Methods for Analyzing Multi-Subject Resting-State Neuroimaging Time Series Data

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE SCHOOL OF PUBLIC HEALTH
OF THE UNIVERSITY OF MINNESOTA
BY

Brian B. Hart

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Mark Fiecas and
Dr. Lynn Eberly, Advisers

May 2019

Acknowledgements

I have many people to thank for contributing to my dissertation in so many different ways. I want to thank my friends and family who have supported me throughout the past five years. I want to especially thank my parents who provided me with the opportunity to pursue my studies as far as I wanted and encouraged me along the way and my wife Erin who patiently allowed me to live as a stay at home dog dad as I wrote my dissertation over the past few years. Thank you to my fellow students who struggled through the course work and waded into the world of research along side me. I would not have gone on to a PhD program without the guidance of Dr. Paul Roback and Dr. Katie Ziegler-Graham at St. Olaf. Thanks to Sally Olander for answering any question and solving any problem that came up over the last five years. Classes taught by Dr. John Hughes, Susan Wei, and Dr. Wei Pan have all be instrumental to how I think about and practice statistics. I have been lucky to work with incredible scientists and mentors through my research assistantships, all of which deserve mention. Dr. Shweta Sharma and Dr. Jim Neaton gave me a taste of the world of international clinical trials, Olga Gurvich introduced my to consulting while working with students from the School of Nursing, and Dr. Katie Cullen and Dr. David Redish challenged me with collaborative work in a new area. Thanks to my co-authors Dr. Ivor Cribben and Dr. Michele Guindani who provided guidance and expertise to help our research through the writing and publication process. Thanks to Dr. Steve Malone for providing expertise on the Minnesota Twin Family Study, supplying the data for much of my work, and serving on my dissertation committee. Thanks to Dr. Lynn Eberly for serving on my committee and being a fantastic mentor in terms of both statistics and professionalism. Thanks to Dr. Jim Hodges who was always willing to read through a draft of a paper or proposal and provide thoughtful feedback. Discussion from imaging group meetings, seminars, and your class were

always enlightening and you have certainly helped form my view on the usefulness (or lack there of) of asymptotics. Finally, thank you Dr. Mark Fiecas. You have been a fantastic advisor and mentor over the past three years. You taught me much of what I know about time series and neuroimaging, provided research guidance when asked, gave me freedom to work independently, and introduced me to the larger network of international statisticians.

Dedication

This work is dedicated to my parents who gave me the opportunity to pursue my education and to my wife Erin for supporting me throughout my PhD studies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

1.1.1 Resting State Neuroimaging Data

Resting-state neuroimaging data is increasingly available from a variety of different imaging modalities, of which functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) are two of the most common and the focus of this dissertation. Unlike task based imaging, participants are generally told to relax and do nothing in particular during a resting-state scan. Instead of trying to discern which networks or areas of the brain predominate during certain tasks, resting-state imaging gives researchers the opportunity to capture the underlying intrinsic activity patterns in the brain. This intrinsic brain activity characterizes the typical behavior of the brain and how it changes over time. Resting-state data has the ability to show that different people not only react to and perform tasks in different manners, but that the background activity and fundamental cognitive processing may occur in different manners as well.

Each imaging modality offers different advantages when it comes to resting-state data collection. fMRI offers very high spatial resolution by measuring the blood oxygenation level dependent (BOLD) signal at hundreds of thousands of locations within the brain. The high spatial resolution comes at the cost of temporal resolution,

as fMRI typically only captures a single image of the brain every 2 seconds. Even the latest fMRI technology struggles to sample faster than once every half a second. Another factor that inhibits the temporal resolution of fMRI data is that it measures the oxygenation level as a proxy for brain activity, the idea being that more active areas of the brain will receive increased blood flow to account for the extra demand for oxygen. Because the BOLD signal measures blood oxygenation, the signal can only move as fast as blood and may not fully capture the electrical activity of the brain.

Despite these drawbacks, the high spatial resolution of fMRI data makes it an ideal choice for analyses intended to localize brain activity to specific regions of interest. High spatial resolution comes with its own set of challenges for the statistician. A typical resting-state fMRI acquisition protocol may collect upwards of 5 minutes worth of data sampled every 2 seconds. This results in a multivariate time series measuring the BOLD signal at approximately 300,000 locations across 150-250 time points, resulting in very large datasets with complex spatial and temporal dependence structures. To reduce the data to a digestible form, brain atlases are often used to group sets of voxels (3D locations in the brain at which the BOLD signal is captured) into functionally or spatial similar sets known as regions of interest (ROI) prior to analysis. Once the data has been summarized to a set of between 5 and 120 ROI, many methods assume spatial independence between ROI to further simplify the problem. Even at this stage the statistician must account for temporal autocorrelation present in the fMRI data, a main goal of the method presented in Chapter 2.

EEG data offers a solution to the relatively poor temporal resolution of fMRI data. EEG data is typically sampled at a rate between 128 and 1000 Hertz (HZ). Because EEG directly measures electrical activity, it also does not suffer from the drawbacks of the BOLD signal in terms of capturing brain activity at a high temporal resolution. The electrical activity captured by EEG can directly match the pace of the electrical activity of the brain. This impressive temporal resolution comes at the cost of the spatial definition offered by fMRI. Non-invasive EEG is only measured on the surface

of the scalp at between 3 and 256 different locations (often referred to as channels or nodes).

Even in the case of the newer 256 channel EEG devices, because all measurements are from the surface, localizing the signal to its source within the brain offers an additional challenge for the statistician. Additionally, the spatial and temporal dependencies present in fMRI data also occur in EEG data. The size of the data also becomes an issue with EEG data. If 5 minutes of resting-state EEG data is collected from 3 different channels at 128 HZ, as is the case with the Minnesota Twin Family Study (MTFS) data used in Chapters 3 and 4, the resulting dataset has 115,200 observations per participant. When considering hundreds of participants at once, the size of the data again becomes unmanageable and data reduction techniques are required prior to analysis.

1.1.2 Varied Analysis Goals and Approaches

Given resting-state data for a particular imaging modality, a variety of research questions can be addressed depending on the type of analysis chosen by the statistician. Chapters 2, 3, and 4 each address distinct research questions by viewing the time series in a different light and selecting different types of time series analyses.

Chapter 2 uses fMRI data to study what is known as functional connectivity (FC), defined as temporal dependence, measured through cross-correlations, in the BOLD signals of different brain regions (Friston et al., 1993). Resting-state FC attempts to estimate the intrinsic network present in the resting brain given a set of ROI. We examine FC networks from participants of the Alzheimer's Disease Neuroimaging Initiative (ADNI), assessing the differences in FC between patients with Alzheimer's disease (AD) and their cognitively normal (CN) counterparts.

Chapter 3 uses EEG data to perform spectral time series analysis on data collected from adolescent twins through the MTFS. Spectral time series analysis is a frequency domain technique, which decomposes the time series into a set of waveforms oscillating

at different frequencies and determines which frequencies are driving the variation in the EEG time series. Spectral analysis offers a completely different view of time series data compared to the time domain methods such as the FC analysis in Chapter 2 and the EEG microstate analysis in Chapter 4.

Chapter 4 also uses MTFS resting-state EEG data but to a completely different end. Rather than summarizing the EEG time series' variance characteristics through spectral analysis, Chapter 4 considers the dynamics of the EEG brain activity through a microstate analysis. Microstate analysis shares some similarities with the fMRI FC analysis in Chapter 2, but allows the network of activity patterns to change across the resting-state recording. In this way, microstate analysis does not seek a single predominant pattern but seeks a set of states such that one of these states dominates for a short block of time before switching rapidly to another state.

1.1.3 Accounting for Experimental Design

On top of the challenges posed by the size and dependence structure of the data, resting-state neuroimaging data is often collected as part of a larger study. Whichever avenue of analysis one follows, it is desirable to account for the structure of the data by modeling the dependence between participants or between visits from a single participant.

In the case of Chapter 2, the ADNI data was collected longitudinally and each participant was scanned between 1 and 6 times over the course of a few years. In our analysis, we explicitly model this longitudinal structure to both adequately account for within-subject dependence and to allow inference on the change in FC as the participants age. As we show, simpler methods that fail to account for the longitudinal nature of the data result in inflated type I error rates.

We use the same resting state EEG data from the MTFS for both Chapter 3 and Chapter 4. In this data collected from twins, we found that monozygotic (MZ) twins are generally more similar than dizygotic (DZ) twins, who are generally more similar

than two randomly selected individuals. Chapter 3 tackles this twin study design by sharing information across all participants when performing spectral analysis, but allowing certain participants to be more similar than others. This model finds groups of participants with very similar spectral characteristics and estimates their spectral densities together while allowing separate groups of participants to have vastly different estimates. We fit the EEG microstate analysis in Chapter 4 on the twin pair level in recognition that twins may even share underlying microstates. The model estimates the switching dynamics of these microstates separately for each participant and also learns the amount of shared information within a twin pair to allow different levels of similarity stemming from the MZ or DZ twin relationships.

1.1.4 Bayesian Nonparametrics

Before more thoroughly introducing each chapter, we briefly introduce a class of methods used in Chapters 3 and 4 of this dissertation. The term Bayesian Nonparametrics is used to refer to a class of models which includes the Dirichlet Process (DP), nested DP (Rodriguez et al., 2008), beta process (BP) (Fox et al., 2014), and Indian buffet process (IBP) (Ghahramani & Griffiths, 2006) among others. While their applications and interpretations are varied, this collection of methods is generally used to solve two problems: 1) Estimate the mixture weights in a potentially infinite mixture model and 2) Cluster similar observations together where the number of clusters is unknown. We use the first of these in Chapter 3 when mixing together beta probability density functions (PDFs) to find estimated spectral density curves. We use the second in Chapter 3 to cluster participants with similar spectral densities and in Chapter 4 to cluster observations with similar underlying behavior into microstates. While the clustering nature of these Bayesian nonparametric models is a useful tool to explain the heterogeneity within the samples, one should be wary of using them for inference on the true number of clusters as discussed in Chapter 4.

We now introduce the three remaining chapters of the dissertation with a brief

abstract for each before moving on to full presentations of these models.

1.2 Estimating Functional Connectivity in Longitudinal fMRI Data

Many neuroimaging studies collect functional magnetic resonance imaging (fMRI) data in a longitudinal manner. However, the current fMRI literature lacks a general framework for analyzing functional connectivity (FC) networks in fMRI data obtained from a longitudinal study. In this work, we build a novel longitudinal FC model using a variance components approach. First, for all subjects' visits, we account for the autocorrelation inherent in the fMRI time series data using a non-parametric technique. Second, we use a generalized least squares approach to estimate 1) the within-subject variance component shared across the population, 2) the baseline FC strength, and 3) the FC's longitudinal trend. Our novel method for longitudinal FC networks seeks to account for the within-subject dependence across multiple visits, the variability due to the subjects being sampled from a population, and the autocorrelation present in fMRI time series data, while restricting the number of parameters in order to make the method computationally feasible and stable. We develop a permutation testing procedure to draw valid inference on group differences in the baseline FC network and change in FC over longitudinal time between a set of patients and a comparable set of controls. To examine performance, we run a series of simulations and apply the model to longitudinal fMRI data collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Overall, we found no difference in the global FC network between Alzheimer's disease patients and healthy controls, but did find differing local aging patterns in the FC between the left hippocampus and the posterior cingulate cortex.

1.3 Spectral Analysis of EEG Data from Twins

Electroencephalography (EEG) is a non-invasive neuroimaging modality that captures electrical brain activity many times per second. We seek to estimate power spectra from EEG data that was gathered for 557 adolescent twin pairs through the Minnesota Twin Family Study (MTFS). Typically, spectral analysis methods treat time series from each subject separately, and independent spectral densities are fit to each time series. Since the EEG data was collected on twins, it is reasonable to assume that the time series have similar underlying characteristics, so borrowing information across subjects can significantly improve estimation. We propose a Nested Bernstein Dirichlet Prior model to estimate the power spectrum of the EEG signal for each subject by smoothing periodograms within and across subjects while requiring minimal user input to tuning parameters. Furthermore, we leverage the MTFS twin study design to estimate the heritability of EEG power spectra. The method also facilitates heritability analyses on features of the estimated spectral density curves such as peak frequency and frequency band power. Through simulation studies designed to mimic the MTFS, we show our method out-performs a set of other popular methods.

1.4 EEG Microstate Analysis on Twins

EEG microstate analysis is an investigation into the collection of distinct temporal blocks that characterize the electrical activity of the brain. Brain activity within each of these microstates is stable, but it can switch rapidly between different microstates in a non-random way. We propose a Bayesian nonparametric model that concurrently estimates the number of microstates and their underlying behavior. We use a Markov switching vector autoregressive (VAR) framework, where a hidden Markov model (HMM) controls the non-random state switching dynamics of the EEG activity and a VAR model defines the behavior of all time points within a given state. We analyze

resting state EEG data from twin pairs collected through the Minnesota Twin Family Study, consisting of 70 epochs that each correspond to 2 seconds of EEG data for a total of 140 seconds of data per participant. We fit our model at the twin pair level, sharing information within epochs from the same participant and within epochs from the same twin pair. We capture within twin pair similarity by using a Beta process Bernoulli process to consider an infinite library of microstates and allowing each participant to select a finite number of states from this library. The state spaces of highly similar twins may completely overlap while dissimilar twins could select completely distinct state spaces. In this way, our flexible Bayesian nonparametric model defines a sparse set of states which describe the EEG data. All epochs from a single participant use the same set of states and are assumed to adhere to the same state switching dynamics in the HMM model, enforcing within-participant similarity.

Chapter 2

A Longitudinal Model for Functional Connectivity Networks Using Resting-State fMRI

2.1 Introduction

Resting-state functional magnetic resonance imaging (fMRI) captures a series of images of the brain in subjects who are not given a particular task to perform while in the scanner. The scanner repeatedly captures blood oxygenation level dependent (BOLD) signals at hundreds of thousands of locations within the brain, creating a time series of images of the brain. By capturing the BOLD signal of the resting brain, resting-state fMRI provides an opportunity for researchers to examine the functional connectivity (FC) within a set of regions not tied to a particular task. We define FC as the temporal dependence, measured through cross-correlations, in the BOLD signals between brain regions (Friston et al., 1993). Identifying group differences in FC can help better understand the underlying neurological process of a disease and its progression. Observed group differences can also potentially form biomarkers to be used for early detection and treatment of neurological disorders (Fox & Raichle, 2007).

Previous works have demonstrated the utility of FC analysis. For example, past research has identified altered FC between healthy aging patients and those with

Alzheimer's disease (AD). Even among cognitively normal (CN) individuals, FC demonstrates aging effects that are heterogeneous between different brain regions (Chen et al., 2016). Chase (2014) and Hafkemeijer et al. (2012) showed altered FC patterns beyond healthy aging in patients with dementia and AD. Others, including Wang et al. (2007) have noted abnormal FC in various stages of AD. Wang et al. (2012) even demonstrated the impact of family history of AD on FC. In addition, Xiang et al. (2013) showed decreased FC from CN patients to mild cognitively impaired (MCI) patients to AD patients, and Li et al. (2015) found decreased FC for CN patients who progressed to MCI over the following 24 months compared to non-progressers. These previous works, however, used *cross-sectional models*, which only consider data from a single time point. Ren et al. (2016), using longitudinal data, showed abnormal FC in various stages of AD and Staffaroni et al. (2018) analyzed default mode network (DMN) connectivity longitudinally in patients with AD, but summarized to a single connectivity strength metric of the network instead of considering all pair-wise comparisons.

Of the studies mentioned above, only Ren et al. (2016) and Staffaroni et al. (2018) used truly longitudinal fMRI data. Aging effects are often measured by comparing young and elderly groups rather than following one group of subjects over time. A comprehensive longitudinal model that tests for differences in baseline and trend is needed to verify and expand on the previous results. Zhu et al. (2015) performed a longitudinal FC analysis for concussion patients, but their method was ad hoc and specific to their unique dataset. Finn & Constable (2016) demonstrated that CN patients have distinct brain signatures in fMRI images, implying that separate scans from a single individual exhibit dependence, and Ge et al. (2017) demonstrated the heritability of fMRI FC in a longitudinal study. These findings can be leveraged in a longitudinal framework to better model aging effects.

Methods for both longitudinal imaging data and cross-sectional fMRI FC exist, but, to our knowledge, no modeling framework exists for fMRI FC collected in a longitudinal manner. Recent work by Fiecas et al. (2017) developed a model to carry

out hypothesis tests on the difference between FC networks, but this method was developed for a cross-sectional study. Taking another approach, Simpson et al. (2013) provided an overview of graphical network analysis techniques for fMRI connectivity. The methods they described measure graphical network traits such as small-worldness and graph centrality, but also fail to account for any longitudinal dependence present in the data. Other methods have been developed for analyzing longitudinal data from other modalities, though these methods are not appropriate for modeling longitudinal FC networks. For instance, Gertheiss et al. (2013) used longitudinal diffusion tensor imaging (DTI) data to model health outcomes. Unfortunately, the differing nature of DTI and fMRI data (i.e. fMRI data consists of a time series for each visit) make the direct application of this method impossible. Guillaume et al. (2014) also proposed a massive univariate longitudinal model for neuroimaging data, but it also was not designed to account for the autocorrelation present in fMRI time series.

A wealth of literature exists on methods for longitudinal data analysis to account for the dependence between two visits from a single participant (Laird, 2004), but these methods have not yet been integrated with current fMRI FC models. In this paper, we fill this large gap in the literature by proposing a novel longitudinal fMRI FC network model and inference procedure that considers the set of all possible pairwise groupings of the chosen ROI in resting-state fMRI data. Our longitudinal variance components FC network model accounts for within-subject dependence across multiple visits, variability due to subjects being sampled from a population, and any autocorrelation present in individual fMRI time series. We also propose an efficient permutation-based inference procedure that allows for valid hypothesis testing of group differences in baseline FC and FC aging effects. We show that our method is superior to the mass univariate linear mixed effects (LME) model commonly used for longitudinal neuroimaging data. Our work will build on previous results on the clinical utility of FC as a potential biomarker for AD. It should be noted that in this paper we use the terms FC and FC network interchangeably, but note that a distinction is sometimes drawn between the two. In particular, our model focuses

on inference on the strength of the edges in the network and not on the trends in the topological properties of the overall network such as the methods of Simpson & Laurienti (2015) and Simpson & Laurienti (2016). We adopt the FC network term for our method to emphasize our ability to perform valid inference on the set of all pairwise connections in a pre-defined network instead of simply performing marginal inference on each ROI pair.

Our contributions to the field in this article are 1) a novel general framework for longitudinal analysis of fMRI FC networks, and 2) a novel application of the permutation testing procedure of Ter Braak (1992) to fMRI FC analysis to allow valid inference at the local and global level. The remainder of the paper is organized as follows. Section 2 formally introduces the model, including the estimation and inference procedures. It also explains the design of the simulation study and describes the application of our model to data from the Alheimer's Disease Neuroimaging Initiative (ADNI). Section 3 presents the results of the simulation study and ADNI data analysis. We close with a discussion of the simulation study and ADNI data analysis results along with proposals for future work in Section 4 and a conclusion in Section 5. R code for the methods proposed in this paper may be found at `https://github.com/mfiecas/longitudinalFC`.

2.2 Materials and Methods

2.2.1 Model Specification

Suppose we have a cohort of N individuals and let P denote the number of ROI selected for a FC network analysis. We collect a P-variate, fMRI time series of length T from the preprocessed fMRI images of each of the N subjects at each visit. Let the subscripts i and j denote subject and visit, respectively. Subject i returns for J_i total visits, and the cohort has a total of $J = \sum_{i=1}^{N} J_i$ visits. Let $\boldsymbol{y}_i$ represent the vector of sample correlation coefficients for subject i of length QJ_i, where $Q = P(P-1)/2$ is

the number of ROI pairs. Within $\boldsymbol{y}_i$, the Q correlations from the first visit, $\boldsymbol{y}_{i1}$, are followed by the Q correlations from the second visit, $\boldsymbol{y}_{i2}$, and so on until the J_i-th visit. The full response vector $\boldsymbol{y}$ is formed by stacking the N different $\boldsymbol{y}_i$ vectors and is, thus, of length QJ. Our longitudinal model for FC is a linear model with baseline effect $\boldsymbol{\beta}_0$ and longitudinal trend $\boldsymbol{\beta}_1$, where each of these model parameters is a vector of length Q. We denote the time at visit j for subject i as v_{ij}. The vector $\boldsymbol{v}_i$ is formed by stacking the J_i distinct $v_{ij} \otimes \mathbb{1}_Q$ vectors for subject i, where $\mathbb{1}_Q$ is a vector of ones of length Q and $\otimes$ is the Kronecker product. Likewise $\boldsymbol{v}$ is formed by stacking the N distinct $\boldsymbol{v}_i$ vectors. Depending on the nature of the data and the research questions at hand, v_{ij} can be set to the visit number, the time since baseline, or the patient's age. Then, denoting element-wise multiplication with $*$, our model has the following linear form:

$$\boldsymbol{y} = \mathbb{1}_J \otimes \boldsymbol{\beta}_0 + \boldsymbol{v} * (\mathbb{1}_J \otimes \boldsymbol{\beta}_1) + \boldsymbol{\varepsilon}, \text{ where Var } (\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} + \boldsymbol{\Psi}. \tag{2.1}$$

The key element in our longitudinal linear model is the variance structure of the error term. We separate the error variance into two components, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$, each of dimension $QJ \times QJ$. $\boldsymbol{\Sigma}$ accounts for the within-visit variance and the temporal autocorrelation in the fMRI time series, and $\boldsymbol{\Psi}$ accounts for the variability and covariability arising from the heterogeneity across subjects and the within-subject covariation coming from the longitudinal design. $\boldsymbol{\Sigma}$ is block diagonal where each $Q \times Q$ block, $\boldsymbol{\Sigma}_{ij}$, accounts for the within-visit variance present in visit j for subject i for the Q pairs of ROI. $\boldsymbol{\Psi}$ is also block diagonal with a $QJ_i \times QJ_i$ block for participant i. These diagonal blocks do not differ between subjects except through their dimensions, which depend on the number of visits for each subject. Let $\boldsymbol{\Psi}_{diag}$ be an arbitrary diagonal block of $\boldsymbol{\Psi}$. We then further break $\boldsymbol{\Psi}_{diag}$ into two components, $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$. $\boldsymbol{\Psi}_0$ is a $Q \times Q$ block that is repeated along the diagonal of each $\boldsymbol{\Psi}_{diag}$. This term models the within-visit covariability not captured by $\boldsymbol{\Sigma}$. $\boldsymbol{\Psi}_1$ is a $Q \times Q$ block that populates the off diagonal blocks of $\boldsymbol{\Psi}_{diag}$, modeling the within-subject,

across-visit covariability coming from the longitudinal design.

We write Equation 2.1 in the form of a linear model with a vector response, allowing us to use existing methods for estimating the parameters and for statistical inference. To this end, our model can also be written in the standard linear model form with a design matrix $\boldsymbol{X}$. Let $\boldsymbol{X}_{ij} = [1 \quad v_{ij}] \otimes \boldsymbol{I}_Q$, where $\boldsymbol{I}_Q$ is the $Q \times Q$ identity matrix. To form $\boldsymbol{X}_i$, the portion of the design matrix specific to subject i, we stack the J_i individual $\boldsymbol{X}_{ij}$. Likewise, to form $\boldsymbol{X}$ we stack the N individual $\boldsymbol{X}_i$. If we define $\boldsymbol{\beta}$ as a vector of length $2Q$ where the first Q elements are $\boldsymbol{\beta}_0$ and the last Q elements are $\boldsymbol{\beta}_1$, then Equation 2.1 can be written as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Figure 2.1 shows a diagram of the model layout for a single subject.

2.2.2 Estimating Within Visit Covariance

We start by estimating the sample correlation coefficient for all ROI pairs for all visits and their corresponding variances and covariances. Let $(w_{1t}, \ldots, w_{Pt})'$ for $t \in \{1, 2, \ldots, T\}$ be the time series of preprocessed BOLD signals from P ROI for a single visit, so that w_{pt} indicates the t-th time point from the p-th ROI. Then for the p-th and q-th ROI, $r_{pq} = \sum_{t=1}^{T}(w_{pt} - \bar{w}_p)(w_{qt} - \bar{w}_q) / \sqrt{\sum_{t=1}^{T}(w_{pt} - \bar{w}_p)^2 \sum_{t=1}^{T}(w_{qt} - \bar{w}_q)^2}$. We now address how to estimate the variance (or standard error) of sample correlations and the covariation between pairs of sample correlations, whenever these sample correlations are obtained from data that exhibit some degree of autocorrelation. To this end, we follow the approach described by Roy (1989) and Melard et al. (1991). First, define $\hat{\gamma}_{pq}(u) = \sum_{t=1}^{T-u}(w_{pt} - \bar{w}_p)(w_{q,t+u} - \bar{w}_q)/T$. Then, letting $h(\cdot)$ be the modified Bartlett window with bandwidth $b(T)$, we set $\hat{\Theta}(p, q, p', q') = \sum_{u=-T+1}^{T-1} h^2(u)\hat{\gamma}_{pq}(u)\hat{\gamma}_{p'q'}(u)$. Using $\hat{\gamma}$ and $\hat{\Theta}$, we then let $\hat{\Delta}(p, q, p', q') = \hat{\Theta}(p, q, p', q') / \sqrt{\hat{\gamma}_{pp}(0)\hat{\gamma}_{qq}(0)}$ Finally, we can obtain an estimate of the variance and covariance using the following

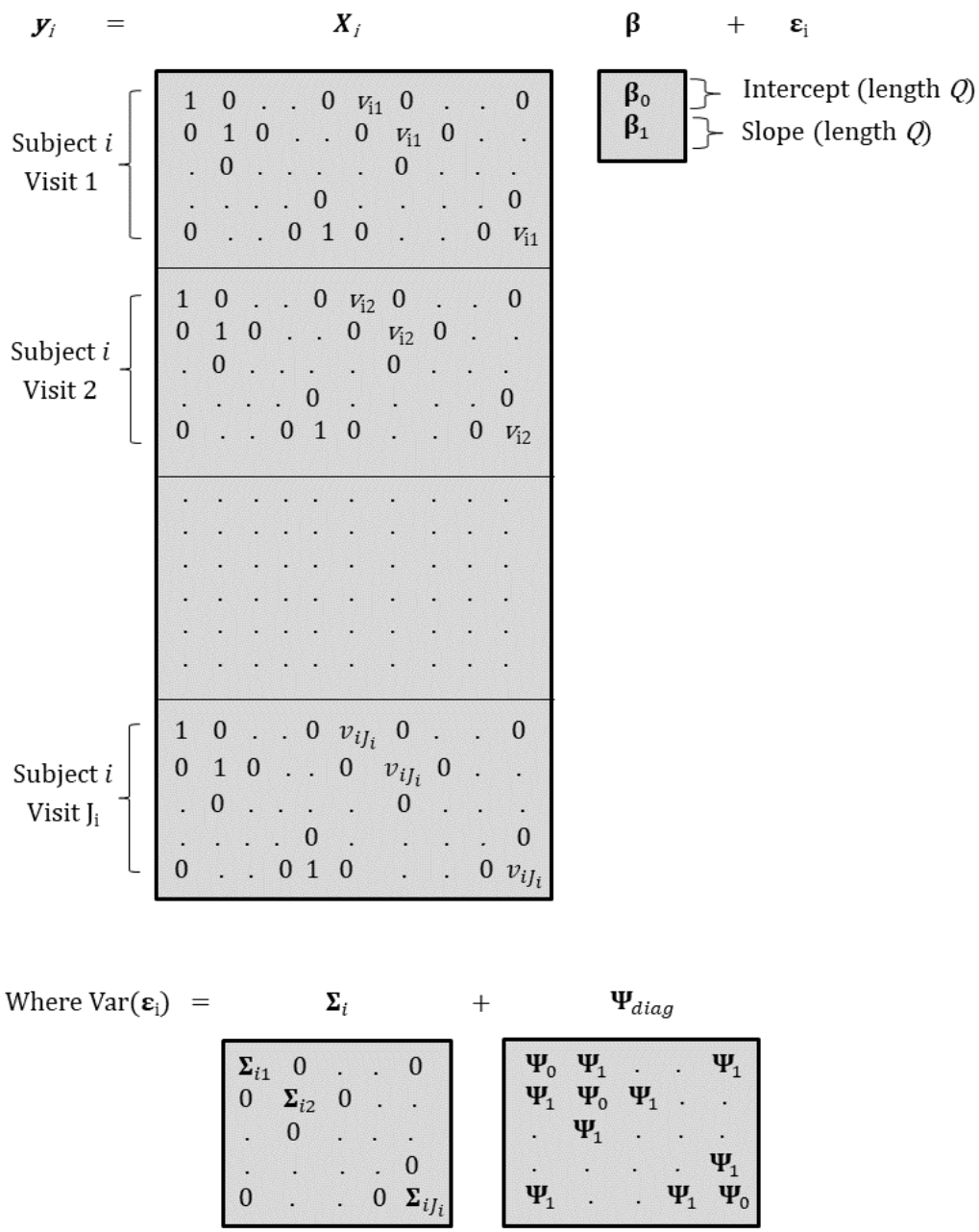


Figure 2.1: A diagram of the model for a single subject.

formula:

$$\begin{aligned}\widehat{Cov}(r_{pq}, r_{p'q'}) = [0.5 r_{pq} r_{p'q'} \{&\hat{\Delta}(p, p', p, p') + \hat{\Delta}(p, q', p, q') + \hat{\Delta}(q, p', q, p') + \hat{\Delta}(q, q', q, q')\} \\ &- r_{pq}\{\hat{\Delta}(p, p', p, q') + \hat{\Delta}(q, p', q, q')\} - r_{p'q'}\{\hat{\Delta}(q, p', p, p') + \hat{\Delta}(q, q', p, q')\} \\ &+ \hat{\Delta}(p, p', q, q') + \hat{\Delta}(q, p', p, q')]/T, \end{aligned} \tag{2.2}$$

where setting $p = p'$ and $q = q'$ gives us the variance of the sample correlation between the p-th and q-th ROI time series. We use Equation 2.2 to populate each Σ_{ij} block to get our estimates $\hat{\boldsymbol{\Sigma}}_{ij}$. Roy (1989) derived the large sample covariance of two correlation coefficients from autocorrelated time series, and Melard et al. (1991) showed that Equation 2.2 consistently estimates this large sample variance.

2.2.3 Estimating Between Subject Covariance, $\boldsymbol{\Psi}$, and $\boldsymbol{\beta}$

Using a generalized least squares (GLS) approach, we now proceed with the estimation of the between subject covariance $\boldsymbol{\Psi}$ and the regression coefficients $\boldsymbol{\beta}$, conditional on the previously estimated within-visit covariances, $\hat{\boldsymbol{\Sigma}}_{ij}$. We use consistent estimators from the work of Laird (2004). Although the framework allows for many different structures for $\boldsymbol{\Psi}$, we assume a block compound symmetry structure so that all diagonal blocks, $\boldsymbol{\Psi}_0$, are equal and all off diagonal blocks, $\boldsymbol{\Psi}_1$, are equal. The block compound symmetry assumption keeps the parameter space to a reasonable size, but note that one could easily consider other forms of $\boldsymbol{\Psi}$, such as an autoregressive structure, with minimal modification to the estimation procedure. We use the ordinary least squares estimator $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ to provide a good starting estimate of $\boldsymbol{\beta}$. We then update the two components of $\boldsymbol{\Psi}$ using method of moments style estimators. These estimators resemble the empirical variance estimates. To estimate $\boldsymbol{\Psi}_0$ we find the sum of squared errors for each visit and subtract the previously estimated $\hat{\boldsymbol{\Sigma}}_{ij}$ for each visit. We then sum all of these terms and divide the remaining covariance matrix by the total number of visits. This estimator is then an empirical estimate of the variance remaining after accounting for the already calculated $\boldsymbol{\Sigma}_{ij}$ terms averaged

across all visits for all subjects. We estimate $\boldsymbol{\Psi}_1$ in a similar fashion, but here we no longer have to subtract any $\boldsymbol{\Sigma}$ terms since $\boldsymbol{\Sigma}$ is set to zero for the off diagonal blocks which $\boldsymbol{\Psi}_1$ occupies. $\boldsymbol{\Psi}_1$ then empirically estimates the average covariance between any two visits from a single participant. Exact formulas for $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$ are shown below:

$$\hat{\boldsymbol{\Psi}}_0 = \frac{1}{\sum_{i=1}^{N} J_i}\left\{ \sum_{i=1}^{N}\sum_{j=1}^{J_i}(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}\hat{\boldsymbol{\beta}})(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}\hat{\boldsymbol{\beta}})' - \hat{\boldsymbol{\Sigma}}_{ij} \right\}, \quad \text{and} \tag{2.3}$$

$$\hat{\boldsymbol{\Psi}}_1 = \frac{1}{\sum_{i=1}^{N} \frac{J_i(J_i-1)}{2}} \sum_{i=1}^{N}\sum_{j\neq k}(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}\hat{\boldsymbol{\beta}})(\boldsymbol{y}_{ik} - \boldsymbol{X}_{ik}\hat{\boldsymbol{\beta}})'. \tag{2.4}$$

To increase model parsimony, different structures can be considered for $\boldsymbol{\Sigma}_{ij}$, $\boldsymbol{\Psi}_0$, and $\boldsymbol{\Psi}_1$. For example, to enforce a diagonal structure, set all off-diagonal elements to 0, or to enforce a compound symmetry structure, set all diagonal elements to the average of the diagonal elements and likewise for the off-diagonal elements (Laird, 2004). Changing the form of $\boldsymbol{\Sigma}_{ij}$, $\boldsymbol{\Psi}_0$, and $\boldsymbol{\Psi}_1$ allows the model to be fit with flexible variance assumptions as is often done in traditional generalized least squares linear models. We recommend selection of a parsimonious structure for the variance components to stabilize estimation and allow ROI pairs to borrow information from each other in the estimation of their variance terms.

With an estimate of $\boldsymbol{\Psi}$ obtained using Equations 2.3 and 2.4, we can now use the standard GLS formula to update the regression coefficients as follows: $\hat{\boldsymbol{\beta}} = \{\boldsymbol{X}'(\hat{\boldsymbol{\Sigma}}+\hat{\boldsymbol{\Psi}})^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}'(\hat{\boldsymbol{\Sigma}}+\hat{\boldsymbol{\Psi}})^{-1}\boldsymbol{y}$. At this point we have two choices: iteratively update $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\beta}}$ until convergence (full convergence), or accept the estimates (one-step) and proceed with the inferential procedure. As we will see later, we will use a permutation test for inference, making it vital that we have a computationally efficient way to estimate the effects $\boldsymbol{\beta}$ and variance component $\boldsymbol{\Psi}$. The one-step estimator provides a significant advantage in computing time as $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\beta}}$ must be estimated for each permutation of the inference procedure (Ganjgahi et al., 2015). One-step

GLS estimators are not new and have many desirable statistical properties, including maintaining consistency (Amemiya, 1977).

We estimate $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Psi}},$ and $\hat{\boldsymbol{\Sigma}}$ for each group (CN and AD for ADNI) separately using this estimation procedure. Superscripts on the parameter estimates denote the group (e.g. $\hat{\boldsymbol{\beta}}^{G_1}, \hat{\boldsymbol{\Psi}}^{G_1}$, and $\hat{\boldsymbol{\Sigma}}^{G_1}$ are the estimates for group 1).

2.2.4 Inference

We consider two general hypothesis tests in our longitudinal FC model: the group difference in the baseline FC, and the group difference in the longitudinal trend in FC; other tests are possible with slight modifications to the procedure that we describe below. For each hypothesis, we would like to test the group difference in both the global FC network of pre-defined ROI and the local ROI pair FC. We refer to the vector wide test of a difference in the parameter vector $\boldsymbol{\beta}_0$ or $\boldsymbol{\beta}_1$ as a global test and refer to a test of a group difference in a single element of $\boldsymbol{\beta}_0$ or $\boldsymbol{\beta}_1$ as a local test. To accomplish our hypothesis testing objectives, we use the Wald statistic, $\{\boldsymbol{C}(\hat{\boldsymbol{\beta}}^{G1} - \hat{\boldsymbol{\beta}}^{G2})\}'[\boldsymbol{C}\{\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}^{G1}) + \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}^{G2})\}\boldsymbol{C}']^{-1}\boldsymbol{C}(\hat{\boldsymbol{\beta}}^{G1} - \hat{\boldsymbol{\beta}}^{G2})$, and adjust the contrast matrix, $\boldsymbol{C}$, depending on the hypothesis of interest. For instance, to test for a global difference in $\boldsymbol{\beta}_0$, we replace all $\boldsymbol{\beta}$ terms with the $\boldsymbol{\beta}_0$ vector for the proper group and set the contrast matrix, $\boldsymbol{C}$, to the $Q \times Q$ identity matrix.

We estimate the variance of each group's regression coefficients using $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \{\boldsymbol{X}'(\hat{\boldsymbol{\Sigma}} + \hat{\boldsymbol{\Psi}})^{-1}\boldsymbol{X}\}^{-1}$. Because the standard χ^2 statistical leads to very high type I error levels when more than 3 ROI are selected, we resort to the following permutation testing procedure:

1. Calculate residuals from the fitted model for each subject: $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}^{G}$ for subject i in group G.

2. Permute group assignments of $\boldsymbol{e}_i$.

3. Add the nuisance signal back to $\boldsymbol{e}_i$ based on new permuted group assignments G^*. For the main effect (intercept) tests we add in the longitudinal trends by

setting $\boldsymbol{e}^*_{ij} = \boldsymbol{e}_{ij} + v_{ij}\hat{\boldsymbol{\beta}}_1^{G^*}$. Likewise, for the interaction (slope) tests we set $\boldsymbol{e}^*_{ij} = \boldsymbol{e}_{ij} + \hat{\boldsymbol{\beta}}_0^{G^*}$.

4. Refit the model on $\boldsymbol{e}^*$, the permuted, adjusted, and stacked residuals from step 3.

5. Calculate a new Wald statistic for the fitted values of $\hat{\boldsymbol{\beta}}^{G^*}$ and $\hat{\boldsymbol{\Psi}}^{G^*}$.

We repeat steps 2 through 5 a large number of times to create a permutation distribution to be used as a reference distribution of the originally calculated test statistic. Although not immediately evident, step 3 in this permutation procedure is essential for valid inference. If we are testing for a difference in baseline FC, we must add in the longitudinal trend of the new permuted group assignment and, likewise, must add in the permuted group baseline FC when testing for differences in longitudinal trend. This step ensures that we are controlling for potential group differences in $\boldsymbol{\beta}_0$ when testing for difference in $\boldsymbol{\beta}_1$ and vice versa. The number of permutations determines the precision of the p-value and should be chosen to be large enough to offer sufficient precision after any multiple comparisons adjustment. In our particular data example using ADNI data with 10 ROI we chose to run 10,000 permutations. Because the obtained p-values are estimated discrete values, we additionally use a permutation p-value correction procedure, the necessity of which was shown by Phipson et al. (2010). To account for the fact that $2Q$ local hypotheses are tested simultaneously, we then apply the false discovery rate (FDR) controlling procedure of Benjamini & Hochberg (1995) to the corrected p-values from the local tests. The permutation p-value correction helps avoid unadjusted p-values with value 0 which may improperly maintain significance after a multiple comparisons correction.

Chung & Romano (2013) showed that studentized test statistics, such as the proposed Wald statistic, allow for valid inference in many permutation test settings. A recent comparison of the performance of different permutation strategies by Winkler et al. (2014) showed that the Ter Braak permutation testing procedure we use maintains nominal Type I error and is fairly robust (Ter Braak, 1992). This method

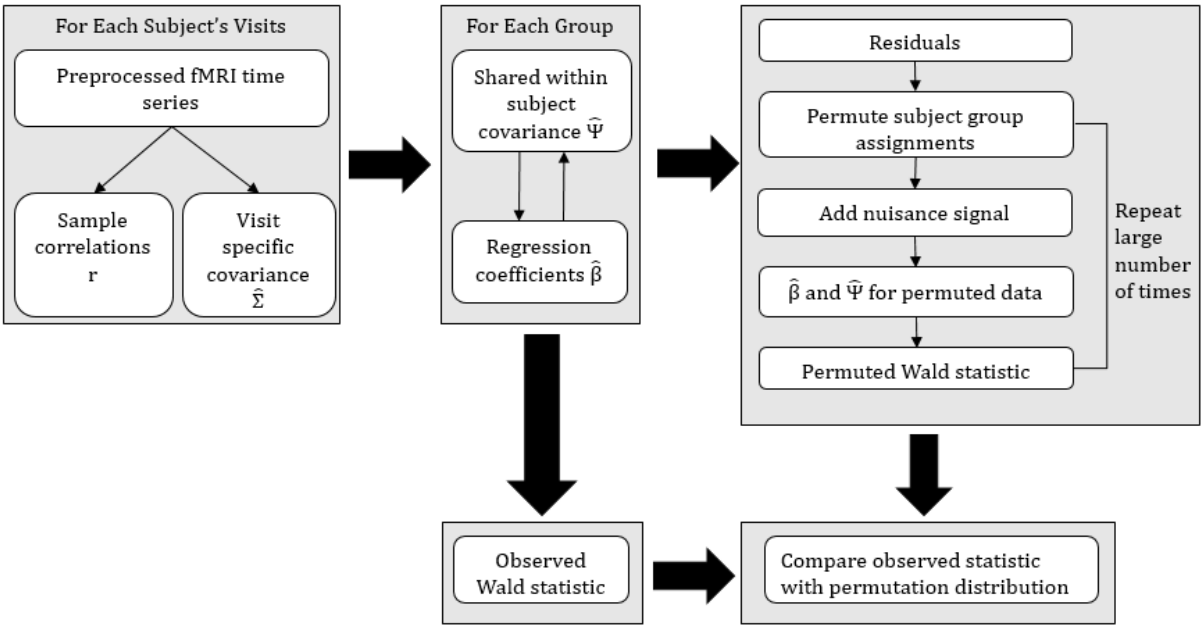


Figure 2.2: A workflow chart of the estimation and inferential procedure of our variance components model.

offers the additional advantage that the data only needs to be permuted once and the model only fit twice at each iteration of the permutation test to test all local and global hypotheses. Testing all hypotheses under a single permutation schedule greatly reduces the computational burden of the testing procedure.

Figure 2.2 shows a workflow chart of the previously described procedures used to estimate the model parameters and test hypotheses.

2.2.5 Simulation Study Setup

A series of simulations were designed with different data generating mechanisms to assess model performance. In all scenarios each time series contained 120 time points and had an autocorrelation structure that followed a first-order autoregressive process with an AR parameter of 0.3. A multivariate time series was simulated for each subject at three visits. For each visit, the Q correlations were simulated from a multivariate normal distribution where the mean and variance varied by group based on the simulation setting. For group 1, the mean vector was always assumed to

be 0 and the covariance matrix was the same across all simulation settings. The simulations used P of either 3, 5, or 10 as the dimension of the multivariate normal distribution. For the 3 and 5 dimension settings only the first element of the group 2 mean vector was allowed to vary by simulation setting, while the other elements were set to match group 1. For the 10 dimension settings the first 5 elements of the group 2 mean vector varied by simulation setting and the other elements were again set to match group 1. 1,000 Monte Carlo simulations were run for all simulation settings with 3 and 5 dimensions, and 500 simulations were run for the 10 dimensional simulation settings. Group sizes of 15 and 30 were considered. The true variance of the correlations was either equal for the two groups or the group 2 variance was double the group 1 variance. 500 permutations were used for the permutation test for all settings. Although more than 500 permutations would likely be desired in practice, the average effect across all simulations will remain the same with a reduced number of permutations with the advantage of a significant savings in computing time. The group size, number of visits, and time series length were selected to reflect values found in typical fMRI studies. The effect size of 0.1 is realistic for fMRI longitudinal scenarios. In the ADNI data an effect size of 0.1 or larger was observed for 20% of the baseline effects and a third of the trend effects over the range of ages under study. Considering a smaller effect size would lead to very similar conclusions with slight decreases in power across all models considered. A summary of the simulation settings used can be found in Table 2.1.

We chose to fit three versions of our model with different variance assumptions and estimation methods so they can be compared to each other. The first model considered was a full convergence model that iterated between $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\beta}}$ until convergence. It assumed an unstructured $\boldsymbol{\Sigma}_{ij}$ and compound symmetry for $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$. This model is referred to as the full convergence full variance model. The second was a one-step model which stops after one iteration of solving for $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\beta}}$. It also assumed an unstructured $\boldsymbol{\Sigma}_{ij}$ and compound symmetry for $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$ and is referred to as the one-step convergence full variance model. The third model was a one-step

Simulation Setting	Group Size	Variance	$\boldsymbol{\beta}_0$	$\boldsymbol{\beta}_1$
1	15	Equal	0	0
2	15	Equal	0.1	0.1
3	15	Equal	0.1	0
4	15	Equal	0	0.1
5	30	Equal	0	0
6	30	Equal	0.1	0.1
7	30	Equal	0.1	0
8	30	Equal	0	0.1
9	15	Group 2 Double	0	0
10	15	Group 2 Double	0.1	0.1
11	15	Group 2 Double	0.1	0
12	15	Group 2 Double	0	0.1
13	30	Group 2 Double	0	0
14	30	Group 2 Double	0.1	0.1
15	30	Group 2 Double	0.1	0
16	30	Group 2 Double	0	0.1

Table 2.1: Simulation settings used in the simulation study. Each setting was used to simulate 3, 5, and 10 ROI data.

model which assumed a diagonal structure for $\boldsymbol{\Sigma}_{ij}$ and scaled identity structures for $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$. This model is referred to as the one-step convergence reduced variance model. For a comparison to common practice, we also consider a massive univariate linear mixed effects (LME) model on the correlation coefficients with a random intercept per subject. Unfortunately, because the LME is a massive univariate approach, we can only run local hypothesis tests on ROI pair effects. For a comparison of global hypothesis tests results, we also fit a multivariate analysis of covariance (MANCOVA) model (Johnson & Wichern, 2002).

2.2.6 ADNI Data

Data used in the preparation of this article were obtained from the ADNI database (`http://adni.loni.usc.edu`). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment

can be combined to measure the progression of MCI and early AD. For up-to-date information, see `www.adni-info.org`.

We preprocessed the ADNI data using both FSL (version 5.0.9, `https://fsl.fmrib.ox.ac.uk/`) and AFNI (version AFNI_17.0.15, `https://afni.nimh.nih.gov/`). The preprocessing steps were as follows. We 1) applied motion correction to the images using FSL's mcflirt (rigid body transform; cost function normalized correlation; reference volume the middle volume) and then 2) normalized the images into the Montreal Neurological Institute space using FSL's flirt (affine transform; cost function correlation ratio). We used FSL's fast to 3) obtain a probabilistic segmentation of the brain to obtain white matter and cerebrospinal fluid (CSF) probabilistic maps, thresholded at 0.75. Using FSL's fslmaths, we 4) spatially smoothed the volumes using a Gaussian kernel with FWHM=5 mm. We used AFNI's 3dDetrend to 5) remove nuisance signals, namely the six motion parameters, white matter and CSF signals, and the global signal. Finally, 6) the linear trend was removed from each time series using linear regression and a 4th order Butterworth low-pass filter with a 0.1 Hertz cutoff was applied to each fMRI time series.

A subset of the ADNI data was used to demonstrate a practical application of our model. The data consists of longitudinal resting-state fMRI images collected at baseline, 3 months from baseline, 6 months from baseline, 12 months from baseline, and annually thereafter. There are two groups of interest, the CN group and the AD group. We focused our attention on late-onset AD and included only patients who were 65 years of age or older at baseline (van der Flier et al., 2011; Holland et al., 2012). To better separate the AD and CN groups, only patients who remained in one group for the entirety of the follow-up were considered in our analysis. The remaining CN group consists of 111 visits from 30 patients (17 females and 13 males) with each patient having between 1 and 6 visits. The AD group consists of 79 visits from 26 patients (11 females and 15 males) with each patient having between 1 and 5 visits. The average age was 75.9 for the CN group with a range of 65.2 to 95.7, while the AD group average age was 76.7 with a range of 66.5 to 88.6.

Our new method analyzes fMRI data at the region of interest (ROI) level. Thus, before fitting the model, the investigator must select a number of ROI to include in the FC network analysis. We used the Automated Anatomical Label (AAL) atlas to subdivide the brain into 116 anatomical regions (Tzourio-Mazoyer et al., 2002). We define the ROI level time series for a given region as the average of the time series from each voxel (3D location) within that region of the brain. We then selected $P = 10$ ROI for analysis of the ADNI data based on previous literature which has shown differences in FC between AD and CN patients in the DMN and hippocampi (Supekar et al., 2008; Greicius et al., 2004; Sorg et al., 2007). The ten regions we selected were the left and right hippocampus (HC), parahippocampus (PHC), posterior cingulate (PCC), precuneus (PQ), and prefrontal cortex (PFC). In all results that follow an l suffix for an ROI denotes the left side of the brain and an r suffix denotes the right side. Because a full brain analysis of all 116 regions is not currently feasible using our method, investigators should select a set of ROI for their particular dataset and research question based on expert knowledge and literature review.

Four models were fit to the ADNI data with differing assumptions. Model 1 is a one-step estimation model which assumes compound symmetry structure for $\mathbf{\Psi}_0$ and $\mathbf{\Psi}_1$ and unstructured $\mathbf{\Sigma}_{ij}$. Model 2 makes the same assumptions for $\mathbf{\Psi}_0$, $\mathbf{\Psi}_1$, and $\mathbf{\Sigma}_{ij}$ but uses the full convergence estimator. Model 3 is a one-step estimation model assuming scaled identity structures for $\mathbf{\Psi}_0$ and $\mathbf{\Psi}_1$ and a diagonal structure for $\mathbf{\Sigma}_{ij}$. Finally, Model 4 uses one-step estimation, assumes a diagonal structure for Ψ_0, sets all elements of $\mathbf{\Psi}_1$ to 0, and assumes a diagonal structure for $\mathbf{\Sigma}_{ij}$. This final model is similar to a massive univariate approach which ignores the within-subject dependence. 10,000 permutations were run for all models fit to the ADNI data. The intercept of each model represents the FC strength of each group at age 65. Due to failure to control Type I error rates in the simulation study, we did not fit the LME/MANCOVA method to the ADNI data.

2.3 Results

2.3.1 Simulation Study

Tables 2.2 and 2.3 show the simulation study results. Table 2.2 shows the global and local Type I error for the main effect (difference in baseline FC) and interaction (difference in change in FC across longitudinal time) across all simulations. The reported global test results are the average global Type I errors across 500 or 1,000 Monte Carlo runs. The local test results are the average Type I errors of the unadjusted p-values for all null hypotheses across the 500 or 1,000 Monte Carlo runs. While the local p-values would be adjusted in practice, the numbers in the table provide easy reference to a nominal Type I error of 0.05. Table 2.3 shows the average global power and average local power using false discovery rate adjusted p-values. All permutation p-values were corrected in accordance with Phipson et al. (2010). Additional simulation study results can be found in Table A.1 in the Appendix. Table 2.4 shows the simulation results from fitting the one-step convergence full variance model to the 10 ROI simulated data. The final section of each of these tables shows the results for the LME/MANCOVA approach. In these sections, all local test results come from LME model and all global test results come from the MANCOVA model.

Table 2.2 shows roughly nominal Type I error rates for all three specifications of our model. The LME model also controls Type I error at a nominal level. The MANCOVA global tests show highly inflated Type I error for the main effect test and deflated Type I error rates for the interaction test. While there was some slight inflation in all three specifications of our model, especially for the 10 ROI simulations, the inflation was attenuated by the increase in sample size from 15 to 30 per group. Table 2.3 demonstrates adequate power, both locally and globally for all three specifications of our model. The LME model showed decreased power for local tests, especially when considering the 5 ROI scenario. This reduced power for the 5 ROI scenarios shows the LME model's decreased performance in higher dimensional

Convergence:		Full		One-Step		One-Step		LME Local	
Variance:		Full		Full		Reduced		MANCOVA Global	
	Setting	**3 ROI**	**5 ROI**	**3 ROI**	**5 ROI**	**3 ROI**	**5 ROI**	**3 ROI**	**5 ROI**
Main Effect Global Test	**1**	0.061	0.052	0.057	0.055	0.056	0.054	0.159	0.232
	4	0.065	0.065	0.067	0.071	0.067	0.063	0.822	0.700
	5	0.045	0.049	0.046	0.050	0.043	0.047	0.132	0.257
	8	0.046	0.050	0.054	0.047	0.048	0.049	0.975	0.948
	9	0.058	0.049	0.055	0.053	0.056	0.048	0.172	0.301
	12	0.068	0.058	0.069	0.056	0.070	0.060	0.718	0.684
	13	0.057	0.054	0.058	0.053	0.058	0.060	0.175	0.297
	16	0.049	0.066	0.055	0.059	0.058	0.053	0.943	0.902
Main Effect Local Tests	**1**	0.071	0.061	0.073	0.061	0.066	0.062	0.049	0.047
	4	0.064	0.067	0.067	0.069	0.064	0.067	0.052	0.052
	5	0.050	0.059	0.048	0.058	0.048	0.057	0.048	0.047
	8	0.053	0.058	0.051	0.058	0.052	0.057	0.048	0.050
	9	0.064	0.063	0.061	0.062	0.064	0.064	0.049	0.048
	12	0.068	0.062	0.069	0.063	0.065	0.063	0.050	0.049
	13	0.063	0.055	0.062	0.055	0.060	0.056	0.049	0.047
	16	0.054	0.059	0.055	0.058	0.053	0.057	0.053	0.046
Interaction Global Test	**1**	0.062	0.054	0.052	0.057	0.049	0.052	0.021	0.009
	3	0.046	0.054	0.044	0.049	0.044	0.049	0.017	0.010
	5	0.054	0.059	0.058	0.061	0.056	0.058	0.015	0.007
	7	0.052	0.067	0.049	0.063	0.051	0.051	0.021	0.011
	9	0.057	0.067	0.050	0.062	0.051	0.069	0.012	0.005
	11	0.079	0.071	0.078	0.061	0.073	0.063	0.012	0.014
	13	0.068	0.065	0.071	0.061	0.075	0.057	0.012	0.001
	15	0.056	0.041	0.057	0.049	0.059	0.040	0.013	0.006
Interaction Local Tests	**1**	0.059	0.062	0.058	0.062	0.059	0.064	0.052	0.048
	3	0.055	0.060	0.054	0.057	0.057	0.060	0.048	0.046
	5	0.050	0.055	0.049	0.054	0.049	0.055	0.054	0.048
	7	0.054	0.059	0.053	0.058	0.058	0.059	0.050	0.049
	9	0.063	0.065	0.065	0.061	0.064	0.063	0.045	0.048
	11	0.074	0.067	0.071	0.061	0.074	0.064	0.044	0.047
	13	0.061	0.059	0.062	0.057	0.063	0.060	0.047	0.050
	15	0.057	0.055	0.057	0.055	0.058	0.055	0.053	0.044

Table 2.2: Type I error rates for simulation study for all globally null simulation settings. Type I errors for the main effect (group difference in intercepts) and interaction effect (group difference in longitudinal slopes) are reported both globally and locally. The global Type I errors are averaged across all simulations. The local Type I errors reported are unadjusted and averaged across all simulations and all null ROI pairs. In the LME Local, MANCOVA Global columns all local test results come from LME model and all global test results come from the MANCOVA model. The standard errors for the Type I error rate across all null local tests and an average of 0.008 with a maximum of 0.023 and were comparable across all models.

Convergence: Variance:		Full Full		One-Step Full		One-Step Reduced		LME Local MANCOVA Global	
Setting		3 ROI	5 ROI	3 ROI	5 ROI	3 ROI	5 ROI	3 ROI	5 ROI
Main Effect Global Test	2	0.382	0.223	0.369	0.229	0.372	0.211	0.999	0.995
	3	0.389	0.214	0.390	0.203	0.391	0.204	0.799	0.730
	6	0.674	0.491	0.679	0.489	0.682	0.469	1.000	1.000
	7	0.684	0.468	0.686	0.470	0.670	0.456	0.972	0.953
	10	0.300	0.196	0.294	0.185	0.303	0.185	0.996	0.987
	11	0.328	0.180	0.321	0.170	0.316	0.177	0.727	0.647
	14	0.582	0.383	0.581	0.385	0.579	0.367	1.000	1.000
	15	0.582	0.383	0.580	0.380	0.572	0.371	0.932	0.902
Main Effect Local Tests	2	0.375	0.215	0.370	0.176	0.388	0.251	0.290	0.109
	3	0.315	0.172	0.314	0.138	0.324	0.200	0.239	0.081
	6	0.700	0.531	0.706	0.504	0.712	0.528	0.703	0.509
	7	0.622	0.419	0.621	0.409	0.629	0.454	0.570	0.292
	10	0.288	0.155	0.283	0.132	0.283	0.187	0.228	0.067
	11	0.258	0.135	0.244	0.102	0.261	0.148	0.180	0.057
	14	0.583	0.421	0.578	0.388	0.592	0.408	0.581	0.315
	15	0.515	0.355	0.515	0.317	0.515	0.346	0.444	0.193
Interaction Global Test	2	0.757	0.525	0.765	0.492	0.771	0.550	0.605	0.271
	4	0.767	0.537	0.753	0.494	0.783	0.557	0.624	0.253
	6	0.983	0.881	0.983	0.877	0.986	0.888	0.928	0.668
	8	0.977	0.899	0.980	0.894	0.977	0.906	0.930	0.660
	10	0.706	0.502	0.694	0.444	0.719	0.515	0.478	0.175
	12	0.959	0.460	0.958	0.422	0.954	0.483	0.845	0.185
	14	0.959	0.842	0.958	0.831	0.954	0.838	0.845	0.541
	16	0.950	0.839	0.952	0.821	0.945	0.828	0.863	0.525
Interaction Local Tests	2	0.752	0.527	0.737	0.410	0.780	0.645	0.712	0.398
	4	0.719	0.470	0.692	0.373	0.734	0.572	0.669	0.340
	6	0.989	0.919	0.987	0.875	0.992	0.943	0.981	0.917
	8	0.967	0.863	0.973	0.817	0.973	0.902	0.961	0.814
	10	0.687	0.442	0.667	0.360	0.693	0.555	0.603	0.273
	12	0.645	0.417	0.617	0.335	0.648	0.515	0.568	0.274
	14	0.965	0.869	0.967	0.819	0.957	0.892	0.945	0.822
	16	0.938	0.817	0.935	0.770	0.932	0.848	0.932	0.742

Table 2.3: The power calculations for the simulation study. Power results for the main effect (group difference in intercepts) and interaction effect (group difference in longitudinal slopes) are reported both globally and locally. The global power results are averaged across all simulations. The local power results reported are FDR adjusted. In the LME Local, MANCOVA Global columns all local test results come from LME model and all global test results come from the MANCOVA model.

	Type I Error Rate				Power			
	Main Effect		Interaction		Main Effect		Interaction	
Setting	**Global**	**Local**	**Global**	**Local**	**Global**	**Local**	**Global**	**Local**
1	0.066	0.066	0.072	0.067	-	-	-	-
2	-	0.069	-	0.073	0.482	0.086	0.742	0.164
3	-	0.067	0.076	0.071	0.466	0.037	-	-
4	0.064	0.069	-	0.073	-	-	0.728	0.130
5	0.066	0.058	0.078	0.059	-	-	-	-
6	-	0.054	-	0.054	0.902	0.305	0.942	0.514
7	-	0.054	0.070	0.056	0.884	0.103	-	-
8	0.066	0.056	-	0.057	-	-	0.936	0.430
9	0.054	0.071	0.070	0.072	-	-	-	-
10	-	0.067	-	0.069	0.356	0.066	0.726	0.125
11	-	0.064	0.090	0.067	0.386	0.026	-	-
12	0.068	0.069	-	0.073	-	-	0.668	0.094
13	0.062	0.057	0.072	0.060	-	-	-	-
14	-	0.056	-	0.056	0.836	0.241	0.922	0.454
15	-	0.056	0.068	0.058	0.800	0.060	-	-
16	0.078	0.059	-	0.060	-	-	0.890	0.363

Table 2.4: Type I error rates for 10 ROI simulation scenarios for all globally null simulation settings for the one-step convergence full variance model. Type I errors for the main effect (group difference in intercepts) and interaction effect (group difference in longitudinal slopes) are reported both globally and locally. The global Type I errors are averaged across all simulations. The local Type I errors reported are unadjusted and averaged across all simulations and all null ROI pairs. Power results are reported both globally and locally. The global power results are averaged across all simulations. The local power results reported are FDR adjusted and averaged across all simulations and all non-null ROI pairs. The average and maximum standard error for the local tests were 0.014 and 0.016 for Type I error rates and 0.008 and 0.015 for power calculations.

scenarios. The MANCOVA global test had very high power in the main effect tests where it failed to control Type I error and relatively poor power in the interaction test where it demonstrated deflated Type I error. As expected, across all models power increased with larger group size and decreased with a larger true group 2 variance. Overall, for balancing Type I error and power, our models outperform the LME at the local level and the MANCOVA at the global level.

2.3.2 ADNI Results

We motivate the challenges of longitudinal FC analysis with a preliminary examination of the ADNI data. Figure 2.3 shows spaghetti plots of the FC between the preprocessed fMRI time series obtained from two ROI pairs for the AD and CN groups. The clustering of points within each line shows the within-subject dependence. In addition, there is considerable within-subject and within-group noise present in the estimates of FC. What is not evident from the figure is that the time series from which these correlations were obtained exhibit autocorrelation that contributes to the overall variability in FC. To add another level of complication, the figure depicts the marginal relationship between two ROI, but to properly model all of the selected ROI we need a joint model that considers the set of all possible pairwise groupings of the chosen ROI.

Table 2.5 shows results from the global hypothesis tests and all local hypothesis tests that were significant before p-value adjustment for all four models. Neither the overall main effect or interaction term were found to be significant in the global tests for any of the four models considered. The only ROI pair level differences that remained significant after p-value adjustment and correction in any of the models were the differences in the CN and AD group longitudinal slopes in the FC between the left HC and the right and left PCC in Models 1 and 2. These two analyses conclude that the FC between HCl and PCC declines at a significantly quicker rate in the AD population than in their CN counterparts. The estimated Model 1 and Model 4 group

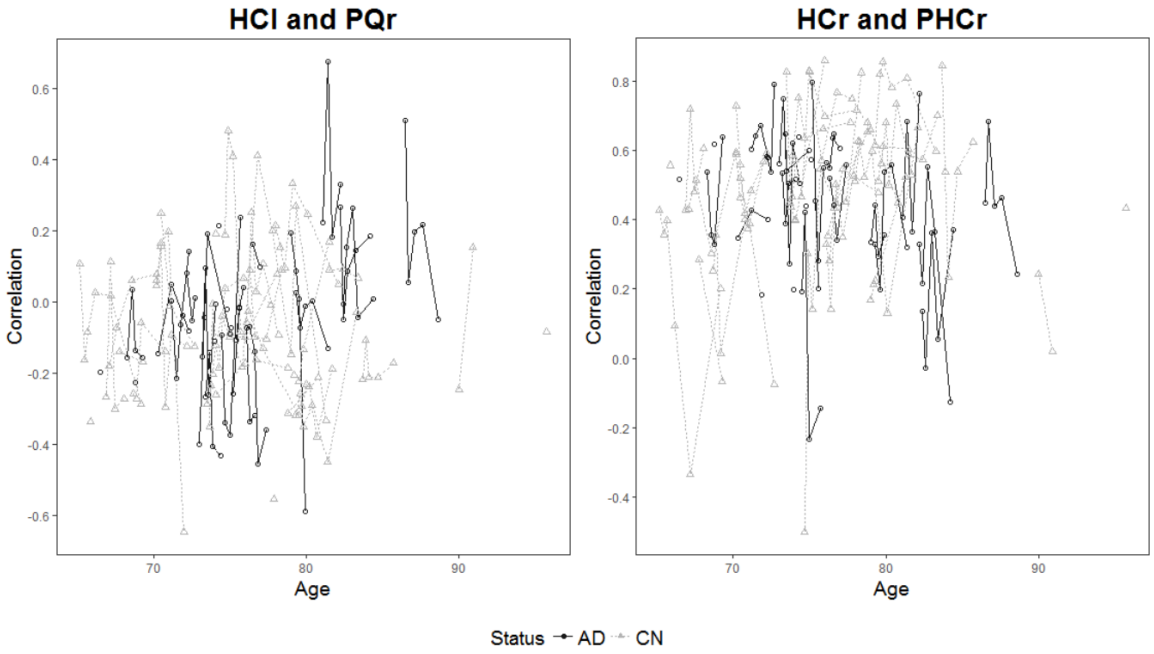


Figure 2.3: Spaghetti plots of the correlation between two ROI against age. Each point represents a visit, and each line represents a subject. The ROI represented in these plots are the left and right hippocampus (HCl and HCr), right precuneus (PQr), and right parahippocampus (PHCr).

intercepts, group longitudinal trends, group differences in intercepts and longitudinal trends, and $-\log_{10}$ p-values after correction and adjustment from local hypothesis tests are presented graphically in Figures 2.4 and 2.5 respectively. Similar figures for Models 2 and 3 can be found in Figures A.1 and A.2 in the Appendix.

2.4 Discussion and Conclusion

2.4.1 Simulation Study

All three specifications of our model out-performed the LME/MANCOVA model. The LME model controlled Type I error rates at a nominal level, but saw decreased power, particularly as the dimension of the data increased. The MANCOVA global tests performed very poorly, failing to control Type I error at a 0.05 level. Our models' improved performance in comparison to the LME/MANCOVA approach shows the utility of our method for both local and global hypothesis testing.

Some interesting results also arise from comparison of the three specifications of our model. The two full variance settings match the true model of the simulated data, yet the reduced variance model did not suffer in comparison, even winning in certain scenarios. The reduced variance model may have offered similar performance because the smaller parameter space allowed for improved estimation. The reduced variance model did not capture the full true variance, but it still performed well by allowing the FC for each ROI pair to be correlated across multiple visits for a given subject. Performance may also change if stronger within-subject correlation was assumed for the simulations.

For the chosen simulation settings, the full convergence, full variance model showed a moderate increase in power for local hypothesis tests compared to the one-step, full variance model. The slight advantage for full convergence was less evident in the 3 ROI setting and did not translate to either global hypothesis test. Another primary difference in the three models was the computational time. The full convergence

	$\boldsymbol{\beta}_{CN}$	$\boldsymbol{\beta}_{AD}$	Test Statistic	Unadjusted p-value	Adjusted p-value
Model 1: One-step, Compound Symmetry $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$, and Unstructured $\boldsymbol{\Sigma}$					
Main Effects			*33.92*	*0.392*	
HCl and PCCl	0.025	0.198	3.98	0.015	0.222
HCl and PCCr	0.033	0.228	5.02	0.004	0.114
HCr and PCCr	-0.005	0.166	3.87	0.012	0.222
Interactions			*33.46*	*0.327*	
HCl and PCCl	0.005	-0.013	7.05	<0.001	0.027
HCl and PCCr	0.004	-0.014	7.33	<0.001	0.027
HCr and PCCr	0.005	-0.008	4.08	0.013	0.222
PHCl and PCCr	0.008	-0.003	3.19	0.039	0.501
Model 2: Full Convergence, Compound Symmetry $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$, and Unstructured $\boldsymbol{\Sigma}$					
Main Effects			*33.10*	*0.399*	
HCl and PCCl	0.024	0.198	3.95	0.015	0.230
HCl and PCCr	0.033	0.227	4.88	0.005	0.141
HCr and PCCr	-0.006	0.165	3.82	0.013	0.230
Interactions			*32.89*	*0.326*	
HCl and PCCl	0.005	-0.013	7.02	<0.001	0.032
HCl and PCCr	0.004	-0.014	7.15	<0.001	0.032
HCr and PCCr	0.005	-0.008	4.02	0.015	0.230
PHCl and PCCr	0.008	-0.003	3.13	0.041	0.531
Model 3: One-step, Scaled Identity $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$, and Diagonal $\boldsymbol{\Sigma}$					
Main Effects			*39.10*	*0.483*	
HCl and PCCl	0.041	0.243	5.02	0.007	0.167
HCl and PCCr	0.052	0.275	6.13	0.003	0.078
HCr and PCCr	0.030	0.167	2.32	0.029	0.432
PCCl and PFCr	0.397	0.215	4.30	0.020	0.356
Interactions			*45.68*	*0.284*	
HCl and PCCl	0.003	-0.014	6.30	0.002	0.078
HCl and PCCr	0.001	-0.017	6.75	0.001	0.078
PHCl and PCCr	0.006	-0.005	2.65	0.045	0.506
PCCl and PFCr	-0.005	0.008	3.55	0.037	0.476
Model 4: One-step, Scaled Identity $\boldsymbol{\Psi}_0$, Zero $\boldsymbol{\Psi}_1$, and Diagonal $\boldsymbol{\Sigma}$					
Main Effects			*62.65*	*0.568*	
HCl and PCCr	0.078	0.281	8.41	0.013	0.585
PHCl and PFCr	-0.116	-0.237	2.84	0.013	0.585
PHCr and PQl	0.040	-0.162	6.61	0.040	0.726
PCCl and PFCr	0.402	0.223	4.35	0.042	0.726
Interactions			*68.63*	*0.421*	
HCl and PCCl	0.000	-0.013	5.60	0.038	0.726
HCl and PCCr	-0.001	-0.018	9.11	0.007	0.585

Table 2.5: Hypothesis tests on the ADNI data. Global tests and all local tests with unadjusted p-values of < 0.05 are shown for Models 1-4. The numbers in italics are from the global hypothesis tests.

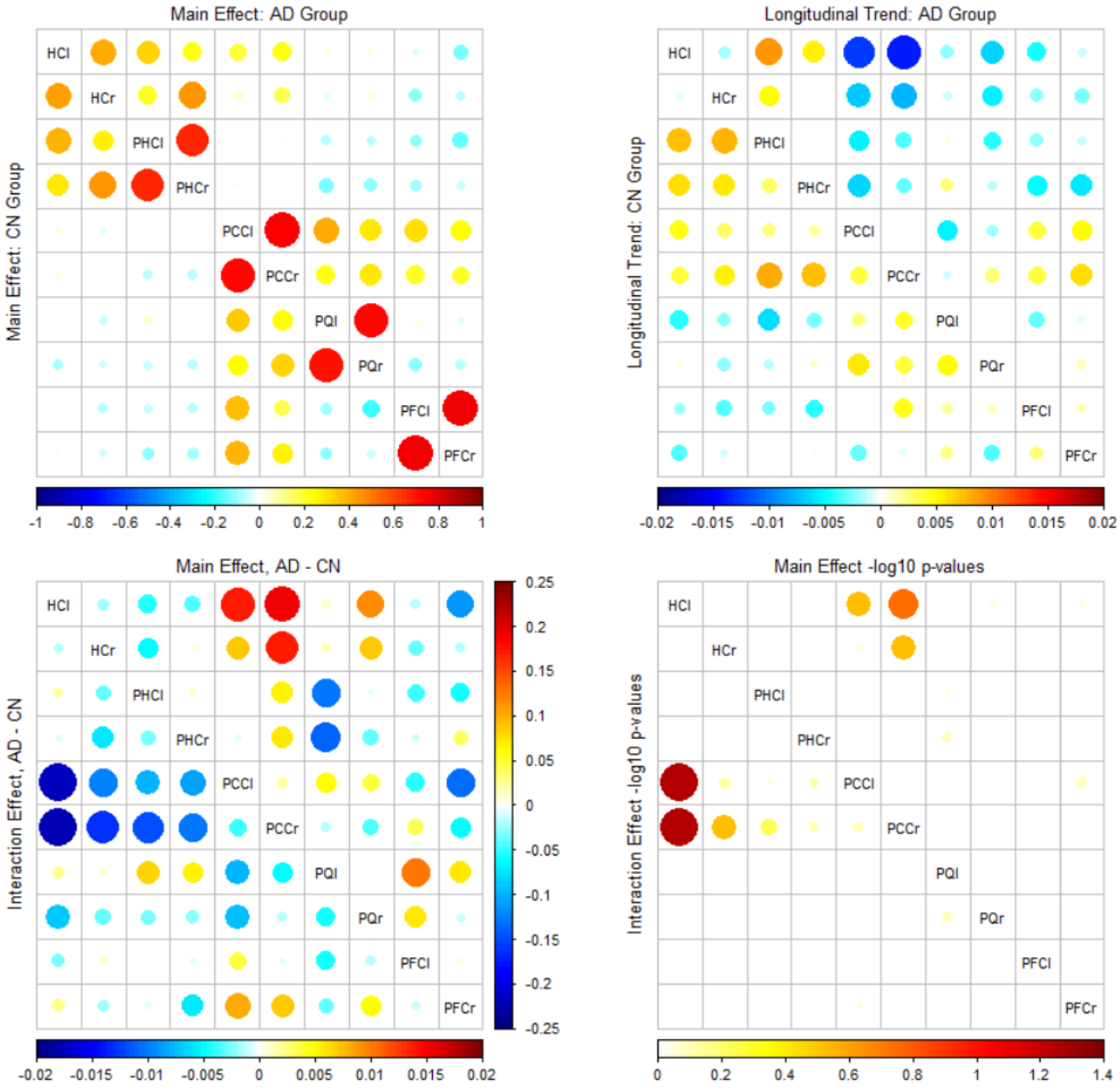


Figure 2.4: Model 1 results. Top left: A plot of the estimated intercept terms for the CN group (bottom left triangle) and AD group (top right triangle). Top right: A plot of the estimated slope terms for the CN group (bottom left triangle) and AD group (top right triangle). Bottom left: a plot of the group differences (AD estimates - CN estimates) for the estimated intercepts (top right triangle) and slopes (bottom left triangle). Bottom right: A plot of the $-\log_{10}$ corrected and adjusted p-values from all local hypothesis tests of group differences (AD estimates - CN estimates) for the estimated intercepts (top right triangle) and slopes (bottom left triangle).

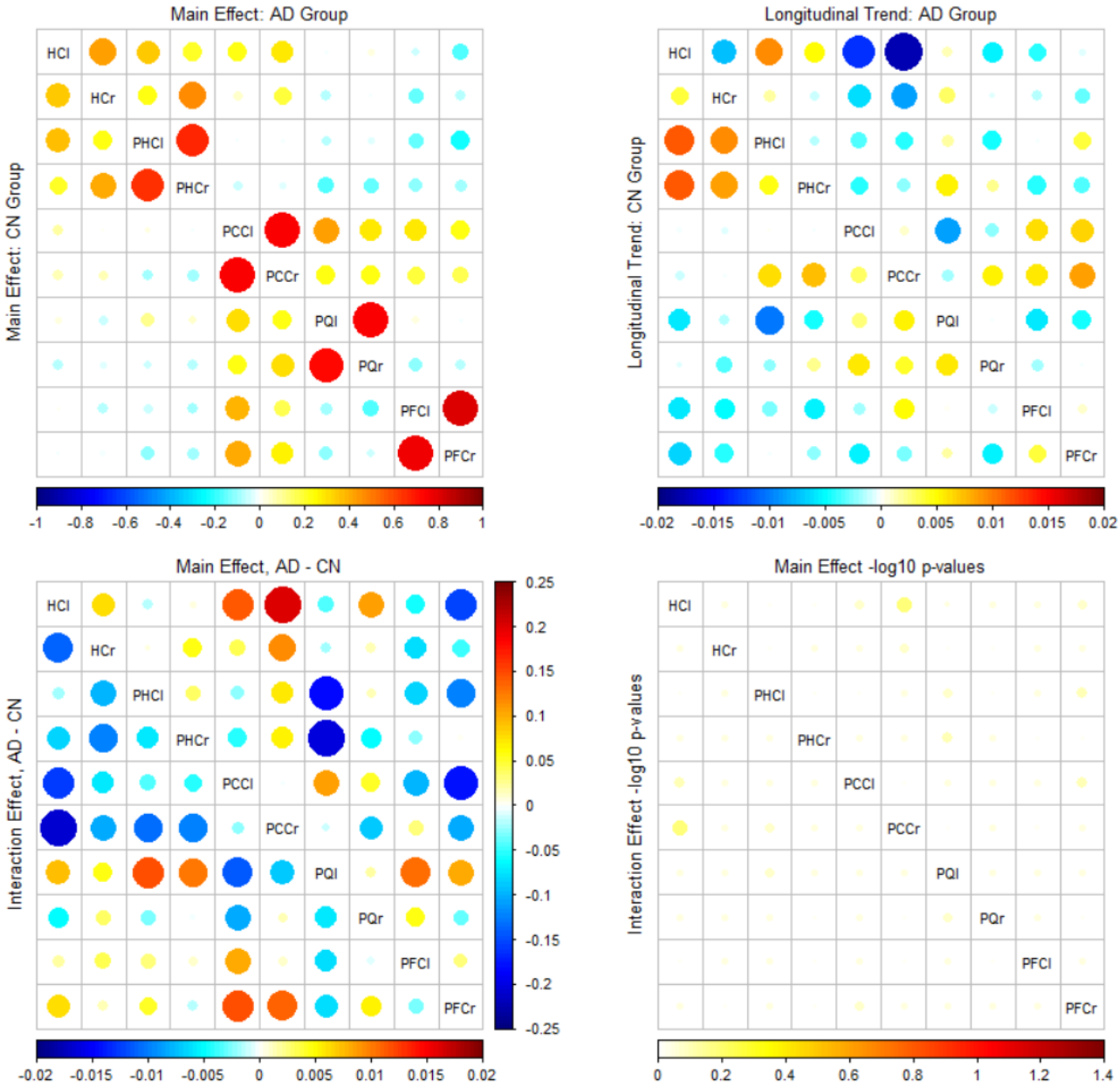


Figure 2.5: Model 4 results. Top left: A plot of the estimated intercept terms for the CN group (bottom left triangle) and AD group (top right triangle). Top right: A plot of the estimated slope terms for the CN group (bottom left triangle) and AD group (top right triangle). Bottom left: a plot of the group differences (AD estimates - CN estimates) for the estimated intercepts (top right triangle) and slopes (bottom left triangle). Bottom right: A plot of the $-\log_{10}$ corrected and adjusted p-values from all local hypothesis tests of group differences (AD estimates - CN estimates) for the estimated intercepts (top right triangle) and slopes (bottom left triangle).

model took, on average, over 2.5 times longer to run. For reference, using a 3.7 GHz Quad-Core Intel Xeon with 16GB ram, the average times to fit the one-step full variance model with 30 subjects per group for 3, 5, and 10 ROI were 3.8 seconds, 54.9 seconds, and 87.2 minutes, respectively. These results show that the time increases quickly with the dimension of the model. The computational time is largely driven by the permutation procedure. Thus, if a larger number of permutations is desired for the testing procedure, then the computational time will see a corresponding increase. The one-step model may serve as an adequate replacement for the full-convergence model in certain computationally demanding scenarios.

2.4.2 ADNI

The four models fit to the ADNI data present slightly different results. The difference between Model 1 and Model 2 is minimal. The nearly identical results show that the full and one-step convergence models can lead to very comparable results. With nearly identical resulting estimates and inference, the one-step estimator should be preferred in this case due to its significant computational advantage

Some more pronounced differences in results arise when Model 1 and 2 are compared with Models 3 and 4. All of the models produce similar estimates for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$, but Model 4 saw a large reduction in power. When comparing Figures 2.4 and 2.5 three of the four plots look very similar, yet the plots displaying $-\log_{10}$ p-values demonstrate reduced power in Model 4. As mentioned earlier, Model 4 is essentially a univariate approach which does not account for the dependence within-visit between ROI pairs or the dependence between visits. Model 3 models the between visit dependence and gets a considerable boost in power compared to Model 4 in the analysis of the ADNI data. Although the results are somewhat mixed, Model 1 and Model 2 see a further slight boost in power over Model 3 for certain ROI pairs by accounting for the dependence between ROI pairs within a visit. The increased power of Model 1 and Model 2 to detect group differences over Model 4 exemplifies the utility of our

novel longitudinal approach to fMRI FC network analysis.

Although the output from Models 1 and 2 differs slightly from that of Model 3, they share some common patterns. In these models many of the local hypotheses that were significant prior to the FDR correction appear between the HC/PHC and the PCC. These group differences strengthen the one local hypothesis that is significant after FDR correction from Models 1 and 2, which shows a significantly larger decrease in FC between the HCl and PCCr in the AD group than in the CN group. While the significant results become marginal after FDR correction for Model 3, the fact that many other HC/PHC connections with the PCC show a similar pattern helps to indicate differing baseline and longitudinal trend effects in the FC of the two groups. This clustering of group differences can be seen in Figure 2.4 with the smallest p-values (red and orange circles) appearing between the HC/PHC and PCC. Wang et al. (2006), Sorg et al. (2007), and Greicius et al. (2004) all noted decreased FC between the HC and PCC in patients with AD in analyses of cross-sectional data. Similar results from Supekar et al. (2008) showed decreased clustering coefficients for the HC. Our analysis confirms these results with the addition of a longitudinal component to the analysis. Our results not only conclude that AD and CN patients have differing FC between the HC and PCC, as the previous works have shown, but we also more clearly describe the group differences in FC across ages between these two regions.

2.4.3 Limitations and Future Work

Our familiar linear model framework allows for easy adoption and understanding of the model and its results. Additionally, the linear model framework offers many natural extensions. One could easily include terms for additional covariates such as scanner effect or sex. Different structures for the variance components could also be implemented to capture a wider range of possible correlation structures.

Our current method has the advantage of allowing joint modeling of FC network

between a set of ROI rather than taking a massive univariate approach. We see this joint modeling as a significant step forward, but complete brain analyses are still not yet feasible due to high computational demands of a model fit to many ROI and the limited sample size of many fMRI studies. Here we have fit models to 10 ROI, but many brain atlases include more than 100 regions. Our longitudinal FC model is likely only feasible for networks containing up to 15 ROI. In the future, dimension reduction techniques, such as regularization, could be introduced into the model to allow analysis of an entire brain atlas using sparse FC networks. The ability to run a full brain FC analysis would alleviate the problems arising from ROI selection but may also make interpretation more difficult as there would be thousands of pairwise connections to consider.

The selection of the proper structure for the variance components deserves more attention. While a block compound symmetry structure for $\boldsymbol{\Psi}$ has a natural interpretation similar to that of a random intercept, there are certainly other viable structures. Choosing between structures is not a trivial task. One way to alleviate the model selection dilemma is to introduce a more robust sandwich type estimator of $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})$, in which case incorrect specification of the variance would lead to valid inference with only a reduction in power.

Finally, GLS estimators such as ours typically require a missing completely at random (MCAR) assumption. While MCAR may not be reasonable for many datasets, our method relies less on this assumption because we opt for a permutation procedure. Our estimators maintain consistency in the missing at random scenario and our permutation testing procedure does not rely on distributional assumptions for the error terms required when using asymptotic results for inference (Laird, 2004).

2.5 Conclusion

We have introduced a novel variance components longitudinal model to estimate and draw inference on the group differences in FC networks using resting-state fMRI

data. The model properly accounts for the correlation inherent in repeated measures data and the autocorrelation present in fMRI time series from which we construct the FC networks. For statistical inference for global and local tests about FC, we used a computationally efficient permutation testing procedure which out-performs the massive univariate LME/MANCOVA approach. The linear model framework and use of generalized least squares estimators offers great simplicity and a large number of natural extensions. This work fills a current gap in the literature by providing a general framework for estimation and hypothesis testing of longitudinal FC data. As a practical example, we applied the method to resting-state fMRI data from the ADNI database. Our analysis found a faster decline in FC between the HCl and the PCC in AD patients compared to the CN controls. This finding confirms the results of previous studies and helps solidify the central roles of the hippocampus and DMN in AD.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimers Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation;

Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (`www.fnih.org`). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimers Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Chapter 3

A Non-parametric Bayesian Model for Estimating Spectral Densities of Resting-State EEG Twin Data

3.1 Introduction

Electroencephalography (EEG) is a non-invasive neuroimaging modality that captures electrical brain activity many times per second by placing recording electrodes at various locations on the head. With sampling rates as high as 1000 Hertz (Hz), EEG data offers the benefit of very high temporal resolution. We seek to estimate power spectra from EEG data that was gathered for 557 adolescent twin pairs through the Minnesota Twin Family Study (MTFS) (Iacono et al., 1999). The EEG time series data collected from the twin study design of the MTFS calls for new methods that take into account the study design in order to account for the heterogeneity of the data and borrow information within and between twin pairs. We harness the strengths of a twin study design in a resting-state EEG dataset to develop a novel statistical model that will accomplish two primary goals: 1) identify the frequencies that drive the variations in the EEG data, an analysis approach known as *spectral analysis*, and 2) draw inference on spectral features and the proportion of variation in these features that can be attributed to genetic factors, a quantity known as *heritability*.

3.1.1 Spectral Analysis

Spectral analysis, a common approach used to analyze EEG data, decomposes a time series into a set of waves oscillating at different frequencies. The primary tool for spectral analysis is the spectral density function, also known as the power spectrum, which is a density of variances that can be understood as an ANOVA where the spectral density curve shows the proportion of the total variance of a time series that is explained by waveforms oscillating at each frequency (Shumway & Stoffer, 2010). By decomposing an EEG time series in such a manner, the spectral density, assuming weak stationarity of the time series, provides a summary of the variance characteristics of the EEG signal. The resulting estimated density curves provide signatures that describe whether low or high frequency oscillations dominate the variance of the time series.

Many common spectral density estimation methods use parametric forms such as autoregressive models (Shumway & Stoffer, 2010). These parametric methods are often very fast and simple, but do not allow sharing of information across multiple time series, as is desired in our twin study data. An alternative approach to estimate the spectral density, following the lead of Wahba (1980), is to use Bayesian smoothing splines. These approaches use the Whittle likelihood, an approximation of the true spectral likelihood (Whittle, 1953), within Markov chain Monte Carlo (MCMC) algorithms.

Moving beyond the simple scenario of a single time series, some work has begun to address experimental design when estimating EEG spectral densities, beginning with Brillinger (1973) who assessed replicated time series as a manner to increase the signal to noise ratio. Diggle & Al Wasel (1997) and Krafty et al. (2011) expanded on these models with mixed effects models for repeated biomedical time series. More recently, Bruce et al. (2017) and Krafty et al. (2017) modeled covariate modulated spectral densities, while work such as Fiecas & Ombao (2016) considered continuously evolving spectral densities through a learning experiment. Finally, Cadonna et al.

(2018) recently developed a Bayesian method to estimate the spectral densities of multiple time series through mixture models.

Each of these methods improves upon spectral analysis of a single time series by considering the design of the experiment in which the time series were collected. Likewise, our goal in this work is to develop a method that takes advantage of the MTFS twin study design to accurately estimate individual spectral densities in order to improve estimation and inference about characteristics of these densities such as frequency band power and their corresponding heritability.

3.1.2 Endophenotypes

The utility of spectral analysis of EEG data from a twin design and heritability analysis of the power spectra averaged within a frequency band is to establish *endophenotypes*, neurobiological indicators that link psychiatric disorders to genetic risk factors. Iacono et al. (2017) laid out seven different criteria for endophenotypes, one of which is that the feature must be shown to be heritable. Heritability is defined as the percentage of variation of a trait that can be explained by genetic variation. A heritability value of 1.0 signifies that variations in the trait are entirely genetic, while a heritabilty value of 0.0 signifies that variations in the trait are entirely environmental, and thus the trait is not a candidate endophenotype. Heritability estimates help establish how much and through which characteristics genetics contribute to electrical activity in the brain as measured by resting-state EEG.

Iacono et al. (2017) provided a summary of the current state of endophenotype research. In particular, they discussed the difficulty of moving beyond heritability analysis to show significant SNP and gene correlations with traits due to the often modest sample sizes available in EEG studies. In fact, SNP and gene-level analyses have only recently proved fruitful through a genome wide association study performed in Smit et al. (2017). Because of limited sample sizes, we focus our attention on improving and expanding on the heritability measures of the EEG power spectrum.

With the goal of improved and expanded EEG spectral endophenotypes in mind, we build a multi-subject spectral density model and derive a novel heritability estimator for spectral density curves. Compared with traditional heritability estimators that first summarize the power spectrum to a single measure, e.g., by taking the average power within a frequency band, our novel estimator for heritability better accounts for the power spectrum changing within a given frequency band. Indeed, our heritability estimator preserves information from the entire EEG time series without having to resort to band averaging. Furthermore, we allow calculation of the heritability of the entire spectral density curve, which is not possible using existing methods. The heritability of the entire spectral density gives us information on the contribution of genetic factors to the set waveforms that make up the time series data, giving a more complete picture of heritability of the EEG spectral density. By enabling heritability estimation of the full spectrum and estimating this heritability using data from the MTFS, we expand the set of potential EEG endophenotypes.

3.1.3 Minnesota Twin Family Study

The MTFS is a population-based study of same-sex reared-together male and female twins (and their parents), the overarching goal of which is to understand genetic and environmental influences on substance abuse and related psychopathology. Data for this study consisted of resting-state EEG data from 365 monozygotic (MZ) and 192 dizygotic (DZ) twin pairs of approximately 17 years of age. More details on the MTFS data are provided in Section 3.4.1.

Table 3.1 shows estimated heritability for the four frequency bands defined as follows: Delta (1-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), and Beta (12-30 Hz). These estimates come from using standard estimators on the MTFS data. All three channels show high levels of heritability in each frequency band, although the level of heritability does vary from channel to channel and frequency band to frequency band. Given the differences in heritability estimates across frequency bands, it is easy

Channel	Frequency Band	Heritability
Cz	Delta (1-4 Hz)	0.67
	Theta (4-8 Hz)	1.00
	Alpha (8-12 Hz)	0.87
	Beta (12-30 Hz)	0.72
O1	Delta (1-4 Hz)	0.41
	Theta (4-8 Hz)	0.70
	Alpha (8-12 Hz)	0.60
	Beta (12-30 Hz)	0.36
O2	Delta (1-4 Hz)	0.49
	Theta (4-8 Hz)	0.70
	Alpha (8-12 Hz)	0.45
	Beta (12-30 Hz)	0.36

Table 3.1: The estimated heritability for each of the four frequency bands and three EEG recording channels. Heritability estimates were calculated using GCV smoothed periodograms and Falconer's formula.

to imagine heritability as a smooth function over frequencies, which we will develop in the present work.

While many resting-state EEG spectral densities share similar general shapes and peaks, the fact that features of these curves are heritable also suggests that certain curves (i.e., those from twins) will be more similar than others. Thus, in developing a model for spectral densities, we need a framework that allows some sharing of information between all participants, but also allows different levels of similarity to match the different relationships between the participants who are either unrelated, DZ twins, or MZ twins.

The fact that the heritability differs between frequency bands suggests that a strict correlation structure imposed upon twin relationships may not be appropriate in this application. The importance of the twin relationship can vary from twin pair to twin pair and from one frequency band to another within a single twin pair. To learn the heterogeneity structure of the sample and allow flexible joint modeling of the spectral densities, we embed our estimation framework within a Bayesian nested Dirichlet process (DP) structure (Rodriguez et al., 2008). This allows us to group very similar spectral densities and shrinks their estimates towards each other while

allowing different groups to have potentially divergent estimates. In combination with our novel heritability estimator, the nested DP multi-subject spectral density model will account for the twin study design of our data and allow us to accomplish both of our stated goals efficiently.

3.2 Model Specification and Inference

3.2.1 The Single Subject Model

We start by considering a model for the spectral density of a single time series to serve as the base of our multi-subject model. Suppose we observe a univariate time series, Y_t for $t = 1, \ldots, T$, and that the time series has been standardized to have mean zero and unit variance. We model the spectral density function as a mixture of probability density functions (PDFs), ensuring a strictly non-negative estimated spectral density that integrates to $\text{Var}(Y_t) = 1$. Since we will be working in the frequency domain, we use the Whittle likelihood (Whittle, 1953), given by

$$L(\boldsymbol{f} \mid \boldsymbol{Y}) \propto \prod_{\omega \in \Omega} \frac{1}{f(\omega)} \exp\Big(\frac{|d(\omega)|^2}{f(\omega)}\Big), \tag{3.1}$$

where $f(\omega)$ is the spectral density, and $d(\omega) = T^{-0.5} \sum_{t=1}^{T} Y_t \exp(-i2\pi\omega t)$ is the discrete Fourier transform of the time series Y_t. The quantity $|d(\omega)|^2$ is known as the periodogram. Because the spectral density is considered only on an interval from zero to the Nyquist frequency (half the sampling rate), one natural model is a mixture of beta PDFs with domain scaled to lie in $(0, 0.5)$. Such a mixture of beta PDFs is commonly known as a basis of Bernstein polynomials. For ease of notation we assume all frequencies have been scaled to fall within the $(0, 0.5)$ interval. Let the spectral density be

$$f(\omega) = \sum_{\lambda=1}^{\Lambda} G\Big(\frac{\lambda - 1}{\Lambda}, \frac{\lambda}{\Lambda}\Big] \beta(\omega; \lambda, \Lambda - \lambda + 1), \tag{3.2}$$

where Λ is the degree of the Bernstein polynomial, $\beta(\omega; \lambda, \Lambda - \lambda + 1)$ is the beta density evaluated at ω with parameters λ and $\Lambda - \lambda + 1$, and $G\big(\frac{\lambda-1}{\Lambda}, \frac{\lambda}{\Lambda}\big] = G(\frac{\lambda}{\Lambda}) - G(\frac{\lambda-1}{\Lambda})$. We employ a Bayesian non-parametric approach and assume that G is a cumulative distribution function obtained as the realization of a Dirichlet process (Petrone, 1999a). That is, G can be written as a discrete probability distribution, $G(x) = \sum_{l=1}^{\infty} p_l \delta_{z_l}(0, x]$, $x \in [0, 0.5]$, with atoms z_l randomly drawn from a base measure G_0, $z_l \overset{iid}{\sim} G_0$, with support on $(0, 0.5)$ and weights p_l characterized through the stick-breaking representation (Sethuraman, 1994), i.e., let $p_1 = v_1$ and $p_l = v_l \prod_{s=1}^{l-1}(1-v_s)$ where $v_l \sim beta(1, \alpha_v)$ with concentration parameter α_v. It follows that G is a random probability measure, such that $\mathbb{E}(G(x)) = G_0(0, x]$, for all $x \in [0, 0.5]$. The concentration parameter α_v characterizes the variability of the realizations G around the base measure G_0. In symbols, we write $G \sim DP(\alpha_v, G_0)$. In the following, we will consider a finite truncation approximation with truncation level L so that $v_L = 1$ and $G = \sum_{l=1}^{L} p_l \delta_{z_l}$ (Ishwaran & James, 2001). We now re-write our model as

$$f(\omega) = \sum_{\lambda=1}^{\Lambda} \sum_{l=1}^{L} p_l I\Big[\frac{\lambda - 1}{\Lambda} < z_l \leq \frac{\lambda}{\Lambda}\Big] \beta(\omega; \lambda, \Lambda - \lambda + 1). \tag{3.3}$$

This model, henceforth referred to as the Bernstein Dirichlet Prior (BDP) model, was introduced by Petrone (1999a) and applied to spectral density estimation for a single time series by Choudhuri et al. (2004). These two papers along with Petrone (1999b) and Barrientos et al. (2017) demonstrated the utility and theoretical properties of the BDP model for spectral density estimation.

In this BDP model, Λ controls the number of beta PDF mixture components when estimating a single spectral density curve. A higher Λ corresponds to a larger number of Bernstein polynomial components, and thus the ability to capture sharper peaks in the underlying spectral density. The DP, G, then properly assigns weights to each component of the selected basis. Along these lines, Petrone (1999b) offered some nice intuition for the BDP model as a smoothed histogram, where Λ can be viewed as the number of bins in the histogram and G assigns each observation to one

of the available bins.

Note that we do not specify a scalar, referred to as τ by Choudhuri et al. (2004), which only serves to multiply the spectral density by the total variance of the time series. By standardizing the time series to mean zero and unit variance, we remove the need to estimate τ while still capturing the desired information about the shape of $f(\omega)$.

3.2.2 The Multi-Subject Model

Having formulated the BDP model to estimate a single spectral density curve, we now consider a model to estimate many spectral densities from time series collected on a sample consisting of twin pairs. Because the twin relationships induce similarities across spectral density estimates, we propose grouping similar individual spectral densities by nesting the BDP model within a second DP. The partitions enforced by the nested DP explain the heterogeneity between subjects by assigning subjects with very different spectra to separate groups while also allowing very similar subjects to receive accordingly similar spectral density estimates by frequently grouping these subjects together.

More specifically, we assume that the individual spectral density may be assigned to one of K groups, with each group characterized by a specific spectrum profile, i.e., K realizations $\{G_1^*, \ldots, G_K^*\}$ from a BDP as in Equation 3.3. The individual spectral density for subject n is estimated as in Equation 3.2. Let G_n be subject n's draw from the nested DP, then, in formulas, $G_n \sim Q$ with $Q = \sum_{k=1}^{K} \pi_k^* \delta_{G_k^*}$ and each $G_k^* = \sum_{l=1}^{L} p_{kl} \delta_{z_{kl}}$ defined as in Section 3.2.1 for $k = 1, \ldots, K$. The weights π_k describe the proportion of subjects assigned to group $k = 1, \ldots, K$ and we assume they are defined similarly to the p_l, through a stick-breaking construction, $\pi_k = u_k \prod_{s=1}^{k-1}(1-u_s)$ where $u_k \sim beta(1, \alpha_u)$. We refer to our model as the nested Bernstein Dirichlet prior (NBDP) model. Let ζ_n be an allocation variable, such that $\zeta_n = k$ for $k = 1, \ldots, K$ and $n = 1, \ldots, N$, if and only if subject n is assigned to group k. Then the spectral

density for subject n can be written as:

$$f_{\zeta_n}(\omega) = \sum_{\lambda=1}^{\Lambda_{\zeta_n}} \sum_{l=1}^{L} p_{\zeta_n l} I\Big[\frac{\lambda-1}{\Lambda_{\zeta_n}} < z_{\zeta_n l} \leq \frac{\lambda}{\Lambda_{\zeta_n}}\Big] \beta(\omega; \lambda, \Lambda_{\zeta_n} - \lambda + 1), \tag{3.4}$$

where we allow different Λ_{ζ_n} in each group.

Figure 3.1 shows a diagram representing the NBDP model. In the context of multi-subject EEG data, the nested DP models the heterogeneity of the sampled EEG spectral densities by partitioning the set of subjects into homogeneous groups. Meanwhile, the BDP fits a functional curve for each group. We use an MCMC sampling algorithm, where at each iteration we first assign each subject to one of the K available groups with function-level estimate G_k. These G_k BDP functional curve estimates are then updated to best fit the subjects assigned to that group. We stress that we are not concerned with the partitions induced by the model, but only employ the nested DP to account for the heterogeneity across subjects, which allows improved estimates of power spectra and heritability. The flexible nature of the NBDP groupings allows us to capture potentially complex twin relationships that vary across twin pairs and across frequencies.

3.2.3 Estimating the Heritability of the Power Spectrum

Given the posterior distributions of the individual power spectra obtained using our NBDP model, we need a valid method for estimating heritability of these resting-state EEG power spectra using the available twin pair relationships in the MTFS data. As previously mentioned, establishing that spectral characteristics are heritable is essential for developing endophenotypes that tie these neurobiological indicators to their genetic underpinnings. Falconer's formula estimates the heritability of the power spectrum at frequency ω as $h^2(\omega) = 2(r_{MZ}(\omega) - r_{DZ}(\omega))$, where $r_{MZ}(\omega)$ and $r_{DZ}(\omega)$ are the correlation in the estimated individual spectral densities at frequency ω among MZ and DZ twins respectively (Falconer, 1960). Intuitively, MZ twins are

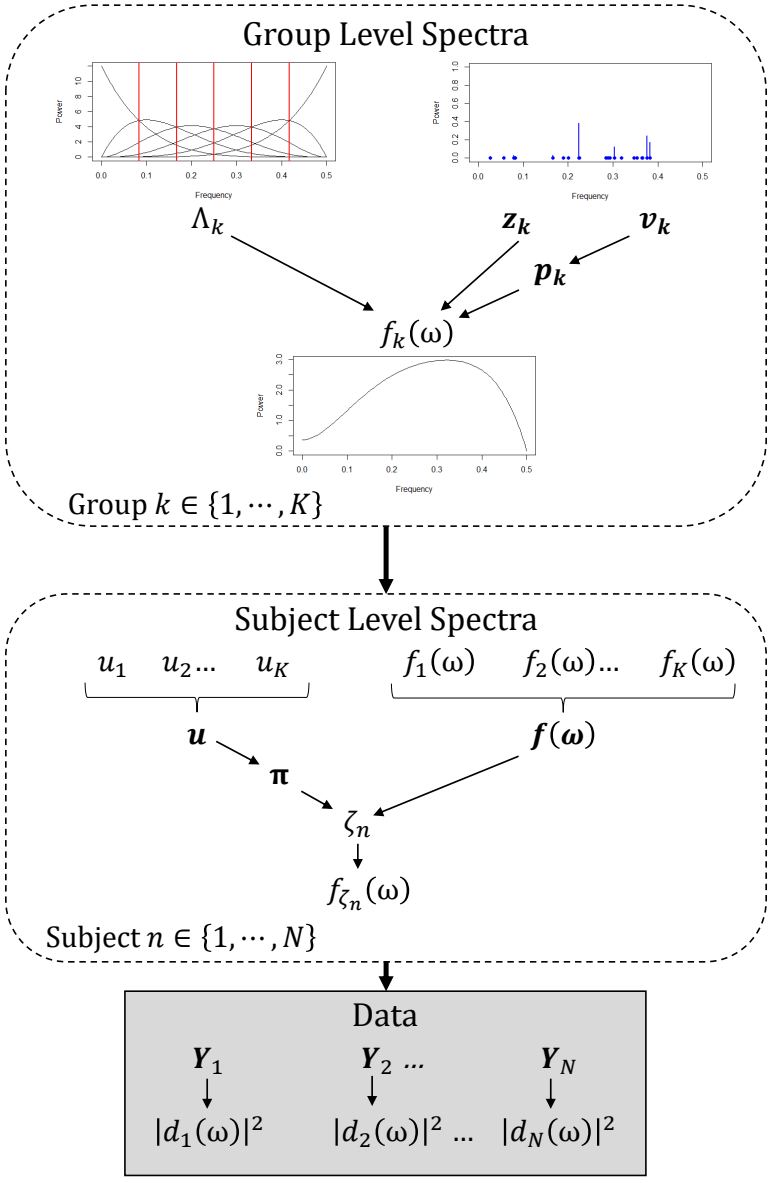


Figure 3.1: A schematic representation of the NBDP model. The BDP is used to estimate a group level spectrum for each of the K different groups. The nested DP then assigns each of the N subjects to one of the K group spectral densities.

genetically identical and DZ twins share 50% of their genetics on average. Falconer's formula estimates heritability by estimating half of the genetic effect as the difference between MZ and DZ correlations and multiplying this by 2 to get the full genetic effect. While this estimator gives us the heritability of a single frequency, it is more scientifically useful to consider neighborhoods of frequencies, say (ω_1, ω_2), known as frequency bands. Findings such as those in Malone et al. (2014) and Rudo-Hutt (2015) generally use frequency band results instead of attempting to interpret each individual frequency. To estimate frequency band heritability we propose a new estimator that integrates across frequencies, weighting the heritability at each frequency by the total variation in the power spectra at that frequency and then dividing by the total integrated variability in the power spectra. Our novel estimator takes the form

$$h^2(\omega_1, \omega_2) = \frac{\int_{\omega_1}^{\omega_2} h^2(\omega)\text{Var}(f(\omega))d\omega}{\int_{\omega_1}^{\omega_2} \text{Var}(f(\omega))d\omega}. \tag{3.5}$$

We estimate $\text{Var}(f(\omega))$ by taking the sample variance at each frequency across all of the curves from our posterior distribution. This novel heritability estimator allows us to calculate heritability for any frequency band without having to first reduce the spectral density curves to a single power estimate. Note that we can compute the heritability of the entire spectral density curve, which we call the full spectrum heritability, by setting the bounds of integration in Equation (3.5) to (0, Nyquist frequency). This full spectrum heritability could not be calculated using existing methods, and so it allows the introduction of a new set of endophenotypes based on the full spectrum of an EEG time series.

3.2.4 MCMC Sampling Algorithm

Given $\Omega = \{1/T, 2/T, \ldots, (\lfloor T/2 \rfloor - 1)/T\}$, the Whittle likelihood for subject n is $L(\boldsymbol{f}_n|\boldsymbol{Y}_n) \propto \prod_{\omega\in\Omega} exp[|d_n(\omega)|^2/f_{\zeta_n}(\omega)]/f_{\zeta_n}(\omega)$, and the posterior density of our

NBDP model is proportional to

$$\Big[\prod_{n=1}^{N}\prod_{\omega\in\Omega}\frac{1}{f_{\zeta_n}(\omega)}exp\Big(\frac{|d_n(\omega)|^2}{f_{\zeta_n}(\omega)}\Big)\Big]\times\Big[\prod_{k=1}^{K}\prod_{l=1}^{L-1}\alpha_v(1-v_{kl})^{\alpha_v-1}\Big]\times\Big[\prod_{k=1}^{K}\prod_{l=1}^{L}g_0(z_{kl})\Big]$$
$$\times\Big[\prod_{k=1}^{K}\rho(\Lambda_k)\Big]\times\Big[\prod_{k=1}^{K-1}\alpha_u(1-u_k)^{\alpha_u-1}\Big], \tag{3.6}$$

where $\rho(\Lambda_k)$ is the prior on the degree of the Bernstein polynomial. Typical vague prior specifications set α_v and α_u equal to 1, let G_0 be a uniform on the interval from zero to the nyquist frequency, and specify $\rho(\Lambda_k)$ as a discrete uniform prior on integers from 1 to some large integer (we use 300). Note that this prior specification significantly simplifies the posterior distribution to be proportional to

$$\prod_{n=1}^{N}\prod_{\omega\in\Omega}\frac{1}{f_{\zeta_n}(\omega)}exp\Big(\frac{|d_n(\omega)|^2}{f_{\zeta_n}(\omega)}\Big). \tag{3.7}$$

Our MCMC procedure takes the following steps:

1. For $k\in(1,\ldots,K-1)$, sample the stick breaking weight for the population-level DP, u_k, from its conditional posterior distribution, $u_k|\Lambda_k,\boldsymbol{\zeta},\boldsymbol{v}_k,\boldsymbol{z}_k\sim beta(1+\sum_{n=1}^{N}I(\zeta_n=k),\alpha_u+\sum_{n=1}^{N}I(\zeta_n>k))$.

2. Sample the group assignment, ζ_n, for each subject from a multinomial distribution where
$Pr(\zeta_n=k|\Lambda_k,u_k,\boldsymbol{v}_k,\boldsymbol{z}_k)\propto\pi_k\sum_{\lambda=1}^{\Lambda_k}\sum_{l=1}^{L}p_{kl}I[(\lambda-1)/\Lambda_k<z_{kl}\leq\lambda/\Lambda_k]\beta(\omega;\lambda,\Lambda_k-\lambda+1)$.

3. For $k\in(1,\ldots,K)$, sample the degree of the Bernstein Polynomial, Λ_k, using a Metropolis-Hastings step with Poisson proposal density with the mean equal to the last value of Λ_k.

4. For $k\in(1,\ldots,K)$ and $l\in(1,\ldots,L-1)$, sample the function-level DP stick breaking weight, v_{kl}, using a Metropolis-Hastings step with a uniform proposal

on a sample space extending ε_l on either side of the previous sample and taking the proposed value modulus 1.

5. For $k \in (1, \ldots, K)$ and $l \in (1, \ldots, L-1)$, sample the function-level DP atom, z_{kl}, using a Metropolis-Hastings step with the same proposal used for v_{kl}.

We use a truncation approximation at both levels. Given the truncation levels are set sufficiently high, the truncation approximation has minimal impact on the resulting estimates and increasing the truncation level only increases computing time of the MCMC sampler.

3.3 Simulation Study

In this section, we use simulated data to show the gains in estimation performance when using our NBDP model relative to existing approaches typically used in the time series community for spectral analysis. We point out that, while we can simulate time series data with a twin correlation structure, we cannot obtain the analytical form for the heritability $h^2(\omega)$ and the band heritability $h^2(\omega_1, \omega_2)$. Thus, in this section, we evaluate performance only with respect to estimating the spectral density and the power within certain frequency bands.

3.3.1 Data Simulation Process

Simulation experiments were designed to mimic the MTFS data as closely as possible. We considered four different scenarios, each with 1000 simulated subjects. In two of these scenarios we simulated the data in pairs to replicate the effects of the twin relationships, using 300 MZ twin pairs and 200 DZ twin pairs in the simulations. As in the MTFS data each simulated time series represents 8 seconds of data sampled at 60 Hz, resulting in 480 time points and 239 non-zero frequencies. The MTFS data was sampled at a higher rate, but only 239 non-zero frequencies fell below the 30 Hz cut-off of the low-pass filter.

Five separate autoregressive (AR) processes with spectral densities that visually resembled those estimated from the MTFS data were selected with orders between 5 and 18. For each simulated subject, one of the five AR processes was selected and random noise was added to the AR parameters. Scenario 1 simulated a grouped twin scenario by considering all 5 AR processes, assigning twins to the same AR process, and enforcing correlation in the random noise of the AR parameters for MZ twins. Scenario 2 assigned an AR process and generated the AR parameter noise independently for each subject. Scenarios 3 and 4 only considered a single AR process with random noise increased relative to scenarios 1 and 2. Scenario 3 also induced twin structure by making the AR parameter random noise moderately correlated for DZ twins and highly correlated for MZ twins. The simulated true spectra from the four different scenarios are shown in Figure 3.2.

3.3.2 Model Comparison

For model comparisons, we chose three competitor models. The first was the generalized cross validation span selection periodogram smoother of Ombao et al. (2001). For this method, henceforth referred to as GCV, we considered spans between 3 and 100 for each subject. The second was the stationary version of Rosen et al. (2009), which uses Bayesian smoothing splines, fit to each subject separately. We considered versions of this model that used $J = 10$ and $J = 20$ spline basis components. These models will be referred to as Spline10 and Spline20, respectively. The final competitor we considered is the BDP model (Choudhuri et al., 2004), which forms the base of our NBDP model but does not allow borrowing of information across subjects. A discrete uniform prior was used for the degree of the Bernstein polynomial, Λ, for each subject in the BDP model and a truncation level of $L = 20$ was used. 50,000 posterior samples were collected for the Spline10, Spline20, and BDP samplers. We fit our NBDP method to each scenario with the same priors chosen for the BDP model. We additionally specified the truncation level of the subject-partitioning DP

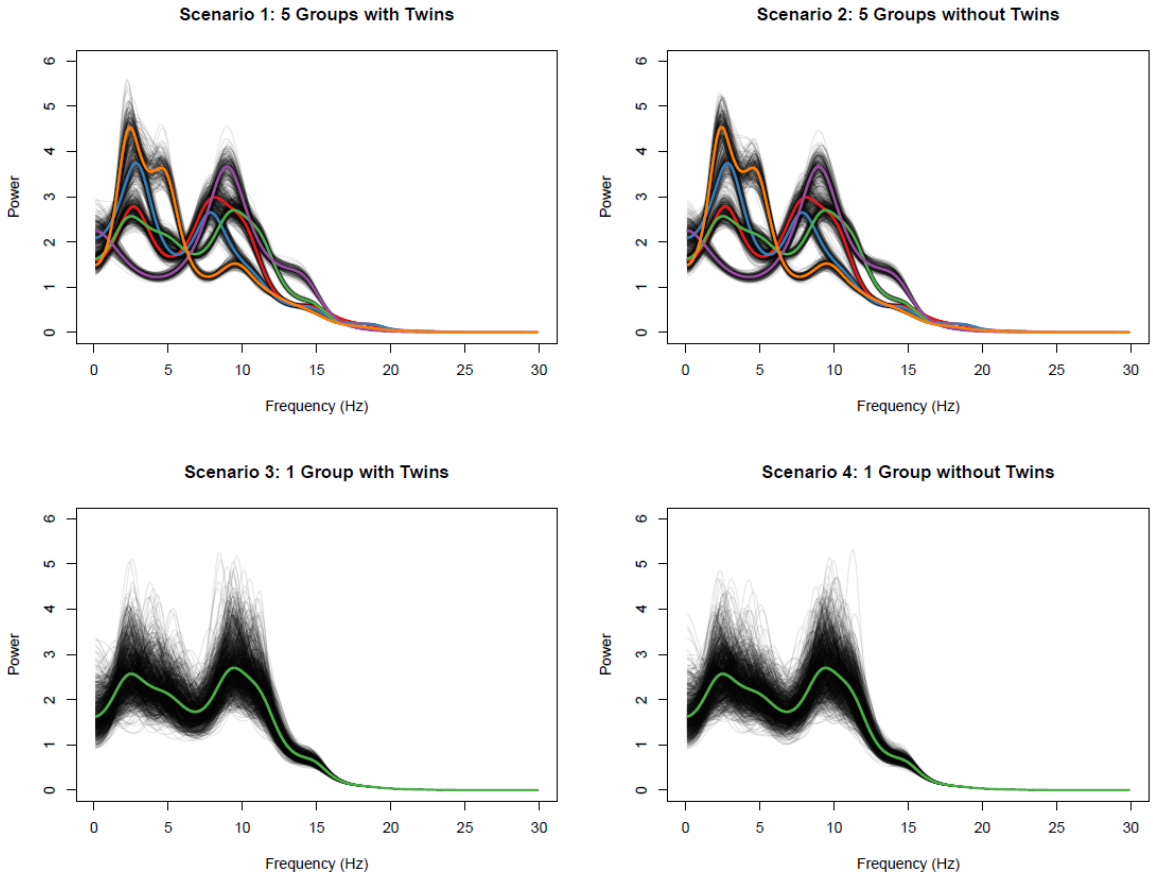


Figure 3.2: Simulated spectra from each of the four simulation scenarios. Each black line represents the true spectrum for a single simulated subject. Each colored line represents the true spectra of the AR groups without any random noise added to the AR coefficients.

at $K = 100$ and collected 100,000 posterior samples from the NBDP model.

To compare the accuracy of the resulting estimated spectral density curves we calculated the integrated absolute error (IAE) for each subject. For a given estimated spectral density $\hat{f}(\omega)$ and true density $f(\omega)$, IAE $= \int_\omega |\hat{f}(\omega) - f(\omega)| \partial\omega$.

3.3.3 Simulation Study Results

Scenario	Method	Delta	Theta	Alpha	Beta	Full Spectrum
Scenario 1: 5 Groups with Twins	**GCV**	0.36 (0.20)	0.38 (0.17)	0.36 (0.19)	0.23 (0.10)	1.44 (0.40)
	Spline10	0.37 (0.22)	0.37 (0.19)	0.33 (0.17)	0.20 (0.09)	1.38 (0.40)
	Spline20	0.42 (0.22)	0.45 (0.19)	0.45 (0.21)	0.27 (0.10)	1.70 (0.41)
	BDP	0.35 (0.19)	0.52 (0.23)	0.40 (0.19)	0.20 (0.09)	1.62 (0.36)
	NBDP	**0.28 (0.16)**	**0.32 (0.18)**	**0.33 (0.16)**	**0.17 (0.07)**	**1.22 (0.32)**
Scenario 2: 5 Groups without Twins	**GCV**	0.36 (0.21)	0.38 (0.17)	0.35 (0.17)	0.22 (0.10)	1.42 (0.38)
	Spline10	0.37 (0.22)	0.36 (0.18)	**0.33 (0.17)**	0.20 (0.09)	1.36 (0.39)
	Spline20	0.42 (0.23)	0.45 (0.20)	0.44 (0.22)	0.27 (0.10)	1.69 (0.41)
	BDP	0.35 (0.19)	0.52 (0.23)	0.39 (0.19)	0.19 (0.09)	1.60 (0.36)
	NBDP	**0.28 (0.17)**	**0.34 (0.20)**	0.36 (0.18)	**0.17 (0.07)**	**1.28 (0.33)**
Scenario 3: 1 Group with Twins	**GCV**	0.33 (0.16)	0.35 (0.15)	0.43 (0.18)	0.25 (0.10)	1.45 (0.35)
	Spline10	0.34 (0.17)	0.35 (0.17)	0.42 (0.19)	0.21 (0.09)	1.41 (0.38)
	Spline20	0.39 (0.19)	0.43 (0.18)	0.51 (0.21)	0.26 (0.10)	1.69 (0.38)
	BDP	0.33 (0.16)	0.51 (0.24)	0.49 (0.22)	0.20 (0.08)	1.67 (0.44)
	NBDP	**0.30 (0.15)**	**0.34 (0.15)**	**0.39 (0.20)**	**0.18 (0.06)**	**1.35 (0.35)**
Scenario 4: 1 Group without Twins	**GCV**	0.33 (0.16)	0.36 (0.15)	0.45 (0.19)	0.25 (0.10)	1.48 (0.37)
	Spline10	0.34 (0.17)	0.36 (0.17)	0.44 (0.19)	0.21 (0.09)	1.44 (0.38)
	Spline20	0.39 (0.19)	0.43 (0.18)	0.52 (0.22)	0.26 (0.09)	1.71 (0.42)
	BDP	0.33 (0.16)	0.50 (0.24)	0.51 (0.22)	**0.20 (0.08)**	1.68 (0.43)
	NBDP	**0.30 (0.15)**	**0.33 (0.15)**	**0.42 (0.20)**	0.20 (0.07)	**1.37 (0.35)**

Table 3.2: Mean (sd) mean integrated absolute error (MIAE) across the 1,116 MTFS subjects. **Bold** numbers indicate the best performing model in each scenario and frequency band.

Table 3.2 contains the mean and standard deviation of the mean IAE for each model and simulation scenario. We also split the results into the four frequency bands commonly considered in EEG analysis. Our NBDP method had the best mean IAE in all four simulation scenarios when considering the entire spectrum. This advantage diminished as the amount of between subject similarity decreased through the four

simulation scenarios, but the NBDP model maintained the lowest mean IAE and the lowest standard deviation of IAE across all four scenarios.

The Spline20 method showed the worst performance in all four scenarios, as it likely over-fit the data with too many spline components. The spline method received a considerable boost in performance when $J = 10$ basis components were used, finishing second in all scenarios behind our NBDP model. While the spline model with $J = 10$ performed relatively well, the number of spline components must be selected a priori in this model, whereas the BDP and NBDP models adaptively select the proper order of Bernstein polynomial to use based on the data. Comparison of the 10 and 20 component spline methods shows that using the incorrect number of components can seriously impact the performance of the model.

The BDP method showed the second worst performance of the models under comparison in all scenarios. The large improvement from the BDP model to our NBDP model demonstrates the added value of a nested model that shares information between subjects. The decreasing advantage of the NBDP model from scenarios 1 to 4 is to be expected due to increased between-subject heterogeneity, and is reflected in the pattern in the Alpha band IAE. The only frequency bands where the NBDP model did not perform best were the Alpha band in scenario 2, where the Spline10 model had an advantage, and the Beta band in scenario 4, where the BDP model very slightly prevailed. The Spline10 model's small advantage in mean IAE in scenario 2 may come from its ability to better estimate a sharp alpha peak for certain subjects, but the NBDP model was still competitive. The decreased mean IAE in the Delta and Theta bands for the NBDP model in scenario 2 still resulted in a near 6% reduction in mean IAE for the full spectrum in comparison with the Spline10 model.

In EEG spectral density analysis the frequency with the highest spectral power over the entire spectrum (i.e., the peak frequency) and within the Alpha band (i.e., the Alpha peak frequency) are of interest. We calculated the error in the peak frequency and Alpha peak frequency for each simulated spectral density in each simulation scenario. When considering peak frequency, our NBDP method struggled when the

true spectral density had two peaks of relatively similar height, as is the case for roughly 40% of the subjects in scenarios 1 and 2 and all of the subjects in scenarios 3 and 4. The NBDP method favored the lower frequency peak when choosing between the two local maxima. This trait of the NBDP model is inherited from the BDP model, however, when considering Alpha peak frequency the NBDP method was much improved. It still exhibited a slightly skewed distribution, but the median error was near zero for all four scenarios and the performance was comparable to the GCV and spline methods. Importantly, the NBDP method performed much better than the BDP method for Alpha peak estimation, another example of the gains from jointly fitting the spectral densities for all subjects. By using the nested DP model, borrowing information across subjects allowed us to overcome the deficiencies of the BDP model in estimating Alpha peak frequency.

Note that the Bayesian machinery of the NBDP model makes inference on features of the spectral density curves such as peak frequency and alpha peak frequency very straight-forward using functions of the posterior samples. The GCV model does not offer a similar general approach to inference on features of the estimated spectral densities and each spectral feature requires the use large-sample results or the bootstrap.

Finally, we note that the methods compared do differ in terms of computing time. For example, MCMC iterations while fitting the model to a single simulation scenario took 6.2 seconds on average for the NBDP model compared to an average of 1.4 seconds for the Spline10 model. While the NBDP model is computationally expensive, the improved spectral density estimates shown in the simulation results justify the additional time required to fit the model.

3.4 MTFS Analysis

Our analysis of the MTFS data had two primary goals: 1) conduct spectral analysis in the resting-state EEG data to quantify the power within frequency bands of in-

terest, and 2) compute the band heritability for each frequency band, as well as the heritability of the entire spectral density.

3.4.1 Data Description

We applied our method to resting-state EEG data collected from the MTFS which constituted part of the molecular-genetic studies of Malone et al. (2014) and Smit et al. (2017). Specifically, the data used here were from the intake assessment of male and female twins from the age-17 cohort of MTFS twins and consisted of data from 3 electrodes common to both sex cohorts (Cz, O1 and O2), collected with a sampling rate of 128 Hz and 12 bits. The Cz electrode is located at the center midline on the top of the head and the O1 and O2 electrodes are located on either side of the midline at the rear of the head. Considering each channel separately, we applied our method to 8 seconds of data resulting in time series of length 1024. In total, our sample included 1116 adolescents consisting of participants from 365 distinct MZ twin pairs and 192 DZ twin pairs. The sample had approximately equal numbers of individuals of each sex with 565 females and 551 males.

While twin participants sat comfortably in a darkened room with their eyes closed, EEG signals were recorded by means of identical Grass 12 Neurodata systems, with a pass band from 1 to 30 Hz (amplifier rolloff, 6 dB/octave). Notes recorded when the data were originally collected guided identification of data that needed to be excluded because of recording problems. Subjects who reported having fallen asleep or were noted to have fallen asleep were excluded. EEG segments containing transient artifacts and excessively small or large voltage deflections were tagged for exclusion by a computer algorithm written in Matlab. Multivariate outliers across the 3 electrodes were identified using a robust version of Mahalanobis distance from the robustbase package in the R statistical programming environment and visually reviewed for contamination by high-frequency noise, other artifacts (e.g., electrocardiogram), or signs of sleepiness. Individual recording sites were excluded from analyses if fewer than 45

2-second artifact-free sweeps were available.

3.4.2 Analysis Setup

For analysis of the MTFS data we consider our NBDP model with truncation levels of $L = 20$ for the function-level DP and $K = 100$ for population-level DP. We limited the analysis to frequencies below 30 Hz. We chose a discrete uniform prior on integers from 1 to 300 for the prior distribution on the degree of the Bernstein polynomials for each group, $\rho(\Lambda_k)$. Putting a flat prior on the number of Bernstein polynomials allows the data to choose the amount of smoothing to apply to the periodogram. We set both concentration parameters, α_v and α_u, equal to 1 and the base distribution G_0 was assumed to be uniform. We used the ε_l values suggested by Choudhuri et al. (2004) for the Metropolis-Hastings steps to sample z_{kl} and p_{kl}. 100,000 posterior samples were collected from a single chain for the NBDP model. These same specifications were used for analysis of the time series from each of the three channels.

3.4.3 MTFS Results

Figure 3.3 shows, for each channel, the estimated power spectrum for each of the 1116 subjects in the MTFS. The somewhat clustered nature of our NBDP model is visible in these estimated curves, as many individual subjects' estimated curves fall nearly on top of each other, creating a darker black line. These groupings are defined by different characteristics such as the Alpha band peak visible around 10 Hz. Most subjects have some sort of peak between 7 and 12 Hz, but the exact frequency and power of this peak changes by group. The nested DP often detected the twin structure in the data, clustering MZ twins Cz power spectra together in 26.4% of posterior samples compared with 8.9% for DZ twins and only 4.6% for non-twins.

The bottom plot in Figure 3.3 shows, for each channel, Falconer's estimate of heritability for each frequency of the spectrum along with 95% point-wise credible intervals. Although it differed somewhat by channel, we found high levels of heritability

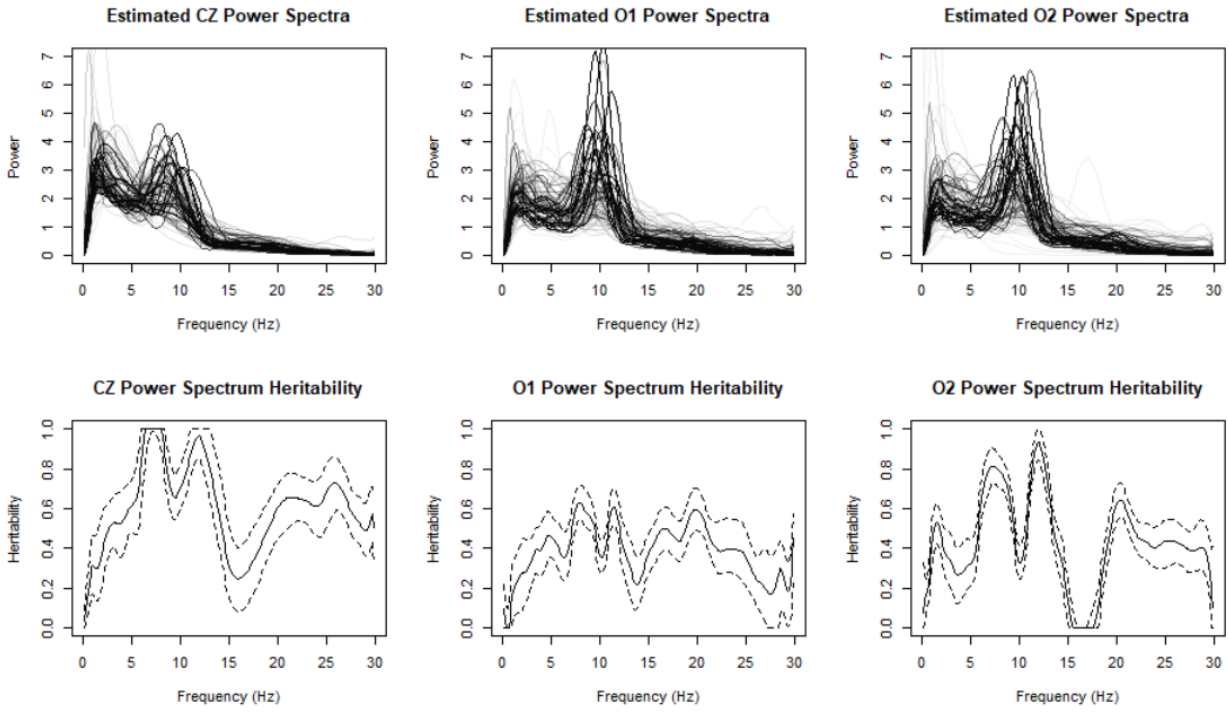


Figure 3.3: Top row: The estimated spectral density curves for each channel and each of the 1116 participants in the MTFS. Each line represents a single subject. Bottom row: The estimated heritability and 95% point-wise credible interval across the power spectrum calculated using Falconer's formula.

across the spectrum for all three channels, particularly in the Theta and Alpha bands. The highest levels of heritability correspond with the Alpha band peak, suggesting that this feature of the data may be largely genetically driven. All three heritability curves demonstrate similar shapes, and the Cz and O2 curves share a particularly strong resemblance to one another.

Table 3.3 reports the median and inter-quartile range of frequency band power and peak frequency estimates for the 1116 MTFS subjects along with the estimated sample heritability for each feature. For the frequency band heritability measurements we used the novel form of Equation 3.5. Compared to the Cz channel, the O1 and O2 channels exhibited larger Alpha peak power, Alpha band power, and Beta band power along with smaller Delta and Theta band power. Features such as the Delta and Alpha band peaks, which are clear when visually inspecting the estimated curves in Figure 3.3, are not as apparent after aggregating power across bands. Note that the full spectrum power is 1 for all subjects due to the nature of the NBDP model and the fact that the original time series were standardized to unit variance. As with the heritability curve in Figure 3.3, Table 3.3 shows low heritability in the Delta band, high heritability in the Theta and Alpha bands, and somewhat diminished heritability in the Beta band. While the Cz channel Beta band heritability is substantial, it should be noted that there is relatively little variation in the power spectra above 12 Hz, which denotes the beginning of the Beta band. Since heritability is the percentage of variation attributable to genetics, it is important to consider the overall variability of the data. Our novel heritability estimator also estimated full spectrum heritabilities of 0.68, 0.45, and 0.56 for the Cz, O1, and O2 channels respectively. Our establishment of the heritability of the full spectral density curve for these three channels paves the way for future research to establish endophenotypes by connecting the spectral density curve with psychiatric disorders.

Table 3.3 also shows the heritability estimates for the peak frequency and Alpha peak frequency. Both peak frequency and Alpha peak frequency showed high levels of heritability, though the Alpha peak appeared to be more heritable.

Channel	Feature	Median (IQR)	Heritability (95% CI)
Cz	Delta Band Power (1-4 Hz)	0.25 (0.23 - 0.32)	0.43 (0.33, 0.53)
	Theta Band Power (4-8 Hz)	0.28 (0.25 - 0.31)	0.88 (0.84, 0.92)
	Alpha Band Power (8-12 Hz)	0.29 (0.23 - 0.33)	0.77 (0.69, 0.87)
	Beta Band Power (12-30 Hz)	0.13 (0.12 - 0.18)	0.65 (0.55, 0.77)
	Full Spectrum	1.00	0.68 (0.62, 0.74)
	Peak Frequency	2.25 (1.63 - 8.63)	0.51 (0.35, 0.66)
	Alpha Peak Frequency	8.63 (8.13 - 9.63)	0.81 (0.70, 0.91)
O1	Delta Band Power (1-4 Hz)	0.18 (0.15 - 0.21)	0.30 (0.18, 0.40)
	Theta Band Power (4-8 Hz)	0.21 (0.16 - 0.24)	0.49 (0.43, 0.57)
	Alpha Band Power (8-12 Hz)	0.39 (0.30 - 0.45)	0.48 (0.44, 0.54)
	Beta Band Power (12-30 Hz)	0.19 (0.14 - 0.26)	0.39 (0.32, 0.46)
	Full Spectrum	1.00	0.45 (0.41, 0.50)
	Peak Frequency	9.50 (8.63 - 10.25)	0.37 (0.21, 0.54)
	Alpha Peak Frequency	9.56 (9.00 - 10.25)	0.52 (0.35, 0.67)
O2	Delta Band Power (1-4 Hz)	0.17 (0.15 - 0.21)	0.40 (0.32, 0.47)
	Theta Band Power (4-8 Hz)	0.21 (0.17 - 0.26)	0.67 (0.60, 0.73)
	Alpha Band Power (8-12 Hz)	0.38 (0.30 - 0.45)	0.58 (0.53, 0.63)
	Beta Band Power (12-30 Hz)	0.21 (0.15 - 0.26)	0.56 (0.50, 0.60)
	Full Spectrum	1.00	0.56 (0.52, 0.60)
	Peak Frequency	9.63 (8.50 - 10.38)	0.25 (0.13, 0.38)
	Alpha Peak Frequency	9.63 (8.88 - 10.38)	0.86 (0.69, 1.00)

Table 3.3: The median and inter-quartile range of different spectral density features across the 1116 MTFS subjects along with the heritability calculated from the sample for each feature and channel.

We investigated the sensitivity of our results with respect to the hyperparameters of our model. Overall, the results are qualitatively similar to the ones we present here. More details are in Appendix B.

3.5 Discussion and Conclusions

We developed a novel Bayesian non-parametric model for estimating the spectral densities of multi-subject, resting-state EEG data and their associated heritability. Our model embeds the BDP within a nested DP to share information across subjects with similar spectral densities, which led to large improvements in estimation as we showed in our simulation study. We applied our method to resting-state EEG data from the MTFS and showed that the resulting estimated spectral densities were highly heritable, especially in the Theta and Alpha bands. Additionally, the peak frequency and Alpha peak frequency were also heritable. These findings are consistent with Malone et al. (2014) and Smit et al. (2005). We extend these works by allowing inferential statements about the features of the spectral density, such as peak frequency and power within frequency bands, using our rigorous Bayesian non-parametric framework. Proper inference on features of the spectral densities and their heritability is straightforward once the posterior samples have been collected, and such inference does not rely on asymptotic results for these features.

Furthermore, our novel heritability estimator also allows the calculation of the heritability of the entire power spectrum, and we estimated the full spectrum heritability for the Cz, O1, and O2 channels to be 0.68, 0.45, and 0.56, respectively. These findings alone point to the genetic underpinnings of resting-state EEG signals and their oscillatory behavior measured in the frequency domain. Our finding that a large proportion of the variability in the spectral density can be attributed to genetic factors leads to the possibility of the entire spectral density being an endophenotype.

Once a trait has been established as heritable, it still remains to show this genetic risk factor is related to the psychiatric disorder under study. Past findings have shown

increased or decreased power within defined frequency bands to be associated with conditions such as alcoholism, depression, and ADHD (Rudo-Hutt, 2015). Quantitative genetics analysis approaches applied to EEG spectral densities highlight their potential for linking psychiatric disorders with genetic risk factors (Malone et al., 2014). To establish the full spectral density as an endophenotype we would need to establish its association with psychiatric conditions through techniques such as functional regression, which is beyond the scope of this work.

The nested DP allows us to account for the heterogeneity between subjects, while simultaneously accounting for the potentially complex correlation between subjects that may arise due to the twin design of the study. This allows us to model the heterogeneity in the spectral features of the data, which has potential clinical implications. For instance, Harper et al. (2018) showed the association between spectral power and behavioral disinhibition. Furthermore, Lizio et al. (2011) showed potential clinical utility for Alpha peak frequency, and Grandy et al. (2013) showed that variation in Alpha peak frequency is associated with cognitive ability.

One possible extension of our model would be to include covariate information in the BDP portion of our model through dependent Dirichlet processes (MacEachern, 1999). Barrientos et al. (2017) proposed and compared methods to make the weights and atoms of the BDP dependent on a matrix of covariates. Incorporation of covariate information, such as age or disease status, at the BDP level could, at the cost of increased computation, improve functional curve estimation and allow more flexibility in the estimation procedure, and it would allow subjects assigned to the same group at a single MCMC iteration to have different estimated spectral densities based on their covariate values. By including psychiatric disorders as covariates, a dependent DP extension of our model may be able to establish endophenotypes in a more unified framework. Bruce et al. (2017) offers a method for modeling spectral densities conditional on covariates, but not within a DP framework.

Edwards et al. (2017) formulated a B-spline prior model, which can be viewed as a generalization of the BDP model, and showed that it captured sharp peaks

in spectral densities better than the BDP model at the cost of a 2-3 fold increase in computing time. There is potential for improved performance of our model by replacing the Bernstein polynomial basis with a B-spline basis within the nested DP, however, the significant increase in computational burden would make the method very cumbersome with datasets as large as the MTFS.

Chapter 4

A Grouped Beta Process Model for Multivariate Resting-State EEG Microstate Analysis on Twins

4.1 Introduction

4.1.1 EEG Microstates and Dynamic Functional Connectivity

In the last two decades a type of analysis known as electroencephalography (EEG) microstate analysis has become popular with resting-state EEG data. In microstate analysis, a multivariate EEG time series is segmented into blocks where a certain time series pattern or EEG topography explains the behavior within each block. These defined EEG patterns or topographic maps are commonly referred to as microstates and it is generally assumed that a limited number of microstates (typically between 4 and 13) explain each EEG time series (Khanna et al., 2015). At each time point, the EEG activity is generated from one of these states, which are stable, persist as the active state anywhere from tens to hundreds of milliseconds, and repeat across long ranges of time within a single EEG recording (Lehmann et al., 1987). In EEG data, the transitions between these states have been shown to be sudden and nonrandom, where the order in which the states appear is potentially important (Betzel et al.,

2012; Lehmann et al., 2005).

The early work of Koenig & Lehmann (1996) showed differing topographic microstates for the production of nouns compared to vowels and Stevens & Kircher (1998) found that elderly demented patients experienced shorter durations in each microstate. Koenig et al. (2002) defined four topographic maps that are commonly used in microstate analyses, taking a spatial first-order view of microstates. These four microstates, also referred to as the canonical microstates, have been shown to mediate trial-by-trial risk-taking (Pedroni et al., 2017), associate with different mental tasks such as visualizing and verbalizing (Milz et al., 2016), and differ between schizophrenic and healthy aging adults (Andreou et al., 2014; Lehmann et al., 2005). Brodbeck et al. (2012) showed slight differences in the four canonical states in different stages of wakefulness and sleep and Schlegel et al. (2012) showed the usage of these states to differ based on belief in paranormal activity. One common theory is that these four microstates represent atoms of thought that combine to form human cognition (Lehmann et al., 1998). Khanna et al. (2015) gave a thorough review of the EEG microstate literature.

Other very similar concepts exist alongside the topographic EEG microstate literature. Van de Ville et al. (2010) provided evidence that EEG microstate sequences in resting-state imaging were scale-free, and thus could be connected with the vast amount of research done in exploring dynamic functional connectivity (FC) in functional magnetic resonance imaging (fMRI). Since then Allen et al. (2014) verified the connection between fMRI FC and EEG microstates and Allen et al. (2018) showed distinctions between the EEG spectra associated with time points concurrent to 5 identified fMRI FC networks. The literature on fMRI dynamic FC has also revealed differential behavior in schizophrenic patients, adding to the EEG literature proporting similar claims (Damaraju et al., 2014). Thorough overviews of the vast fMRI dynamic FC literature for resting-state and task-based fMRI can be found in Hutchison et al. (2013) and Gonzalez-Castillo & Bandettini (2018), respectively.

Through all of this research, it is clear that much can be learned about the brain

through study of its dynamic spatio-temporal networks. The modalities used to study these microstates or dynamic networks may vary and there are a plethora of different manners of quantifying these patterns (Sakkalis, 2011), but each approach to quantifying the dynamic activity of the resting brain may have something to offer to the bigger picture.

4.1.2 Twin Microstate Analysis

Here we analyze data from the Minnesota Twin and Family Study (MTFS) (Iacono et al., 1999). The MTFS is a population-based study of twins, the overarching goal of which is to understand genetic and environmental influences on substance abuse and related psychopathology. We analyze a set of twin pairs from the MTFS, treating each epoch, which is a short segment of the EEG data, collected from a participant as a separate time series. More details on the MTFS data are provided in Section 4.5.1.

While there is a substantial literature on EEG microstates and fMRI dynamic FC, the heritability of microstates and dynamic networks is still largely an open question (Pluta et al., 2018). Works such as Ge et al. (2017) and Fu et al. (2015) have analyzed the heritability of fMRI FC, but have yet to extend the results to dynamic networks. In the EEG microstate literature, only Vidaurre et al. (2017) has addressed this issue, finding that certain EEG microstate characteristics exhibit heritability. The twin study design of the MTFS gives us an opportunity to build on this finding and continue the important work of gaining insight into the genetic basis of brain activity.

Vidaurre et al. (2017) not only showed that EEG microstates demonstrate heritability, but that they also exhibit subject-specific characteristics. Thus, it is desirable that, although we treat each epoch as a separate time series, information about the underlying state sequences and transitions between those states is shared between epochs from the same participant. With the heritability in mind, we choose to build an EEG microstate model at the twin pair level as opposed to the individual level

to share information between twins in the estimation of the microstate characteristics and dynamics. By jointly modeling the EEG microstates of a twin pair, we take advantage of the fact that twins will likely be more similar than two unrelated participants. Existing EEG microstate modeling approaches, however, do not account for the similarity between cotwins. The goal of this work is to develop a model that will 1) allow for different levels of information sharing within a twin pair since MZ twins will likely share more microstate features than DZ twins, and 2) explicitly allow for sharing of information across epochs from a single participant in recognition of subject-specific EEG microstate traits and across twins in recognition of the heritability of these microstates.

This paper is organized as follows. In Section 2, we give a conceptual overview of the different modeling components that we will use for our model. In Section 3, we formally define our model using the different components from the previous section. In Section 4, we illustrate the performance of our model using simulated data, and then we present our analysis of the MTFS data in Section 5. Finally, in Section 6 we give concluding remarks.

4.2 Model Building Blocks

Before giving the details of our proposed EEG microstate model, we present a collection of different models which will each serve as part of the final proposed model. This section serves as a conceptual introduction to these modeling components and Section 4.3 formally specifies the model and it's notation.

4.2.1 The MS-VAR Model

Using the findings of previous studies of EEG microstates and dynamic FC, we propose a Markov switch vector auto-regressive (MS-VAR) EEG microstate model (Stoffer & Shumway, 2006). In this MS-VAR model, the underlying EEG microstate sequence changes across time. The dynamics of this state sequence follow a hidden

Markov model (HMM) framework. In an HMM model a person's EEG signal is in one of K available states and a transition matrix defines the probability of transitioning from that state to each of the $K-1$ other available states. These transition probabilities differ based on the current state of the EEG time series, but do not change over time. In this way, the HMM framework is able to capture a set of stable microstates such as those of Koenig et al. (2002) and allow sudden and nonrandom transitions between these microstates as described in Betzel et al. (2012). We enhance the HMM model by using the sticky HMM of Fox et al. (2011b) which adds a sticky parameter that makes it more likely that transitions will result in successive time points being assigned to the same state more likely.

Conditional on the state sequence, the behavior of the EEG time series at a given time point in an MS-VAR model reduces to a vector autoregressive (VAR) model. In a VAR model a set of parameters describes the behavior of the time series as a regression on a set of previous time points with an additive Gaussian noise. Each of the K available microstates will have a distinct corresponding set of VAR parameters, which are referred to as the emission parameters of the HMM. These emission parameters describe the behavior of the time points in each latent state by defining the lag/lead relationships between different EEG recording channels and the amount of noise seen within a state.

4.2.2 Latent State Models

Given a sticky MS-VAR structure, we now discuss ways to model the different microstates through latent state and feature models. Based on Koenig et al. (2002), four microstates are frequently used in EEG microstate analyses, but other research has suggested there are as many as 13 microstates commonly used (Khanna et al., 2015). By allowing the state space of our model to shrink or grow as dictated by the data, we allow models with any number of states and are able to perform inference on the number of latent states used during an EEG time series, a potentially scientifically

interesting parameter.

A few Bayesian nonparametric models for time series with switching dynamics have been proposed, overviews of which can be found in Lehman et al. (2015) and Fox et al. (2010). Fox et al. (2011a) and Fox et al. (2011b) introduced HMM models with switching dynamics that fit into the framework we seek for a single time series by using the hierarchical Dirichlet process (HDP) of Teh et al. (2005).

4.2.3 Feature Models and The Indian Buffet Process

Because our MTFS data comes from twins and EEG microstates have been shown to be heritable (Vidaurre et al., 2017), we seek a model to share information within a twin pair. Fox et al. (2014) offered another Bayesian nonparametric framework that expands on Fox et al. (2011b) to jointly model a set of time series using a feature-based model. This model is known as the Beta process autoregressive hidden Markov model (BP-AR-HMM or BP for short). In the BP-AR-HMM model, a Beta process Bernoulli process model replaces the HDP of Fox et al. (2011b). Feature models such as the Beta process Bernoulli process model the potentially overlapping latent state spaces of multiple different time series simultaneously while determining the degree of the overlap. For each time series considered, one can imagine flipping a coin for each of a countably infinite number of microstates where the probability of heads is determined by the Beta process. If the coin lands on heads then that feature (microstate) is turned on and can be used in the state sequence of that time series. In this manner, each time series selects a set of features from a feature space that is shared between all of the time series being modeled, allowing a single EEG microstate to appear in the state-sequence for both twins in a pair. The model still retains the flexibility to allow a single participant to exhibit unique microstates not present in the state-sequence of their co-twin. Also note that in the BP-AR-HMM model transition probabilities are estimated separately for each time series. Even if twins share features, they need not have the same microstate dynamics.

The Beta process Bernoulli process model is defined by two parameters, denoted α and c, where α controls the total number of features that appear in the N different time series and c controls the number of features unique to each time series. Ghahramani & Griffiths (2006) provided insight into this distribution which is also known as a two-parameter Indian Buffet Process (IBP). Note that when $c = 1$ we are left with the typical one-parameter IBP. The generative model for the two-parameter IBP can be thought of as a set of random draws from Poisson distributions which determine the number of previously unused features selected by each time series. The number of features available to the first time series is drawn from a Poisson distribution with mean α. The second time series then flips a coin for each feature to determine which of the first time series' features are also available to the second time series, where the probability of the feature being available to the second time series is $1/(c+1)$. The second time series then draws from another Poisson distribution with mean $\alpha c/(c+1)$. This process continues so that, for the n-th time series, the previously sampled features are selected with probability $m_k/(c+n-1)$, where m_k is the number of previous time series which selected feature k, and a new draw from a Poisson distribution with mean $\alpha c/(c+n-1)$ determines the number of new features selected by time series n.

To develop intuition for this generative process and the effect each parameter has on the total number of features selected by this model and the amount of sharing of features between time series it is helpful to consider the extreme values of c. As c approaches 0, the number of new features selected by each new time series also approaches 0, thus the features drawn from the Poisson(α) distribution of the first time series will be used to describe all of the time series being jointly model. As c grows toward infinity, the likelihood of selecting a previously selected feature shrinks towards 0 and the mean of each new Poisson draw approaches α until each time series will select a unique set of features with no shared features between time series. The IBP model then allows any amount of sharing of microstates between time series while trying to determine a sparse set of microstates that adequately describe the EEG signals.

4.2.4 The GBP-AR-HMM Model

While the BP-AR-HMM model is sufficient to model a single time series from a single individual, we have 70 epochs which may or may not be contiguous from each individual in a twin pair. We thus want to model each of these epochs as a separate time series, creating 140 time series coming from two different participants which we would like to jointly model. To accomplish this goal we propose an extension of the BP-AR-HMM model which we call the grouped BP-AR-HMM model (GBP-AR-HMM or GBP for short). In the GBP-AR-HMM model, each time series is assigned to a group a-priori. The feature selection and transition distributions are then modeled at the group level as opposed to the time series level. When applied to our MTFS EEG data, we define each participant as a group. In this way, the 70 epochs from each participant are grouped together with a shared a set of underlying microstates and state dynamics.

Forcing all 70 epochs to share a set of features and transition distributions ensures that epochs from the same participant exhibit consistent microstate dynamics across the EEG recording. The transition distributions will be estimated separately for each participant, but because we fit the model at the twin pair level, twins will share the same library of available features to sample from, allowing the analysis to learn the amount of shared information between twins. The flexibility of the GBP model means you can jointly fit epochs from participants whom you expect to have differing amounts of similarity. For example, MZ twins will likely be more similar than DZ twins which will be more similar than two randomly selected participants, but the model will be able to learn the overlap in the state-space in any of those situations.

4.3 Model and Prior Specification

With the GBP-AR-HMM model and its individual components introduced conceptually in the previous section, we now proceed to formally define these components

and the overall GBP model.

4.3.1 MS-VAR Model

In the simplest case of a univariate time series following a first order auto-regressive model, the observation at time point t, y_t, is modeled as $y_t = \mathrm{A}y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ and ε_t are independent across t. Extending this to an order R model, denoted VAR(R), we have $y_t = \sum_{r=1}^{R} \mathrm{A}_r y_{t-r} + \varepsilon_t$. In a multivariate time series this definition of a VAR model still holds, except that for a p-dimensional time series, y_t is now a p-dimensional vector for any t, $\mathbf{A}_i$ is a $p \times p$ matrix, and ε_t is a p-dimensional vector drawn from a normal distribution with mean $\mathbf{0}$ and variance $\mathbf{\Sigma}$.

Now to move from a VAR to a MS-VAR, we introduce a latent state, z_t, defined at each time t. Indexing the autoregression coefficients $\mathbf{A}$ and error variance $\mathbf{\Sigma}$ by the underlying state, the switching-VAR model can be written as $\mathbf{y}_t = \sum_{r=1}^{R} \mathbf{A}_r^{(z_t)} \mathbf{y}_{t-r} + \varepsilon_t$ where $\varepsilon_t \sim \mathcal{MVN}(0, \mathbf{\Sigma}^{(z_t)})$ and $\mathcal{MVN}$ denotes the multivariate normal distribution and ε_t is independent across t. As mentioned previously, the sequence of latent states in our MS-VAR model is determined through the HMM portion of the model. Together we denote the emission parameters for microstate k as $\theta^{(k)} = \{\mathbf{A}^{(k)}, \mathbf{\Sigma}^{(k)}\}$ where we have dropped the subscripts on $\mathbf{A}^{(k)}$ when referring to the concatenated set of R autoregression coefficient matrices. For both the simulation study and MTFS analysis we use $R = 1$ to keep the dimensionality of the parameter space to a reasonable size.

With consideration of an efficient MCMC sampler in mind, we use a matrix-normal inverse-Wishart (MNIW) prior for $\theta^{(k)}$ (West & Harrison, 2006). Under the MNIW prior, $\mathbf{A}^{(k)} \sim \mathcal{MN}(\mathbf{M}, \mathbf{V}, \mathbf{L})$ where $\mathcal{MN}$ is a matrix normal distribution with mean matrix $\mathbf{M}$ and left and right covariance matrices $\mathbf{L}^{-1}$ and $\mathbf{V}$. In general we set $\mathbf{M}$ to be a matrix of zeros, $\mathbf{V}$ to the $p \times p$ identity matrix, and $\mathbf{L}$ to a scaled $p \times p$ identity matrix. The MNIW prior also assumes $\mathbf{\Sigma}^{(k)} \sim \mathrm{IW}(n_0, \mathbf{S}_0)$, where IW is an inverse-Wishart distribution with n_0 degrees of freedom and scale matrix $\mathbf{S}_0$. We

chose $n_0 = p + 2$ and follow the recommendation of Fox et al. (2011a) in setting $\mathbf{S}_0$ equal to 0.75 times the empirical covariance of the first difference of the time series. The conjugacy of this prior allows efficient parameter updating and a sampler that includes both birth-death and split-merge sampling steps. These birth-death and split-merge steps vastly improve the efficiency of the sampler (Fox et al., 2014).

The sticky HMM model used for the Markovian dynamics requires specification of a prior for each transition probability. Let $g(n)$ be the group assigned to time series n. In our MTFS EEG application, time series n is a single recorded epoch and $g(n)$ is the participant recorded during that epoch. For time series group g we define a random variable that is proportional to the transition probability from state j to k as $\eta_{jk}^{(g)}|\gamma, \kappa \sim \text{gamma}(\gamma + \kappa\delta(j,k), 1)$, where $\delta(j,k) = 1$ if $j = k$ and $\delta(j,k) = 0$ otherwise. Then γ controls the number of pseudo-observations the prior assigns to each possible state transition. The second parameter κ then controls the amount of stickiness in the HMM model by determining the number of extra pseudo-observations to assign to self transitions. We use gamma priors on both of these parameters so that $\gamma \sim \text{gamma}(a_\gamma, b_\gamma)$ and $\kappa \sim \text{gamma}(a_\kappa, b_\kappa)$. In our models we set $a_\gamma = 1, b_\gamma = 1, a_\kappa = 1$, and $b_\kappa = 0.01$.

4.3.2 Feature Model

Our GBP-AR-HMM model considers an infinite number of potential features (microstates), but during a single stage in the sampler there are only K available features. We keep track of whether each available feature is turned on or off for each time series in a feature matrix F, where $f_k^{(g)} = 1$ if the k-th feature is selected by the g-th time series group and $f_k^{(g)} = 0$ otherwise. The vector of feature indicators for the g-th time series group is denoted as $f^{(g)}$. The feature selection process is controlled by the two-parameter IBP model discussed earlier. We place gamma priors on both hyper parameters so that $\alpha \sim \text{gamma}(a_\alpha, b_\alpha)$ and $c \sim \text{gamma}(a_c, b_c)$. In both our MTFS analysis and our simulation studies we set $a_\alpha = 1, b_\alpha = 1, a_c = 1,$ and $b_c = 1$.

Again, note that the time series groups must be specified before fitting the model. In the case of our MTFS data, we fit a single twin pair at a time and specify two time series groups, each containing the 70 epochs from one of the twins.

As previously mentioned, one can equivalently view the Beta process Bernoulli process model as a countably infinite number of coin flips for each time series group to determine which features are selected. Given this view of the model we draw from a beta process with base measure B_0, $B \sim BP(c, B_0)$ to get $B = \sum_{k=1}^{\infty} \omega_k \delta_{\theta^{(k)}}$. We set B_0 to a uniform distribution on the unit interval. To connect this generative process to the two-parameter IBP it is crucial to notice that $\sum_{k=1}^{\infty} \omega_k = \alpha$. The ω_k represent the coin flipping probabilities and they sum to α, resulting in an average of α features selected for each time series group.

Because not all features are available to all of the time series groups in this feature model, we most translate each transition distribution into a feature constrained transition distribution. Let $\eta^{(g)}$ be the matrix containing the unconstrained transition distributions, $\eta_{jk}^{(g)}$. If we denote the column vector containing $\eta_{jk}^{(g)}$ for all k and a single given j as $\eta_j^{(g)}$, then the feature constrained transition distribution for time series group g and state j is $\pi_j^{(g)} = \eta_j^{(g)} f^{(g)} / (\sum_k^{\infty} \eta_{jk}^{(g)} f^{(g)})$. Given $f^{(g)}, \gamma$, and κ, this specification for π is equivalent to $\pi_j^{(g)} | f^{(g)}, \gamma, \kappa \sim Dir([\gamma, \ldots, \gamma, \gamma + \kappa, \gamma, \ldots] f^{(g)})$.

4.3.3 The GBP-AR-HMM Model

We now put the parts of the GBP-AR-HMM model together to describe the full generative model.

1. Draw from beta process with base measure B_0, $B \sim BP(c, B_0)$ to get $B = \sum_{k=1}^{\infty} \omega_k \delta_{\theta^{(k)}}$ where δ is the Kronecker delta

2. For each time series group $g \in 1, \ldots, G$:

   (a) Draw feature vector $f^{(g)} | B \sim BeP(B)$

(b) Draw feature constrained transition distributions $\pi_k^{(g)}|f^{(g)} \sim Dir([\gamma, \ldots, \gamma, \gamma+\kappa, \gamma, \ldots]f^{(g)})$ for each feature k such that $f_k^{(g)} = 1$

(c) For each time series n with $g(n) = g$

i. For time $t \in \{1, \ldots, T\}$ draw the state sequence $z_t^{(n)}|z_{t-1}^{(n)} \sim \pi^{(g)}_{z_{t-1}^{(n)}}$

ii. For time $t \in \{1, \ldots, T\}$ draw observations $\mathbf{y}_t^{(n)}|z_t^{(n)} \sim \mathcal{N}(\sum_{r=1}^{R} \mathbf{A}_r^{(z_t^{(n)})}\mathbf{y}_{t-r}^{(n)}, \mathbf{\Sigma}^{(z_t^{(i)})})$

Note that the grouping of time series allows us to share information about microstates and their dynamics between epochs from the same participant.

4.3.4 MCMC Sampling Algorithm

Our MCMC sampling algorithm largely follows the algorithm by Fox et al. (2014) for the BP-AR-HMM model, and so for a more detailed description of posterior derivations and sampling steps the reader is referred to the supplemental material from Fox et al. (2014). Here we give a general overview of the MCMC sampler and discuss the changes required to adapt the sampler from the BP-AR-HMM to our proposed GBP-AR-HMM model.

The MCMC procedure takes the following steps:

1. For each time series group, g, sample the feature vector $f^{(g)}$.

2. For each time series, n, sample the state, $z_t^{(n)}$, at each time point, t, from available features k with $f_k^{(g)} = 1$.

3. For each time series group, g, perform a birth-death step on features that are unique to that group and sample new state-sequences to match proposed birth or death if accepted.

4. Perform a split-merge step to attempt to join two features into a single feature or split a single feature into two features and sample new state-sequences if the move is accepted.

5. For each available feature (i.e., state), k, sample new emission parameters, $\mathbf{A}^{(k)}$ and $\mathbf{\Sigma}^{(k)}$, from the MNIW conjugate form.

6. For each time series group g, sample new transition distribution parameters $\eta^{(g)}$.

7. Update the HMM and IBP hyper-parameters γ, κ, α, and c.

Note that, compared to the MCMC procedure of Fox et al. (2014) for the BP-AR-HMM model, our MCMC procedure for the GBP-AR-HMM model considers the likelihood of all time series within a group in steps 1, 3, and 6. Indeed, in the non-grouped model, each time series has its own transition distribution and feature allocation, so these steps are performed for each time series instead of each time series group. On the other hand, in the grouped model proposed here, all time series within the same group share transition distributions and feature allocations.

Through the use of split-merge and birth-death types of moves, this sampler is able to efficiently move through the large parameter space (Dahl, 2005; Jain & Neal, 2004). The use of annealing during burn-in ensures that the model quickly finds a state space with high posterior density. With long time series, computational underflow becomes an issue, necessitating the steps of Scott (2002) to ensure the stability of the sampler. R code for the methods proposed in this paper may be found at `https://github.com/mfiecas/GBP`.

4.4 Simulation Study

4.4.1 Data Simulation process

We considered four different simulation scenarios, each with a different data size. In Scenario 1, each simulated participant had 50 epochs with 128 time points each. Scenario 2 also had 50 epochs but only 64 time points in each epoch. Scenarios 3 and

4 only had 20 epochs with 128 and 64 time points respectively. The four scenarios allow us to examine the performance of the model on EEG datasets of differing sizes.

In each scenario data was simulated for 200 twin pairs. A true library of 4 microstates was designed so that $\mathbf{A}^{(k)}$ and $\mathbf{\Sigma}^{(k)}$ mimicked output from fitting the model to the real MTFS data. The state self-transition probabilities for each twin pair were simulated from a bivariate normal distribution with the MZ twins having a higher correlation than the DZ twins. Each simulated participant was then randomly selected to have one of four different transition matrix structures. The first transition matrix visited all four microstates, the second visited only the first three, the third visited the first, second, and fourth microstates, and the final transition matrix structure only visited the first two microstates. Randomly selecting from these four different transition matrix structures resulted in twin pairs with a potentially different number of states and different overlaps in their state spaces. Together the self-transition probability and the transition matrix structure completely define the transition matrix for each simulated participant. Given the data length, true states, and true transition matrices, data of the appropriate length can be simulated for each of the 200 twin pairs in each scenario. The emission parameters and transition matrix structures used to simulate the data are shown in Table C.1 and C.2 in Appendix **??**.

4.4.2 Model Comparison

We compare the results from using our model with those from fitting the BP-AR-HMM model (Fox et al., 2014) to the set of all epochs from each twin pair without grouping the epochs from each participant. To assess the classification accuracy of the modeling results we needed a way to partition the time points into the correct number of states at each iteration. To accomplish this we turned to the sequentially-allocated latent structure optimization (SALSO) algorithm from the sdols R package (Dahl & Müller, 2017). The SALSO algorithm attempts to find a posterior mode in Bayesian latent state models, and thus provides a method for obtaining a best estimated state

sequence based on the model output. We restricted the SALSO algorithm to the number of true microstates when summarizing both the GBP-AR-HMM and BP-AR-HMM model results. While SALSO is a useful tool for summarizing the output from this model and comparing simulation study results, it is not necessary and could be replaced with another method to select a posterior mode for the state sequence.

To compare model performance, we calculated the percent of time points correctly classified given the SALSO optimal state sequence across all simulated participants. The results are reported for the time series overall and for each of the four true microstates. We also compared the mean squared error (MSE) of the estimated emission parameters, $\theta^{(k)} = \{\mathbf{A}^{(k)}, \mathbf{\Sigma}^{(k)}\}$, averaged across all simulated participants in each scenario. The results presented in Table 4.1 are a percentage improvement in MSE relative to the BP-AR-HMM model results, so a value of 0.6 in the GBP column signifies that the average MSE for the GPB model was 60% lower than the average MSE for the BP model fit to the same scenario. Finally, Table 4.1 displays the percent of the posterior samples that correctly identify the number of true microstates, the MSE in the estimated true number of microstates and the MSE of the average duration of a microstate.

4.4.3 Simulation Study Results

The results in Table 4.1 show the GBP-AR-HMM model to be superior to the BP-AR-HMM model in all four scenarios in terms of classification rate. The non-grouped BP model particularly struggled to classify microstates three and four, which occurred less often in the data. In Scenarios 2 and 4, where the time series were each 64 time points, the BP model missed on nearly all of the time points in these microstates while the GBP model maintained its relatively high level of performance. Interestingly, neither model showed a clear trend of increasing performance with increasing data and even saw slight decreases in performances in many cases.

The GBP model also showed much lower average MSE in emission parameters

	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	GBP	**BP**	**GBP**	**BP**	**GBP**	**BP**	**GBP**	**BP**
Overall Classification %	58.8	47.4	59.5	42.7	62.8	49.2	61.0	43.3
State 1 Classification %	61.4	56.4	61.8	52.2	72.9	51.7	72.0	53.2
State 2 Classification %	50.6	48.0	52.6	53.5	51.6	52.4	52.0	52.7
State 3 Classification %	74.4	37.0	72.2	1.5	70.0	45.2	65.7	4.6
State 4 Classification %	67.3	19.8	66.3	2.1	61.0	29.8	48.5	1.2
Relative $\theta^{(1)}$ Average MSE	0.61	-	0.61	-	0.50	-	0.56	-
Relative $\theta^{(2)}$ Average MSE	0.60	-	0.83	-	0.58	-	0.81	-
Relative $\theta^{(3)}$ Average MSE	0.81	-	0.85	-	0.68	-	0.72	-
Relative $\theta^{(4)}$ Average MSE	0.23	-	0.34	-	0.32	-	0.33	-
% K Correct	1.5	6.0	5.0	44.5	44.0	24.0	56.0	44.0
K MSE	411.9	20.6	179.1	0.6	26.6	15.2	6.6	0.7
Microstate Duration MSE	36	253,640	7	468,568	4	37,820	24	91,367

Table 4.1: Results from fitting the GBP-AR-HMM and BP-AR-HMM models to each of the four different simulation scenarios. The Relative $\theta^{(k)}$ Average MSE lines represent the MSE in the emission parameters averaged across all simulated twin pairs in a given scenario as a percent improvement in the average MSE compared to the BP model. Because the BP model represents the reference group, the relative MSE in those columns is filled by a hyphen (-). % K correct is the percent of posterior samples that correctly identified the number of true microstates, K MSE is the mean squared error in the estimated number of microstates.

for all four true microstates and in all four simulation scenarios compared to the BP model. The average MSE was reduced by at least 23% for the GBP model compared to the BP model, reaching an 85% improvement for microstate 3 in Scenario 2. While the 85% reduction may not be surprising given the BP model's inability to identify time points in microstate 3, the emission parameters for microstate 2 saw a similarly dramatic improvement from the BP model to the GBP model despite similar classification rates across the two models. We again note that neither model improved as the size of the data increased.

The one area where the BP model has an advantage over the GBP model is in the identification of the true number of microstates. The BP model tends to underestimate the number of microstates and thus sees improved performance in these simulations where the true number of microstates is small. The GBP model tends to over-estimate the number of true microstates in attempt to more fully explain the data. As the data size increases the GBP model uses more microstates. This behavior improves the estimation of the emission parameters at the cost of inference on the number of true microstates, an apparent drawback of the model. Because of these findings, we recommend limiting the number of microstates considered in posterior microstate label modes to match the findings in the literature (e.g., between 4 and 13 states). Alternatively, a more informative prior could be put on α to force the model to consider fewer microstates.

Despite the BP model's edge in estimating the number of microstates, its limitations are most apparent when considering the MSE of the average microstate duration. The BP model estimates long segments where the microstate is static in a single state and thus does not mimic the reality of the data. This stems from its tendency to underestimate the number of true microstates, the same trait that appeared to help with identifying the true number of microstates with a limited amount of error. The GBP model is far more able to capture the state switching dynamics compared with the BP model, resulting in much lower microstate duration MSEs.

Figure 4.1 shows the estimated optimal state sequences for a given twin pair along

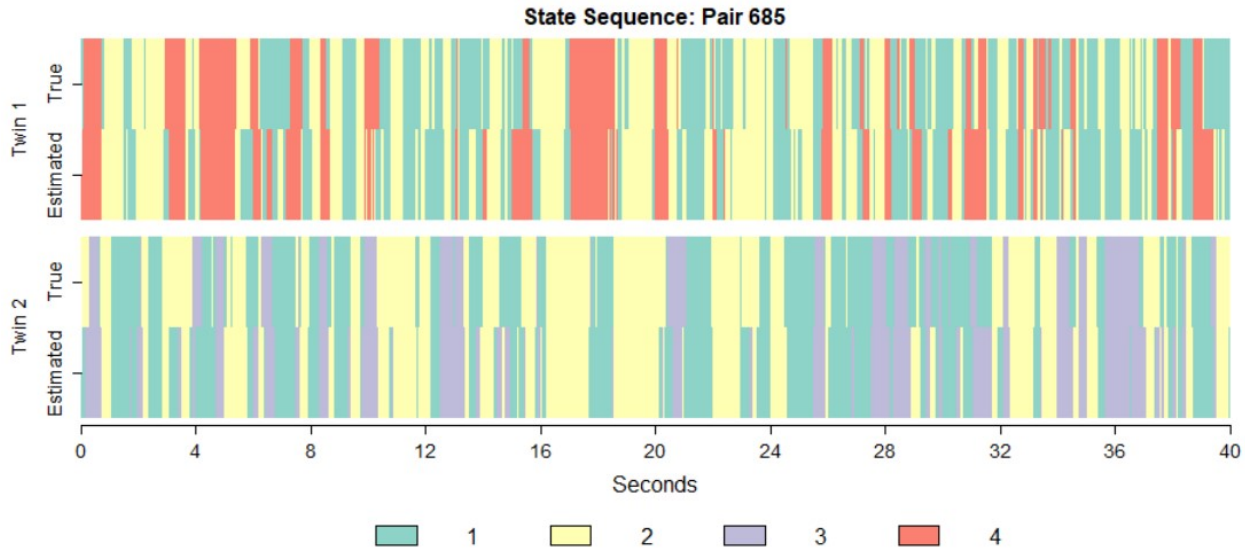


Figure 4.1: Optimal state sequences estimated from the GBP-AR-HMM model output using the SALSO algorithm limited to 4 states for a single simulated twin pair along with the true state sequences for those simulated participants.

with their simulated true state sequences. While the classification isn't perfect, there is a large amount of overlap between the estimated and true state sequences. The model adequately captured the general state distribution, transition distributions, and microstate durations. It was also able to identify the true set of microstates for each participant by estimating that each only has three of the four true microstates.

4.5 MTFS Analysis

4.5.1 Data Description

We applied our method to resting-state EEG data on age-17 female twins collected from 3 electrodes (Cz, O1 and O2) as part of the MTFS. The Cz electrode is located at the center midline on the top of the head and the O1 and O2 electrodes are located on either side of the midline at the rear of the head. The data was collected with a sampling rate of 128 Hz and 12 bits. Considering each epoch as a separate time series and fitting the model at the twin pair level, we applied our method to 70 epochs

of data from each twin with each epoch containing 128 observations collected over 2 seconds. Data for each epoch was demeaned and standardized to unit variance. In total, our sample included 140 seconds of data from 492 adolescents consisting of participants from 167 distinct MZ twin pairs and 79 DZ twin pairs.

While twin participants sat comfortably in a darkened room with their eyes closed, EEG signals were recorded by means of identical Grass 12 Neurodata systems, with a pass band from 1 to 30 Hz (amplifier rolloff, 6 dB/octave). Notes recorded when the data were originally collected guided identification of data that needed to be excluded because of recording problems. Subjects who reported having fallen asleep or were noted to have fallen asleep were excluded. EEG segments containing transient artifacts and excessively small or large voltage deflections were tagged for exclusion by a computer algorithm written in Matlab. Multivariate outliers across the 3 electrodes were identified using a robust version of Mahalanobis distance from the robustbase package in the R statistical programming environment and visually reviewed for contamination by high-frequency noise, other artifacts (e.g., electrocardiogram), or signs of sleepiness. Individual recording sites were excluded from analyses if fewer than 45 2-sec artifact-free sweeps were available. The same priors specified in previous sections were used for the analysis of all 246 twin pairs.

4.5.2 MTFS Results

Table 4.2 shows the results from fitting our model to the MTFS data. The average microstate durations show that most participants exhibited very rapid microstate switching dynamics, only staying in a single microstate for an average of 3.4 time points, which is equivalent to a little over a tenth of a second. Interestingly, MZ twins showed greater within twin pair correlation in their estimated average microstate durations when compared with DZ twins, suggesting the average microstate duration and microstate switching dynamics may be driven in part by genetics. The same cannot be said for the mode of the number of microstates used by each participant,

which were highly correlated within a twin pair but failed to show a difference in correlation between MZ and DZ twins.

Participant Level Parameters	**Median [IQR]**	$\mathbf{r_{MZ}}$	$\mathbf{r_{DZ}}$
Average Microstate Duration	3.4 [2.4, 4.7]	0.54	0.39
Mode Number of Microstates Used	10 [8, 12]	0.86	0.86
Twin Pair Level Parameters	**Median [IQR]**		
Mode Number of Microstates Used	12 [9, 14]		
c	0.05 [0.02, 0.11]		
α	36.5 [28.0, 45.8]		
κ	0.02 [0.01, 0.05]		
γ	0.00004 [0.00002, 0.00020]		

Table 4.2: Results from fitting the GBP-AR-HMM model to each of the 246 female twin pairs in the MTFS data. Results shown are the median and IQR of the estimated parameters at either the individual participant or twin pair level. For the participant level estimates, we have also provided the correlation of these estimates among MZ twins and among DZ twins.

Although our simulation results did not instill confidence in the GBP model's ability to identify to correct number of microstates with data sets as large as the MTFS data, the posterior mode of the number of microstates used by each participant generally fell within the ranges expected after a review of the literature. Although 4 microstates are typically used for microstate analysis, Khanna et al. (2015) showed that between 4 and 13 different microstates have been identified in various microstate EEG analyses. Here we find that half of all participants are estimated to have between 8 and 12 microstates. These results suggest that while the simulations showed the GBP model's tendency to over-estimate the number of true microstates, when applied to the real data the results support the conclusions drawn by other methods. We see the GBP model's ability to capture varying degrees of state space complexity as a strength of the model in comparison with models that are required to set a number of microstates a priori. Note that our results can be reduced to a desired number of states through methods such as SALSO if desired.

The twin pair level parameters also shed light on the behavior of our GBP model.

As expected, the posterior mode of the number of microstates used for both twins in a pair is slightly higher than that of an individual twin. This result suggests some states are unique to each participant within a pair, but the majority of states are shared by both twins. The extremely low estimates of the c parameter also suggest a high amount of overlap in the state spaces of the twins within a pair. As c approaches 0 in the MCMC sampler, the library of state spaces available to a set of twins converges toward a single set for the pair instead of a unique set for each participant. Despite this near complete overlap in state spaces, the estimated number of state spaces used by each participant shows that while all states may have been available to both twins, a participant may have only actually used a subset of the available microstates. This conclusion is supported by the estimates of α, which are roughly three times the number of estimated number of microstates used by a twin pair. The parameter α can be interpreted as the average number of microstates in the finite library of microstates available to a twin pair across the MCMC sampler. We thus conclude that the GBP model has a tendency to select a large library of possible microstates of which only a subset are actually used by the state sequences of the EEG data.

The final two parameters estimated from our MTFS analysis, κ and α are also estimated to be nearly equal to 0. These parameters control the prior on the microstate switching dynamics of the HMM portion of the model. The very low estimates of these parameters suggest that the prior on the state switching dynamics is overwhelmed by the data with time series as long as those in the MTFS, and thus do not contribute to the posterior distribution in a significant way. Therefore, the posterior distribution on the state-switching dynamics is controlled almost entirely by the switches estimated from the data and that the switching dynamics do not exhibit a strong sticky tendency in our dataset.

Figure 4.2 shows the SALSO estimated optimal state sequence for a single MZ twin pair. Similar plots could be created for each twin pair, but we focus on the results from a single twin pair as an example of how the output from the twin pair level models can be summarized and interpreted. Note that because this is resting

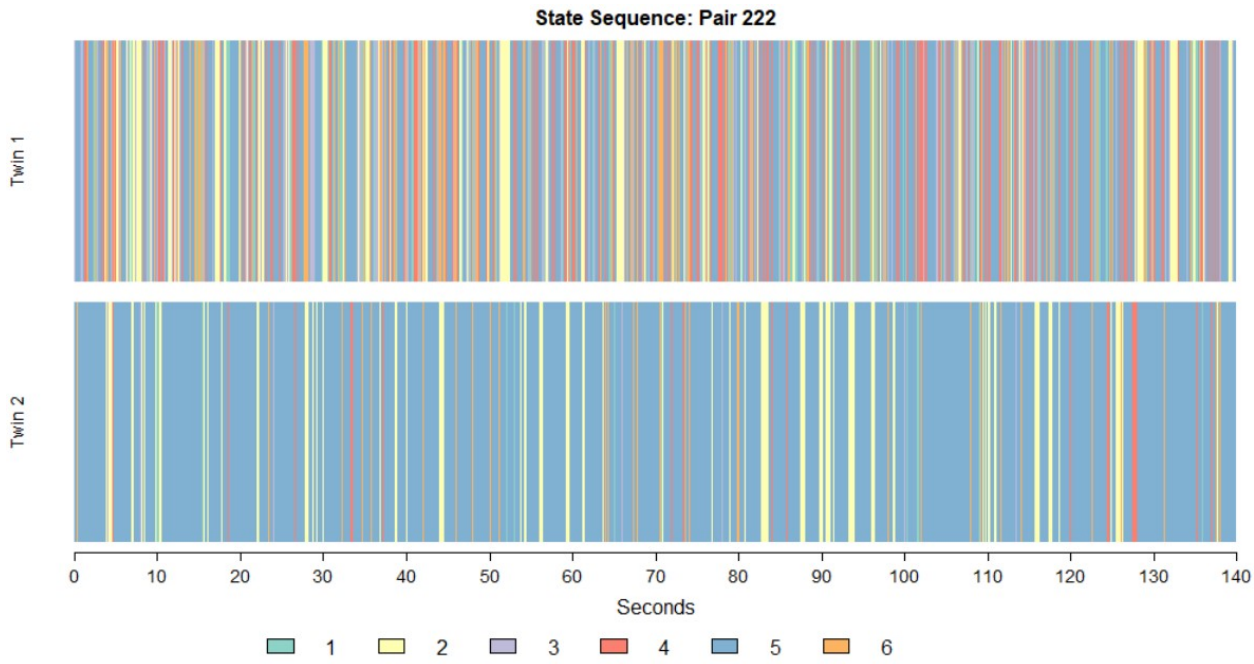
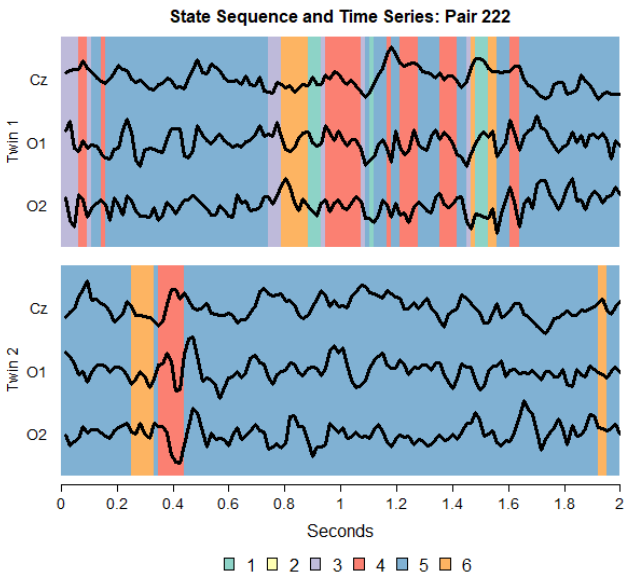


Figure 4.2: Optimal state sequences for twin pair 222 estimated from the GBP-AR-HMM model output using the SALSO algorithm limited to the posterior mode number of states for a single MZ MTFS twin pair.

state data, there is no reason to believe the microstate sequences of the two twins are temporally registered. The state sequence plots are, instead, useful for visualizing the microstate dynamics of the twin pair. Both twins visit all six of the estimated microstates and state five is the most common microstate for each twin. Beyond that, the microstate dynamics of the twins differ greatly. 41.7% of the time points are classified as state 5 for twin 1 compared with 82.1% for twin 2. Additionally, twin 1 has much shorter microstate durations in state five and rapidly transitions among all six states. Twin 2 has very long durations in state five and visits each of the other states for brief stints before switching back to the dominant state five. From this plot, we can see these twins were estimated to share the same set of microstates, justifying our modeling at the twin pair level. At the same time, the switching dynamics are notably different across subjects, supporting the models ability to estimate the HMM transition probability matrices separately for each participant.

Figure 4.3 shows a short segment of the SALSO estimated optimal state sequences



Microstate	$\hat{\mathbf{A}}^{(k)}$	$\hat{\mathbf{\Sigma}}^{(k)}$
1	$\begin{bmatrix} 0.75 & -0.03 & -0.07 \\ 0.23 & 0.40 & 0.24 \\ 0.20 & -0.24 & 0.66 \end{bmatrix}$	$\begin{bmatrix} 0.34 & 0.02 & -0.01 \\ 0.02 & 0.47 & 0.18 \\ -0.01 & 0.18 & 0.75 \end{bmatrix}$
2	$\begin{bmatrix} 0.67 & -0.03 & 0.00 \\ 0.02 & 0.49 & 0.02 \\ -0.01 & 0.00 & 0.50 \end{bmatrix}$	$\begin{bmatrix} 0.63 & 0.03 & -0.04 \\ 0.03 & 0.88 & 0.29 \\ -0.04 & 0.29 & 1.24 \end{bmatrix}$
3	$\begin{bmatrix} 0.71 & 0.09 & 0.00 \\ -0.22 & 0.62 & -0.09 \\ -0.15 & 0.08 & 0.48 \end{bmatrix}$	$\begin{bmatrix} 0.34 & 0.02 & -0.02 \\ 0.02 & 0.48 & 0.17 \\ -0.02 & 0.17 & 0.74 \end{bmatrix}$
4	$\begin{bmatrix} 0.78 & -0.07 & 0.00 \\ 0.04 & 0.75 & -0.34 \\ 0.02 & 0.31 & 0.40 \end{bmatrix}$	$\begin{bmatrix} 0.36 & 0.01 & -0.01 \\ 0.01 & 0.35 & 0.12 \\ -0.01 & 0.12 & 0.49 \end{bmatrix}$
5	$\begin{bmatrix} 0.78 & 0.00 & 0.02 \\ -0.04 & 0.61 & -0.03 \\ -0.04 & 0.01 & 0.61 \end{bmatrix}$	$\begin{bmatrix} 0.38 & 0.02 & -0.02 \\ 0.02 & 0.49 & 0.14 \\ -0.02 & 0.14 & 0.56 \end{bmatrix}$
6	$\begin{bmatrix} 0.81 & 0.07 & 0.02 \\ -0.26 & 0.40 & 0.29 \\ -0.16 & -0.26 & 0.69 \end{bmatrix}$	$\begin{bmatrix} 0.33 & 0.00 & -0.01 \\ 0.00 & 0.40 & 0.14 \\ -0.01 & 0.14 & 0.62 \end{bmatrix}$

Figure 4.3: A two second segment of the optimal state sequences for twin pair 222 shown in Figure 4.2 with the trivariate EEG time series from pair 222 overlaid on top of the state sequence. The table on the right shows the estimated emission parameter for these microstates conditional on the SALSO state labels.

for the same twin pair shown in Figure 4.2. The figure also has the original trivariate EEG time series overlaid on the estimated state sequences along with the estimated emission parameters for each of the six microstates in the SALSO state sequences. The dominance of State 5 is again apparent in these smaller sections of the estimated state sequences. In state 5, the three EEG dimensions are largely desynchronized (i.e., uncorrelated) as the off diagonals of both $\mathbf{A}^{(5)}$ and $\mathbf{\Sigma}^{(5)}$ are near zero. The other five states then correspond to different levels and patterns of synchronization between the three EEG recording channels, primarily between O1 and O2 which are located symmetrically on either side at the back of the head. While similar patterns of synchronization may appear in other twin pairs, we emphasize that this analysis only examines a single twin pair as a proof of the utility of the output of our model.

4.6 Conclusion

We have developed a novel EEG microstate model for resting-state EEG data collected on twins. Our model accounts for the fact that multiple epochs collected from the same participant will likely share the same set of microstate dynamics and also recognizes that some microstates may be shared by twins. We use a non-parametric Bayesian framework to model the state space in our MS-VAR model. When compared to the BP-AR-HMM model, which does not jointly model the transition matrices of epochs from the same participant, via a simulation study our GBP-AR-HMM model showed superior classification rates, emission parameter estimation, and microstate duration estimation while struggling to estimate the true number of microstates.

We also fit our GBP-AR-HMM model to 246 female twin pairs from the MTFS. The estimated parameters suggest a large overlap in the state spaces of many twin pairs and an estimated number of true microstates which falls within the range typically found in other studies. Closer examination of output from the model through algorithms such as SALSO allows for easy visualization of the microstate dynamics of each twin pair. We also found the average microstate duration had a correlation of 0.54 between MZ twins compared to only 0.39 for DZ twins, suggesting the heritability of microstate duration. This finding supports the recent findings of Vidaurre et al. (2017) of the heritability of EEG microstates. While it is beyond the scope of this paper, future work will tie these heritable traits with behavioral characteristics with the overall goal of tying genes, brain, and behavior.

Our GBP-AR-HMM model offers a nice extension of the BP-AR-HMM model which better fits the twin structure of our data, but there are still several potential avenues for further extension. The HMM model could be replaced with a hidden semi-Markov model which would allow for more flexible modeling of microstate duration (Johnson & Willsky, 2013). Such a change would come at the cost of computational time, which may be prohibitive given the large size of the MTFS and other resting-state EEG datasets. Variational Bayes approaches to inference, such as Hughes et al.

(2015), could also be applied to our model to speed up inference and allow the model to be fit to even larger datasets.

Chapter 5

Bibliography

Allen, E., Damaraju, E., Eichele, T., Wu, L., & Calhoun, V. D. (2018). Eeg signatures of dynamic functional network connectivity states. *Brain topography*, *31*, 101–116.

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, *24*, 663–676.

Amemiya, T. (1977). A note on a heteroscedastic model. *Journal of Econometrics*, *6*, 365–370.

Andreou, C., Faber, P. L., Leicht, G., Schoettle, D., Polomac, N., Hanganu-Opatz, I. L., Lehmann, D., & Mulert, C. (2014). Resting-state connectivity in the prodromal phase of schizophrenia: insights from eeg microstates. *Schizophrenia research*, *152*, 513–520.

Barrientos, A. F., Jara, A., & Quintana, F. A. (2017). Fully nonparametric regression for bounded data using dependent bernstein polynomials. *Journal of the American Statistical Association*, (pp. 1–20).

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.

Betzel, R. F., Erickson, M. A., Abell, M., ODonnell, B. F., Hetrick, W. P., & Sporns, O. (2012). Synchronization dynamics and evidence for a repertoire of network states in resting eeg. *Frontiers in computational neuroscience*, *6*, 74.

Brillinger, D. R. (1973). The analysis of time series collected in an experimental design. In *Multivariate Analysis–III* (pp. 241–256). Elsevier.

Brodbeck, V., Kuhn, A., von Wegner, F., Morzelewski, A., Tagliazucchi, E., Borisov, S., Michel, C. M., & Laufs, H. (2012). Eeg microstates of wakefulness and nrem sleep. *Neuroimage*, *62*, 2129–2139.

Bruce, S. A., Hall, M. H., Buysse, D. J., & Krafty, R. T. (2017). Conditional adaptive bayesian spectral analysis of nonstationary biomedical time series. *Biometrics*, .

Cadonna, A., Kottas, A., & Prado, R. (2018). Bayesian spectral modeling for multiple time series. *Journal of the American Statistical Association*, *0*, 1–38.

Chase, A. (2014). Altered functional connectivity in preclinical dementia. *Nature Reviews Neurology*, *10*, 609.

Chen, P.-Y., Chiou, J.-M., Yang, Y.-F., Chen, Y.-T., Hsieh, H.-L., Chang, Y.-L., & Tseng, W.-Y. I. (2016). Heterogeneous aging effects on functional connectivity in different cortical regions: A resting-state functional mri study using functional data analysis. *PLOS ONE*, *11*, e0162028.

Choudhuri, N., Ghosal, S., & Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, *99*, 1050–1059.

Chung, E., & Romano, J. P. (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics*, *41*, 484–507.

Dahl, D. B. (2005). Sequentially-allocated merge-split sampler for conjugate and non-conjugate dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, *11*, 6.

Dahl, D. B., & Müller, P. (2017). *sdols: Summarizing Distributions of Latent Structures*. R package version 1.6.

Damaraju, E., Allen, E. A., Belger, A., Ford, J., McEwen, S., Mathalon, D., Mueller, B., Pearlson, G., Potkin, S., Preda, A. et al. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *NeuroImage: Clinical*, *5*, 298–308.

Diggle, P. J., & Al Wasel, I. (1997). Spectral analysis of replicated biomedical time series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *46*, 31–71.

Edwards, M. C., Meyer, R., & Christensen, N. (2017). Bayesian nonparametric spectral density estimation using b-spline priors. *Statistics and Computing*, (pp. 1–12).

Falconer, D. S. (1960). *Introduction to quantitative genetics*. Oliver And Boyd; Edinburgh; London.

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

Fiecas, M., & Ombao, H. (2016). Modeling the evolution of dynamic brain processes during an associative learning experiment. *Journal of the American Statistical Association*, *111*, 1440–1453.

Finn, E. S., & Constable, R. T. (2016). Individual variation in functional brain connectivity: implications for personalized approaches to psychiatric disease. *Dialogues in Clinical Neuroscience*, *18*, 277–287.

van der Flier, W. M., Pijnenburg, Y. A., Fox, N. C., & Scheltens, P. (2011). Early-onset versus late-onset alzheimer's disease: the case of the missing apoe 4 allele. *The Lancet Neurology*, *10*, 280–288.

Fox, E., Sudderth, E., Jordan, M., & Willsky, A. (2010). Bayesian nonparametric methods for learning markov switching processes. *IEEE Signal Processing Magazine*, *27*, 43–54.

Fox, E., Sudderth, E. B., Jordan, M. I., & Willsky, A. S. (2011a). Bayesian nonparametric inference of switching dynamic linear models. *IEEE Transactions on Signal Processing*, .

Fox, E. B., Hughes, M. C., Sudderth, E. B., Jordan, M. I. et al. (2014). Joint modeling of multiple time series via the beta process with application to motion capture segmentation. *The Annals of Applied Statistics*, *8*, 1281–1313.

Fox, E. B., Sudderth, E. B., Jordan, M. I., & Willsky, A. S. (2011b). A sticky hdp-hmm with application to speaker diarization. *The Annals of Applied Statistics*, (pp. 1020–1056).

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, *8*, 700–711.

Friston, K., Frith, C., Liddle, P., & Frackowiak, R. (1993). Functional connectivity: the principal-component analysis of large (pet) data sets. *Journal of Cerebral Blood Flow & Metabolism*, *13*, 5–14.

Fu, Y., Ma, Z., Hamilton, C., Liang, Z., Hou, X., Ma, X., Hu, X., He, Q., Deng, W., Wang, Y. et al. (2015). Genetic influences on resting-state functional networks: A twin study. *Human brain mapping*, *36*, 3959–3972.

Ganjgahi, H., Winkler, A. M., Glahn, D. C., Blangero, J., Kochunov, P., & Nichols, T. E. (2015). Fast and powerful heritability inference for family-based neuroimaging studies. *NeuroImage*, *115*, 256–268.

Ge, T., Holmes, A. J., Buckner, R. L., Smoller, J. W., & Sabuncu, M. R. (2017). Heritability analysis with repeat measurements and its application to resting-state functional connectivity. *Proceedings of the National Academy of Sciences*, *114*, 5521–5526.

Gertheiss, J., Goldsmith, J., Crainiceanu, C., & Greven, S. (2013). Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics*, *14*, 447–461.

Ghahramani, Z., & Griffiths, T. L. (2006). Infinite latent feature models and the indian buffet process. In *Advances in neural information processing systems* (pp. 475–482).

Gonzalez-Castillo, J., & Bandettini, P. A. (2018). Task-based dynamic functional connectivity: Recent findings and open questions. *Neuroimage*, *180*, 526–533.

Grandy, T. H., Werkle-Bergner, M., Chicherio, C., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2013). Individual alpha peak frequency is related to latent factors of general cognitive abilities. *Neuroimage*, *79*, 10–18.

Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004). Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 4637–4642.

Guillaume, B., Hua, X., Thompson, P. M., Waldorp, L., Nichols, T. E., Initiative, A. D. N. et al. (2014). Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. *NeuroImage*, *94*, 287–302.

Hafkemeijer, A., van der Grond, J., & Rombouts, S. A. (2012). Imaging the default mode network in aging and dementia. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1822*, 431–441.

Harper, J., Malone, S. M., & Iacono, W. G. (2018). Impact of alcohol use on eeg dynamics of response inhibition: a cotwin control analysis. *Addiction biology*, *23*, 256–267.

Holland, D., Desikan, R. S., Dale, A. M., McEvoy, L. K., Initiative, A. D. N. et al. (2012). Rates of decline in alzheimer disease decrease with age. *PLOS ONE*, *7*, e42325.

Hughes, M. C., Stephenson, W. T., & Sudderth, E. (2015). Scalable adaptation of state complexity for nonparametric hidden markov models. In *Advances in Neural Information Processing Systems* (pp. 1198–1206).

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J. et al. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, *80*, 360–378.

Iacono, W. G., Carlson, S. R., Taylor, J., Elkins, I. J., & McGue, M. (1999). Behavioral disinhibition and the development of substance-use disorders: findings from the minnesota twin family study. *Development and psychopathology*, *11*, 869–900.

Iacono, W. G., Malone, S. M., & Vrieze, S. I. (2017). Endophenotype best practices. *International Journal of Psychophysiology*, *111*, 115–144.

Ishwaran, H., & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, *96*, 161–173.

Jain, S., & Neal, R. M. (2004). A split-merge markov chain monte carlo procedure for the dirichlet process mixture model. *Journal of computational and Graphical Statistics*, *13*, 158–182.

Johnson, M. J., & Willsky, A. S. (2013). Bayesian nonparametric hidden semi-markov models. *Journal of Machine Learning Research*, *14*, 673–701.

Johnson, R. A., & Wichern, D. (2002). *Multivariate analysis*. Wiley Online Library.

Khanna, A., Pascual-Leone, A., Michel, C. M., & Farzan, F. (2015). Microstates in resting-state eeg: current status and future directions. *Neuroscience & Biobehavioral Reviews*, *49*, 105–113.

Koenig, T., & Lehmann, D. (1996). Microstates in language-related brain potential maps show noun-verb differences. *Brain and Language*, *53*, 169–182.

Koenig, T., Prichep, L., Lehmann, D., Sosa, P. V., Braeker, E., Kleinlogel, H., Isenhart, R., & John, E. R. (2002). Millisecond by millisecond, year by year: normative eeg microstates and developmental stages. *Neuroimage*, *16*, 41–48.

Krafty, R. T., Hall, M., & Guo, W. (2011). Functional mixed effects spectral analysis. *Biometrika*, *98*, 583–598.

Krafty, R. T., Rosen, O., Stoffer, D. S., Buysse, D. J., & Hall, M. H. (2017). Conditional spectral analysis of replicated multiple time series with application to nocturnal physiology. *Journal of the American Statistical Association*, *112*, 1405–1416.

Laird, N. (2004). Analysis of longitudinal and cluster-correlated data. In *NSF-CBMS Regional Conference Series in Probability and Statistics* (pp. i–155). JSTOR.

Lehman, L., Johnson, M., Nemati, S., Adams, R., & Mark, R. (2015). 11 bayesian nonparametric learning of switching dynamics in cohort physiological time series: application in critical care patient monitoring. *Advanced State Space Methods for Neural and Clinical Data*, (p. 257).

Lehmann, D., Faber, P. L., Galderisi, S., Herrmann, W. M., Kinoshita, T., Koukkou, M., Mucci, A., Pascual-Marqui, R. D., Saito, N., Wackermann, J. et al. (2005). Eeg microstate duration and syntax in acute, medication-naive, first-episode schizophrenia: a multi-center study. *Psychiatry Research: Neuroimaging*, *138*, 141–156.

Lehmann, D., Ozaki, H., & Pal, I. (1987). Eeg alpha map series: brain microstates by space-oriented adaptive segmentation. *Electroencephalography and clinical neurophysiology*, *67*, 271–288.

Lehmann, D., Strik, W., Henggeler, B., König, T., & Koukkou, M. (1998). Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. visual imagery and abstract thoughts. *International Journal of Psychophysiology*, *29*, 1–11.

Li, Y., Wang, X., Li, Y., Sun, Y., Sheng, C., Li, H., Li, X., Yu, Y., Chen, G., Hu, X. et al. (2015). Abnormal resting-state functional connectivity strength in mild cognitive impairment and its conversion to alzheimers disease. *Neural Plasticity*, *2016*.

Lizio, R., Vecchio, F., Frisoni, G. B., Ferri, R., Rodriguez, G., & Babiloni, C. (2011). Electroencephalographic rhythms in alzheimers disease. *International Journal of Alzheimers disease*, *2011*.

MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA proceedings of the section on Bayesian statistical science* (pp. 50–55). Alexandria, Virginia. Virginia: American Statistical Association; 1999.

Malone, S. M., Burwell, S. J., Vaidyanathan, U., Miller, M. B., McGue, M., & Iacono, W. G. (2014). Heritability and molecular-genetic basis of resting eeg activity: A genome-wide association study. *Psychophysiology*, *51*, 1225–1245.

Melard, G., Paesmans, M., & Roy, R. (1991). Consistent estimation of the asymptotic covariance structure of multivariate serial correlations. *Journal of Time Series Analysis*, *12*, 351–361.

Milz, P., Faber, P. L., Lehmann, D., Koenig, T., Kochi, K., & Pascual-Marqui, R. D. (2016). The functional significance of eeg microstatesassociations with modalities of thinking. *Neuroimage*, *125*, 643–656.

Ombao, H. C., Raz, J. A., Strawderman, R. L., & Von Sachs, R. (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika*, *88*, 1186–1192.

Pedroni, A., Gianotti, L. R., Koenig, T., Lehmann, D., Faber, P., & Knoch, D. (2017). Temporal characteristics of eeg microstates mediate trial-by-trial risk taking. *Brain topography*, *30*, 149–159.

Petrone, S. (1999a). Bayesian density estimation using bernstein polynomials. *Canadian Journal of Statistics*, *27*, 105–126.

Petrone, S. (1999b). Random bernstein polynomials. *Scandinavian Journal of Statistics*, *26*, 373–393.

Phipson, B., Smyth, G. K. et al. (2010). Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Stat Appl Genet Mol Biol*, *9*, Article39–Article39.

Pluta, D., Yu, Z., Shen, T., Chen, C., Xue, G., & Ombao, H. (2018). Statistical methods and challenges in connectome genetics. *Statistics & Probability Letters*, *136*, 83–86.

Ren, P., Lo, R. Y., Chapman, B. P., Mapstone, M., Porsteinsson, A., & Lin, F. (2016). Longitudinal alteration of intrinsic brain activity in the striatum in mild cognitive impairment. *Journal of Alzheimer's Disease*, *54*, 69–78.

Rodriguez, A., Dunson, D. B., & Gelfand, A. E. (2008). The nested dirichlet process. *Journal of the American Statistical Association*, *103*, 1131–1154.

Rosen, O., Stoffer, D. S., & Wood, S. (2009). Local spectral analysis via a bayesian mixture of smoothing splines. *Journal of the American Statistical Association*, *104*, 249–262.

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

Rudo-Hutt, A. S. (2015). Electroencephalography and externalizing behavior: a meta-analysis. *Biological psychology*, *105*, 1–19.

Sakkalis, V. (2011). Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg. *Computers in biology and medicine*, *41*, 1110–1117.

Schlegel, F., Lehmann, D., Faber, P. L., Milz, P., & Gianotti, L. R. (2012). Eeg microstates during resting represent personality differences. *Brain topography*, *25*, 20–26.

Scott, S. L. (2002). Bayesian methods for hidden markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association*, *97*, 337–351.

Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica sinica*, (pp. 639–650).

Shumway, R. H., & Stoffer, D. S. (2010). *Time Series Analysis and its Applications: with R Examples*. Springer Science & Business Media.

Simpson, S. L., Bowman, F. D., & Laurienti, P. J. (2013). Analyzing complex functional brain networks: fusing statistics and network science to understand the brain. *Statistics surveys*, *7*, 1.

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

Simpson, S. L., & Laurienti, P. J. (2016). Disentangling brain graphs: a note on the conflation of network and connectivity analyses. *Brain connectivity*, *6*, 95–98.

Smit, D., Posthuma, D., Boomsma, D., & Geus, E. d. (2005). Heritability of background eeg across the power spectrum. *Psychophysiology*, *42*, 691–697.

Smit, D. J., Wright, M. J., Meyers, J., Martin, N., Ho, Y. Y., Malone, S. M., Zhang, J., Burwell, S. J., Chorlian, D. B., de Geus, E. J. et al. (2017). Genome-wide association analysis links multiple psychiatric liability genes to oscillatory brain activity. *bioRxiv*, (p. 232330).

Sorg, C., Riedl, V., Mühlau, M., Calhoun, V. D., Eichele, T., Läer, L., Drzezga, A., Förstl, H., Kurz, A., Zimmer, C. et al. (2007). Selective changes of resting-state networks in individuals at risk for alzheimer's disease. *Proceedings of the National Academy of Sciences*, *104*, 18760–18765.

Staffaroni, A. M., Brown, J. A., Casaletto, K. B., Elahi, F. M., Deng, J., Neuhaus, J., Cobigo, Y., Mumford, P. S., Walters, S., Saloner, R. et al. (2018). The longitudinal trajectory of default mode network connectivity in healthy older adults varies as a function of age and is associated with changes in episodic memory and processing speed. *Journal of Neuroscience*, *38*, 2809–2817.

Stevens, A., & Kircher, T. (1998). Cognitive decline unlike normal aging is associated with alterations of eeg temporo-spatial characteristics. *European archives of psychiatry and clinical neuroscience*, *248*, 259–266.

Stoffer, D. S., & Shumway, R. H. (2006). Time series analysis and its applications: With r examples.

Supekar, K., Menon, V., Rubin, D., Musen, M., & Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in alzheimer's disease. *PLoS Comput Biol*, *4*, e1000100.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2005). Sharing clusters among related groups: Hierarchical dirichlet processes. In *Advances in neural information processing systems* (pp. 1385–1392).

Ter Braak, C. J. (1992). Permutation versus bootstrap significance tests in multiple regression and anova. In *Bootstrapping and related techniques* (pp. 79–85). Springer.

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

Vidaurre, D., Smith, S. M., & Woolrich, M. W. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, *114*, 12827–12832.

Van de Ville, D., Britz, J., & Michel, C. M. (2010). Eeg microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proceedings of the National Academy of Sciences*, (p. 201007841).

Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, *75*, 122–132.

Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., & Jiang, T. (2007). Altered functional connectivity in early alzheimer's disease: a resting-state fmri study. *Human Brain Mapping*, *28*, 967–978.

Wang, L., Roe, C. M., Snyder, A. Z., Brier, M. R., Thomas, J. B., Xiong, C., Benzinger, T. L., Morris, J. C., & Ances, B. M. (2012). Alzheimer disease family history impacts resting state functional connectivity. *Annals of Neurology*, *72*, 571–577.

Wang, L., Zang, Y., He, Y., Liang, M., Zhang, X., Tian, L., Wu, T., Jiang, T., & Li, K. (2006). Changes in hippocampal connectivity in the early stages of alzheimer's disease: evidence from resting state fmri. *NeuroImage*, *31*, 496–504.

West, M., & Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.

Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för matematik*, *2*, 423–434.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397.

Xiang, J., Guo, H., Cao, R., Liang, H., Chen, J. et al. (2013). An abnormal resting-state functional brain network indicates progression towards alzheimer's disease. *Neural Regeneration Research*, *8*, 2789.

Zhu, D. C., Covassin, T., Nogle, S., Doyle, S., Russell, D., Pearson, R. L., Monroe, J., Liszewski, C. M., DeMarco, J. K., & Kaufman, D. I. (2015). A potential biomarker in sports-related concussion: brain functional connectivity alteration of the default-mode network measured with longitudinal resting-state fmri over thirty days. *Journal of neurotrauma*, *32*, 327–341.

Appendix A

Supplementary Materials for Chapter 2

Convergence:		Full		One-Step		One-Step		LME Local	
Variance:		Full		Full		Reduced		MANCOVA Global	
	Setting	**3 ROI**	**5 ROI**	**3 ROI**	**5 ROI**	**3 ROI**	**5 ROI**	**3 ROI**	**5 ROI**
Main Effect Local Tests	**2**	0.050	0.062	0.050	0.067	0.054	0.066	0.041	0.046
	3	0.054	0.060	0.050	0.057	0.056	0.061	0.049	0.046
	6	0.052	0.057	0.053	0.057	0.053	0.059	0.050	0.047
	7	0.048	0.058	0.046	0.057	0.048	0.057	0.044	0.050
	10	0.061	0.064	0.056	0.062	0.057	0.062	0.044	0.050
	11	0.066	0.065	0.068	0.061	0.066	0.063	0.045	0.047
	14	0.054	0.059	0.057	0.057	0.052	0.057	0.045	0.047
	15	0.056	0.059	0.057	0.057	0.060	0.057	0.056	0.048
Interaction Local Tests	**2**	0.060	0.062	0.060	0.068	0.060	0.063	0.049	0.050
	4	0.072	0.062	0.074	0.068	0.072	0.060	0.050	0.049
	6	0.056	0.054	0.054	0.056	0.056	0.057	0.056	0.046
	8	0.058	0.058	0.054	0.057	0.054	0.058	0.050	0.049
	10	0.058	0.067	0.055	0.065	0.061	0.065	0.048	0.054
	12	0.068	0.060	0.070	0.061	0.066	0.062	0.054	0.046
	14	0.062	0.053	0.064	0.055	0.063	0.053	0.046	0.050
	16	0.064	0.054	0.062	0.053	0.062	0.053	0.048	0.050

Table A.1: Type I error rates for simulation study. Type I errors for the main effect (group difference in intercepts) and interaction effect (group difference in slopes) are reported both globally and locally. The global Type I errors are averaged across all models. The local Type I errors reported are unadjusted and averaged across all simulations and all null ROI pairs.

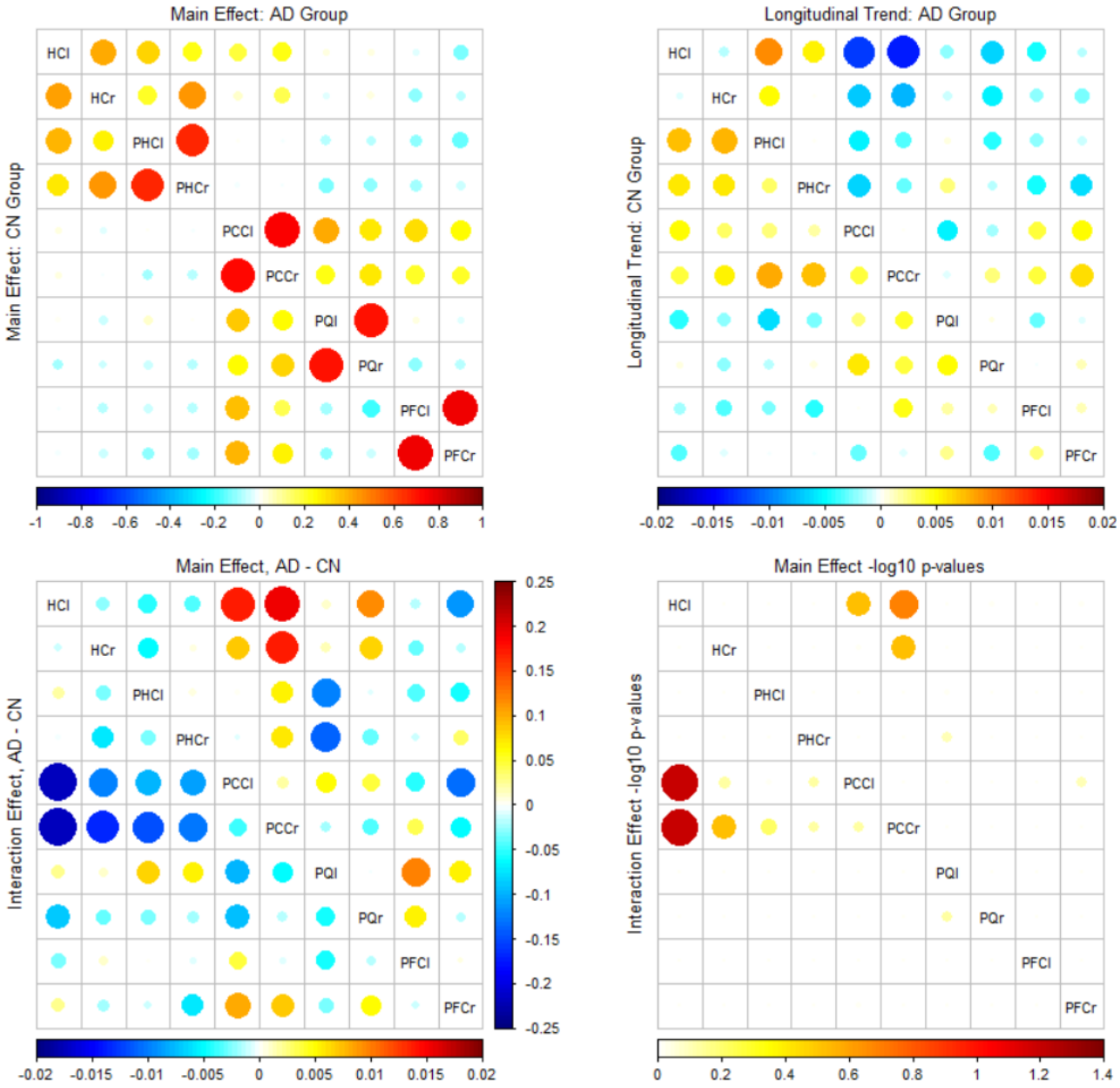


Figure A.1: Model 2 results. Top left: A plot of the estimated intercept terms for the CN group (bottom left triangle) and AD group (top right triangle). Top right: A plot of the estimated slope terms for the CN group (bottom left triangle) and AD group (top right triangle). Bottom left: a plot of the group differences (AD estimates - CN estimates) for the estimated intercepts (top right triangle) and slopes (bottom left triangle). Bottom right: A plot of the $-\log_{10}$ corrected and adjusted p-values from all local hypothesis tests of group differences (AD estimates - CN estimates) for the estimated intercepts (top right triangle) and slopes (bottom left triangle).

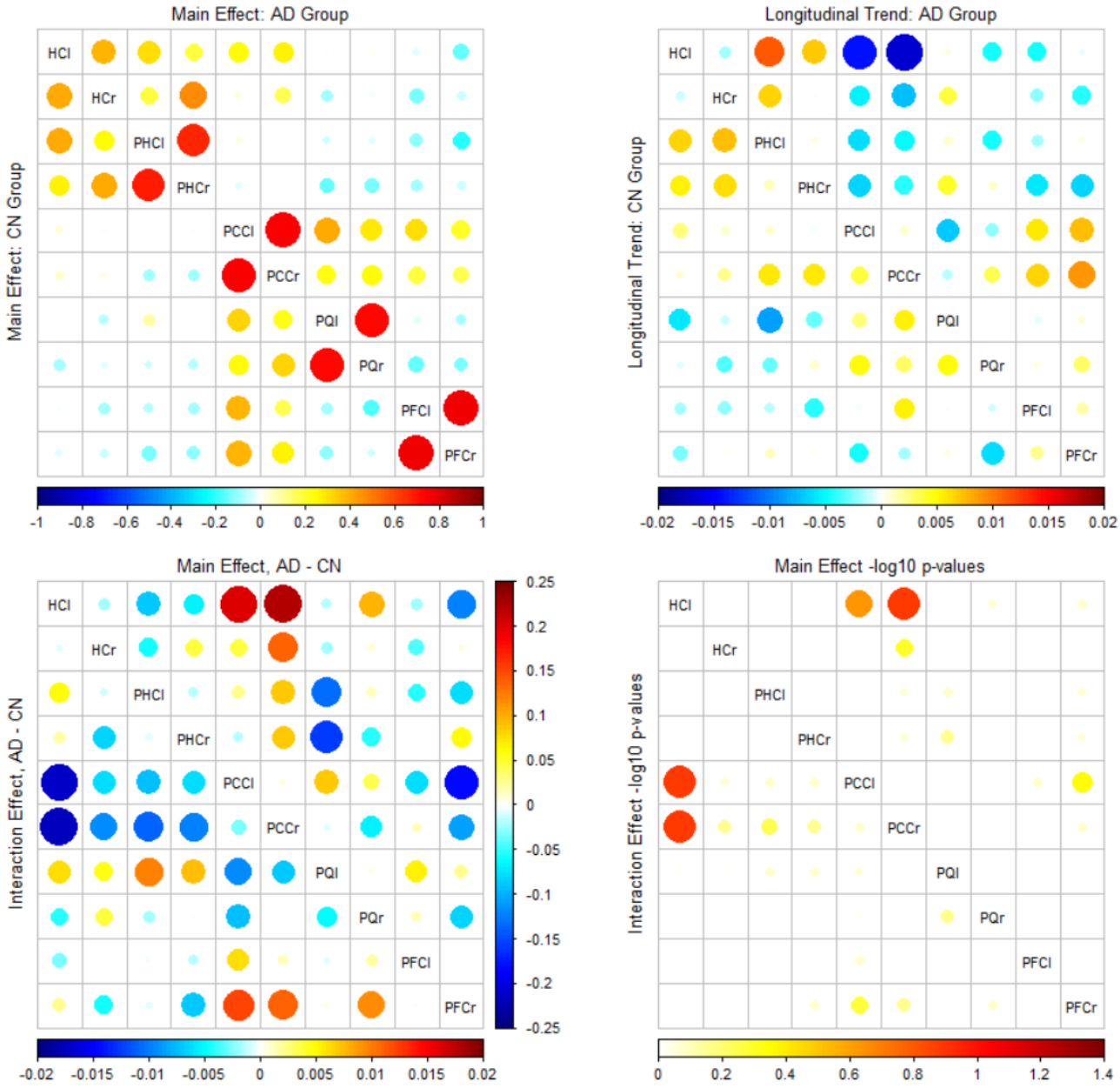


Figure A.2: Model 3 results. Top left: A plot of the estimated intercept terms for the CN group (bottom left triangle) and AD group (top right triangle). Top right: A plot of the estimated slope terms for the CN group (bottom left triangle) and AD group (top right triangle). Bottom left: a plot of the group differences (AD estimates - CN estimates) for the estimated intercepts (top right triangle) and slopes (bottom left triangle). Bottom right: A plot of the $-\log_{10}$ corrected and adjusted p-values from all local hypothesis tests of group differences (AD estimates - CN estimates) for the estimated intercepts (top right triangle) and slopes (bottom left triangle).

Appendix B

Supplementary Materials for Chapter 3

B.1 Sensitivity Analysis

The assess the sensitivity of our model to the truncation levels selected for the nested DP, we performed a sensitivity analysis. We fit a version of the model to the Cz channel data with the upper level truncation level reduced from K=100 to K=75 and a version of the model with the lower level truncation level reduced from L=20 to L=15. The results comparing these two model fits with the original fit where K=100 and L=20 are shown in Figure B.1 and Table B.1. The computation time required to fit the model is linear in both L and K, so the new models with lower truncation levels did run significantly faster. The results wre qualitatively very similar, showing that, beyond a certain point, increasing K and L only increases the computation time and the model is not sensitive to their specific values.

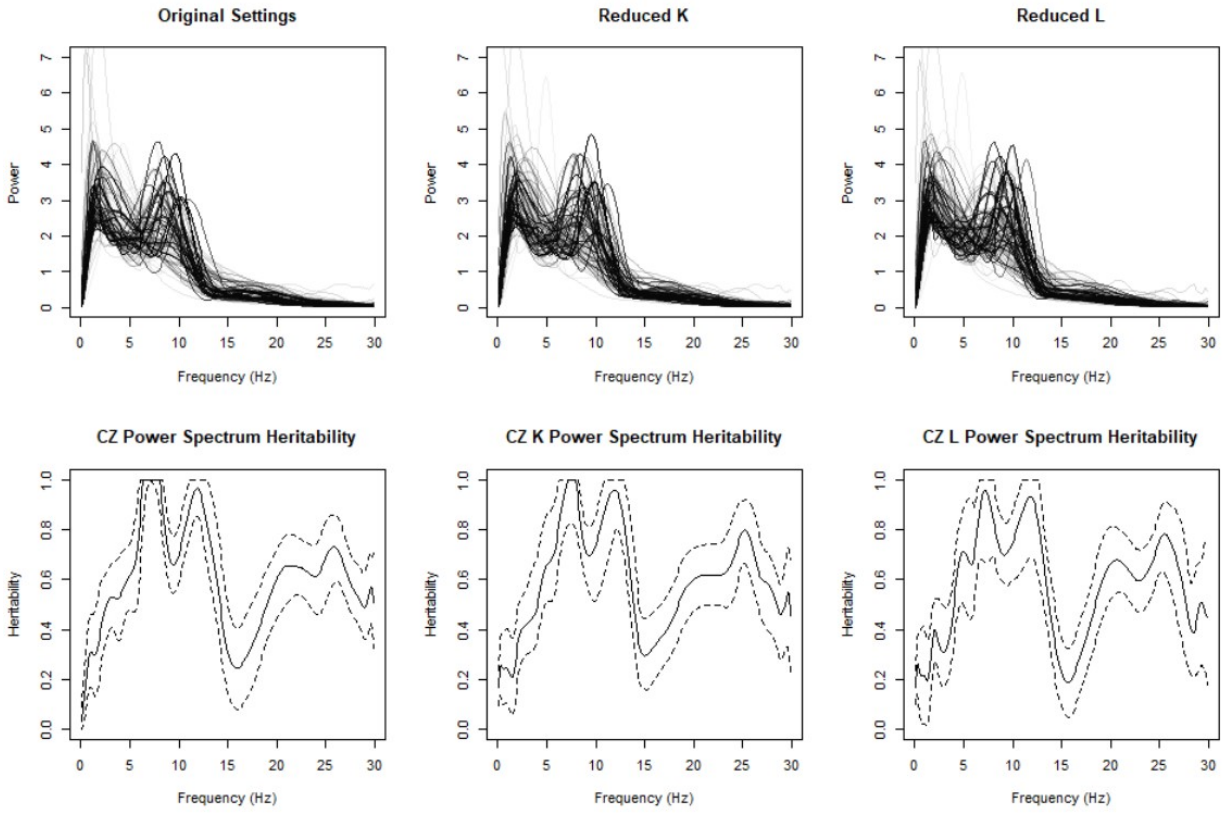


Figure B.1: The top row shows the estimated spectral density curves for the Cz channel for each of the 1116 participants in the MTFS. Each line represents a single subject. The bottom row show the estimated heritability and 95% point-wise credible interval across the power spectrum calculated using Falconer's formula. The left column shows the results from the original analysis with K=100 and L=20, while the middle column shows an the results from the sensitivity analysis with K=75 and L=20 and the right column shows an the results from the sensitivity analysis with K=100 and L=15.

Analysis	Feature	Median (IQR)	Heritability (95% CI)
Original	Delta Band Power (1-4 Hz)	0.25 (0.23 - 0.32)	0.43 (0.33, 0.53)
	Theta Band Power (4-8 Hz)	0.28 (0.25 - 0.31)	0.88 (0.84, 0.92)
	Alpha Band Power (8-12 Hz)	0.29 (0.23 - 0.33)	0.77 (0.69, 0.87)
	Beta Band Power (12-30 Hz)	0.13 (0.12 - 0.18)	0.65 (0.55, 0.77)
	Full Spectrum	1.00	0.68 (0.62, 0.74)
	Peak Frequency	2.25 (1.63 - 8.63)	0.51 (0.35, 0.66)
	Alpha Peak Frequency	8.63 (8.13 - 9.63)	0.81 (0.70, 0.91)
Reduced K	Delta Band Power (1-4 Hz)	0.26 (0.24 - 0.31)	0.35 (0.26, 0.45)
	Theta Band Power (4-8 Hz)	0.28 (0.24 - 0.31)	0.85 (0.76, 0.92)
	Alpha Band Power (8-12 Hz)	0.29 (0.23 - 0.36)	0.79 (0.63, 0.88)
	Beta Band Power (12-30 Hz)	0.13 (0.11 - 0.18)	0.68 (0.60, 0.476)
	Full Spectrum	1.00	0.68 (0.58, 0.74)
	Peak Frequency	7.38 (2.00 - 9.50)	0.50 (0.33, 0.67)
	Alpha Peak Frequency	8.38 (8.13 - 9.75)	0.81 (0.52, 0.96)
Reduced L	Delta Band Power (1-4 Hz)	0.26 (0.24 - 0.32)	0.33 (0.24, 0.43)
	Theta Band Power (4-8 Hz)	0.27 (0.25 - 0.31)	0.82 (0.67, 0.89)
	Alpha Band Power (8-12 Hz)	0.28 (0.22 - 0.35)	0.78 (0.67, 0.85)
	Beta Band Power (12-30 Hz)	0.14 (0.12 - 0.17)	0.63 (0.52, 0.71)
	Full Spectrum	1.00	0.65 (0.56, 0.71)
	Peak Frequency	7.63 (1.88 - 9.25)	0.55 (0.39, 0.72)
	Alpha Peak Frequency	8.75 (8.13 - 9.50)	0.80 (0.49, 0.93)

Table B.1: The median and inter-quartile range of different spectral density features across the 1116 MTFS subjects along with the heritability calculated from the sample for each feature. The orginal analysis represents the fit when K=100 and L=20. The reduced K analysis uses K=75 and the reduced L analysis uses L=15.

Appendix C

Supplementary Materials for Chapter 4

Simulated Microstate	$\mathbf{A}^{(k)}$	$\mathbf{\Sigma}^{(k)}$
1	$\begin{bmatrix} 0.8 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0.35 & 0.00 & 0.00 \\ 0.00 & 0.25 & 0.10 \\ 0.00 & 0.10 & 0.35 \end{bmatrix}$
2	$\begin{bmatrix} 0.9 & -0.3 & 0.0 \\ 0.5 & 0.6 & 0.1 \\ 0.4 & -0.2 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0.30 & 0.10 & 0.00 \\ 0.10 & 0.20 & 0.10 \\ 0.00 & 0.10 & 0.40 \end{bmatrix}$
3	$\begin{bmatrix} 0.7 & 0.1 & -0.1 \\ 0.0 & 1.1 & -0.6 \\ 0.0 & 0.6 & 0.3 \end{bmatrix}$	$\begin{bmatrix} 0.30 & 0.00 & 0.00 \\ 0.00 & 0.20 & 0.10 \\ 0.00 & 0.10 & 0.35 \end{bmatrix}$
4	$\begin{bmatrix} 0.6 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.60 \end{bmatrix}$	$\begin{bmatrix} 0.60 & 0.00 & 0.00 \\ 0.00 & 0.70 & 0.30 \\ 0.00 & 0.30 & 0.80 \end{bmatrix}$

Table C.1: The emission parameters for each of the four microstates used to simulate the data for the simulation study.

Transition Class	$\mathbf{A}^{(k)}$
1	$(1-\rho) * \begin{bmatrix} \rho/(1-\rho) & 0.8 & 0.1 & 0.1 \\ 0.1 & \rho/(1-\rho) & 0.4 & 0.5 \\ 0.5 & 0.5 & \rho/(1-\rho) & 0.0 \\ 0.4 & 0.4 & 0.2 & \rho/(1-\rho) \end{bmatrix}$
2	$(1-\rho) * \begin{bmatrix} \rho/(1-\rho) & 0.7 & 0.3 & 0.0 \\ 0.8 & \rho/(1-\rho) & 0.2 & 0.0 \\ 0.7 & 0.3 & \rho/(1-\rho) & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$
3	$(1-\rho) * \begin{bmatrix} \rho/(1-\rho) & 0.7 & 0.0 & 0.3 \\ 0.8 & \rho/(1-\rho) & 0.0 & 0.2 \\ 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.0 & \rho/(1-\rho) \end{bmatrix}$
4	$(1-\rho) * \begin{bmatrix} \rho/(1-\rho) & 1.0 & 0.0 & 0.0 \\ 1.0 & \rho/(1-\rho) & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$

Table C.2: Given a state self-transition parameter of ρ, each simulated participant was randomly assigned to one of these four true transition matrices with equal probability.