–1048.

Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics* **2**(2), 111–120.

Green, P. J. (1990). On use of the em for penalized likelihood estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* **52**(3), 443–452.

Guo, J., Levina, E., Michailidis, G. and Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika* **98**(1), 1–15.

Gupta, S. D. (1977). Tests on multiple correlation coefficient and multiple partial correlation coefficient. *Journal of multivariate analysis* **7**(1), 82–88.

Hannan, E. J. (1976). The asymptotic distribution of serial covariances. *The Annals of Statistics* **4**(2), 396–399.

Hao, B., Sun, W. W., Liu, Y. and Cheng, G. (2018). Simultaneous clustering and estimation of heterogeneous graphical models. *Journal of Machine Learning Research* **18**, 1–58.

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: A univariate residual cross-correlation approach. *Journal of the American Statistical Association* **71**(354), 378–385.

Hedges, L. V. and Olkin, I. (1983). Joint distributions of some indices based on correlation coefficients. In: *Studies in Econometrics, Time Series, and Multivariate Statistics*. Academic Press, pp. 437–454.

Hill, S. M. and Mukherjee, S. (2013). Network-based clustering with mixtures of l1-penalized gaussian graphical models: an empirical investigation.

Honey, G. D., Pomarol-Clotet, E., Corlett, P. R., Honey, R. A. E., Mckenna, P. J., Bullmore, E. T. and Fletcher, P. C. (2005). Functional dysconnectivity in schizophrenia associated with attentional modulation of motor function. *Brain* **128**(11), 2597–2611.

Hotelling, H. (1931). The generalization of student's ratio. *The Annals of mathematical statistics* **2**(3), 360–378.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification* **2**(1), 193–218.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**(4), 419–426.

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., Üstün, T. B. and Wang, P. S. (2009). The global burden of mental disorders: An update from the who world mental health (wmh) surveys. *Epidemiologia e psichiatria sociale* **18**(1), 23–33.

Kim, J., Wozniak, J. R., Mueller, B. A. and Pan, W. (2015). Testing group differences in brain functional connectivity: Using correlations or partial correlations? *Brain connectivity* **5**(4), 214–231.

Kim, J., Wozniak, J. R., Mueller, B. A., Shen, X. and Pan, W. (2014). Comparison of statistical tests for group differences in brain functional networks. *NeuroImage* **101**, 681–694.

Lauritzen, S. L. (2002). *Thiele: pioneer in statistics*. Oxford ; New York: Oxford University Press.

Lauritzen, S. L. (2004). *Graphical models*, Repr. with corrections.. edition., Oxford statistical science series ; 17. Oxford : New York: Clarendon Press ; Oxford University Press.

Lazar, N. A. (2008). *The statistical analysis of functional MRI data*, Statistics for biology and health. New York: Springer.

Lee, M. H., Smyser, C. D. and Shimony, J. S. (2013). Resting-state fMRI: a review of methods and clinical applications. *American Journal of Neuroradiology* **34**(10), 1866.

Levy, K. J. and Narula, S. C. (1978). Testing hypotheses concerning partial correlations: Some methods and discussion. *International statistical review* **46**(2), 215–218.

Li, K., Guo, L., Nie, J., Li, G. and Liu, T. (2009). Review of methods for functional brain connectivity detection using fMRI. *Computerized Medical Imaging and Graphics* **33**(2), 131–139.

Li, S., Hu, N., Zhang, W., Tao, B., Dai, J., Gong, Y., Tan, Y., Cai, D. and Lui, S. (2019). Dysconnectivity of multiple brain networks in schizophrenia: A meta-analysis of resting-state functional connectivity. *Frontiers in psychiatry* **10**, 482–482.

Liu, H., Roeder, K. and Wasserman, L. (2010). Stability approach to regularization selection (stars) for high dimensional graphical models. In: *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 2*, NIPS'10. USA: Curran Associates Inc. pp. 1432–1440.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature (London)* **453**(7197), 869–878.

Lottman, K. K., Gawne, T. J., Kraguljac, N. V., Killen, J. F., Reid, M. A. and Lahti, A. C. (2019). Examining resting-state functional connectivity in first-episode schizophrenia with 7T fMRI and MEG. *NeuroImage: Clinical* **24**, 101959.

Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J. and Benali, H. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage* **32**(1), 228–237.

Mars, R. B., Jbabdi, S., Sallet, J., O'Reilly, J. X., Croxson, P. L., Olivier, E., Noonan, M. P., Bergmann, C., Mitchell, A. S., Baxter, M. G., Behrens, T. E. J., Johansen-Berg, H., Tomassini, V., Miller, K. L. et al. (2011). Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *The Journal of Neuroscience* **31**(11), 4087–4100.

Meng, X. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**(2), 267.

Michael, A. M., Anderson, M., Miller, R. L., Adali, T. and Calhoun, V. D. (2014). Preserving subject variability in group fMRI analysis: Performance evaluation of GICA vs. IVA. *Frontiers in systems neuroscience* **8**(JUNE), 106–106.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J. and Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron* **77**(3), 586–595.

Murtagh, F. and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Classification* **31**(3), 274–295.

Pettersson-Yeo, W., Allen, P., Benetti, S., Mcguire, P. and Mechelli, A. (2011). Dysconnectivity in schizophrenia: Where are we now? *Neuroscience and Biobehavioral Reviews* **35**(5), 1110–1124.

Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric reviews* **23**(1), 53–70.

Price, R. B., Lane, S., Gates, K., Kraynak, T. E., Horner, M. S., Thase, M. E. and Siegle, G. J. (2017). Parsing heterogeneity in the brain connectivity of depressed and healthy adults during positive mood. *Biological Psychiatry* **81**(4), 347–357.

Pruim, R. H. R., Mennes, M. J. J., Rooij, D. van, Llera, A., Buitelaar, J. K. and Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage* **112**, 267–277.

Qiu, H., Han, F., Liu, H. and Caffo, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society* **78**(2), 487–504.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* **66**(336), 846–850.

Roy, R. (1989). Asymptotic covariance structure of serial correlations in multivariate time series. *Biometrika* **76**(4), 824–827.

Salaman, M. S., Du, Y., Lin, D., Fu, Z., Fedorov, A., Damaraju, E., Sui, J., Chen, J., Mayer, A. R., Posse, S., Mathalon, D. H., Ford, J. M., Van Erp, T. et al. (2019). Group ICA for identifying biomarkers in schizophrenia: 'adaptive' networks via spatially constrained ICA show more sensitivity to group differences than spatio-temporal regression. *NeuroImage: Clinical* **22**, 101747.

Samorodnitsky, G. (2016). *Stochastic processes and long range dependence.*, Springer series in operations research. Cham: Springer.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology* **4**(1), 32–30.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*, Wiley series in probability and mathematical statistics. New York: Wiley.

Shen, H., Wang, L., Liu, Y. and Hu, D. (2010). Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. *NeuroImage* **49**(4), 3110–3121.

Simpson, S. L. and Laurienti, P. J. (2015). A two-part mixed-effects modeling framework for analyzing whole-brain network data. *NeuroImage* **113**, 310–319.

Smith, S., Miller, K., Khorshidi, G., Webster, M. A., Beckmann, C., Nichols, T., Ramsey, J. and Woolrich, M. (2011). Network modelling methods for fMRI. *NeuroImage* **54**, 875–891.

Smitha, K. A., Raja, K. A., Arun, K. M., Rajesh, P. G., Thomas, B., Kapilamoorthy, T. R. and Kesavadas, C. (2017). Resting state fMRI: A review on methods in resting state connectivity analysis and resting state networks. *The Neuroradiology Journal* **30**(4), 305–317.

Student. (1914). The elimination of spurious correlation due to position in time or space. *Biometrika* **10**(1), 179–180.

Substance Abuse and Mental Health Services Administration. (2020). Key substance use and mental health indicators in the united states: Results from the

2019 national survey on drug use and health (hhs publication no. pep20-07-01-001, nsduh series h-55). center for behavioral health statistics and quality, substance abuse and mental health services administration. pp. 1–71.

Tibshirani, R., Walther, G. and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(2), 411–423.

Wager, T. D. and Lindquist, M. A. (2015). *Principles of fMRI*, Chapter 6. Leanpub.

Wang, H. E., Bénar, C. G., Quilichini, P. P., Friston, K. J., Jirsa, V. K. and Bernard, C. (2014). A systematic framework for functional connectivity measures. *Frontiers in neuroscience* **8**(405), 405–405.

Westgate, P. M. (2014). Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data. *Statistics in medicine* **33**(13), 2222–2237.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wilks, S. S. (1932, 08). On the sampling distribution of the multiple correlation coefficient. *The Annals of Mathematical Statistics* **3**(3), 196–203.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* **9**(1), 60–62.

Williams, E. J. (1978). A simple derivation of the distribution of the multiple correlation coefficient. *Communications in Statistics - Theory and Methods* **7**(15), 1413–1420.

Woodward, N. D. and Heckers, S. (2015). Mapping thalamocortical functional connectivity in chronic and early stages of psychotic disorders. *Biological Psychiatry* **79**(12), 1016–1025.

Woodward, N. D., Karbasforoushan, H. and Heckers, S. (2012). Thalamocortical dysconnectivity in schizophrenia. *American Journal of Psychiatry* **169**(10), 1092–1099.

Woolrich, M. W., Ripley, B. D., Brady, M. and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. *NeuroImage* **14**(6), 1370–1386.

Wright, S. J. (2015). Coordinate descent algorithms. *Mathematical programming* **151**(1), 3–34.

Yang, J., Gohel, S. and Vachha, B. (2020). Current methods and new directions in resting state fMRI. *Clinical imaging* **65**, 47–53.

Yoon, J. H., Minzenberg, M. J., Ursu, S., Walters, R., Wendelken, C., Ragland, J. D. and Carter, C. S. (2008). Association of dorsolateral prefrontal cortex dysfunction with disrupted coordinated brain activity in schizophrenia: Relationship with impaired cognition, behavioral disorganization, and global function. *American Journal of Psychiatry* **165**(8), 1006–1014.

Zalesky, A., Fornito, A. and Bullmore, E. T. (2010). Network-based statistic: Identifying differences in brain networks. *NeuroImage* **53**(4), 1197–1207.

Zeng, L., Shen, H., Liu, L. and Hu, D. (2014). Unsupervised classification of major depression using functional connectivity MRI. *Human Brain Mapping* **35**(4), 1630–1641.

Zhang, L., DiLernia, A., Quevedo, K., Camchong, J., Lim, K. and Pan, W. (2020). A random covariance model for bi-level graphical modeling with application to resting-state fMRI data. *Biometrics*.

Zhou, H., Pan, W. and Shen, X. (2009). Penalized model-based clustering with unconstrained covariance matrices. *Electronic Journal of Statistics* **3**(1473-1496), 1473–1496.

Zhou, Y., Shu, N., Liu, Y., Song, M., Hao, Y., Liu, H., Yu, C., Liu, Z. and Jiang, T. (2008). Altered resting-state functional connectivity and anatomical

connectivity of hippocampus in schizophrenia. *Schizophrenia Research* **100**(1-3), 120–132.

Zhu, X., Wang, X., Xiao, J., Liao, J., Zhong, M., Wang, W. and Yao, S. (2012). Evidence of a dissociation pattern in resting-state default mode network connectivity in first-episode, treatment-naive major depression patients. *Biological Psychiatry* **71**(7), 611–617.

Zhu, Y., Shen, X. and Pan, W. (2014). Structural pursuit over multiple undirected graphs. *Journal of the American Statistical Association* **109**(508), 00–00.

# Appendix A

# Chapter 2 Supplementary Materials

Several equations and formulas in Chapter 2 are not trivial and require more justification. This appendix shows some supplementary tables as well as brief derivations.

**Updates for $\pi_g$'s**

With a LaGrange multiplier constraint to force $\sum_{g=1}^{G} \pi_g = 1$, the terms of $Q(\Theta; \Theta^{(r)})$ containing $\pi_g$'s are

$$Q_\pi(\Theta; \Theta^{(r)}) = -2 \sum_{k=1}^{K} \sum_{g=1}^{G} w_{gk}^{(r)} (\log \pi_g) + \gamma \left( \sum_{g=1}^{G} \pi_g - 1 \right),$$

so

$$\frac{\partial Q_\pi(\Theta; \Theta^{(r)})}{\partial \pi_g} = -\frac{2}{\pi_g} \sum_{k=1}^{K} w_{gk}^{(r)} + \gamma = 0 \implies$$

$$\pi_g \gamma = 2 \sum_{k=1}^{K} w_{gk}^{(r)} \implies \pi_g^{(r+1)} = \frac{1}{K} \sum_{k=1}^{K} w_{gk}^{(r)}$$

since $\sum_{g=1}^{G} \pi_g = \sum_{g=1}^{G} w_{gk}^{(r)} = 1$.

**Updates for $\mathbf{\Omega_{0g}}$'s**

The terms of $Q(\Theta;\Theta^{(r)})$ containing $\mathbf{\Omega_{0g}}$'s are

$$Q_{\mathbf{\Omega_{0g}}}(\Theta;\Theta^{(r)}) = -2\sum_{k=1}^{K} w_{gk}^{(r)} \log\Big(p_g(\mathbf{\Omega_k}^{(r)};\lambda_2,\mathbf{\Omega_{0g}})\Big) + \lambda_3||\mathbf{\Omega_{0g}}||_1 =$$

$$-2\sum_{k=1}^{K} w_{gk}^{(r)} \left(-\frac{\lambda_2}{2}\log|\frac{1}{\lambda_2}\mathbf{\Omega_{0g}}| - \frac{1}{2}\mathrm{tr}(\lambda_2 \mathbf{\Omega_{0g}}^{-1}\mathbf{\Omega_k}^{(r)})\right) + \lambda_3||\mathbf{\Omega_{0g}}||_1 =$$

$$\lambda_2\sum_{k=1}^{K} w_{gk}^{(r)} \left(\log|\frac{1}{\lambda_2}\mathbf{\Omega_{0g}}|\right) + \lambda_2\sum_{k=1}^{K} w_{gk}^{(r)}\mathrm{tr}(\mathbf{\Omega_{0g}}^{-1}\mathbf{\Omega_k}^{(r)}) + \lambda_3||\mathbf{\Omega_{0g}}||_1 =$$

$$\mathrm{tr}\left(\lambda_2\sum_{k=1}^{K} w_{gk}^{(r)}\mathbf{\Omega_k}^{(r)}\mathbf{\Omega_{0g}}^{-1}\right) + \lambda_2\sum_{k=1}^{K} w_{gk}^{(r)}\log|\mathbf{\Omega_{0g}}| + \lambda_3||\mathbf{\Omega_{0g}}||_1 \propto$$

$$\mathrm{tr}\left(\frac{\sum_{k=1}^{K} w_{gk}^{(r)}\mathbf{\Omega_k}^{(r)}}{\sum_{k=1}^{K} w_{gk}^{(r)}}\mathbf{\Omega_{0g}}^{-1}\right) + \log|\mathbf{\Omega_{0g}}| + \frac{\lambda_3}{\lambda_2\sum_{k=1}^{K} w_{gk}^{(r)}}||\mathbf{\Omega_{0g}}||_1,$$

which can be solved using the covariance lasso algorithm.

**Updates for $\mathbf{\Omega_k}$'s**

The terms of $Q(\Theta;\Theta^{(r)})$ containing $\mathbf{\Omega_k}$'s are

$$Q_{\Omega_k}(\Theta;\Theta^{(r)}) = \sum_{t=1}^{n_k}\big(-\log|\mathbf{\Omega_k}| + \mathbf{y_{kt}^T}\mathbf{\Omega_k}\mathbf{y_{kt}}\big) - 2\sum_{g=1}^{G} w_{gk}^{(r)}\Big(\log p_g(\mathbf{\Omega_k};\lambda_2,\mathbf{\Omega_{0g}}^{(r)})\Big) + \lambda_1||\mathbf{\Omega_k}||_1 =$$

$$\sum_{t=1}^{n_k}\big(\mathbf{y_{kt}^T}\mathbf{\Omega_k}\mathbf{y_{kt}} - \log|\mathbf{\Omega_k}|\big) - 2\sum_{g=1}^{G} w_{gk}^{(r)}\left(\frac{\lambda_2 - p - 1}{2}\log|\mathbf{\Omega_k}| - \frac{1}{2}\mathrm{tr}(\lambda_2\mathbf{\Omega_{0g}}^{(r)^{-1}}\mathbf{\Omega_k})\right) + \lambda_1||\mathbf{\Omega_k}||_1 =$$

$$\sum_{t=1}^{n_k}\big(\mathrm{tr}(\mathbf{y_{kt}}\mathbf{y_{kt}^T}\mathbf{\Omega_k}) - \log|\mathbf{\Omega_k}|\big) + \sum_{g=1}^{G} w_{gk}^{(r)}\Big(\lambda_2\mathrm{tr}(\mathbf{\Omega_{0g}}^{(r)^{-1}}\mathbf{\Omega_k}) - (\lambda_2 - p - 1)\log|\mathbf{\Omega_k}|\Big) + \lambda_1||\mathbf{\Omega_k}||_1 =$$

$$\text{tr}(n_k S_k \mathbf{\Omega_k}) + \text{tr}(\lambda_2 \sum_{g=1}^{G} w_{gk}^{(r)} \mathbf{\Omega_{0g}}^{(r)^{-1}} \mathbf{\Omega_k}) - \log|\mathbf{\Omega_k}|(n_k + (\lambda_2 - p - 1)) + \lambda_1 ||\mathbf{\Omega_k}||_1,$$

since $\sum_{g=1}^{G} w_{gk}^{(r)} = 1$ and $\mathbf{S_k} = \frac{1}{n_k} \sum_{t=1}^{n_k} \mathbf{y_{kt}} \mathbf{y_{kt}^T}$. Therefore,

$$Q_{\Omega_k}(\Theta; \Theta^{(r)}) = \text{tr}\left( (n_k \mathbf{S_k} + \lambda_2 \sum_{g=1}^{G} w_{gk}^{(r)} \mathbf{\Omega_{0g}}^{(r)^{-1}}) \mathbf{\Omega_k} \right) - \log|\mathbf{\Omega_k}|(n_k + \lambda_2 - p - 1) + \lambda_1 ||\mathbf{\Omega_k}||_1 \propto$$

$$\text{tr}\left( \frac{n_k \mathbf{S_k} + \lambda_2 \sum_{g=1}^{G} w_{gk}^{(r)} \mathbf{\Omega_{0g}}^{(r)^{-1}}}{n_k + \lambda_2 - p - 1} \mathbf{\Omega_k} \right) - \log|\mathbf{\Omega_k}| + \frac{\lambda_1}{n_k + \lambda_2 - p - 1} ||\mathbf{\Omega_k}||_1,$$

which is solved using the glasso algorithm.

**Updates for $\mathbf{w_{gk}}$'s**

$$\begin{aligned} w_{gk}^{(r+1)} &= \frac{\pi_g^{(r+1)} p_g(\mathbf{\Omega_k}^{(r+1)}; \lambda_2, \mathbf{\Omega_{0g}}^{(r+1)})}{\sum_{c=1}^{G} \pi_c^{(r+1)} p_c(\mathbf{\Omega_k}^{(r+1)}; \lambda_2, \mathbf{\Omega_{0c}}^{(r+1)})} \\ &= \frac{\pi_g^{(r+1)} \exp\left( -\frac{\lambda_2}{2} \text{tr}(\mathbf{\Omega_{0g}}^{(r+1)^{-1}} \mathbf{\Omega_k}^{(r+1)}) \right) |\mathbf{\Omega_{0g}}^{(r+1)}|^{-\frac{\lambda_2}{2}}}{\sum_{c=1}^{G} \pi_c^{(r+1)} \exp\left( -\frac{\lambda_2}{2} \text{tr}(\mathbf{\Omega_{0c}}^{(r+1)^{-1}} \mathbf{\Omega_k}^{(r+1)}) \right) |\mathbf{\Omega_{0c}}^{(r+1)}|^{-\frac{\lambda_2}{2}}}. \end{aligned}$$

**Selecting the Number of Clusters**

For selecting the number of clusters, we used a gap statistic as proposed by Tibshirani et al. (2001). Generally, the gap statistic compares the observed change in within-cluster dispersion of a clustering of subjects when specifying different numbers of clusters to what is expected under a corresponding null setting based on the observed data. For the setting in which we have data for $K$ subjects each with $n_k$ observations of $p$ variables, we calculate a gap statistic for the RCCM as follows:

1. Obtain the estimated clustering of subjects using RCCM when specifying $G$ total clusters for $G = 2, \ldots, G_{\max}$ where $G_{\max}$ is the maximum number of clusters to be considered.

2. Calculate the GLasso estimate for the $k^{th}$ subject using a small amount of penalization for $k = 1, 2, \ldots, K$. We used a tuning parameter value of $10^{-16}$ so that the optimal number of clusters was invariant to the choice of tuning parameters.

3. Calculate the log of the within-cluster variability for the observed data when specifying $G$ total clusters for $G = 2, \ldots, G_{\max}$ as

$$V_G = \log\left(\sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{i=1}^{p}\sum_{j=1}^{p}(\omega_{k;i,j} - \bar{\omega}_{g;i,j})^2 \cdot z_{gk}/(G \cdot p^2)\right), \tag{A.1}$$

   where $\log(\cdot)$ is the natural log, $z_{gk} = \mathbf{1}\{\text{subject } k \text{ is in cluster } g\}$ is an indicator of the $k^{th}$ subject belonging to the $g^{th}$ cluster, $\bar{\omega}_{g;i,j} = \sum_{k=1}^{K}(\omega_{k;i,j} \cdot z_{gk})/N_g$ is the average value of the GLasso estimate entry in the $i^{th}$ row and $j^{th}$ column for subjects in the $g^{th}$ cluster, and $N_g = \sum_{k=1}^{K} z_{gk}$ is the number of subjects in the $g^{th}$ cluster. We note that $V_G$ represents the log of the average variance of the precision matrix entries across subjects within each cluster.

4. Generate $B$ reference data sets for $K$ subjects and obtain the estimated clustering of subjects using RCCM when specifying $G$ total clusters for $G = 2, \ldots, G_{\max}$. Specifically, generate $n_k$ observations for the $k^{th}$ subject from a $\mathcal{N}_p(\mathbf{0}, \Omega_{bk}^{-1})$. The entry in the $i^{th}$ row and $j^{th}$ column of $\Omega_{bk}^{-1}$ is generated from a uniform $\left(\min\{\omega_{k;i,j}\}_{k=1}^{K}, \max\{\omega_{k;i,j}\}_{k=1}^{K}\right)$ distribution where $\{\omega_{k;i,j}\}_{k=1}^{K}$ are the GLasso entries for the observed data. We adjust the $\Omega_{bk}^{-1}$ matrix to make it positive definite if needed.

5. Implement the RCCM for each generated data set varying the specified number of clusters, and calculate $V_{G;b}$ for $b = 1, 2, \ldots, B$ and $G = 2, \ldots, G_{\max}$ as described in Equation (A.1).

6. Calculate the estimated gap statistics as

$$\text{Gap}(G) = \frac{1}{B}\sum_{b=1}^{B}(V_{G;b} - V_G) = \bar{V} - V_G.$$

7. Choose the optimal number of clusters as

$$G^* = \arg\min\{G : \text{Gap}(G) \geq \text{Gap}(G+1) - \sigma_{G+1}\},$$

where $\sigma_{G+1} = \sqrt{\frac{\sum_{b=1}^{B}\left(V_{G;b} - \bar{V}\right)^2}{B}} \cdot \sqrt{1 + 1/B}$.

Table A.1: Accuracy of gap statistic across 100 simulations for selecting the correct number of clusters, where $G$ is the true number of clusters, Magnitude indicates whether the true precision matrices had entries high or low in magnitude, and Overlap is the proportion of overlapping edges across clusters.

| $G$ | Magnitude | Overlap | Accuracy |
|---|---|---|---|
| 2 | High | 0.20 | 1.00 |
| | | 0.50 | 1.00 |
| | | 0.80 | 1.00 |
| | Low | 0.20 | 1.00 |
| | | 0.50 | 0.99 |
| | | 0.80 | 0.97 |
| 3 | High | 0.20 | 1.00 |
| | | 0.50 | 1.00 |
| | | 0.80 | 1.00 |
| | Low | 0.20 | 1.00 |
| | | 0.50 | 1.00 |
| | | 0.80 | 0.99 |

Table A.2: Clustering performance of RCCM, Ward, and K-means clustering using 5-fold CV for tuning parameter selection. Results are for $G = 2$ unbalanced groups, true precision matrices with entries large or small in magnitude, and group overlap being 0.20, 0.50, or 0.80 for each of 104 subjects averaged across 100 simulations. Ward clustering was based on a difference matrix of the Frobenius-norm differences between matrix estimates, while K-means clustering was based on vectorized GLasso matrix estimates.

| Magnitude | Overlap | Method | RI | $RI_{adj}$ |
|---|---|---|---|---|
| High | 0.2 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 1.000 (0.000) | 1.000 (0.000) |
| | 0.5 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 1.000 (0.000) | 1.000 (0.000) |
| | 0.8 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 1.000 (0.000) | 1.000 (0.000) |
| Low | 0.2 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 0.998 (0.007) | 0.996 (0.014) |
| | | GLasso & K-means | 0.998 (0.009) | 0.996 (0.018) |
| | 0.5 | RCCM | 1.000 (0.003) | 0.999 (0.005) |
| | | Ward & GGL | 0.996 (0.012) | 0.992 (0.025) |
| | | GLasso & K-means | 0.988 (0.034) | 0.976 (0.067) |
| | 0.8 | RCCM | 0.992 (0.015) | 0.983 (0.031) |
| | | Ward & GGL | 0.950 (0.047) | 0.900 (0.094) |
| | | GLasso & K-means | 0.909 (0.152) | 0.810 (0.324) |

Table A.3: Clustering performance of RCCM, Ward, and K-means clustering using 5-fold CV for tuning parameter selection. Results are for $G = 3$ unbalanced groups, true precision matrices with entries large or small in magnitude, and group overlap being 0.20, 0.50, or 0.80 for each of 104 subjects averaged across 100 simulations. Ward clustering was based on a difference matrix of the Frobenius-norm differences between matrix estimates, while K-means clustering was based on vectorized GLasso matrix estimates.

| Magnitude | Overlap | Method | RI | $\text{RI}_{\text{adj}}$ |
|---|---|---|---|---|
| High | 0.2 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 0.853 (0.124) | 0.691 (0.260) |
| | 0.5 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 1.000 (0.000) | 1.000 (0.000) |
| | | GLasso & K-means | 0.872 (0.124) | 0.733 (0.259) |
| | 0.8 | RCCM | 1.000 (0.000) | 1.000 (0.000) |
| | | Ward & GGL | 0.999 (0.003) | 0.998 (0.007) |
| | | GLasso & K-means | 0.965 (0.088) | 0.928 (0.180) |
| Low | 0.2 | RCCM | 0.999 (0.003) | 0.998 (0.007) |
| | | Ward & GGL | 0.995 (0.009) | 0.989 (0.020) |
| | | GLasso & K-means | 0.934 (0.100) | 0.863 (0.206) |
| | 0.5 | RCCM | 0.999 (0.003) | 0.998 (0.007) |
| | | Ward & GGL | 0.991 (0.012) | 0.980 (0.026) |
| | | GLasso & K-means | 0.929 (0.094) | 0.855 (0.190) |
| | 0.8 | RCCM | 0.989 (0.019) | 0.975 (0.041) |
| | | Ward & GGL | 0.945 (0.040) | 0.880 (0.086) |
| | | GLasso & K-means | 0.896 (0.098) | 0.788 (0.192) |

Table A.4: Comparison of method performances for $G = 2$ unbalanced groups containing 67 and 37 subjects. Results are for the high magnitude setting with observations of $p = 10$ variables and $n = 177$ observations for each subject using 5-fold CV for tuning parameter selection, averaged across 100 simulations. Group-level performance measures for GLasso are missing since the method does not yield cluster-level estimates.

| Overlap | Method | $TPR_g$ | $FPR_g$ | $PPV_g$ | $TPR_k$ | $FPR_k$ | $PPV_k$ |
|---|---|---|---|---|---|---|---|
| 0.2 | RCCM | 0.999 | 0.210 | 0.507 | 1.000 | 0.626 | 0.252 |
| | | (0.007) | (0.041) | (0.046) | (0.002) | (0.029) | (0.014) |
| | Ward & GGL | 1.000 | 0.322 | 0.400 | 1.000 | 0.451 | 0.319 |
| | | (0.000) | (0.049) | (0.037) | (0.005) | (0.010) | (0.011) |
| | GLasso & K-means | | | | 1.000 | 0.452 | 0.318 |
| | | | | | (0.005) | (0.010) | (0.011) |
| 0.5 | RCCM | 1.000 | 0.250 | 0.501 | 1.000 | 0.626 | 0.286 |
| | | (0.000) | (0.018) | (0.019) | (0.000) | (0.056) | (0.016) |
| | Ward & GGL | 1.000 | 0.247 | 0.507 | 1.000 | 0.437 | 0.364 |
| | | (0.000) | (0.044) | (0.044) | (0.000) | (0.008) | (0.004) |
| | GLasso & K-means | | | | 1.000 | 0.438 | 0.363 |
| | | | | | (0.000) | (0.008) | (0.004) |
| 0.8 | RCCM | 0.947 | 0.247 | 0.508 | 1.000 | 0.736 | 0.255 |
| | | (0.000) | (0.021) | (0.021) | (0.000) | (0.071) | (0.020) |
| | Ward & GGL | 0.951 | 0.304 | 0.459 | 1.000 | 0.444 | 0.360 |
| | | (0.013) | (0.043) | (0.037) | (0.000) | (0.007) | (0.004) |
| | GLasso & K-means | | | | 1.000 | 0.446 | 0.359 |
| | | | | | (0.000) | (0.007) | (0.004) |

Table A.5: Comparison of method performances for $G = 2$ unbalanced groups containing 67 and 37 subjects. Results are for the low magnitude setting with observations of $p = 10$ variables and $n = 177$ observations for each subject using 5-fold CV for tuning parameter selection, averaged across 100 simulations. Group-level performance measures for GLasso are missing since the method does not yield cluster-level estimates.

| Overlap | Method | $TPR_g$ | $FPR_g$ | $PPV_g$ | $TPR_k$ | $FPR_k$ | $PPV_k$ |
|---|---|---|---|---|---|---|---|
| 0.2 | RCCM | 0.996 | 0.438 | 0.370 | 0.999 | 0.559 | 0.287 |
| | | (0.015) | (0.236) | (0.168) | (0.006) | (0.188) | (0.083) |
| | Ward & GGL | 0.990 | 0.387 | 0.369 | 0.989 | 0.429 | 0.350 |
| | | (0.023) | (0.141) | (0.130) | (0.015) | (0.151) | (0.128) |
| | GLasso & K-means | | | | 0.803 | 0.212 | 0.443 |
| | | | | | (0.046) | (0.048) | (0.042) |
| 0.5 | RCCM | 0.989 | 0.476 | 0.390 | 0.992 | 0.586 | 0.314 |
| | | (0.023) | (0.245) | (0.157) | (0.013) | (0.186) | (0.081) |
| | Ward & GGL | 0.849 | 0.341 | 0.427 | 0.877 | 0.376 | 0.412 |
| | | (0.066) | (0.160) | (0.151) | (0.046) | (0.174) | (0.151) |
| | GLasso & K-means | | | | 0.682 | 0.200 | 0.464 |
| | | | | | (0.047) | (0.041) | (0.034) |
| 0.8 | RCCM | 0.990 | 0.789 | 0.307 | 0.993 | 0.810 | 0.314 |
| | | (0.032) | (0.139) | (0.043) | (0.008) | (0.105) | (0.029) |
| | Ward & GGL | 0.868 | 0.482 | 0.507 | 0.789 | 0.486 | 0.390 |
| | | (0.150) | (0.338) | (0.275) | (0.093) | (0.168) | (0.062) |
| | GLasso & K-means | | | | 0.634 | 0.226 | 0.514 |
| | | | | | (0.050) | (0.049) | (0.037) |

Table A.6: Comparison of method performances for $G = 3$ unbalanced groups containing 61, 24, and 19 subjects. Results are for the high magnitude setting with observations of $p = 10$ variables and $n = 177$ observations for each subject using 5-fold CV for tuning parameter selection, averaged across 100 simulations. Group-level performance measures for GLasso are missing since the method does not yield cluster-level estimates.

| Overlap | Method | $TPR_g$ | $FPR_g$ | $PPV_g$ | $TPR_k$ | $FPR_k$ | $PPV_k$ |
|---|---|---|---|---|---|---|---|
| 0.2 | RCCM | 0.913 | 0.102 | 0.647 | 1.000 | 0.620 | 0.229 |
| | | (0.000) | (0.005) | (0.010) | (0.000) | (0.007) | (0.002) |
| | Ward & GGL | 0.940 | 0.298 | 0.395 | 1.000 | 0.448 | 0.292 |
| | | (0.025) | (0.030) | (0.025) | (0.000) | (0.007) | (0.003) |
| | GLasso & K-means | | | | 1.000 | 0.449 | 0.291 |
| | | | | | (0.000) | (0.007) | (0.003) |
| 0.5 | RCCM | 1.000 | 0.092 | 0.655 | 0.973 | 0.619 | 0.224 |
| | | (0.000) | (0.006) | (0.016) | (0.005) | (0.007) | (0.002) |
| | Ward & GGL | 1.000 | 0.277 | 0.388 | 0.960 | 0.444 | 0.285 |
| | | (0.000) | (0.033) | (0.028) | (0.005) | (0.007) | (0.003) |
| | GLasso & K-means | | | | 0.960 | 0.445 | 0.284 |
| | | | | | (0.005) | (0.007) | (0.003) |
| 0.8 | RCCM | 0.963 | 0.102 | 0.723 | 0.999 | 0.772 | 0.274 |
| | | (0.045) | (0.025) | (0.060) | (0.005) | (0.040) | (0.019) |
| | Ward & GGL | 0.998 | 0.366 | 0.432 | 0.992 | 0.463 | 0.384 |
| | | (0.012) | (0.056) | (0.034) | (0.026) | (0.019) | (0.028) |
| | GLasso & K-means | | | | 0.997 | 0.463 | 0.385 |
| | | | | | (0.009) | (0.017) | (0.027) |

Table A.7: Comparison of method performances for $G = 3$ unbalanced groups containing 61, 24, and 19 subjects. Results are for the low magnitude setting with observations of $p = 10$ variables and $n = 177$ observations for each subject using 5-fold CV for tuning parameter selection, averaged across 100 simulations. Group-level performance measures for GLasso are missing since the method does not yield cluster-level estimates.

| Overlap | Method | $TPR_g$ | $FPR_g$ | $PPV_g$ | $TPR_k$ | $FPR_k$ | $PPV_k$ |
|---|---|---|---|---|---|---|---|
| 0.2 | RCCM | 0.973 | 0.455 | 0.340 | 0.994 | 0.551 | 0.290 |
| | | (0.076) | (0.204) | (0.128) | (0.012) | (0.170) | (0.075) |
| | Ward & GGL | 0.957 | 0.380 | 0.361 | 0.969 | 0.393 | 0.365 |
| | | (0.103) | (0.124) | (0.109) | (0.026) | (0.128) | (0.116) |
| | GLasso & K-means | | | | 0.696 | 0.182 | 0.446 |
| | | | | | (0.031) | (0.020) | (0.019) |
| 0.5 | RCCM | 0.983 | 0.724 | 0.280 | 0.993 | 0.753 | 0.275 |
| | | (0.035) | (0.140) | (0.044) | (0.010) | (0.111) | (0.029) |
| | Ward & GGL | 0.974 | 0.712 | 0.280 | 0.947 | 0.634 | 0.302 |
| | | (0.047) | (0.107) | (0.041) | (0.035) | (0.121) | (0.033) |
| | GLasso & K-means | | | | 0.763 | 0.249 | 0.470 |
| | | | | | (0.041) | (0.045) | (0.040) |
| 0.8 | RCCM | 0.964 | 0.762 | 0.295 | 0.985 | 0.774 | 0.312 |
| | | (0.046) | (0.126) | (0.036) | (0.013) | (0.108) | (0.029) |
| | Ward & GGL | 0.834 | 0.498 | 0.413 | 0.789 | 0.472 | 0.378 |
| | | (0.145) | (0.252) | (0.191) | (0.079) | (0.124) | (0.045) |
| | GLasso & K-means | | | | 0.602 | 0.207 | 0.509 |
| | | | | | (0.050) | (0.045) | (0.034) |

# Appendix B

# Chapter 3 Supplementary Materials

**Gradient Vector and Hessian Matrix**

Consider the function

$$f(\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} \end{bmatrix}^T) = \frac{\boldsymbol{x}^T\boldsymbol{y}}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y}}} = \frac{\sum_{k=1}^{N} x_k y_k}{\sqrt{\sum_{k=1}^{N} x_k^2 \sum_{k=1}^{N} y_k^2}},$$

such that $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^N$ and $f : \mathbb{R}^{2N} \to \mathbb{R}$. We can approximate this function using a Taylor series expansion. Specifically, a second-order Taylor series approximation of $f(\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} \end{bmatrix}^T)$ around a point $\theta = \begin{bmatrix} \theta_x & \theta_y \end{bmatrix}^T \in \mathbb{R}^{2N}$ is given by

$$f(\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} \end{bmatrix}^T) = f(\theta) + (\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} \end{bmatrix}^T - \theta)^\mathsf{T}\nabla f(\theta) + \frac{1}{2}(\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} \end{bmatrix}^T - \theta)^\mathsf{T}\mathbf{H}(f(\theta))(\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} \end{bmatrix}^T - \theta),$$

where $\nabla f(\theta)$ is the gradient of $f(\theta)$ and $\mathbf{H}(f(\theta))$ is the Hessian matrix of $f(\theta)$. For the gradient of $f(\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} \end{bmatrix}^T)$, we observe that

$$\frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\boldsymbol{x}^T\boldsymbol{y}(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{1}{2}}\right) = \frac{\partial}{\partial x_i}\left(\sum_{k=1}^{N} x_k y_k (\sum_{k=1}^{N} x_k^2 \sum_{k=1}^{N} y_k^2)^{-\frac{1}{2}}\right),$$

$$= y_i(\sum_{k=1}^{N} x_k^2 \sum_{k=1}^{N} y_k^2)^{-\frac{1}{2}} - x_i(\sum_{k=1}^{N} x_k^2 \sum_{k=1}^{N} y_k^2)^{-\frac{3}{2}}(\sum_{k=1}^{N} y_k^2) \sum_{k=1}^{N} x_k y_k,$$

and similarly

$$\frac{\partial f}{\partial y_i} = x_i(\sum_{k=1}^{N} x_k^2 \sum_{k=1}^{N} y_k^2)^{-\frac{1}{2}} - y_i(\sum_{k=1}^{N} x_k^2 \sum_{k=1}^{N} y_k^2)^{-\frac{3}{2}}(\sum_{k=1}^{N} x_k^2) \sum_{k=1}^{N} x_k y_k.$$

Hence, it follows that the gradient of $f(\begin{bmatrix}\boldsymbol{x} & \boldsymbol{y}\end{bmatrix}^T)$ is given by

$$\nabla f(\begin{bmatrix}\boldsymbol{x} \\ \boldsymbol{y}\end{bmatrix}) = \begin{bmatrix}\frac{\partial f}{\partial \boldsymbol{x}} \\ \frac{\partial f}{\partial \boldsymbol{y}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{y}(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{1}{2}} - \boldsymbol{x}(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{3}{2}}(\boldsymbol{y}^T\boldsymbol{y})(\boldsymbol{x}^T\boldsymbol{y}) \\ \boldsymbol{x}(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{1}{2}} - \boldsymbol{y}(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{3}{2}}(\boldsymbol{x}^T\boldsymbol{x})(\boldsymbol{x}^T\boldsymbol{y})\end{bmatrix} \in \mathbb{R}^{2N}.$$

Moreover, the Hessian matrix of $f(\begin{bmatrix}\boldsymbol{x} & \boldsymbol{y}\end{bmatrix}^T)$ is equivalent to the Jacobian matrix of the gradient vector given by

$$\mathbf{H}(f(\begin{bmatrix}\boldsymbol{x} & \boldsymbol{y}\end{bmatrix}^T)) = \begin{bmatrix}\frac{\partial^2 f}{\partial \boldsymbol{x}\partial \boldsymbol{x}^T} & \frac{\partial^2 f}{\partial \boldsymbol{y}\partial \boldsymbol{x}^T} \\ \frac{\partial^2 f}{\partial \boldsymbol{x}\partial \boldsymbol{y}^T} & \frac{\partial^2 f}{\partial \boldsymbol{y}\partial \boldsymbol{y}^T}\end{bmatrix} \in \mathbb{R}^{2N\times 2N}.$$

Letting $\boldsymbol{I}_{N\times N}$ denote the $N \times N$ identity matrix, it follows that

$$\begin{aligned}\frac{\partial^2 f}{\partial \boldsymbol{x}\partial \boldsymbol{x}^T} =& -2(\boldsymbol{x}\boldsymbol{y}^T)(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{3}{2}}(\boldsymbol{y}^T\boldsymbol{y}) + 3(\boldsymbol{x}\boldsymbol{x}^T)(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{5}{2}}(\boldsymbol{y}^T\boldsymbol{y})^2(\boldsymbol{x}^T\boldsymbol{y}) \\ & - \boldsymbol{I}_{N\times N}\cdot(\boldsymbol{x}^T\boldsymbol{y})(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{3}{2}}(\boldsymbol{y}^T\boldsymbol{y}),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial \boldsymbol{y}\partial \boldsymbol{x}^T} = \frac{\partial^2 f}{\partial \boldsymbol{x}\partial \boldsymbol{y}^T} =& -(\boldsymbol{x}\boldsymbol{x}^T)(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{3}{2}}(\boldsymbol{y}^T\boldsymbol{y}) + (\boldsymbol{x}\boldsymbol{y}^T)(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{3}{2}}(\boldsymbol{x}^T\boldsymbol{y}) \\ & - (\boldsymbol{y}\boldsymbol{y}^T)(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{3}{2}}(\boldsymbol{x}^T\boldsymbol{x}) + \boldsymbol{I}_{N\times N}\cdot(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{1}{2}},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial \boldsymbol{y}\partial \boldsymbol{y}^T} =& -2(\boldsymbol{x}\boldsymbol{y}^T)(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{3}{2}}(\boldsymbol{x}^T\boldsymbol{x}) + 3(\boldsymbol{y}\boldsymbol{y}^T)(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{5}{2}}(\boldsymbol{x}^T\boldsymbol{x})^2(\boldsymbol{x}^T\boldsymbol{y}) \\ & - \boldsymbol{I}_{N\times N}\cdot(\boldsymbol{x}^T\boldsymbol{y})(\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{y})^{-\frac{3}{2}}(\boldsymbol{x}^T\boldsymbol{x}).\end{aligned}$$

**Proof of Theorem 1**

*Proof.* We observe that since $\mathbf{x}(t)$ is an ergodic stationary stochastic process, $\mathbf{e}(t)$ is also a stationary ergodic process (Samorodnitsky, 2016). Moreover, since the $\{\mathbf{x}_i\}_{i=1}^{p}$ are normally distributed with square integrable spectral density functions, it follows that the OLS residuals are normally distributed, and thus all corresponding fourth-order cumulants are 0 (Lauritzen, 2002), so condition 1 is satisfied for $\mathbf{e}(t)$. Furthermore, since a linear combination of square integrable functions is also square integrable, it follows that the spectral density functions for the OLS residuals are also square integrable, so condition 2 is met for $\mathbf{e}(t)$ as well. Similarly, a linear combination of a linear process with finite variances is also a linear process with finite variances, so condition 3 is satisfied for $\mathbf{e}(t)$. Lastly, since the OLS residuals are normally distributed, condition 4 is satisfied. Hence, by the result of Roy (1989), the empirical partial correlations $\{r_{ij\cdot(ij)}\}_{i\neq j}$, which are equivalent to the marginal correlations between the OLS residuals, are jointly asymptotically normal with mean the population partial correlation. □

**Proof of Theorem 2**

*Proof.* To derive an asymptotic covariance estimator for the $q = \binom{p}{2}$ unique partial correlations, we use a Taylor series approximation of the sample partial correlations and properties of quadratic forms of multivariate normal random vectors. We approximate $f(\cdot)$ in Equation (3.1) using a second-order Taylor series expansion around $\boldsymbol{\theta}_{ij} = \begin{bmatrix}\boldsymbol{\theta}_i & \boldsymbol{\theta}_j\end{bmatrix}^T \in \mathbb{R}^{2N}$ given by

$$f(\boldsymbol{e}_{ij}) = f(\boldsymbol{\theta}_{ij}) + (\boldsymbol{e}_{ij} - \boldsymbol{\theta}_{ij})^{\mathsf{T}} \nabla f(\boldsymbol{\theta}_{ij}) + \frac{1}{2}(\boldsymbol{e}_{ij} - \boldsymbol{\theta}_{ij})^{\mathsf{T}} \mathbf{H}[f(\boldsymbol{\theta}_{ij})](\boldsymbol{e}_{ij} - \boldsymbol{\theta}_{ij}),$$

where $\nabla f(\boldsymbol{\theta}_{ij}) = \mathbb{E}[f(\boldsymbol{e}_{ij})] \in \mathbb{R}^{2N}$ is the expected value of the gradient of $f(\boldsymbol{e}_{ij})$ and $\boldsymbol{H}[f(\boldsymbol{\theta}_{ij})] = \mathbb{E}[H(\boldsymbol{e}_{ij})] \in \mathbb{R}^{2N\times 2N}$ is the expected value of the Hessian matrix of $f(\boldsymbol{e}_{ij})$. Details regarding the form of the gradient vector and Hessian matrix are provided in Appendix B. Note that $\nabla f(\boldsymbol{\theta}_{ij}) = \mathbf{0}$, and thus the first term of the Taylor series is 0. Therefore, a second-order approximation of the variance of $f(\boldsymbol{e}_{ij}) = r_{ij\cdot(ij)}$ is $\text{Var}(r_{ij\cdot(ij)}) \approx 1/4\text{Var}\left[(\boldsymbol{e}_{ij} - \boldsymbol{\theta}_{ij})^T \boldsymbol{H}[f(\boldsymbol{\theta}_{ij})](\boldsymbol{e}_{ij} - \boldsymbol{\theta}_{ij})\right]$. Moreover, since $\boldsymbol{e}_{ij} - \boldsymbol{\theta}_{ij}$ is

normal with mean $\mathbf{0}$, it follows by the variance of quadratic forms of multivariate normal random vectors that $\text{Var}(r_{ij\cdot(ij)}) \approx \frac{1}{2}\left[\text{tr}\left(\boldsymbol{H}[f(\boldsymbol{\theta}_{ij})]\boldsymbol{\Sigma}_{ij}\boldsymbol{H}[f(\boldsymbol{\theta}_{ij})]\boldsymbol{\Sigma}_{ij}\right)\right]$, where $\text{tr}(\cdot)$ denotes the trace function and $\boldsymbol{\Sigma}_{ij} = \text{Cov}(\boldsymbol{e}_{ij}) \in \mathbb{R}^{2N\times 2N}$ is

$$\boldsymbol{\Sigma}_{ij} = \begin{bmatrix} \text{Cov}(\boldsymbol{e}_{i\cdot(ij)}) & \text{Cov}(\boldsymbol{e}_{i\cdot(ij)}, \boldsymbol{e}_{j\cdot(ij)}) \\ \text{Cov}(\boldsymbol{e}_{j\cdot(ij)}, \boldsymbol{e}_{i\cdot(ij)}) & \text{Cov}(\boldsymbol{e}_{j\cdot(ij)}) \end{bmatrix}.$$

Extending this result to obtain the asymptotic covariance matrix for any pair of partial correlations, we let $\boldsymbol{e}_{ijkm} = \begin{bmatrix} \boldsymbol{e}_{i\cdot(ij)}^T & \boldsymbol{e}_{j\cdot(ij)}^T & \boldsymbol{e}_{k\cdot(km)}^T & \boldsymbol{e}_{m\cdot(km)}^T \end{bmatrix}^T \in \mathbb{R}^{4N}$, and we consider the function

$$g(\boldsymbol{e}_{ijkm}) = \begin{bmatrix} f(\begin{bmatrix} \boldsymbol{e}_{i\cdot(ij)}^T & \boldsymbol{e}_{j\cdot(ij)}^T \end{bmatrix}^T) \\ f(\begin{bmatrix} \boldsymbol{e}_{k\cdot(km)}^T & \boldsymbol{e}_{m\cdot(km)}^T \end{bmatrix}^T) \end{bmatrix} = \begin{bmatrix} r_{ij\cdot(ij)} \\ r_{km\cdot(km)} \end{bmatrix},$$

such that $g : \mathbb{R}^{4N} \to \mathbb{R}^2$. Through a similar process as in the single partial correlation case, we observe that the asymptotic covariance matrix of $g(\boldsymbol{e}_{ijkm})$ is

$$\text{Cov}(\begin{bmatrix} r_{ij\cdot(ij)} & r_{km\cdot(km)} \end{bmatrix}^T) \approx$$

$$\frac{1}{4}\begin{bmatrix} \text{Var}(\boldsymbol{e}_{ij}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{ij})\right]\boldsymbol{e}_{ij}) & \text{Cov}(\boldsymbol{e}_{ij}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{ij})\right]\boldsymbol{e}_{ij}, \boldsymbol{e}_{km}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{km})\right]\boldsymbol{e}_{km}) \\ \text{Cov}(\boldsymbol{e}_{km}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{km})\right]\boldsymbol{e}_{km}, \boldsymbol{e}_{ij}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{ij})\right]\boldsymbol{e}_{ij}) & \text{Var}(\boldsymbol{e}_{km}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{km})\right]\boldsymbol{e}_{km}) \end{bmatrix}.$$

We have already shown the diagonal entries in the single partial correlation case. For the off-diagonal terms, we observe that

$$\text{Cov}(\boldsymbol{e}_{ij}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{ij})\right]\boldsymbol{e}_{ij}, \boldsymbol{e}_{km}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{km})\right]\boldsymbol{e}_{km}) =$$

$$\frac{1}{2}\left(\text{Var}(\boldsymbol{e}_{ijkm}^T\boldsymbol{H}_{ijkm}\boldsymbol{e}_{ijkm}) - \text{Var}(\boldsymbol{e}_{ij}^T\boldsymbol{H}[f(\boldsymbol{e}_{ij})]\boldsymbol{e}_{ij}) - \text{Var}(\boldsymbol{e}_{km}^T\boldsymbol{H}[f(\boldsymbol{e}_{km})]\boldsymbol{e}_{km})\right),$$

where $\boldsymbol{H}_{ijkm} = \begin{bmatrix} \boldsymbol{H}(f(\boldsymbol{e}_{ij})) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}(f(\boldsymbol{e}_{km})) \end{bmatrix} \in \mathbb{R}^{4N\times 4N}$. This is since

$$\begin{aligned} \text{Var}(\boldsymbol{e}_{ijkm}^T\boldsymbol{H}_{ijkm}\boldsymbol{e}_{ijkm}) &= \text{Var}\left(\boldsymbol{e}_{ij}^T\boldsymbol{H}[f(\boldsymbol{e}_{ij})]\boldsymbol{e}_{ij} + \boldsymbol{e}_{km}^T\boldsymbol{H}[f(\boldsymbol{e}_{km})]\boldsymbol{e}_{km}\right), \\ &= \text{Var}(\boldsymbol{e}_{ij}^T\boldsymbol{H}[f(\boldsymbol{e}_{ij})]\boldsymbol{e}_{ij}) + \text{Var}(\boldsymbol{e}_{km}^T\boldsymbol{H}[f(\boldsymbol{e}_{km})]\boldsymbol{e}_{km}) \\ &\quad + 2\text{Cov}(\boldsymbol{e}_{ij}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{ij})\right]\boldsymbol{e}_{ij}, \boldsymbol{e}_{km}^T\boldsymbol{H}\left[f(\boldsymbol{\theta}_{km})\right]\boldsymbol{e}_{km}). \end{aligned}$$

Again, by the variance of a quadratic form of a multivariate normal random vector,

$$\text{Var}(\boldsymbol{e}_{ijkm}^T\boldsymbol{H}_{ijkm}\boldsymbol{e}_{ijkm}) = 2\text{tr}\left(\boldsymbol{H}_{ijkm}\boldsymbol{\Sigma}_{e_{ijkm}}\boldsymbol{H}_{ijkm}\boldsymbol{\Sigma}_{e_{ijkm}}\right),$$

where $\boldsymbol{\Sigma}_{e_{ijkm}} = \mathrm{Cov}(\boldsymbol{e}_{ijkm}) = \begin{bmatrix} \mathrm{Cov}(\boldsymbol{e}_{ij}) & \boldsymbol{\Sigma}_{e_{ijkm}12} \\ \boldsymbol{\Sigma}^T_{e_{ijkm}12} & \mathrm{Cov}(\boldsymbol{e}_{km}) \end{bmatrix} \in \mathbb{R}^{4N \times 4N}$, and

$$\boldsymbol{\Sigma}_{e_{ijkm}12} = \begin{bmatrix} \mathrm{Cov}(\boldsymbol{e}_{i\cdot(ij)}, \boldsymbol{e}_{k\cdot(km)}) & \mathrm{Cov}(\boldsymbol{e}_{i\cdot(ij)}, \boldsymbol{e}_{m\cdot(km)}) \\ \mathrm{Cov}(\boldsymbol{e}_{j\cdot(ij)}, \boldsymbol{e}_{k\cdot(km)}) & \mathrm{Cov}(\boldsymbol{e}_{j\cdot(ij)}, \boldsymbol{e}_{m\cdot(km)}) \end{bmatrix} \in \mathbb{R}^{2N \times 2N}.$$

Thus, the asymptotic covariance between $r_{ij\cdot(ij)}$ and $r_{km\cdot(km)}$ is
$\frac{1}{2}\mathrm{tr}\left(\mathbf{H}[f(\boldsymbol{\theta}_{ij})]\boldsymbol{\Sigma}_{e_{ijkm}12}\mathbf{H}[f(\boldsymbol{\theta}_{km})]\boldsymbol{\Sigma}^T_{e_{ijkm}12}\right)$. □