AN APPLICATION OF SEMIPARAMETRIC MODEL ESTIMATION UNDER SHAPE INVARIANCE TO FMRI DATA

by

CHRISTOPHER J. HELMS

(Under the Direction of Cheolwoo Park)

ABSTRACT

Tasks become easier with practice and we all can subjectively assess this. We propose a method for measuring this difference by comparing fMRI activations during saccadic task trials before and after practice sessions. We treat each subject's pre-practice and post-practice activations as functions. We estimate the functions using local polynomial regression with kernel smoothing, establish a shape invariant model for the relevant function estimates and use evolutionary algorithms to estimate shape invariant model parameters. We apply hypothesis testing on the estimates of the shape invariant parameters to decide if attenuation is significant for different practice groups. We see significant attenuation where we expect to see it and have similar results to those of a different method.

INDEX WORDS: Attenuation, Functional magnetic resonance imaging data, Local polynomial smoothing, Semiparametric estimation, Shape invariance

An Application of Semiparametric Model Estimation Under Shape Invariance to FMRI Data

by

Christopher J. Helms

B.S., Mathematics, The University of Georgia, 2014
B.S., Statistics, The University of Georgia, 2014

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Master of Science

Athens, Georgia

2018

An Application of Semiparametric Model Estimation Under Shape Invariance to FMRI Data

by

Christopher J. Helms

Approved:

Major Professor:	Cheolwoo Park
Committee:	Nicole Lazar Jennifer McDowell

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2018

Acknowledgments

Thank you to Lynne Seymour for encouraging me to apply to the University of Georgia Statistics program and all the fun chats, to Nicole Lazar for always pushing me to do my best, to Cheolwoo Park for all your patience and for letting me into a class against your better judgment and also to Jennifer McDowell and Kara Dyckman for providing and explaining so much interesting data. This work was supported by National Science Foundation under Grant No. NSF IIS-1607919

TABLE OF CONTENTS

Page

CHAPTER

APPENDIX

List of Figures

LIST OF TABLES

Chapter 1

Introduction

Magnetic resonance imaging (MRI) is proposed and invented by Paul Lauterbur in 1972-1973 as a non-invasive method of *in vivo* observation images of tissue and structures such as the lungs, brain, etc (Lauterbur, 1973; Huettel et al., 2008). MRI machines have been used to study static structures by emitting magnetic waves into tissue, where it gets absorbed by hydrogen nuclei which then emit energy detected by the MRI machine as in Figure 1.1 (Huettel et al., 2008). Later Ogawa et al. (1990a) and Ogawa et al. (1990b) study changes in deoxygenated hemoglobin in rodent brains by adjusting oxygen concentration and call them blood oxygen level dependent contrasts (BOLD). These changes in the BOLD signals are measurements of local changes in blood flow that correlated with neuronal activity, but not the activity itself (Huettel et al., 2008; Toga and Mazziotta, 2000). Because the BOLD contrasts or signal is measured over time whereas the MRI is static, the functional MRI (fMRI) measures blood flow in the brain over time which presents a four-dimensional set of information about a brain, shown in Figure 1.2.

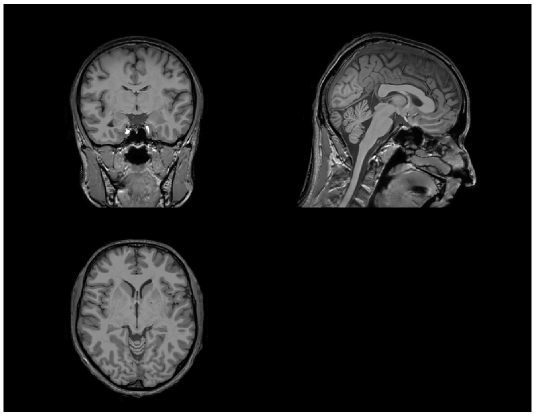
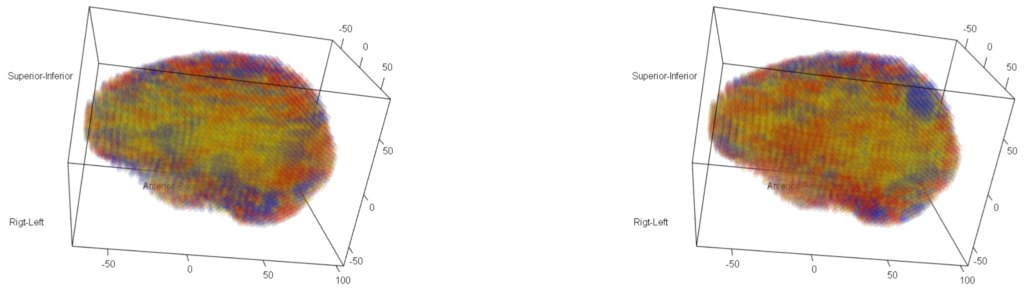
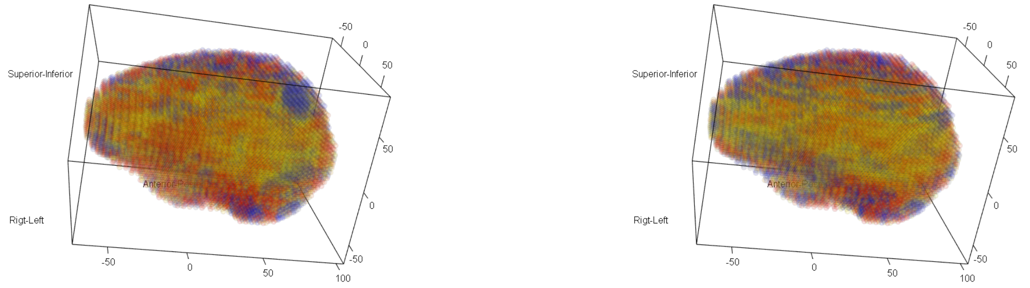


Figure 1.1: Two-dimensional cross-sections of human skull from MRI machine



(a) Time = 1

(b) Time = 6



(c) Time = 11

(d) Time = 16

Figure 1.2: fMRI BOLD Activations in the Brain at Four Times (low to high activation runs from blue to red)

Interest in studying fMRI data is a growing interdisciplinary venture involving psychologists, physicists, statisticians and many more. Researchers are commonly interested in finding locations of activation within the brain that are associated with certain tasks, exploring networks of locations and predicting psychological or diseases conditions. Decisions in all these settings can be accomplished with appropriate statistical techniques (Ashby, 2011). These locations are associated with sets of time series and present significant challenges for statisticians, including various types of noise from sources of heat, mechanical processes, biological processes and motion. Additionally, the recorded data are both temporally and spatially correlated (Lazar, 2008).

Many researchers are interested in comparing changes in a particular location within a group or individual with respect to some sort of treatment, i.e. visual or auditory practice (Panouillres et al., 2012; Jncke et al., 2001; Nelles et al., 2009). Our motivating example is the cognitive control task data studied by Dyckman (2007), Lee et al. (2013),and Lee (2013) where three different practice groups are compared while they perform the antisaccade task (glancing away from a target) before and after a week-long practice. As people get better at a task, the activity in the neural circuitry may decrease as fewer neural resources are necessary for the well-practiced task, called attenuation. Identifying how the brain changes in response to learning could have important implications for understanding how to support or enhance learning. Lee et al. (2013) use a bootstrap approach, and Lee (2013) uses a mixed effect model with feature extraction to test the practice effect, reduction in activation, across different practice groups in various brain regions.

In this work, we propose a semiparametric approach for comparing the different groups of study subjects to quantify changes in activation and assess their statistical significance. In identifying brain regions that show the differences between two sessions, the registration issue could be problematic in fMRI data due to the variation among subjects. The proposed semiparametric approach under shape invariance (Kneip and Engel, 1995) account for this issue by utilizing a horizontal shifting parameter while the main test for the differences of multiple curves is determined by the vertical scale parameter. We capture the amplitude of brain signals for each subject through the scale parameter while estimating the common BOLD signal using nonparametric smoothing. Then, we compare these scale parameters before and after practice for three different groups in pre-determined eleven regions of interest inside the brain.

This thesis is organized so that Chapter 2 includes a discussion of the terminology, collection and structure of the fMRI data. In Chapter 3, we discuss the relevant statistical methods we use in our analysis. In Chapter 4, we discuss our simulation of fMRI data and the results. In Chapter 5, we discuss the analysis of the real fMRI data. In Chapter 6, we discuss conclusions and suggestions for further analysis.

Chapter 2

Data Description

2.1 Introduction to FMRI Data

All of our bodies contain blood, and blood contains hemoglobin which contains iron. Blood flows through the body in a cycle, where it is oxygenated by the lungs, sent out into the body where it becomes deoxygenated and returns to pass through the lungs once again to be oxygenated. It is known that oxygenated blood has different magnetic properties than deoxygenated blood. In particular, iron is paramagnetic, meaning that if it is placed into a magnetic field that is powerful enough, the iron atoms will line up with the magnetic field. It is known that deoxygenated blood has a greater magnetic susceptibility than oxygenated blood. The MRI machine emits a strong magnetic field that resonates with sets of atoms that are affected by the field. The change in the ratio of oxygenated to deoxygenated blood is called a haemodynamic response, and this is the BOLD signal that we measure (Lazar, 2008).

2.2 Experiment

All methods in this section are results of Dyckman (2007)'s research and dissertation. We have the data containing the BOLD activation signals of 37 right-handed University of Georgia (UGA) undergraduate women with a mean age of 19.5 years and standard deviation of 1.8. All participants self-report to be free of psychiatric illnesses, previous head injuries and substance abuse and provide written consent to the UGA Institutional Review Board, approved UGA IRB #1037.

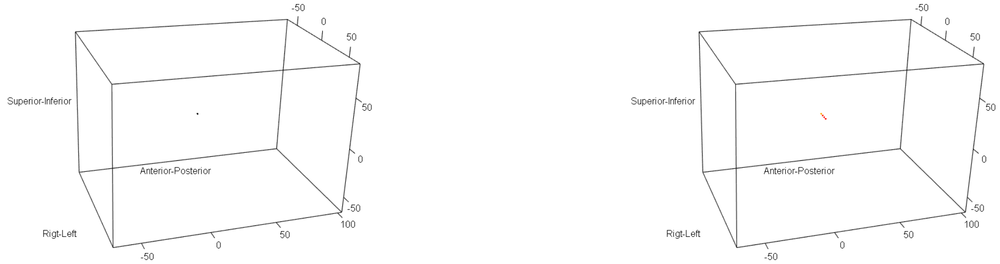
To begin, a subject receives task instructions and then lies on a gurney and is placed inside of a 1.5T GE Signal Horizon LX Scanner (GE Medical Systems, Waukesha, Wisconsin). Her head is stabilized with foam padding and a strap across her forehead. A rear-projection screen is placed near her feet and pointed at a mirror above her head. This mirror allows her to see the what is projected on the screen. An fMRI-compatible iView X System (SensoMotoric Instrument, Inc.; 60 Hz sampling rate) eye tracking system records eye movements. At the beginning of a session, two loacalizer images are taken to guarantee optimal brain coverage. The images are recorded under SPGR-protocol: sagittal, 2 NEX 0.9375 × 1.5 mm, 124 slices, TE 2.8 ms, TR = 10.8 ms, flip angle = 20°, scan time = 5 min 41 sec. Then the functional imaging is done using a spiral scan with 2 interleaves, 24 continuous axial slices, 3.75 × 3.75 × 4 mm, TE = 40 ms, TR = 1912 MS, flip angle = 77° and a scan time of 5 min 1 sec.

The tasks are presented in 13 blocks that are 22.4 sec each and alternate between fixation (maintaining a gaze) and antisaccade tasks. To begin, the subject is presented with purple colored dot in the center of the projection and this lasts 22.4 s. Next a blue dot is presented in place of the purple dot and the subjects fixates on this dot for 200 ms. Next, the projection is blank for 200 ms, then a blue dot manifests in the right or left half of the projection for 1000 ms. During this time, the subject looks at the opposite half of the projection. There are seven blocks of antisaccade tasks.

After the initial run, the subjects are split into three practice groups. Each practice group member practices their assigned task once a day, for four days, excluding weekends. The subjects use training devices called a FixTrain. The antisaccade group have FixTrains programmed to display a star in the center of the screen, 70 ms later a letter "T" is displayed for 100 ms. The antisaccade group is instructed to look away from the star and at the "T", then press an arrow that corresponded with its orientation. The prosaccade (glancing towards a target) group is given FixTrains programmed to display a "T" in the center of the screen and then jump to one side in a different orientation, after which they press an arrow key that corresponds with the new orientation. The fixation group is presented with the "T" in the center of the screen, it changes orientations two to five times and the subject presses an arrow key that corresponds to final orientation of the "T." Once the four practice days are completed, all subjects repeat antisaccade tasks in an MRI machine.

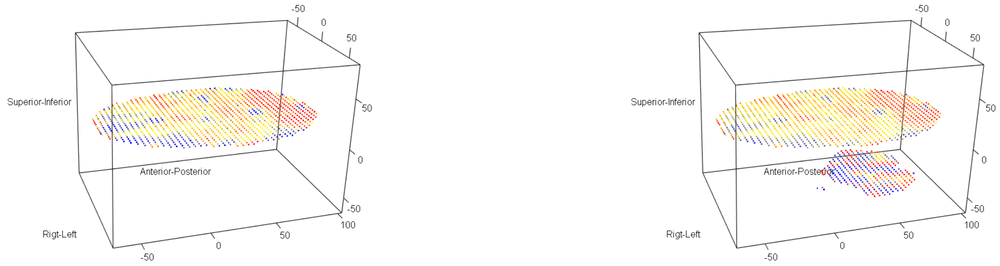
2.3 Structure of Data

One scan of the brain consists of a $40 \times 48 \times 24$ array, or 24 lateral slices. The set of fMRI data for one subject consists of time series of BOLD signals at different points in the brain. The point is called a voxel and shown in Figure 2.1[1]. Each voxel has a BOLD activation time series from $t = t_{start}, \ldots, t_{stop}$.



(a) One Voxel

(b) Four Voxels



(c) One Slice

(d) Two Slices

Figure 2.1: FMRI BOLD activations in the brain for four sets of voxels

We analyze eleven particular regions of interest (ROI) related to cognitive circuitry displayed in Table 2.1 and Figure 2.2 defined by Dyckman (2007). Each ROI is a collection

---

[1] The brain data displayed in here and in the previous chapters are not the brain data we use, but are for demonstration purposes.

of adjacent voxels or collections of adjacent voxels. We want to compare pre-practice and post-practice activations across different practice groups. We use a shape-invariant, semi-parametric method inspired by Kneip and Engel (1995) which is described in the following chapter.

Table 2.1: Regions of interest described by Dyckman (2007).

Region	Abbreviation	Voxel Size
Supplementary eye field	SEF	36
Frontal eye field	FEF	113
Superior parietal lobe	SPL	63
Cuneus	Cuneus	59
Thalamus	Thalamus	56
Inferior parietal lobule	IPL	59
Left PFC	PFC-L	28
Right PFC	PFC-R	36
Striatum (basal & ganglia)	Striatum	67
Middle occipital gyrus	MOG	64
Right inferior frontal cortex	IFC	35

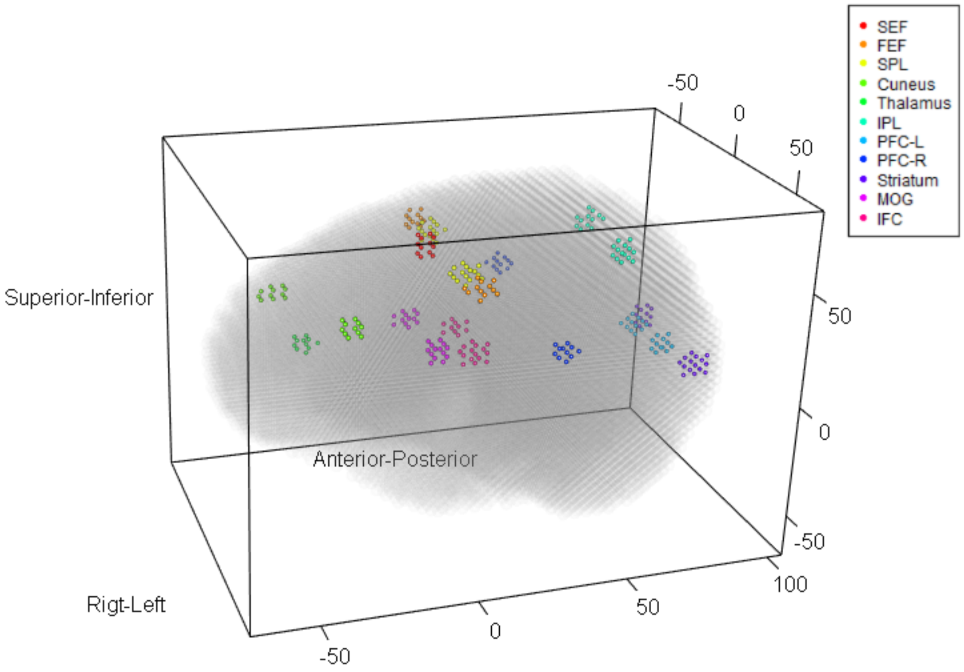


Figure 2.2: Locations of Dyckman (2007)'s regions of interest within the brain.

2.4 MOTIVATING EXAMPLE

In this section, we introduce a motivating example to motivate our research. We consider the supplementary eye field (SEF) region of the brain of the data described in Section 2.3. We take the average of all the signals at each time point over subjects and voxels in that region for pre- and post-practice sessions for three different groups. We plot these average signals for SEF in Figure 2.3. Since the antisaccade practice group practiced the antisaccade task, we expect them to have the most attenuation of the three groups for some regions of the brain, and we see in Figure 2.3 that the post-practice session average activation for the antisaccade group tends to be lower than the pre-practice session average for SEF. This difference appears to be larger for the antisaccade practice group than for the other practice groups. Therefore, the main objective in this work is to determine which groups show the attenuation at the post-practice session in each ROI. In the next chapter, we propose a semiparametric approach that measures and statistically tests these differences.

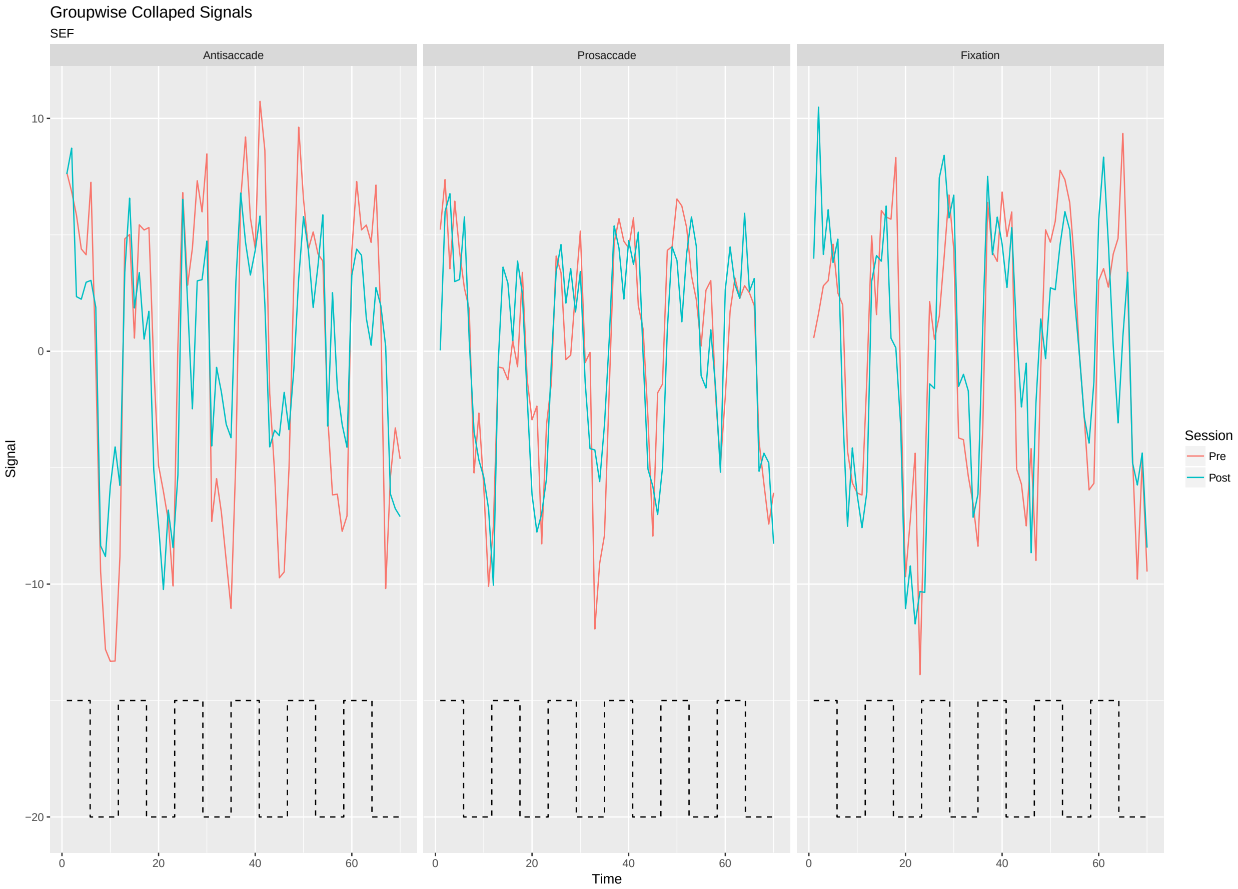


Figure 2.3: Collapsed signals for each group and each session from the supplementary eye field region. The dashed line represents the antisaccade and fixation blocks.

Chapter 3

Methods

In this section, we illustrate the proposed process including detrending, collapsing, local polynomial smoothing, a shape-invariant model in a regression, generalized simulated annealing method, and testing procedures.

3.1 Detrending

Effects of the MRI machine contribute to both a linear and quadratic trend in time-course of a BOLD signal measurement (Lazar, 2008). The methods of centering, linear-only, quadratic-only and quadratic are shown in Equations (3.1a) - (3.1d). We detrend our signals using the residuals $Y_j - \hat{Y}_j$ from Equation (3.1d) (Shumway and Stoffer, 2011):

$$\hat{Y}_j = \hat{\beta}_0 \tag{3.1a}$$

$$\hat{Y}_j = \hat{\beta}_0 + \hat{\beta}_1 t \tag{3.1b}$$

$$\hat{Y}_j = \hat{\beta}_0 + \hat{\beta}_2 t^2 \tag{3.1c}$$

$$\hat{Y}_j = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2. \tag{3.1d}$$

We have examples of a raw signal:

$$y(t) = t^2 - 10\cos(\pi t) + 5t, \qquad t \in \{\frac{-32\pi}{16}, \frac{-31\pi}{16}, \ldots, 0, \ldots, \frac{31\pi}{16}, 2\pi\}, \tag{3.2}$$

and the results of the detrending methods in Equations (3.1a) - (3.1d) in Figure 3.1. The motivation for this detrending process is to remove the variation in the data that comes from the MRI machine and not the subject (Lazar, 2008).

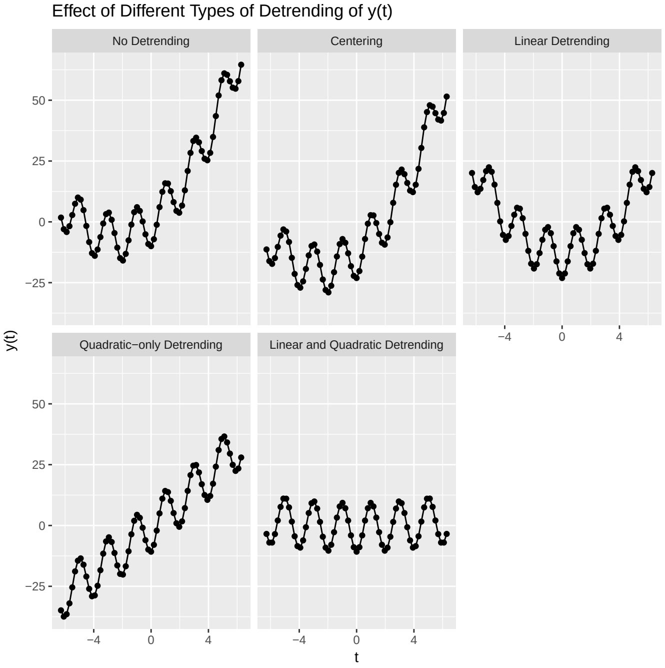


Figure 3.1: Results of detrending Equation (3.2).

3.2 Collapsing

Consider three functions:

$$f_1(t) = \cos(t), f_2(t) = 2\sin(t), f_3(t) = 3(\sin(t) + \cos(t)),$$

for $t = \{\frac{\pi}{10}, \frac{2\pi}{10}, \ldots, 4\pi\}$. We can *collapse* these functions by taking the average $\bar{f}_j$ of each f at each time j, as in Figure 3.2. In the context of a region of interest, we collapse all the signals in the region to one common signal for each session and for each practice group.

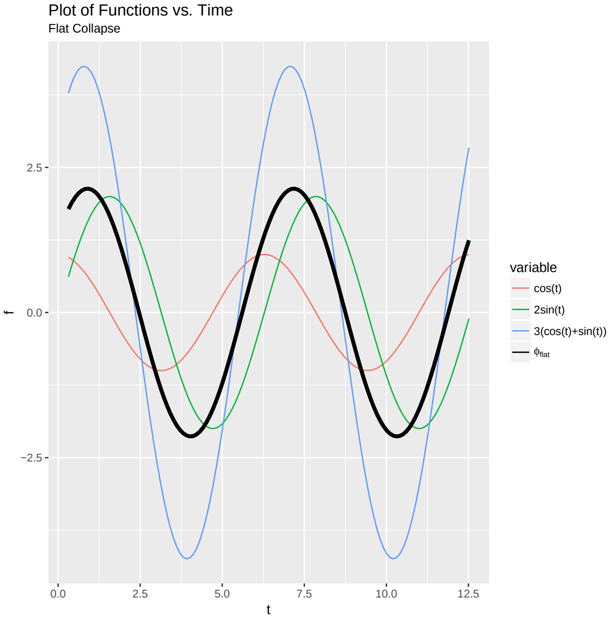


Figure 3.2: Plot of three signals and the collapsed common function.

3.3 Local Polynomial Smoothing

Polynomial regression is an application of the ordinary least squares approach using a Taylor series

$$f(x) = \lim_{p\to\infty} \sum_{n=0}^{p} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \tag{3.3}$$

where $f(x)$ is a smooth function and $f^n(x_0)$ denotes the n^{th} derivative at point x_0, as discussed in Wand and Jones (1995). For a fixed p, Equation (3.3) results in

$$\begin{aligned} f(x) \approx \tilde{f}(x) &= f(x_0) + \frac{f^{(1)}(x_0)}{1!}(x - x_0) + \cdots + \frac{f(x_0)(x_0)}{p!}(x - x_0)^p \\ &= \beta_0 + \beta_1(x - x_0) + \cdots + \beta_p(x - x_0)^p. \end{aligned}$$

We use local quadratic regression in our estimation. Letting $p = 2$, we approximate $f(x)$ close to a point x_0 as

$$\begin{aligned} \tilde{f}(x_i) &= f(x_0) + \frac{f'(x_0)}{1!}(x_i - x_0) + \frac{f''(x_0)}{2!}(x_i - x_0)^2 \\ &= \beta_0 + \beta_1(x_i - x_0) + \beta_2(x_i - x_0)^2 \end{aligned} \tag{3.4}$$

and use the ordinary least squares (OLS) method to get estimates of β_0, β_1 and β_2 in Equation (3.4). Since the approximation of f is centered at x_0, we can estimate f locally near any point x_0. By (3.4), we can use the OLS method for

$$Y_i = \tilde{f}(x_i) = \beta_0 + \beta_1(x_i - x_0) + \beta_2({}_ix - x_0)^2 + \epsilon_i,$$

and minimize $\sum_{i=1}^{n} \epsilon_i^2$. We can assign weight the influence of points x_i based on their proximity to x_0 by including a kernel $K_h(x_i - x_0)$ where $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ and h is a bandwidth. Consider the Gaussian kernel $K(t) \propto e^{-\frac{1}{2}t^2}$ where $t = \frac{x-x_0}{h}$ and $b \in \mathbb{R}_+$; now we minimize $\sum_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{(x_i-x_0)^2}{h}\right)}\epsilon_i^2$.

We consider the following nonparametric regression setting:

$$Y_i = f(x_i) + \epsilon_i,\ i = 1, \dots, n$$

where the ϵ_i's are identically and independently distributed. For the one-dimensional case, we minimize

$$\sum_{i=1}^{N} \{Y_i - \beta_0 - \cdots - \beta_p(x_i - x_0)^p\}^2 K_h(x_i - x_0).$$

Let

$$\mathbf{X}_{x_0} = \begin{bmatrix} 1 & (x_1 - x_0) & \dots & (x_1 - x_0)^p \\ 1 & (x_2 - x_0) & \dots & (x_2 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (x_n - x_0) & \dots & (x_n - x_0)^p \end{bmatrix}$$

$$\mathbf{W}_{x_0} = \begin{bmatrix} K_{h_\lambda}(x_1 - x_0) & 0 & \dots & 0 \\ 0 & K_{h_\lambda}(x_2 - x_0) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & K_{h_\lambda}(x_n - x_0) \end{bmatrix}$$

and minimize

$$(\mathbf{Y} - \mathbf{X}_{x_0}\boldsymbol{\beta})^T \mathbf{W}_{x_0} (\mathbf{Y} - \mathbf{X}_{x_0}\boldsymbol{\beta})$$

and if $\mathbf{X}_{x_0}^T \mathbf{W}_{x_0} \mathbf{X}_{x_0}$ is invertible, then we get the result

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_{x_0}^{\mathbf{T}} \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^{\mathbf{T}} \mathbf{W}_{x_0} \mathbf{Y}. \tag{3.5}$$

It is important to recall that this is an estimation of $Y_i = Y_{x_0}$ at x_0 and will produce different estimates at each choice of x_0, hence locally polynomial. Recall Equation (3.4), and if we want $f(x_0)$, then we multiply $[\mathbf{1}, \mathbf{0}, \ldots, \mathbf{0}]_{1\times(p+1)}$ by Equation (3.5) to estimate $f(x_0)$ or apply this technique to estimate $f_i(\theta_{i,2}t + \theta_{i,3})$ in Equation (3.7) (Rencher, 2000; Wand and Jones, 1995).

3.4 A Shape-invariant Model

we propose a semiparametric method under shape invariance because the fMRI time series of interest share the box-car shaped function due to the nature of the experiment, and thus each time series can be regarded as horizontal and/or vertical shifting and/or scaling of this common function. The shape invariant model assumes sets of shapes f that can all be considered permutations of some common shape ϕ. For our purposes, we assume our shapes to be signals f and estimate their permutations $\boldsymbol{\theta}$ from some common ϕ. We propose an application of shape invariant semiparametric regression model proposed by Kneip and Engel (1995). For a given ROI and a given practice group, there are multiple voxels for each subject. We first detrend those signals and then obtain a single curve for each subject by averaging the voxels within the same ROI. For a given practice group, suppose that there are S subjects and each BOLD signal has a length of T. Then, given an experimental design $t_{ij} = t_j$ for $i = 1, \ldots, S$ and $j = i, \ldots, T$, we have

$$Y_{ij} = f(t_j, \boldsymbol{\theta}_i) + \epsilon_{ij} \tag{3.6}$$

where ϵ_{ij} are i.i.d errors. Under shape invariance, Kneip and Engel (1995) consider the following form for f:

$$f(t, \boldsymbol{\theta}_i) = \theta_{i,1}\phi\left(\frac{t - \theta_{i,3}}{\theta_{i,2}}\right) + \theta_{i,4},$$

with $[\theta_{i,1}, \theta_{i,2}, \theta_{i,3}, \theta_{i,4}]^{\mathbf{T}} = \boldsymbol{\theta}_i$ and the postulated shape-invariant model

$$f_i(\theta_{i,2}t + \theta_{i,3}) = \theta_{i,1}\phi(t) + \theta_{i,4}, \tag{3.7}$$

where $\phi(t)$ is a reference function. For identifiability purposes the authors also impose the constraint that $\frac{1}{s}\sum_{i=1}^{S}\theta_{i,1} = \frac{1}{s}\sum_{i=1}^{S}\theta_{i,2} = 1$ and $\frac{1}{S}\sum_{i=1}^{S}\theta_{i,3} = \frac{1}{s}\sum_{i=1}^{S}\theta_{i,4} = 0$.

We have proposed the following algorithm for estimating $[\theta_{i,1}, \theta_{i,2}, \theta_{i,3}, \theta_{i,4}]^{\mathbf{T}} = \boldsymbol{\theta}_i$:

Step 1. For each i, estimate f_i using local smoothing estimation. Next choose or calculate $\hat{\boldsymbol{\theta}}^{(0)}_{i=1,\ldots,S}$ as the initial estimate of $\boldsymbol{\theta}_{i=1,\ldots,S}$, then compute an initial estimate for the common function $\hat{\phi}^{(0)}(t) = \frac{1}{S}\sum_{i=1}^{S}\hat{f}_i(\hat{\theta}^{(0)}_{i,2}t + \hat{\theta}^{(0)}_{i,3})$.

Step 2. Iterate until convergence from $l = 1, 2, \ldots, L$:

- (i) We estimate $\boldsymbol{\theta}_i$ by solving

$$\min_{\boldsymbol{\theta}_i}\int\left[\hat{f}_i(\hat{\theta}^{(0)}_{i,2}t + \hat{\theta}^{(0)}_{i,3}) - (\hat{\theta}^{(0)}_{i,1}\phi^{(l-1)}(t) + \hat{\theta}^{(0)}_{i,4})\right]^2 dt$$

- (ii) Impose constraints on $\boldsymbol{\theta}^{(l)}_i$
- (iii) Update $\hat{\phi}^{(l)}(t) = \frac{1}{S}\sum_{i=1}^{S}\hat{f}_i(\hat{\theta}^{(l)}_{i,2}t + \hat{\theta}^{(l)}_{i,3})$

Step 3. At iteration L let the final estimates be $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(L)}$ and $\hat{\phi}(t) = \hat{\phi}(t)^{(L)}$.

The idea being that when visual stimulation is provided, the BOLD signal should increase, and then return to a lower activation when the stimulation is removed. Since all the subjects are exposed to visual stimulation at the same intervals, we do not expect $\theta_{i,2}$ to be different from one for all the subjects and set $\theta_{i,2} = 1$. In addition, we make the assumption that $\theta_{i,4}$ would be zero for all subjects, since we detrended all of their signals. Therefore, we only focus on estimating $\theta_{i,1}$ and $\theta_{i,3}$ in our work. Here, the parameters $\theta_{i,3}$ govern horizontal shift, and plays an important role in the frequently observed lag issue due to the variation among subjects in fMRI data. Therefore, these parameters enable alignment of all the curves in the same phase. The parameter $\theta_{i,1}$ controls vertical scale. Especially, $\theta_{i,1}$ is our main concern because we aim to show statistical evidence for $\theta_{1,pre} > \theta_{1,post}$, which implies the attenuation at post-practice particularly for the antisaccade group.

Kneip and Engel (1995) recommend the use of a nonparametric kernel estimation of each f_i, with the constraint that the same bandwidth be used for each estimation of f_i. In this work, instead of choosing a single bandwidth, we use a wide range of bandwidths.

3.5 Optimization

We illustrate the estimation of the parameters and function in Equation (3.7). We do not use to use a traditional optimization algorithm because we cannot assume a convex optimization function. Instead, we employ a generalized simulated annealing algorithm as described by Tsallis and Stariolo (1995), Xiang et al. (2000), and Xiang et al. (2013), and Particle Swarm Optimization 2011 (PSO2011) (Kennedy and Eberhart, 1995; Swarm Intelligence Symposium, 2007; Evolutionary Computation, 2013; Zambrano-Bigiarini and Rojas, 2013), for a range of different bandwidths. We choose to use two different optimization algorithms in order to provide a check on the both of them. We expect to see similar results for both, if they work as expected.

3.5.1 Generalized Simulated Annealing

Generalized simulated annealing (GenSA) is a meta-heuristic optimization algorithm (Mandal et al., 2015). Meta-heuristic algorithms are probabilistic global optimization techniques that only require that a global maximum or minimum exists. While there are not mathematical proofs that these algorithms will converge, they have been demonstrated to converge to the optimum (Mandal et al., 2015; Yusup et al., 2012).

Simulated annealing optimization gets its name from the metalurgic process of heating a metal to a state that allows atoms to migrate the crystal lattice of the metal more freely. As the metal cools and returns to a state of equilibrium, the metal becomes hard and the atoms migrate less freely. The simulated annealing optimization technique considers an optimization function $\Psi(x)$, for some array x. We choose a temperature T to probabilisticly heat our search up, set an initial state for x and evaluate Ψ. This first evaluation of Ψ becomes our global optimum, and another x_{new} is chosen by perturbing x_{old} using some realization from a probability distribution. We evaluate Ψ again and observe if $\Psi(x_{new})$ has a better value than $\Psi(x_{old})$, if so, we keep $\Psi(x_{new})$ as our global optimum and record x_{new}. If $\Psi(x_{new})$ is worse than $\Psi(x_{old})$, it can be accepted with a probability dependent on temperature T. The value of T decreases as the number of iterations (time) increases (Tsallis and Stariolo, 1995; Mandal et al., 2015). We apply a Generalized Simulated Annealing (GenSA) as described by Xiang et al. (2000), Xiang et al. (2013), and Yang Xiang et al. (2013). They use a distorted Cauchy distribution as a visiting distribution

$$g_{q_v}(\Delta x(t)) \propto \frac{[T_{q_v}(t)]^{-\frac{D}{3-q_v}}}{\left[1+(q_v-1)\frac{(\Delta x(t))^2}{[T_{q_v}(t)]^{\frac{2}{3-q_v}}}\right]^{\frac{1}{q_v-1}+\frac{D-1}{2}}}$$

with temperature

$$T_{q_v}(t) = T_{q_v}(1)\frac{2^{q_v-1}-1}{(1+t)^{q_v-1}-1}$$

and a generalized Metropolis algorithm

$$p_{q_a} = \min\{1, [1-(1-q_a)\beta\Delta E]^{\frac{1}{1-q_a}}\}$$

for the acceptance probability (Xiang et al., 2000; Tsallis and Stariolo, 1995).

3.5.2 Particle Swarm Optimization

Particle swarm optimization (PSO) is a method for optimizing continuous, nonlinear functions (Kennedy and Eberhart, 1995). PSO searches the solution space of Psi to find a global optimum. In order to avoid being "trapped" in a local optimum, PSO uses a "swarm" of particles or candidate solutions $\Psi(x_i^t)$[1], each with a position X_i and velocity V_i. Position is defined as a linear combination of previous position and velocity

$$\mathbf{X}_i^{t+1} = \mathbf{X}_i^t + \mathbf{V}_i^{t+1},$$

---

[1] This i is separate from previous uses of i, and the superscripts t and $t+1$ denote iterations at times t and $t+1$, not exponentiation.

and velocity is defined as

$$\mathbf{V}_i^{t+1} = \omega \mathbf{V}_i^t + c_1 \mathbf{U}_1^t \odot (\mathbf{P}_i - \mathbf{X}_i^t) + c_2 \mathbf{U}_2^t \odot (\mathbf{L}^t - \mathbf{X}_i^t),$$

where $\omega \in (0, 1)$ is an inertial weight to adjust influence of $\mathbf{V}_i^t$ and prevent a swarm explosion, $\mathbf{P}_i$ is personal best (optimal) position for particle i to date $\Psi(P_i)$, $\mathbf{L}^t$ is global best (optimal) position for all particles to date $\Psi(\mathbf{L}^t)$, c_1 is cognitive learning factor of the particle, c_2 is the social learning factor of the swarm, $\mathbf{U}_1, \mathbf{U}_2 \sim (0, 1)$, and $\odot$ is the Hadamard product (Evolutionary Computation, 2013). This description of PSO is known in the canonical PSO described by Kennedy (Kennedy and Eberhart, 1995; Evolutionary Computation, 2013).

A more modern version of PSO is known as Standard PSO 2011 (SPSO-2011) is developed by Evolutionary Computation (2013), Zambrano-Bigiarini et al. (2013), and Zambrano-Bigiarini and Rojas (2014). They define, for a particle i at time t, a center of Gravity $\mathbf{G}_i^t$, defined by the points $\mathbf{X}_i^t$, $\mathbf{p}_i^t$ and $\mathbf{l}_i^t$. The best previous personal position $\mathbf{p}_i^t$, a point $\mathbf{l}_i^t$ beyond $\mathbf{p}_i^t$ in neighborhood of $\mathbf{p}_i^t$ and are defined

$$\begin{aligned} \mathbf{p}_i^t &= \mathbf{X}_i^t + c_1 \mathbf{U}_1^t \odot (\mathbf{P}_i^t - \mathbf{X}_i^t), \\ \mathbf{l}_i^t &= \mathbf{X}_i^t + c_2 \mathbf{U}_2^t \odot (\mathbf{L}^t - \mathbf{X}_i^t), \\ \mathbf{G}_i^t &= \frac{\mathbf{X}_i^t + \mathbf{p}_i^t + \mathbf{l}_i^t}{3}, \end{aligned}$$

and now velocity becomes

$$\mathbf{V}_i^{t+1} = \omega_t \mathbf{V}_i^t + \mathcal{H}_i(\mathbf{G}_i^t, ||\mathbf{G}_i^t - \mathbf{X}_i^t||) - \mathbf{X}_i^t,$$

where $\mathcal{H}_i(\mathbf{G}_i^t, ||\mathbf{G}_i^t - \mathbf{X}_i^t||)$ is a hypersphere (Evolutionary Computation, 2013).

3.6 Testing

We construct test statistics to compare the vertical scale parameter $\theta_{1,i}$ that characterizes our primary interest for each ROI. First, we examine the attenuation before and after practice for each group using three pairwise tests and determine which group shows attenuation for each ROI. Second, we test whether there exists a difference in estimates of the vertical scale parameter using ANOVA and Kruskal-Wallis tests.

3.6.1 PAIRWISE ATTENUATION TEST

We test for significant attenuation for a given practice group

$$H_0 : \theta_{1,pre} = \theta_{1,post}$$
$$H_1 : \theta_{1,pre} > \theta_{1,post}$$

using the paired t-test, Wilcoxon signed-rank test and the sign test. The t-test is more powerful than the Wilcoxon test, and the Wilcoxon test is more powerful than the sign test. However, the assumptions for the t-test are the most strict, then Wilcoxon and the sign test assumes the least. It is important to note that for the sign tests, we can have the same test statistics and p-values for different samples if the rankings are the same.

3.7 GROUP ATTENUATION TEST

We test for the group difference:

$$H_0 : \text{Group attenuations are the same}$$
$$H_1 : \text{At least one group's attenuation is different.}$$

We take $\hat{\theta}_{i,1,pre} - \hat{\theta}_{i,1,post} = \hat{\theta}_{i,1,D}$ to estimate attenuation for each subject in each group. Then, we test

$$H_0 : \theta_{1,D,anti} = \theta_{1,D,pro} = \theta_{1,D,fix},$$
$$H_1 : \text{At least one of } \theta_{\cdot,1,D} \text{ are different}$$

using the parametric one-way ANOVA and the nonparametric Kruskal-Wallis ANOVA.

Chapter 4

Simulation

In section we describe our simulation settings and report both pairwise and group testing results.

4.1 Design of Simulated Data

We simulate a single bold signal $x(t)$ by modeling the box-car pattern of on-off stimulation we describe in Section 2.2. We choose a more simple design of nine time points on and nine time points off using a square wave Fourier series and approximation shown in Figure 4.1 and Equation (4.1). We choose this function because it is easily mapped to a horizontal shift and vertical scale for simulating θ_1 and θ_3.

$$x(t) = \lim_{n\to\infty} \frac{4}{\pi} \sum_{\kappa=1}^{n} \frac{\sin(2\pi(2\kappa-1)\omega t_i)}{2\kappa-1},$$

For practical reasons we choose $n = 10^3$ and frequency $\omega = 0.05$ to get our nine time points on and nine time points off, with one point separating the two states.

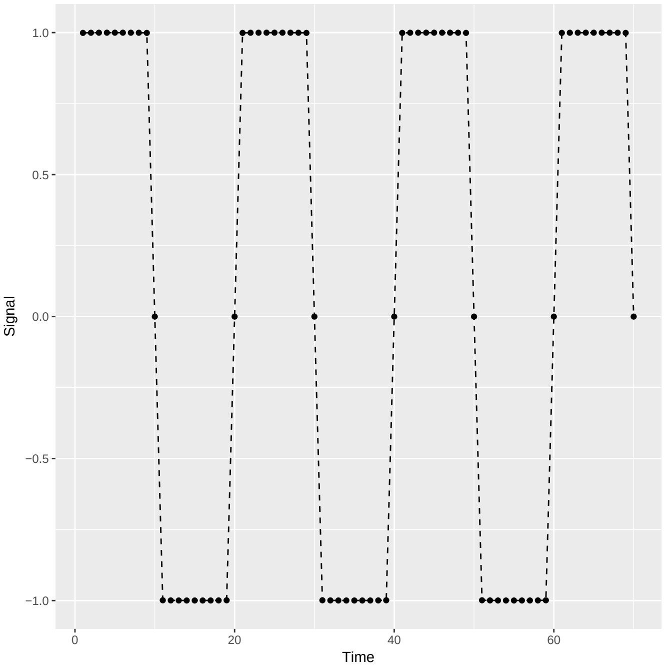


Figure 4.1: Plot of square wave similar to the box-car pattern of the saccade and fixation blocks.

For any subject i in group k ($k = 1, \ldots, K = 3$), we apply an amplitude a_k as a scalar to control signal amplitude later estimated by $\hat{\theta}_{i,1}$, and we can shift time t for any subject by $d_{jk} \overset{iid}{\sim} U(-2, 2)$ uniform distribution, and later estimated by $\hat{\theta}_{i,3}$. All the signals are over domain $t_j = j \in \{1, \ldots, 70\}$.

$$x_{ik}(t_j) = a_k \frac{4}{\pi} \sum_{\kappa=1}^{10^3} \frac{\sin(2\pi(2\kappa - 1)0.05(t_j + d_{j,k}))}{2\kappa - 1},$$

We let each vector $\boldsymbol{x}_{ik} = [x_{ik}(t_1), \ldots, x_{ik}(t_{70})]^T$ so that we have

$$\boldsymbol{X}_{SK \times 70} \quad = \begin{bmatrix} \boldsymbol{x}_{11}^T \\ \vdots \\ \boldsymbol{x}_{SK}^T \end{bmatrix},$$

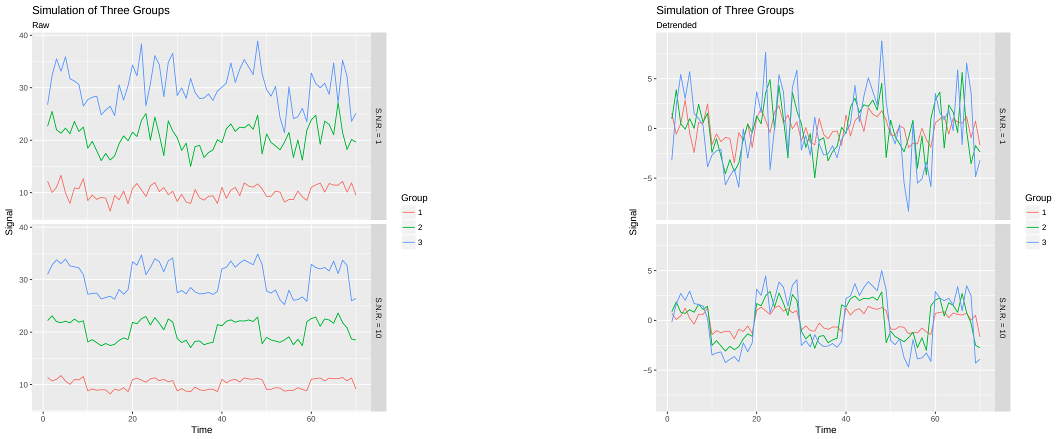
and then find standard deviation $\sigma_{signal\ ik}$ for each x_{ik} and calculate $\sigma_{noise\ ik}$ for a given signal to noise ratio (SNR):

$$SNR = \frac{\sigma^2_{signal\ ik}}{\sigma^2_{noise\ ik}}.$$

We generate noise vectors $\boldsymbol{\mathfrak{s}}_{ik} = [\boldsymbol{\mathfrak{s}}_{i1}, \ldots, \boldsymbol{\mathfrak{s}}_{i70}]^T$ where $\boldsymbol{\mathfrak{s}}_{ij} \sim N(0, \sigma^2_{noise\ ik})$ and calculate a noise matrix:

$$\boldsymbol{\mathfrak{S}}_{SK\times 70} = \begin{bmatrix} \boldsymbol{\mathfrak{s}}_{11}^T \\ \vdots \\ \boldsymbol{\mathfrak{s}}_{SK}^T \end{bmatrix},$$

and get our final simulated data $\boldsymbol{\mathfrak{D}} = \boldsymbol{X}_{SK\times 70} + \boldsymbol{\mathfrak{S}}_{SK\times 70}$. Once we have $\boldsymbol{\mathfrak{D}}$, we can apply our linear and quadratic detrending techniques, which results in centering each $\boldsymbol{x}_{ik}$. For example, let $[a_1, a_2, a_3]^T = [1, 2, 3]^T$, for three groups each with one subject for SNR values of 1 and 10 as in Figure 4.2.



(a) Raw

(b) Detrended

Figure 4.2: Simulations for three groups, with amplitudes 1,2,3 and one subject in each group.

The simulation procedure begins with an SNR from $\{0.5, 1.0, 1.5, 2.0\}$, then five amplitude sets to generate five different sets of data as in Figure 4.3. We take a set $\boldsymbol{\mathfrak{D}}_{(\cdot)}$ from the generated $\{\boldsymbol{\mathfrak{D}}_1, \boldsymbol{\mathfrak{D}}_2, \boldsymbol{\mathfrak{D}}_3, \boldsymbol{\mathfrak{D}}_4, \boldsymbol{\mathfrak{D}}_5\}$ and consider twenty equally spaced bandwidth values h from 0.5 to 5 on the natural log scale. This log-scale method results in more bandwidth choices on the low end as in Figure 4.4. This single procedure of generating data is repeated 100 times.

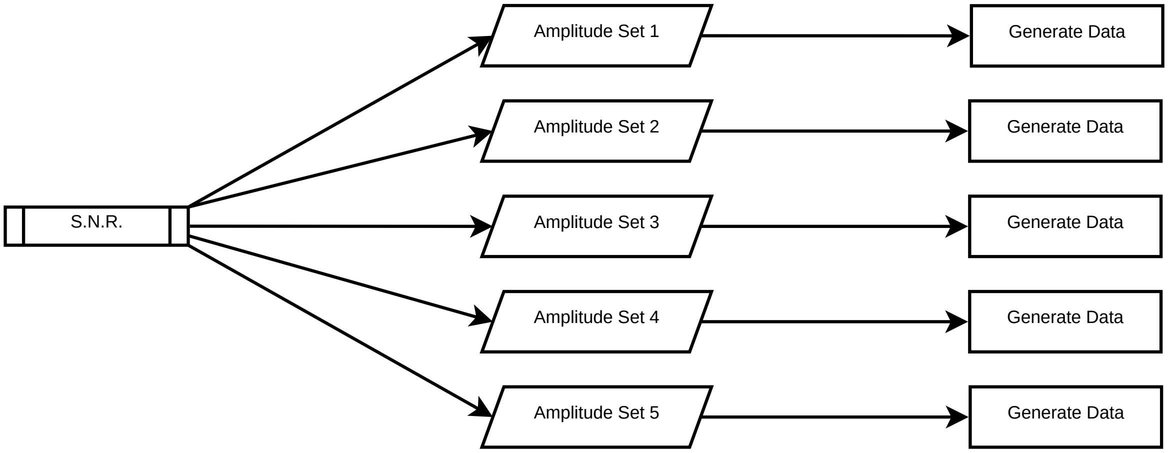


Figure 4.3: Micro view of simulation. This process is done for each signal to noise ratio and repeated 100 times.

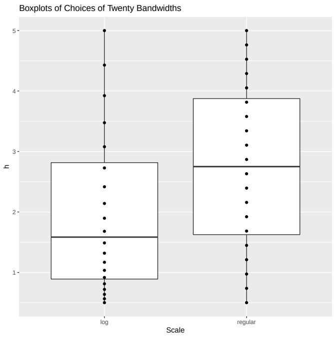


Figure 4.4: Boxplots of choices of bandwidth on different scales. From this plot, it is easy to see the concentration of lower bandwidths by using the log scale.

4.2 Evaluation of Fitness Function ψ

For a given $\mathfrak{D}_{(\cdot)}$ and h, we begin with a candidate set $\hat{\boldsymbol{\theta}}_3^{(l)} = [\theta_{31}^{(l)}, \ldots, \theta_{3S}^{(l)}]^T$ for some iteration l, then normalized[1] such that $\frac{1}{S}\mathbf{1}_{1\times S}\hat{\boldsymbol{\theta}}_3^{(l)} = 0$. We then adjust each time t_{ij} for a subject i

---

[1] Normalization is done before the measure of fitness to use the properties of our algorithms that avoid looking in areas that produce similar results.

such that $t_{ij} \to (t_{ij} + \hat{\theta}_{3i}^{(l)})$. We now estimate $\boldsymbol{f}$ where

$$\boldsymbol{f}^{(l)} = \begin{bmatrix} f_1^{(l)} \\ \vdots \\ f_S^{(l)} \end{bmatrix}$$

with a local quadratic regression and Gaussian kernel smoother to estimate values for each f_i on time domain $\{1, 2, \ldots, 70\}$. We then collapse our estimation $\hat{\boldsymbol{f}}^{(l)}$ by taking the mean column-wise to get an estimation $\hat{\phi}(t_j)$ of the common function. We evaluate $\hat{\boldsymbol{\theta}}_3^{(l)}$ and $\hat{\boldsymbol{f}}^{(l)}$ by considering our fitness function

$$\psi^{(l)} = \sum_{i=1}^{S} \sum_{j=4}^{67} \left(\hat{f}_i(t_j)^{(l)} - \hat{\phi}(t_j)^{(l)} \right)^2.$$

We choose to trim off the first and last three time points in our evaluation of $\psi^{(l)}$ because the boundaries of the $\hat{f}_i^{(l)}$ tend toward relatively extreme values that are a product of the local regression and not representative of the actual data.

Our initial choice of $\hat{\boldsymbol{\theta}}_3^{(1)}$ is set to all zeros. Any $\hat{\boldsymbol{\theta}}_3^{(l)}$ that produces a $\psi^{(l)}$ smaller than a $\psi^{(l-1)}$ gets stored $\psi^{(*)}$ as our best, and similarly store $\hat{\boldsymbol{\theta}}_3^{(*)}$. We bound our choice of $\hat{\boldsymbol{\theta}}_3^{(1)}$ between $-2\mathbf{1}_{S\times 1}$ and $2\mathbf{1}_{S\times 1}$. We decide convergence of fitness function ψ when it does not increase by a very small amount for an extended number of iterations[2] and our final estimate of $\hat{\boldsymbol{\theta}}_3^{(*)}$ is stored as $\hat{\boldsymbol{\theta}}_3^{(L_{\theta_3})}$.

We similarly repeat the process above to find $\hat{\boldsymbol{\theta}}_3^{(L_{\theta_1})}$, but use the $\boldsymbol{f}^{(L_{\theta_3})}$ from before for each $\hat{\boldsymbol{\theta}}_1^{(1)}$, initialize at $\mathbf{0}_{S\times 1}$ and limit our search to be between $10^{-4}\mathbf{1}_{S\times 1}$ and $30\mathbf{1}_{S\times 1}$. Our fitness function now becomes

$$\psi^{(l)} = \sum_{i=1}^{S} \sum_{j=4}^{67} \left(\hat{\theta}_{1i}^{(l)} \hat{f}_i(t_j)^{(L_{\theta_3})} - \hat{\phi}(t_j)^{(l)} \right)^2.$$

We choose to estimate $\boldsymbol{\theta}_3$ and $\boldsymbol{\theta}_1$ separately for sake of computation time, i.e. we search $\mathbb{R}^S$ once under the local regression procedure and then again by simply evaluating $\phi^{(l)}$ in another $\mathbb{R}^S$ rather than performing the entire procedure together in an $\mathbb{R}^{2S}$ space.

---

[2] For details see references for GenSA and HydroPSO packages and discussion in the Appendix.

CHAPTER 5

NUMERICAL RESULTS

In this chapter, we discuss the results of both our simulated and real data analyses. We present the pairwise test results first and the group-wise results second.

5.1 PAIRWISE ATTENUATION TEST RESULTS

We perform pairwise attenuation tests using our estimation of $\theta_{1,pre}$ and $\theta_{1,post}$ to decide significant attenuation in the post-practice for each group and each ROI.

5.1.1 SIMULATION

We generate data as in Section 4.1 for twelve subjects as pre and post practice groups using the amplitudes in Table 5.1 to test and estimate $\theta_{3,pre}$ and $\theta_{3,post}$ using the same common function ϕ, and likewise for $\theta_{1,pre}$ and $\theta_{1,post}$.

Table 5.1: Pairwise Simulation

	Amplitudes	
	Pre	Post
a_1	3.0	3.0
a_2	3.0	2.5
a_3	3.0	2.0
a_4	3.0	1.5
a_5	3.0	1.0

We see a clear separation between pre and post estimates of θ_1 as the difference in amplitudes increases and decrease in variation of the estimates as SNR increases in Figures A.11 - A.12. The choice of bandwidth does not appear to be associated with the estimation of attenuation θ_1 over the 100 repetitions of the simulation. The pairwise test results in Figures 5.1 - 5.2 are what we expect for amplitude set a_1 and the significance becomes stronger as the post amplitudes increases. We also see an increase in significance as SNR increases. The bandwidth selection does not appear to be associated with significance aver all the simulations. Recall that when using rank testing in both the Wilcoxon and Sign paired tests, we have a finite number of test statistics and that is why we see repeated p-values for both of these methods.

Figure 5.1: Boxplots of p-values from t, Wilcoxon and Sign tests for significant attenuation. The testing is performed on estimates from the Generalized Simulated Annealing method.

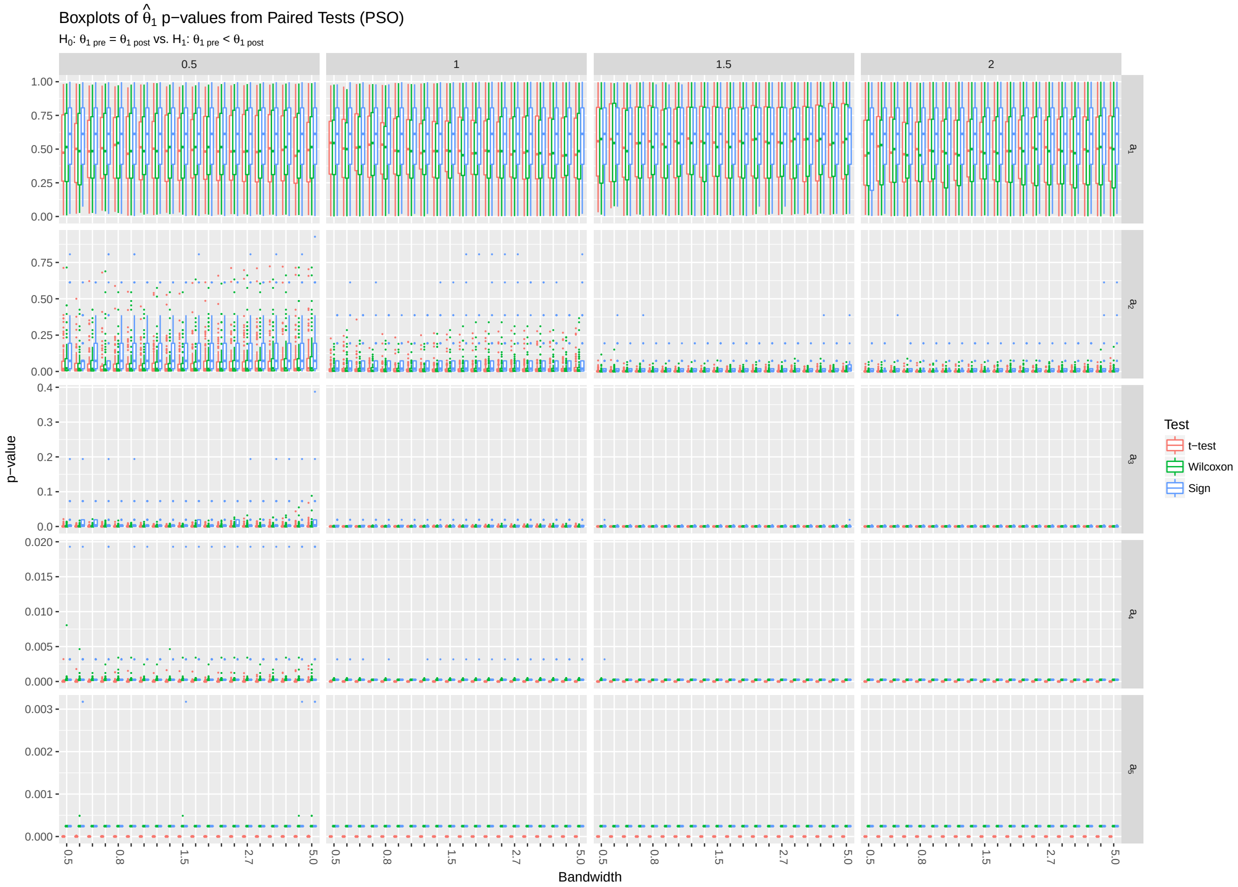


Figure 5.2: Boxplots of p-values from t, Wilcoxon and Sign tests for significant attenuation. The testing is performed on estimates from the Particle Swarm Optimization method.

5.1.2 Real Data

We apply the same analysis as in Section 5.1.1 and use separate common functions ϕ for each practice group. We analyze the SEF, FEF, SPL, Cuneus, Thalamus, IPL, PFC-L, PFC-R, Striatum, MOG and IFC regions of interest in the three practice groups in Section 2.2. Lee et al. (2013) find that attenuation occurs only for the antisaccade group in SEF, FEF, precuneus, and cuneus using a bootstrap-based approach, and Lee (2013) find that SEF, FEF, SPL and Cuneus regions have significant differences in attenuation between the practice groups using a mixed model at $\alpha = 0.10$ with feature extraction. We have significant attenuation ($\alpha = 0.05$) for the antisaccade practice group only in the SEF, FEF, IPL, PFC-

L and PFC-R regions under the GenSA estimation in Figure 5.3. The significant results tend to be in the lower bandwidth regions. While we do have significant attenuation for the antisaccade group in the Thalamus and PFC-R regions, it is only under the Sign test and it is for only two bandwidths in the Thalamus and only one bandwidth in the PFC-R. We also have a significant result for the fixation practice group at the maximum bandwidth of 5, but this is the only significant result for the fixation group. The pro-saccade practice group does not have any significant results under the GenSA estimation.

The PSO estimation produces results similar to the GenSA estimation. However, we do see some fairly consistent significance for the cuneus region for the antisaccade practice group. We also see significant attenuation in the fixation group for maximum bandwidth. Additionally, we find a single significant attenuation in the prosaccade practice group in the PFC-R region, but it is only for one small bandwidth.

We also see in Section A.2.2 that the variation in our estimates of θ_1 is not consistently larger for the pre- or post-practice session for the antisaccade group. The same is true for the other two practice groups.

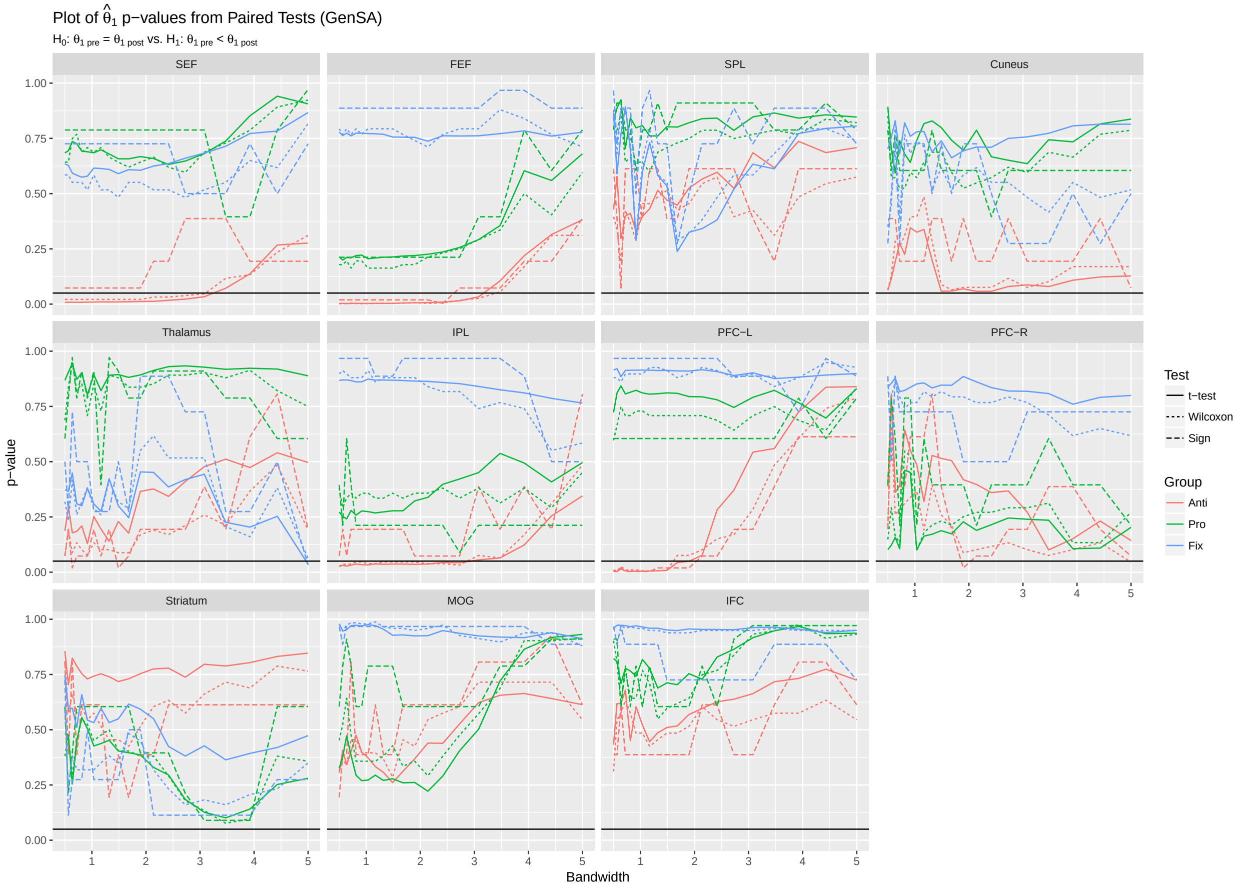


Figure 5.3: Pairwise test results for the three practice groups using Generalized Simulated Annealing. Each practice group is estimated separately. The horizontal line is drawn at our significance level 0.05.

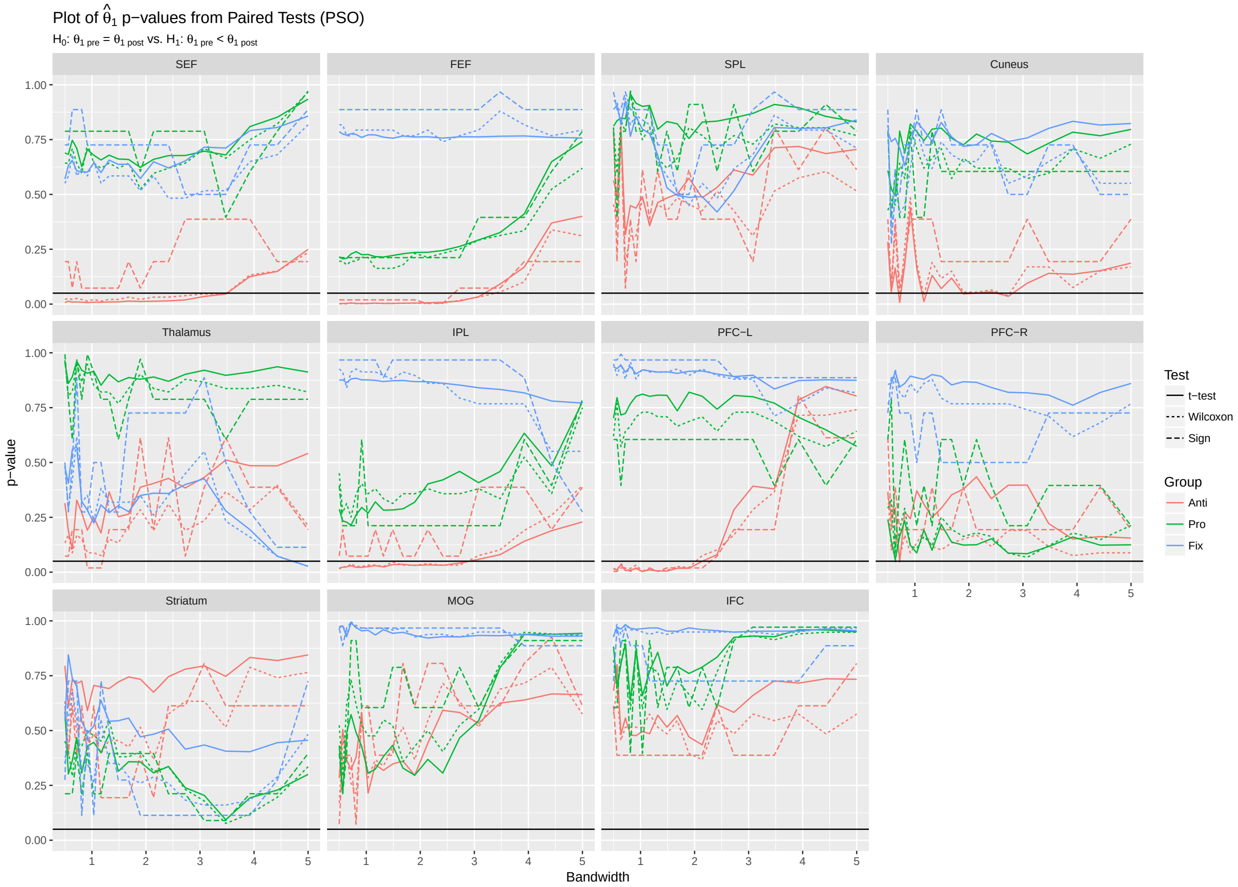


Figure 5.4: Pairwise test results for the three practice groups using Particle Swarm Optimization. Each practice group is estimated separately. The horizontal line is drawn at our significance level 0.05.

5.2 Group Difference Test Results

To simulate data for three-way testing, we generate data for three groups containing 12, 14 and 11 subjects that each have an amplitude of 3 for the pre-practice session and varying amplitudes for the post-practice session in Table 5.2.

5.2.1 Simulation

We simulate the data with the sets of amplitudes in Table 5.2. The results in Figures 5.5 and 5.6 show that as SNR and signal amplitudes increase, the groupwise difference in attenuation is more significant. Both the parametric ANOVA and the Kruskal-Wallis tests appear to produce similar results. We can also see in Figures A.21 and A.22 that the separation in the estimates of attenuation becomes clearly distinct as SNR and signal amplitudes increase. Bandwidth choice does appear to affect the variation in the significance of attenuation for larger SNR and post-practice amplitude differences.

Table 5.2: Sets of amplitudes differences in the simulation post-practice session. All amplitudes in the simulation pre-practice session are 3.00.

Amplitude Set	Group 1	Group 2	Group 3
a_1	3.00	3.00	3.00
a_2	2.97	2.98	3.00
a_3	2.75	2.87	3.00
a_4	2.00	2.50	3.00
a_5	1.00	2.00	3.00

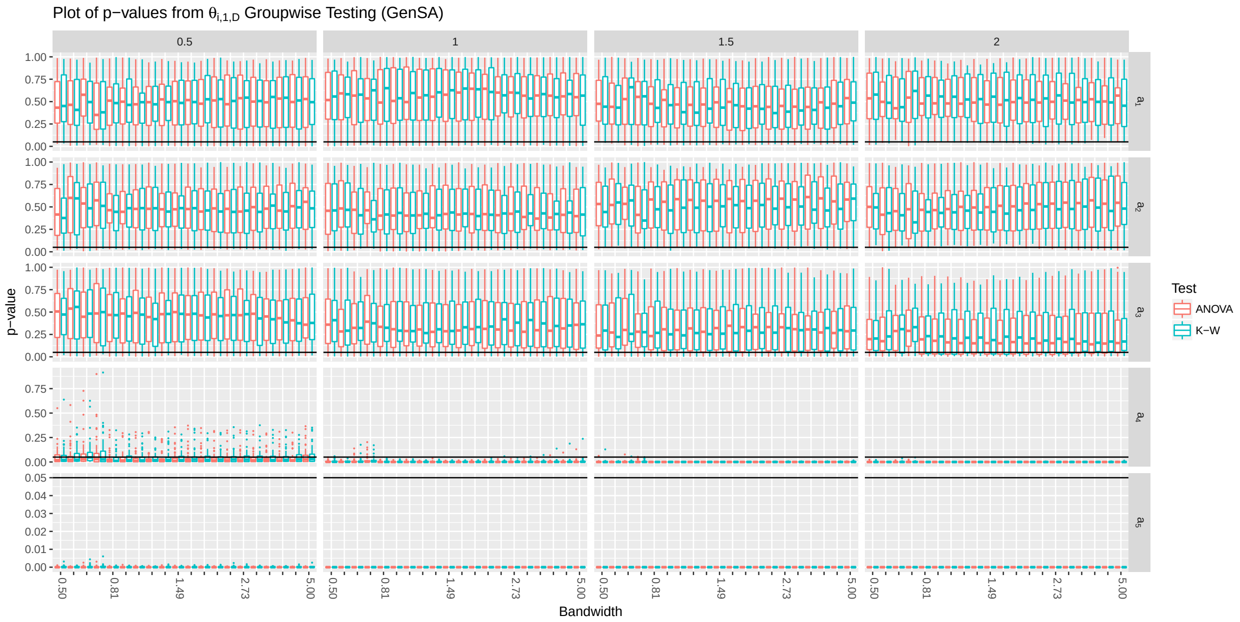


Figure 5.5: Boxplots of p-values to test for groupwise differences in $\hat{\theta}_{i,1,pre} - \hat{\theta}_{i,1,post} = \hat{\theta}_{i,1,D}$ from Generalized Simulated Annealing optimization.

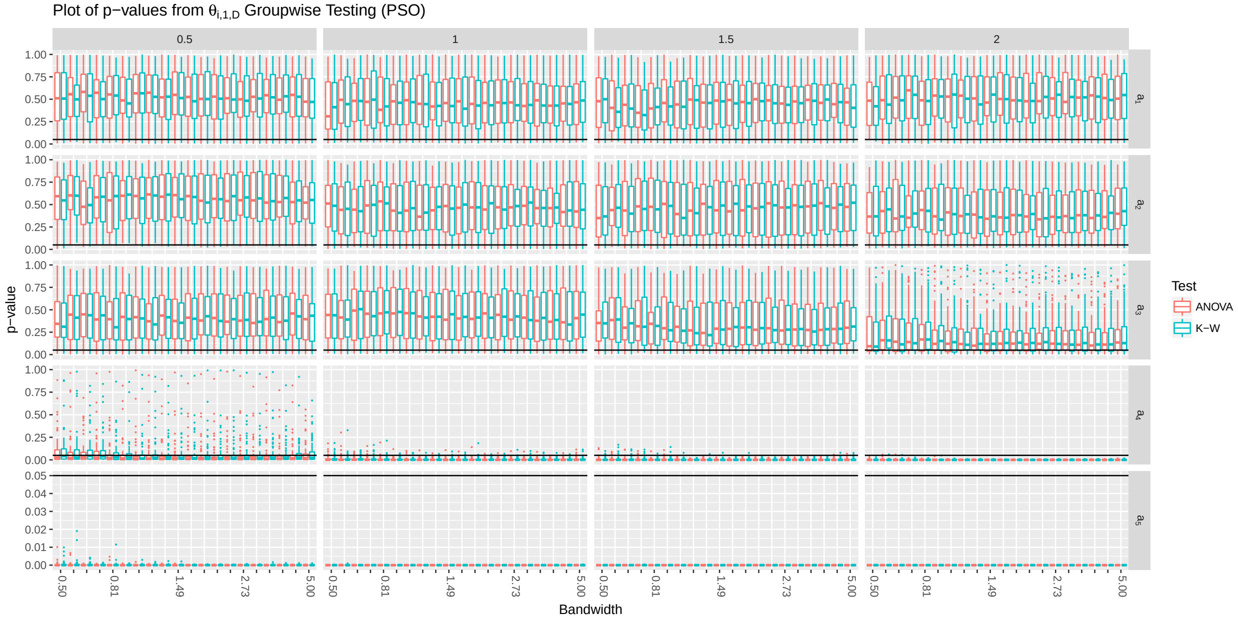


Figure 5.6: Boxplots of p-values to test for groupwise differences in $\hat{\theta}_{i,1,pre} - \hat{\theta}_{i,1,post} = \hat{\theta}_{i,1,D}$ from Particle Swarm Optimization.

5.2.2 Real Data

We see significant attenuation difference between the practice groups for Thalamus, IPL, PFC-L, MOG and IFC regions for both the PSO and GenSA estimations in Figures 5.7 - 5.8. The results of the parametric ANOVA and the Kruskal-Wallis are mostly in agreement, except for the Thalamus, IPL and IFC regions. Note that the parametric ANOVA testd the means and the Kruskal-Wallis compares medians to establish ranks. We can see in Figures A.23 - A.24 that the medians are much more separated than the inferred means for the more significant attenuation differences for the Thalamus, but the opposite is true for the IPL and IFC regions.

Table 5.3: The ANOVA results of Lee (2013) for a mixed effects model with feature extraction using both mean and median ($\alpha = 0.10$).

ROIs	Mean p-value	Median p-value
SEF	0.0511	0.0737
FEF	0.0230	0.0220
SPL	0.0022	0.0054
Cuneus	0.0024	<.0001
Thalamus	0.1403	0.1309
IPL	0.7872	0.8402
PFC-L	0.8883	0.9603
PFC-R	0.4824	0.7247
Striatum	0.3433	0.2090
MOG	0.7726	0.7682
IFC	0.3944	0.2339

While our results are not similar to those of Lee (2013) in Table 5.3 at out chosen $\alpha = 0.05$, they are similar for the SEF and FEF regions for a range of bandwidths. We can see in Figures A.23 and A.24 that the attenuation estimation for the antisaccade group does appear to be different, but the variation in the estimates is very large with a lot of interquartile range overlap.

Both the IPL and PFC-L regions have consistent test results for multiple bandwidths, and this is consistent with what we see in Figures 5.3 and 5.4 in Section 5.1.2. Similarly, we have some significant results for small bandwidths in the MOG and IFC regions in Figures 5.7 and 5.8 and we can see that the attenuation significance for the fixation group in Figures 5.3 and 5.4 is consistently low and distinct from those of the antisaccade and prosaccade practice groups.

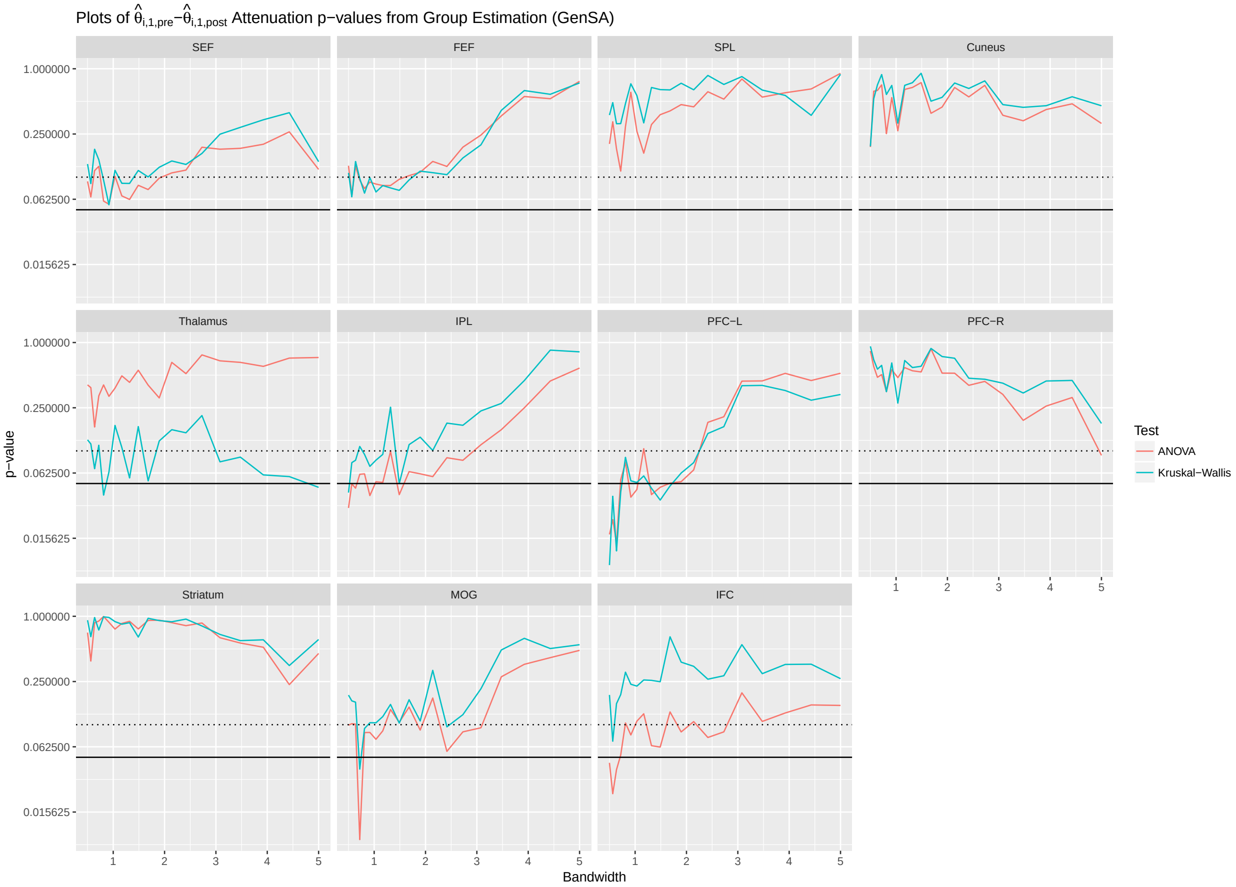


Figure 5.7: Plot of p-values for the three group attenuation estimation and the GenSA optimization. The solid black line is at 0.05 and the dotted line is at 0.10.

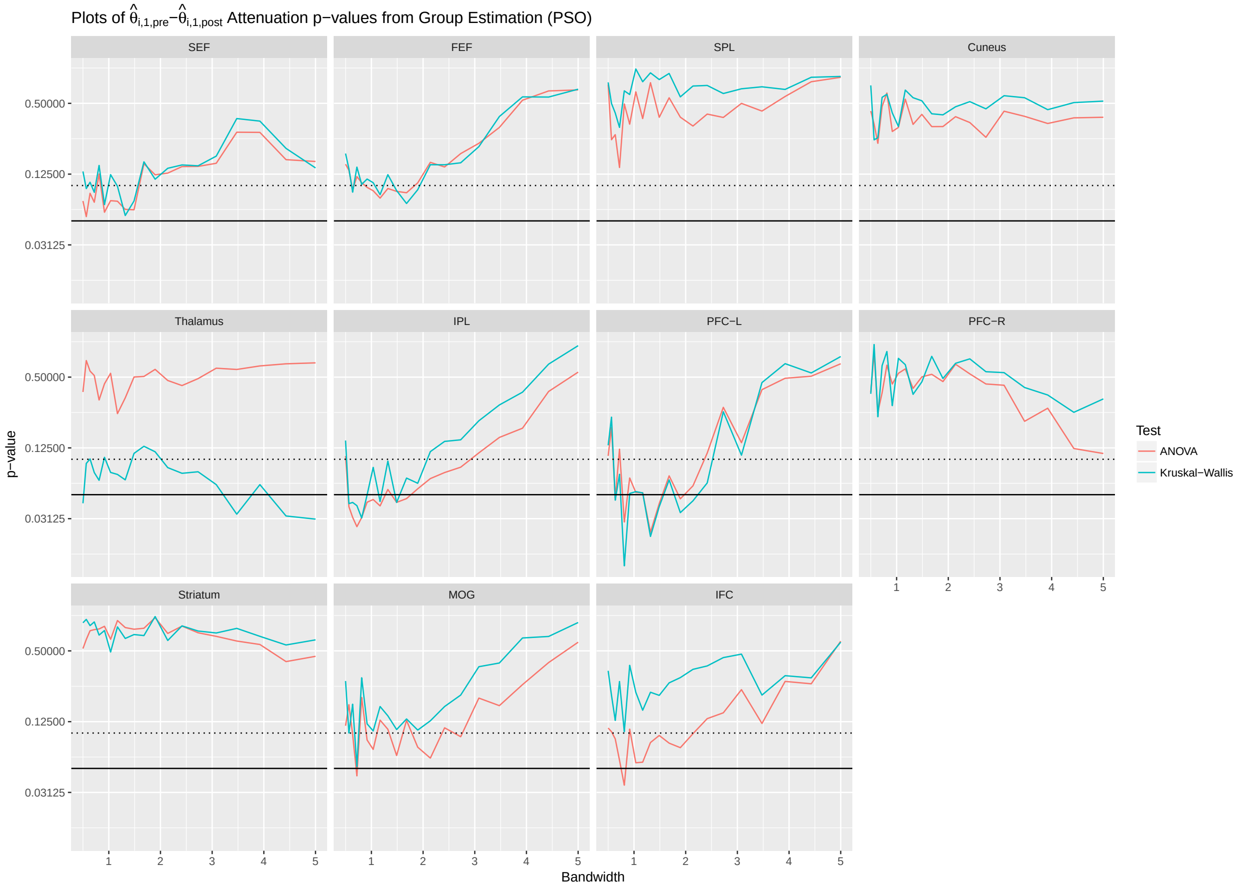


Figure 5.8: Plot of p-values for the three group attenuation estimation and the PSO optimization. The solid black line is at 0.05 and the dotted line is at 0.10.

CHAPTER 6

CONCLUSION AND FUTURE WORK

The shape-invariant estimation method appears to work well. We see in Section 5 that we have consistent results for both GenSA and PSO estimations for each test, and also reasonable agreement between the tests themselves when the mean and median are close together. Our three group results in Section 5.2.2 are similar to those of Lee (2013) using the mixed effects with feature extraction method for the SEF and FEF regions. However, we do not see the same significance in the SPL and Cuneaus regions, and we do have significant attenuation difference in the IPL and PFC-L regions which agree with our pairwise results in Section 5.1.2. Our pair-wise results in Section 5.1.2 show that there is convincing evidence of significant attenuation in the subjects that practice antisaccade tasks in the SEF, FEF, Cuneus, IPL and PFC-L regions.

Moving forward, we plan to apply different smoothing techniques such as splines or wavelets, develop a bootstrap test statistic, and determine an optimal bandwidth. The objective function does not have to be based on the L_2 norm, but maybe another distance such as L_p or correlation. Doing so could produce estimates of attenuation with smaller variation. We choose to estimate attenuation for a collapsed time course using a single parameter θ_1, but we could have estimated $\theta_{1,b}$ where b is the current block of time. Furthermore, we could use the shape-invariance technique to establish a common surface across an ROI for a subject to avoid collapsing with an average. Using such a surface, we would increase our sample size from one-per-subject to the number of voxels in a given ROI.

BIBLIOGRAPHY

Ashby, F. G. (2011). *Statistical Analysis of FMRI Data.* The MIT Press.

Dyckman, K. A. (2007). *FMRI/EEG study of changes in behavior and brain activity during and after saccade practice. [electronic resource].* 2007.

Evolutionary Computation (2013). Standard particle swarm optimisation 2011 at cec-2013: A baseline for future pso improvements. *2013 IEEE Congress on Evolutionary Computation, Evolutionary Computation (CEC), 2013 IEEE Congress on*, page 2337.

Huettel, S. A., Song, A. W., and McCarthy, G. (2008). *Functional magnetic resonance imaging.* Sunderland, Mass. : Sinauer Associates, c2008.

Jncke, L., Gaab, N., Wstenberg, T., Scheich, H., and Heinze, H.-J. (2001). Short communication: Short-term functional plasticity in the human auditory cortex: an fmri study. *Cognitive Brain Research*, 12:479 – 485.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 4, pages 1942–1948 vol.4.

Kneip, A. and Engel, J. (1995). Model estimation in nonlinear regression under shape invariance. *The Annals of Statistics*, (2):551.

Lauterbur, P. C. (1973). Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(5394):190.

Lazar, N. A. (2008). *The statistical analysis of functional MRI data. [electronic resource].* Statistics for biology and health. New York ; London : Springer, c2008.

Lee, J. (2013). *Functional magnetic resonance imaging data clustering. [electronic resource].* 2013.

Lee, J., Park, C., Dyckman, K., Lazar, N., Austin, B., Li, Q., and McDowell, J. (2013). Practice-related changes in neural activation patterns investigated via wavelet-based clustering analysis. *HUMAN BRAIN MAPPING*, 34(9):2276 – 2291.

Mandal, A., Wong, W., and Yu, Y. (2015). chapter Algorithmic Searches for Optimal Designs, pages 755–784. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Chapman and Hall/CRC. 0.

Nelles, G., Pscherer, A., de Greiff, A., Forsting, M., Gerhard, H., Esser, J., and Diener, H. C. (2009). Eye-movement training-induced plasticity in patients with post-stroke hemianopia. *Journal of Neurology*, 256(5):726–733.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990a). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, (24):9868.

Ogawa, S., Lee, T.-M., Nayak, A. S., and Glynn, P. (1990b). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, 14(1):68.

Panouillres, M., Neggers, S. F., Gutteling, T. P., Salemme, R., Stigchel, S. v., van der Geest, J. N., Frens, M. A., and Plisson, D. (2012). Transcranial magnetic stimulation and motor plasticity in human lateral cerebellum: Dual effect on saccadic adaptation. *HUMAN BRAIN MAPPING*, 33(7):1512 – 1525.

Rencher, A. C. (2000). *Linear models in statistics.* Wiley series in probability and statistics. New York : John Wiley, c2000.

Shumway, R. H. and Stoffer, D. S. (2011). *Time series analysis and its applications.* Springer Texts in Statistics. Springer, New York.

Swarm Intelligence Symposium (2007). Defining a standard for particle swarm optimization. *2007 IEEE Swarm Intelligence Symposium, Swarm Intelligence Symposium, 2007. SIS 2007. IEEE*, page 120.

Toga, A. W. and Mazziotta, J. C. (2000). *Brain mapping : the systems.* San Diego : Academic Press, c2000.

Tsallis, C. and Stariolo, D. A. (1995). Generalized simulated annealing.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing.* Monographs on statistics and applied probability: 60. London ; New York : Chapman & Hall, 1995.

Xiang, Y., Gubian, S., Suomela, B., and Hoeng, J. (2013). Generalized simulated annealing for global optimization: The gensa package. *R JOURNAL*, 5(1):13 – 28.

Xiang, Y., Sun, D. Y., and Gong, X. G. (2000). Generalized simulated annealing studies on structures and properties of nin (n = 2-55) clusters. *JOURNAL OF PHYSICAL CHEMISTRY A*, 104:2746 – 2751.

Yang Xiang, Gubian, S., Suomela, B., and Hoeng, J. (2013). Generalized simulated annealing for efficient global optimization: the GenSA package for R. *The R Journal Volume 5/1, June 2013.*

Yusup, N., Zain, A. M., and Hashim, S. Z. M. (2012). Review: Evolutionary techniques in optimizing machining parameters: Review and recent applications (20072011). *Expert Systems With Applications*, 39:9909 – 9927.

Zambrano-Bigiarini, M., Rojas, and R. (2013). A model-independent particle swarm optimisation software for model calibration. *Environmental Modelling & Software*, 43:5–25.

Zambrano-Bigiarini, M. and Rojas, R. (2013). A model-independent particle swarm optimisation software for model calibration. *Environmental Modelling and Software*, 43:5 – 25.

Zambrano-Bigiarini, M. and Rojas, R. (2014). *hydroPSO: Particle Swarm Optimisation, with focus on Environmental Models.* R package version 0.3-4 — For new features, see the 'NEWS' file (on CRAN, rforge or the package source).

Appendix A

Appendix

A.1 Average Signals for Practice Groups

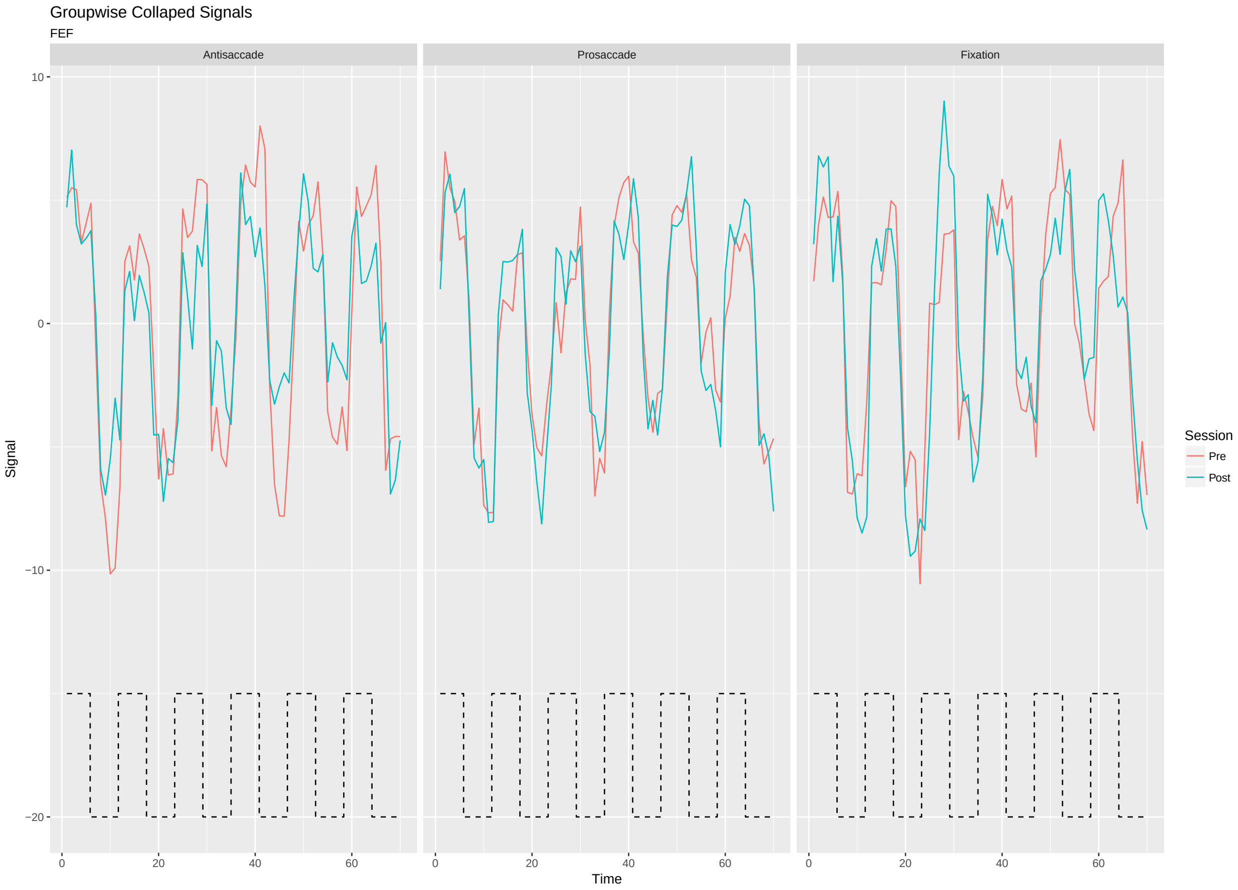


Figure A.1: Collapsed signals for each group and each session from the frontal eye field region. The dashed line represents the antisaccade and fixation blocks.

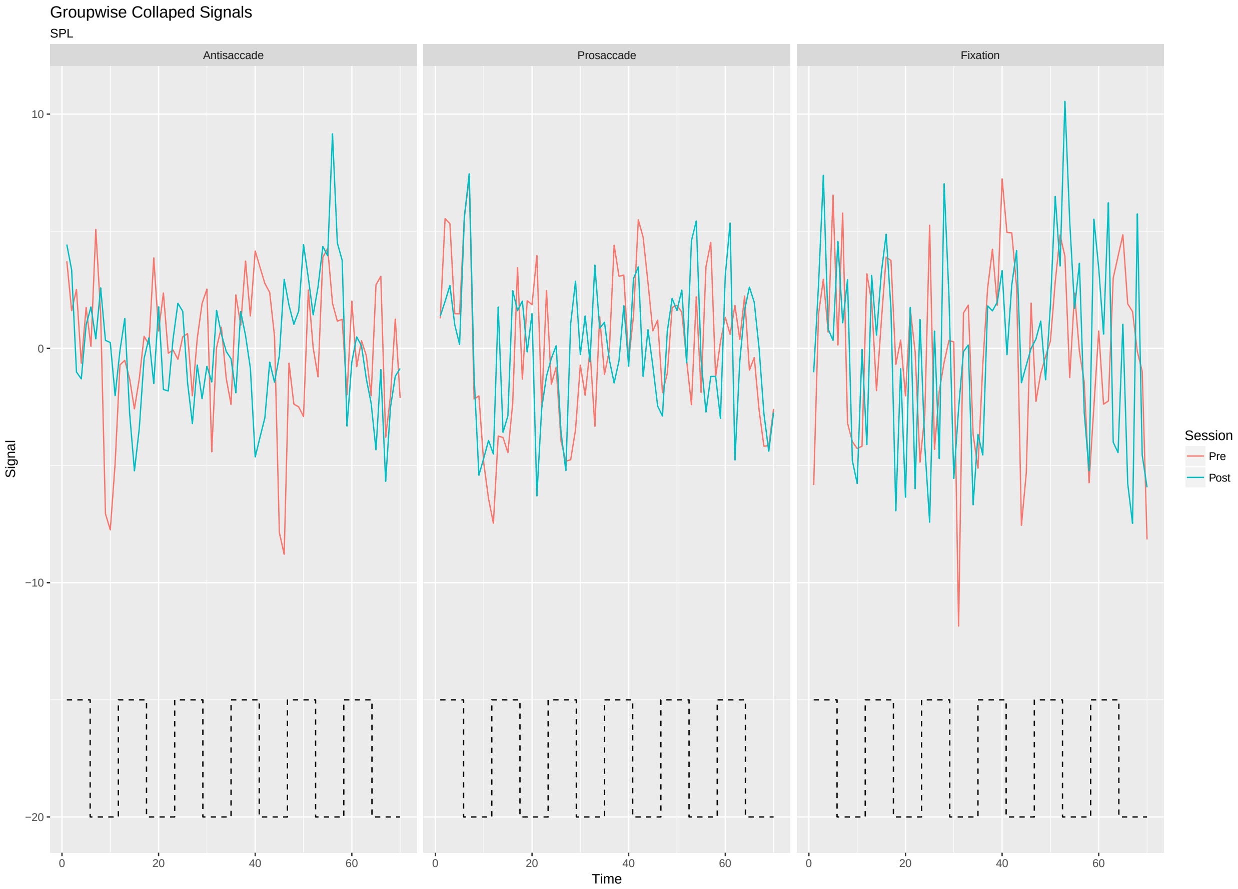


Figure A.2: Collapsed signals for each group and each session from the supplementary eye fields region. The dashed line represents the antisaccade and fixation blocks.

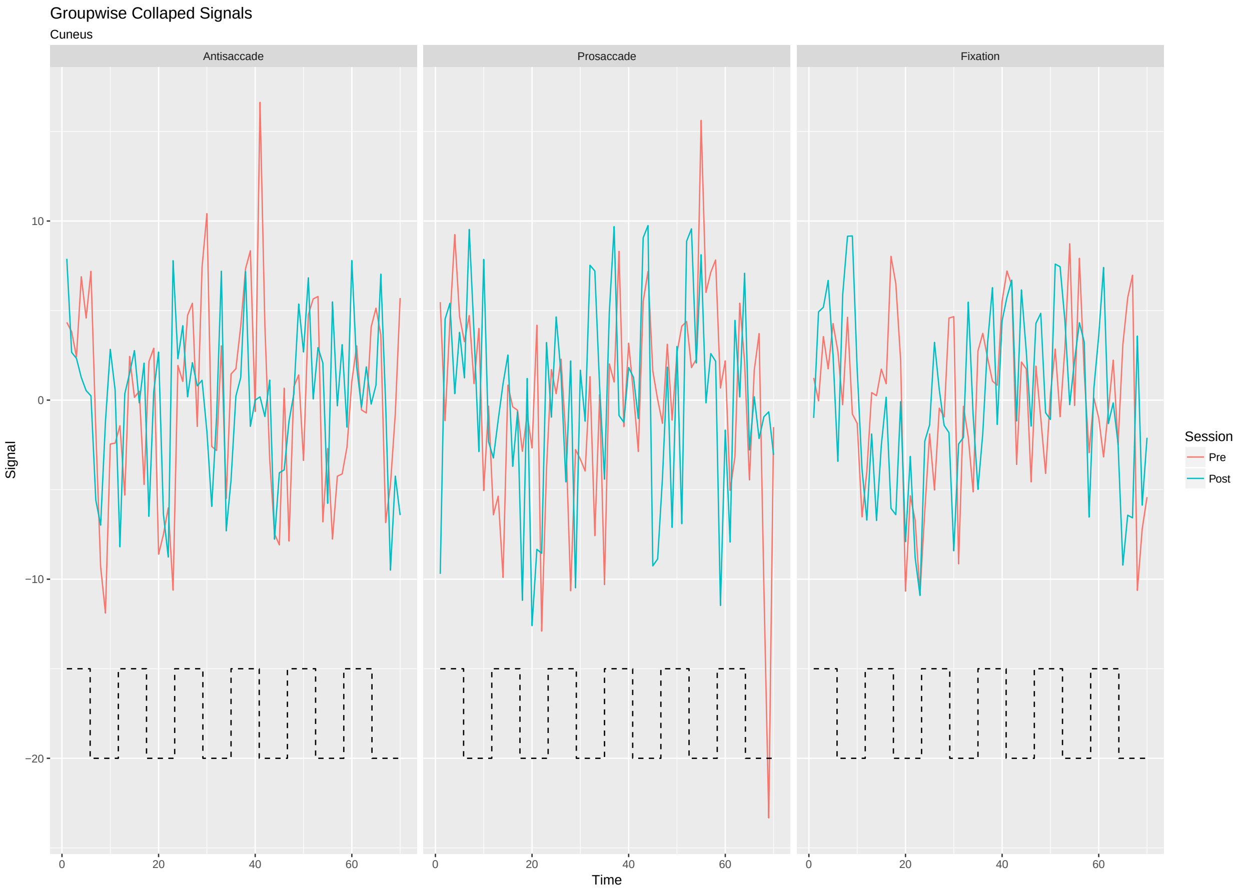


Figure A.3: Collapsed signals for each group and each session from the cuneus region. The dashed line represents the antisaccade and fixation blocks.

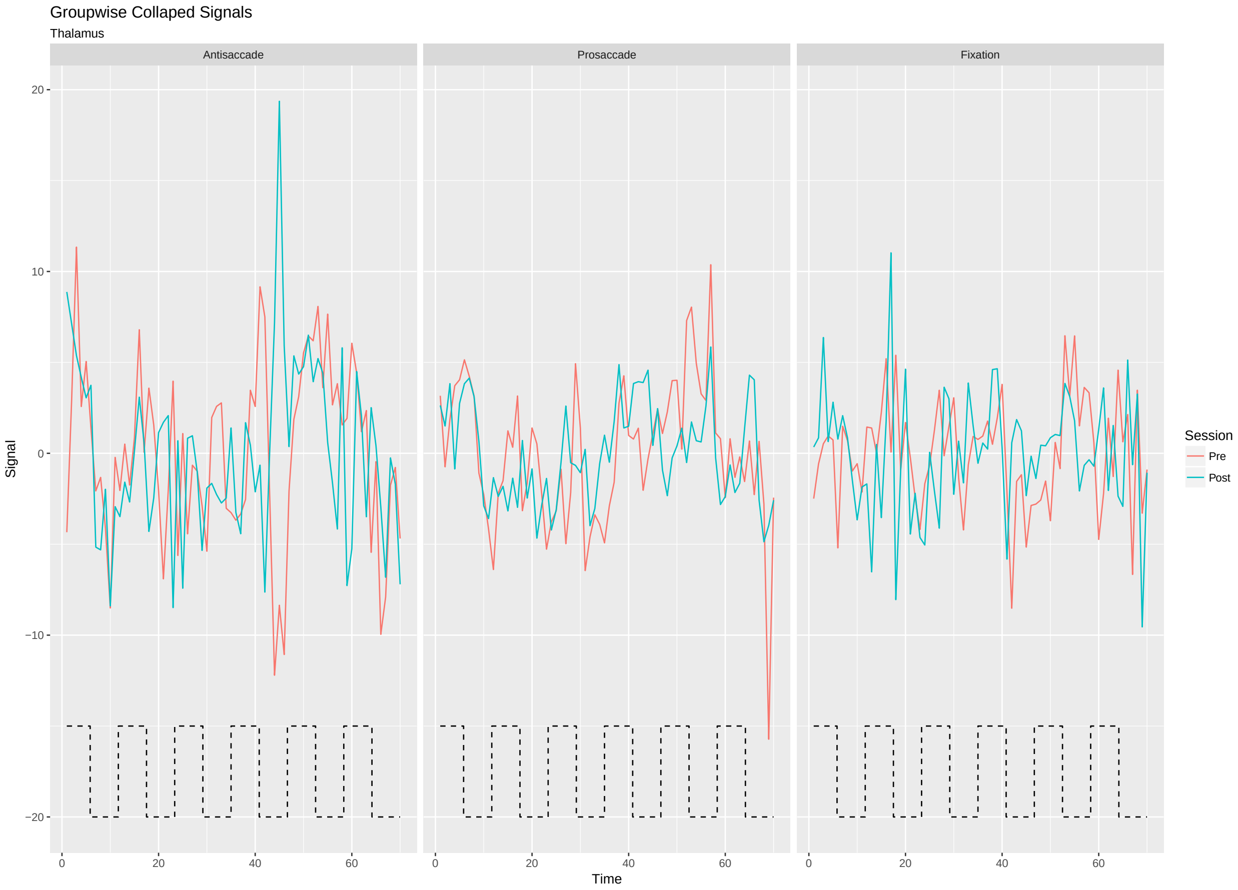


Figure A.4: Collapsed signals for each group and each session from the thalamus region. The dashed line represents the antisaccade and fixation blocks.

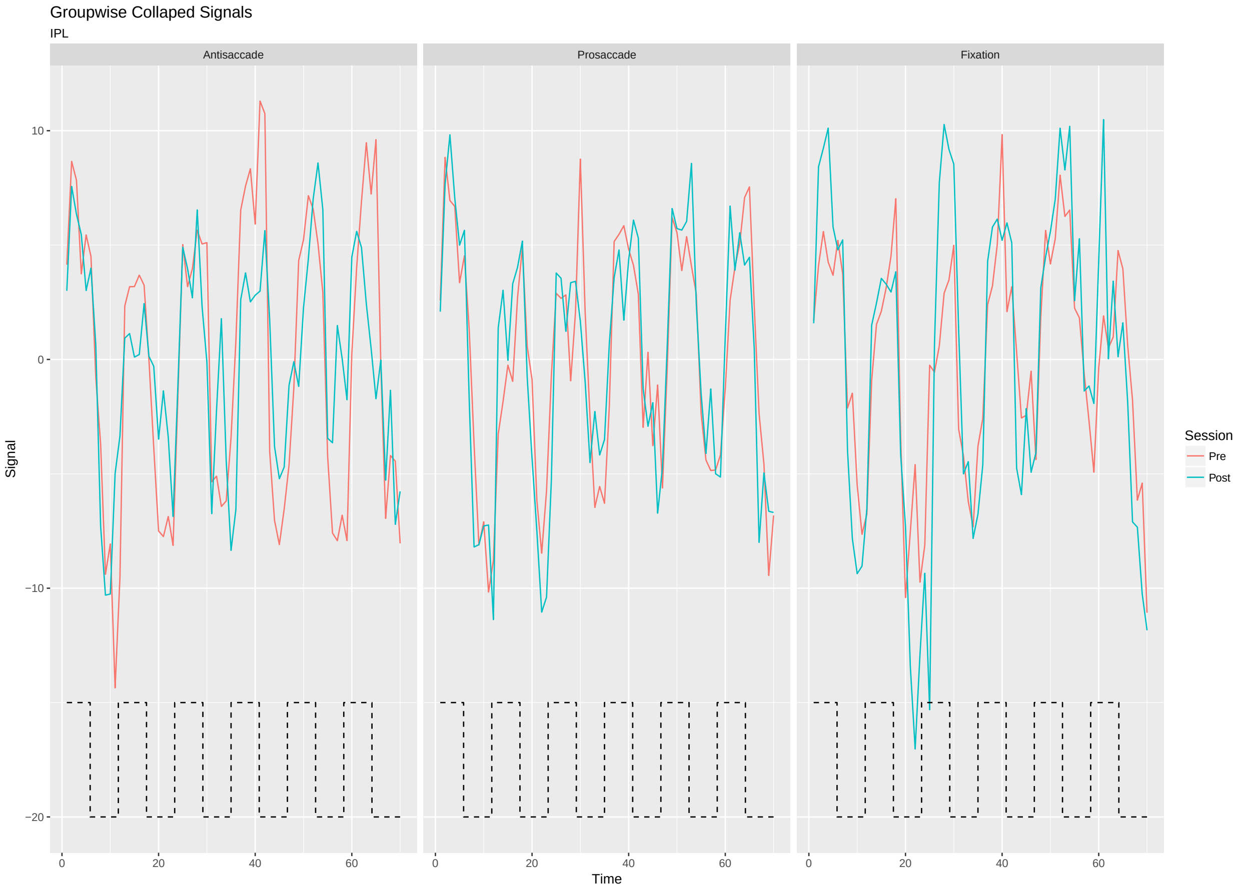


Figure A.5: Collapsed signals for each group and each session from the inferior parietal lobule region. The dashed line represents the antisaccade and fixation blocks.

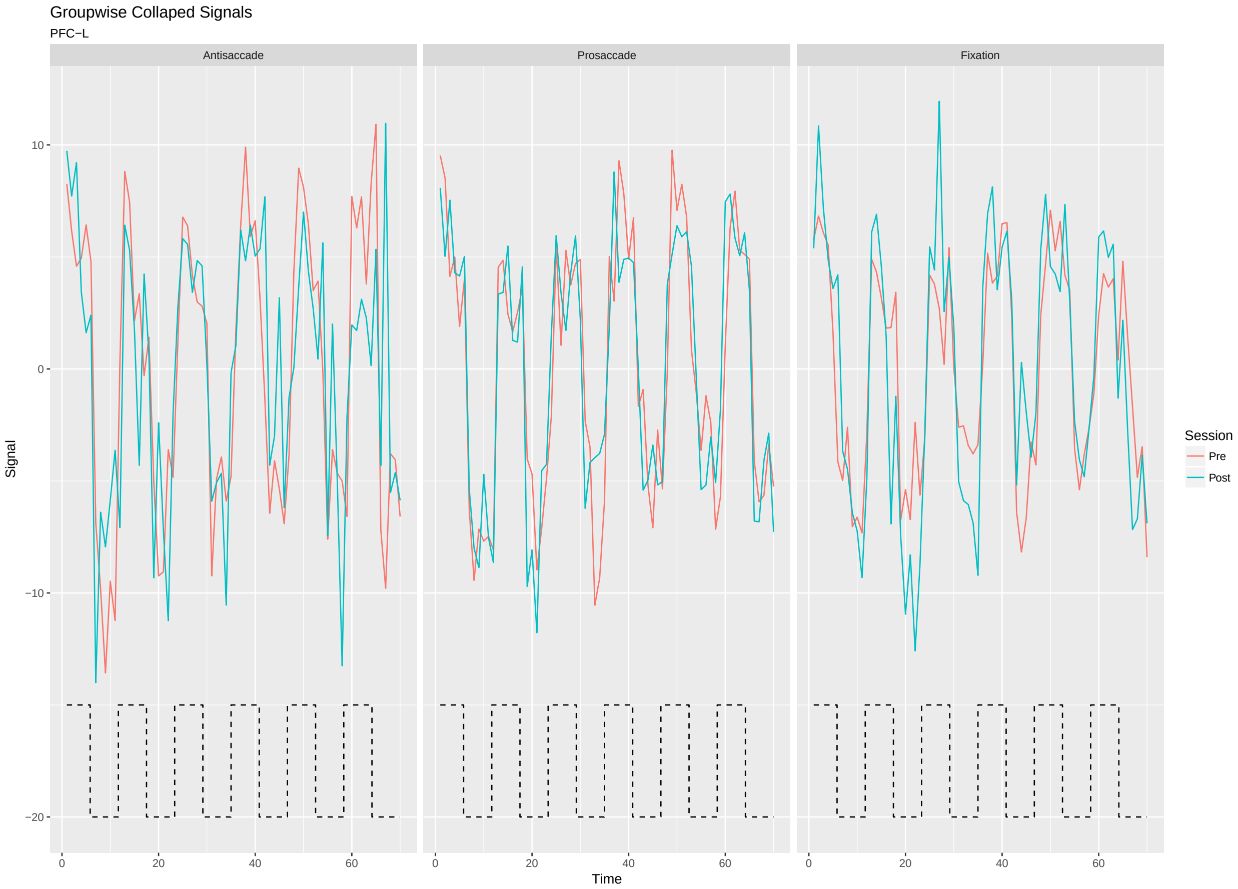


Figure A.6: Collapsed signals for each group and each session from the left prefrontal cortex region. The dashed line represents the antisaccade and fixation blocks.

Figure A.7: Collapsed signals for each group and each session from the right prefrontal cortex region. The dashed line represents the antisaccade and fixation blocks.

Figure A.8: Collapsed signals for each group and each session from the middle occipital gyrus region. The dashed line represents the antisaccade and fixation blocks.

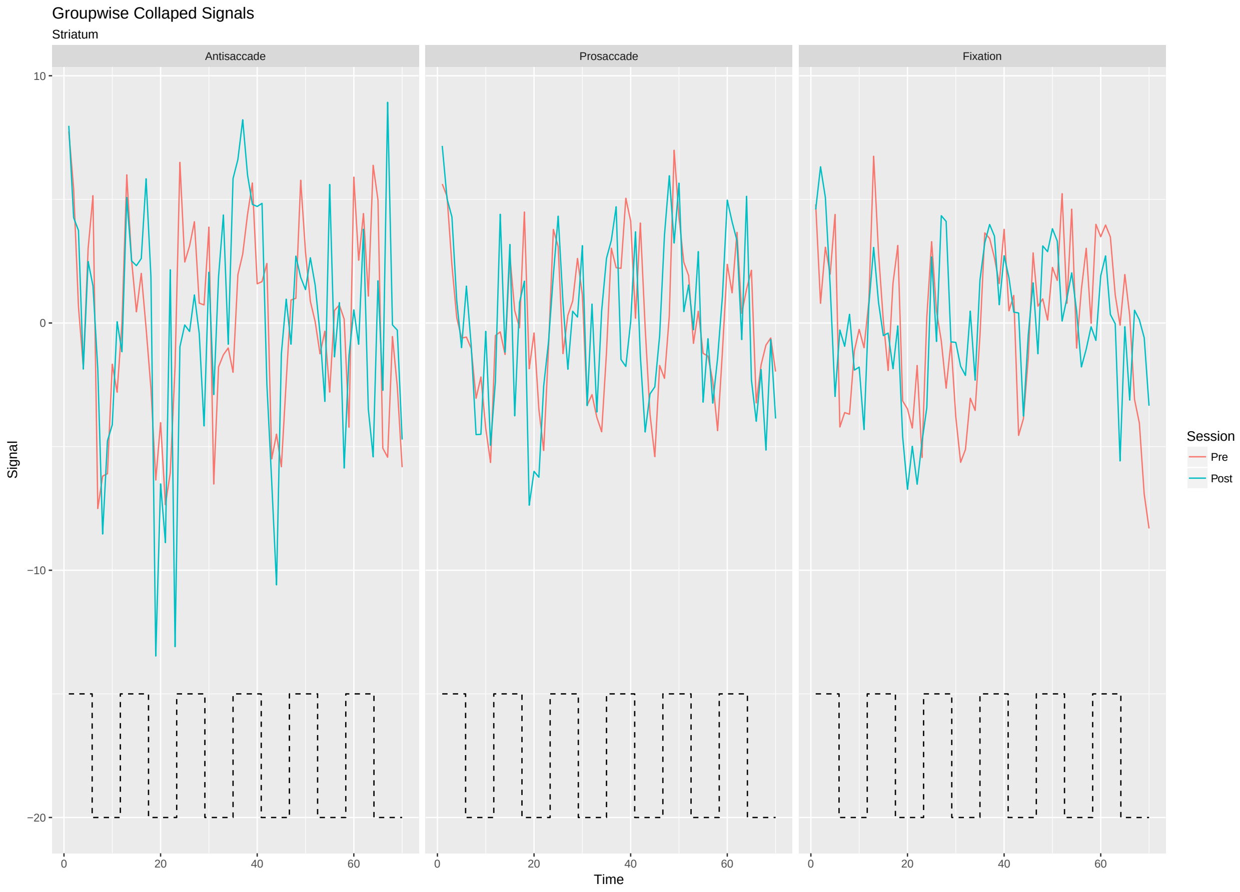


Figure A.9: Collapsed signals for each group and each session from the striatum (basal ganglia) region. The dashed line represents the antisaccade and fixation blocks.

Figure A.10: Collapsed signals for each group and each session from the right inferior frontal cortex region. The dashed line represents the antisaccade and fixation blocks.

A.2 PAIRWISE

A.2.1 SIMULATION

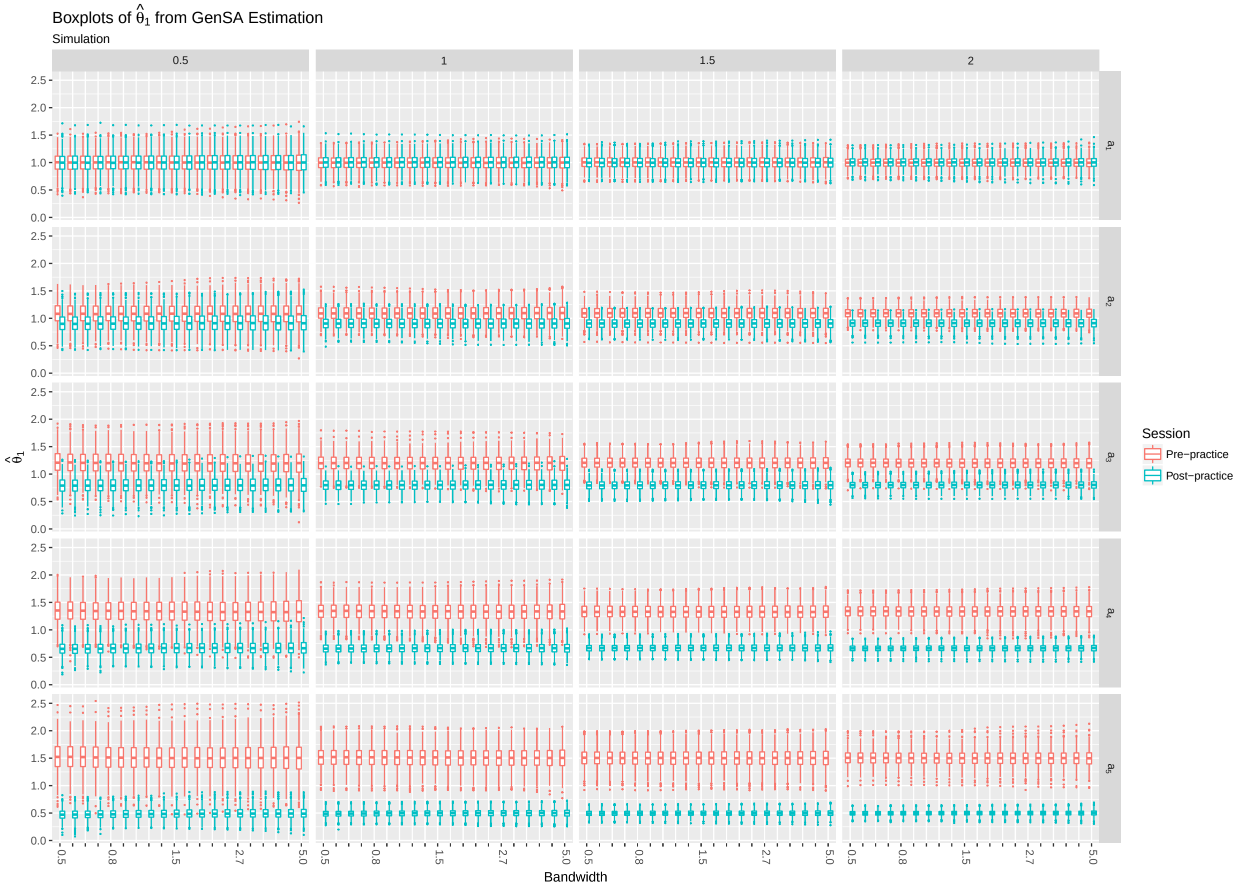


Figure A.11: Boxplots of $\theta_{i,1}$ for pre and post sessions using the Generalized Simulated Annealing method.

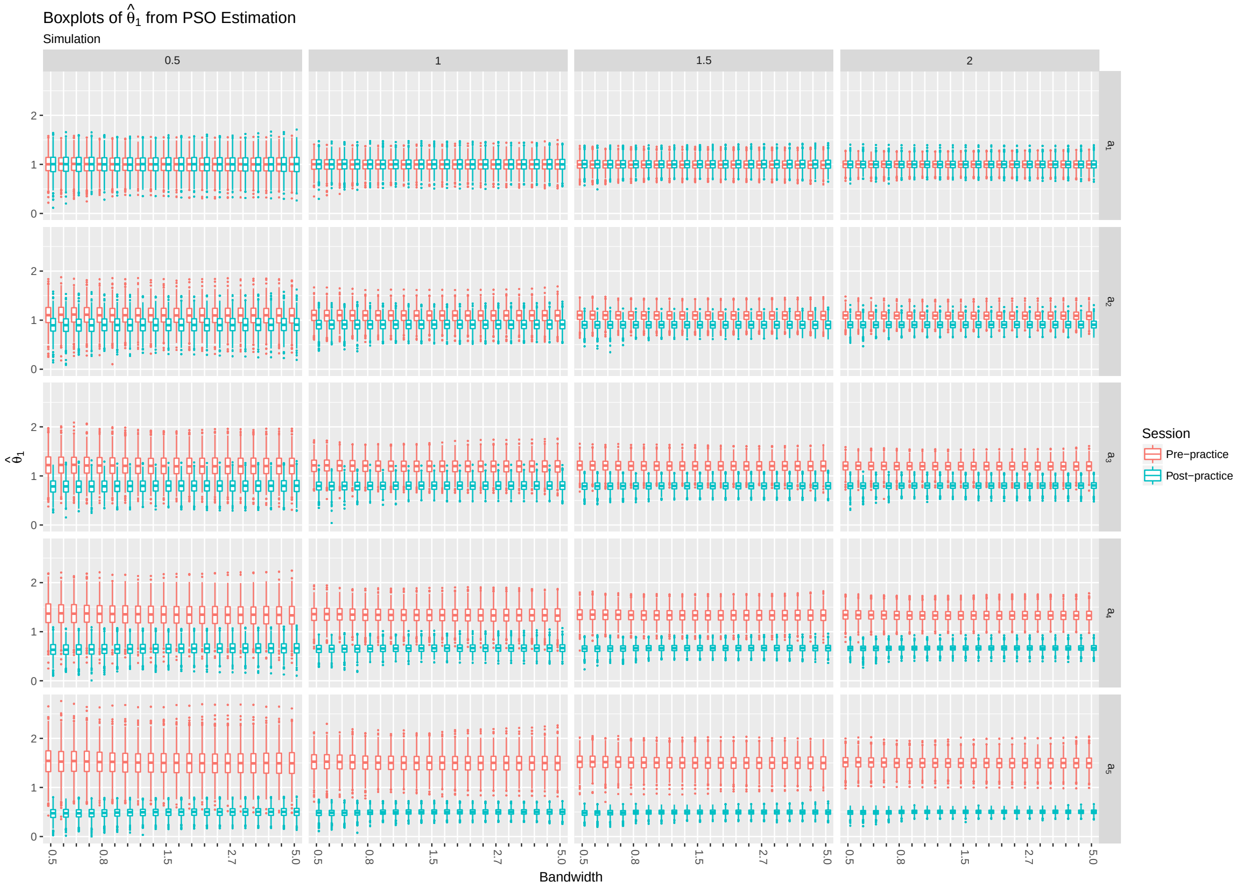


Figure A.12: Boxplots of $\theta_{i,1}$ for pre and post sessions using the Particle Swarm Optimization method.

A.2.2 REAL DATA

GENSA

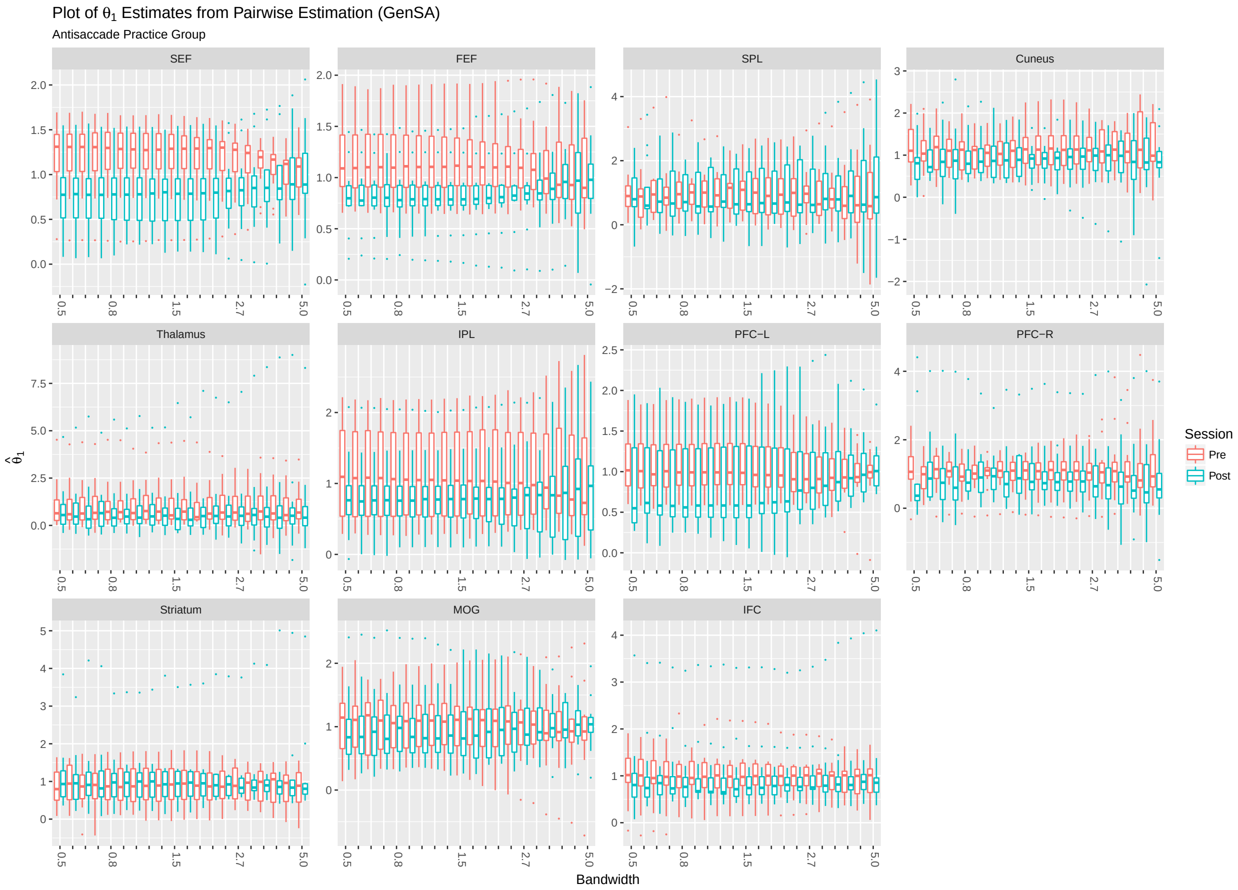


Figure A.13: Boxplots of pairwise estimation of attenuation for the antisaccade practice group using the Generalized Simulated Annealing method.

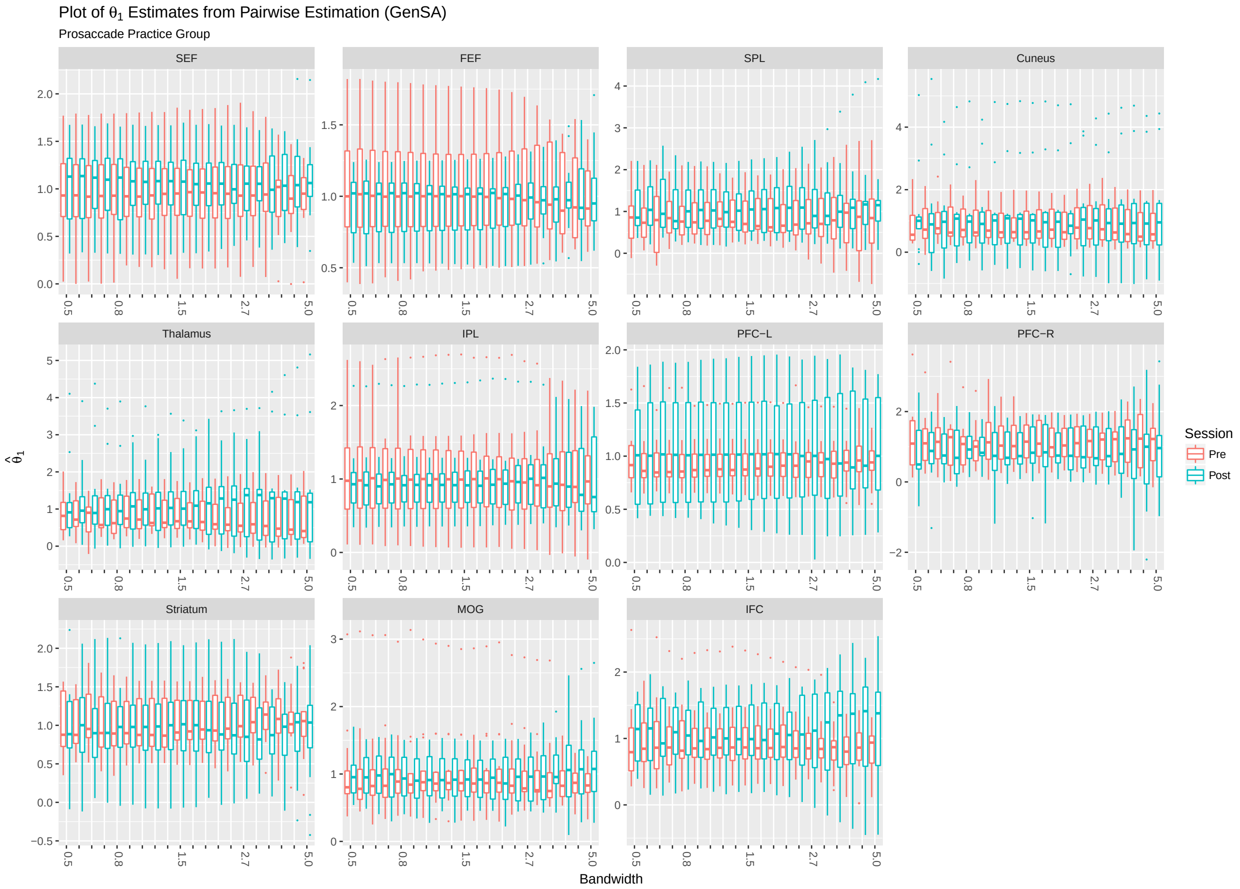


Figure A.14: Boxplots of pairwise estimation of attenuation for the prosaccade practice group using the Generalized Simulated Annealing method.

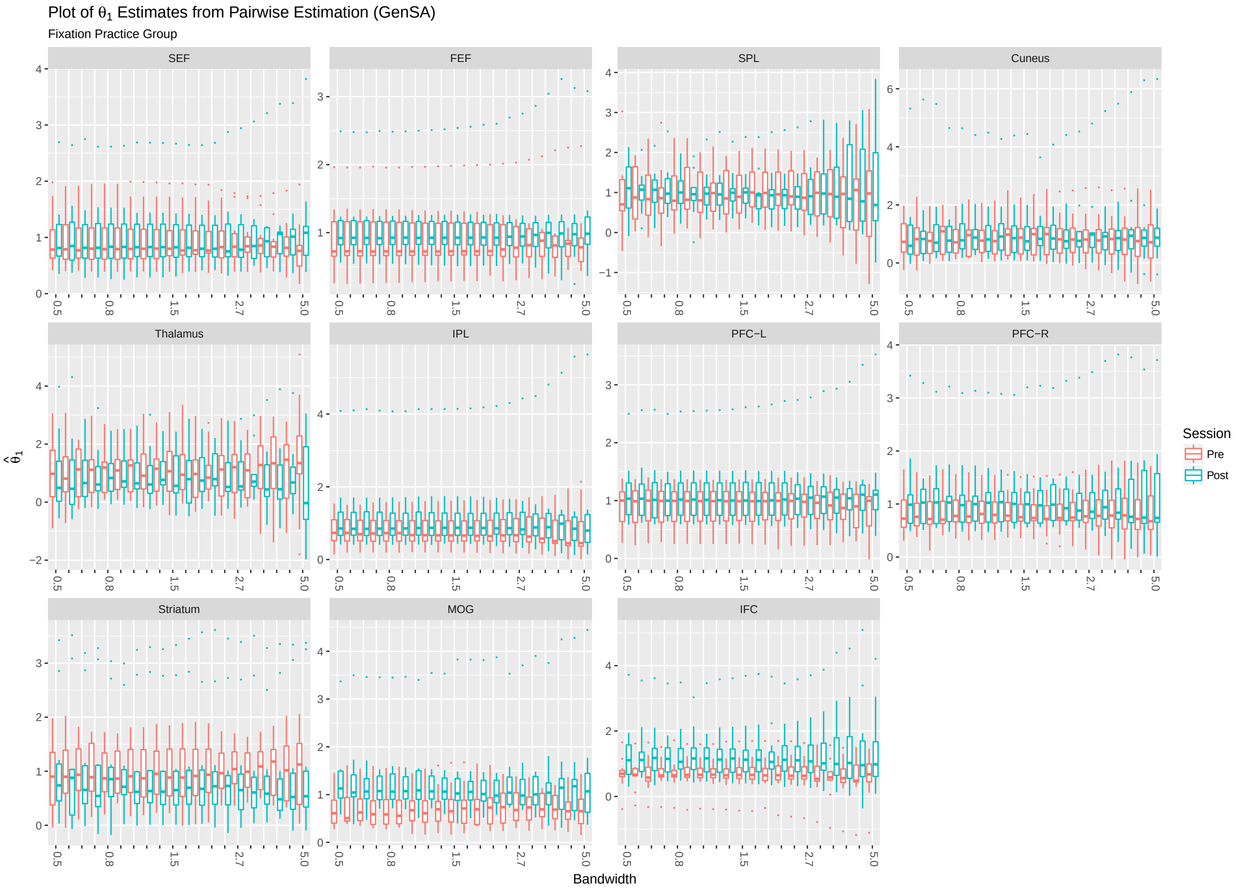


Figure A.15: Boxplots of pairwise estimation of attenuation for the fixation practice group using the Generalized Simulated Annealing method.

PSO

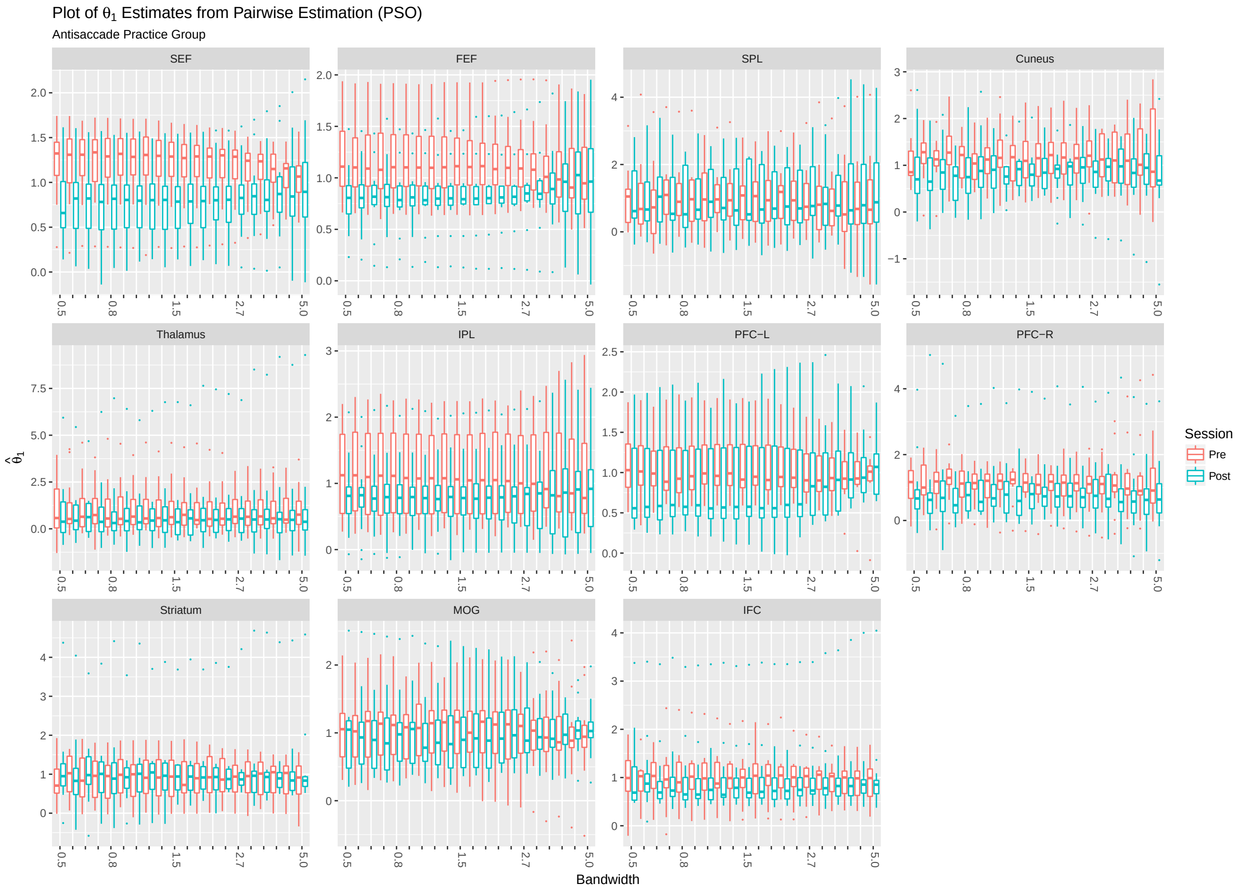


Figure A.16: Boxplots of pairwise estimation of attenuation for the antisaccade practice group using the Particle Swarm Optimization method.

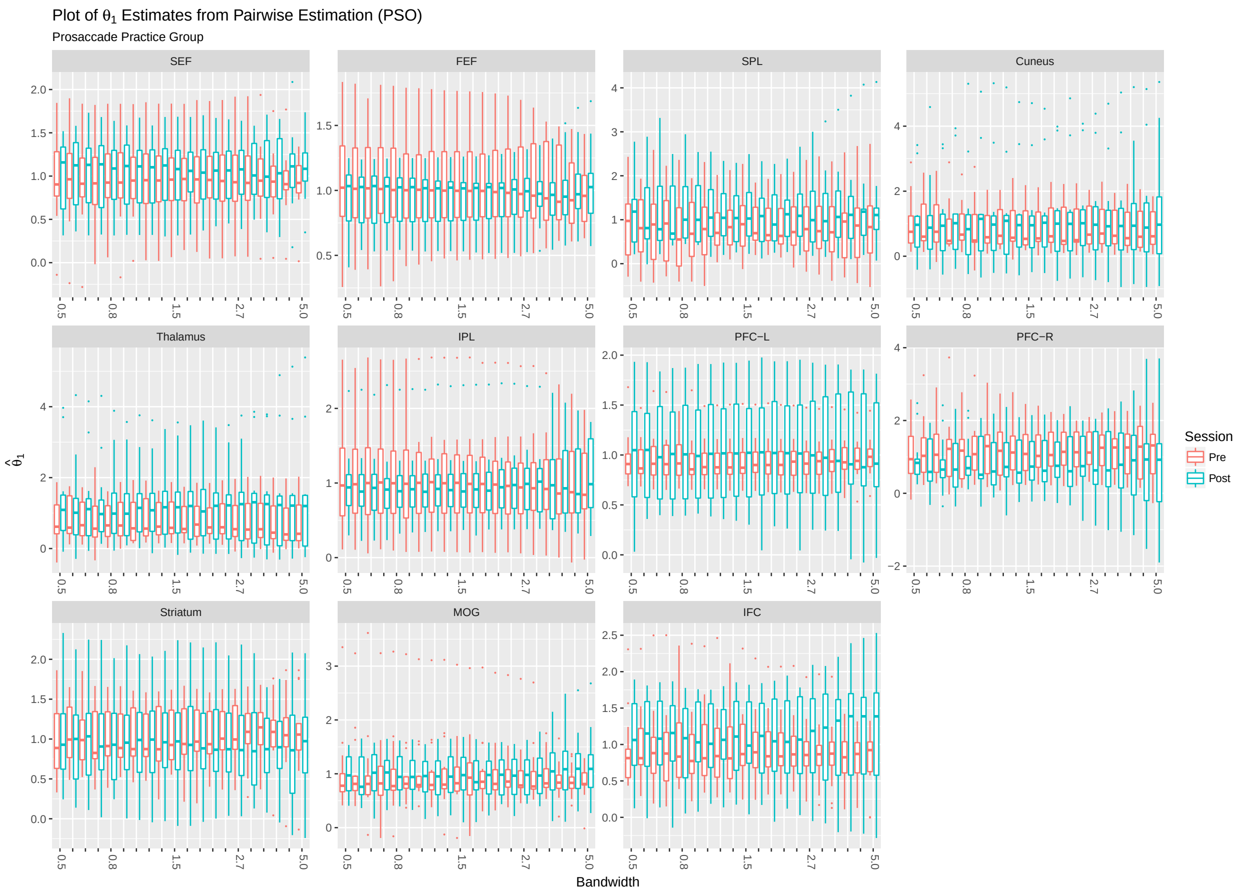


Figure A.17: Boxplots of pairwise estimation of attenuation for the prosaccade practice group using the Particle Swarm Optimization method.

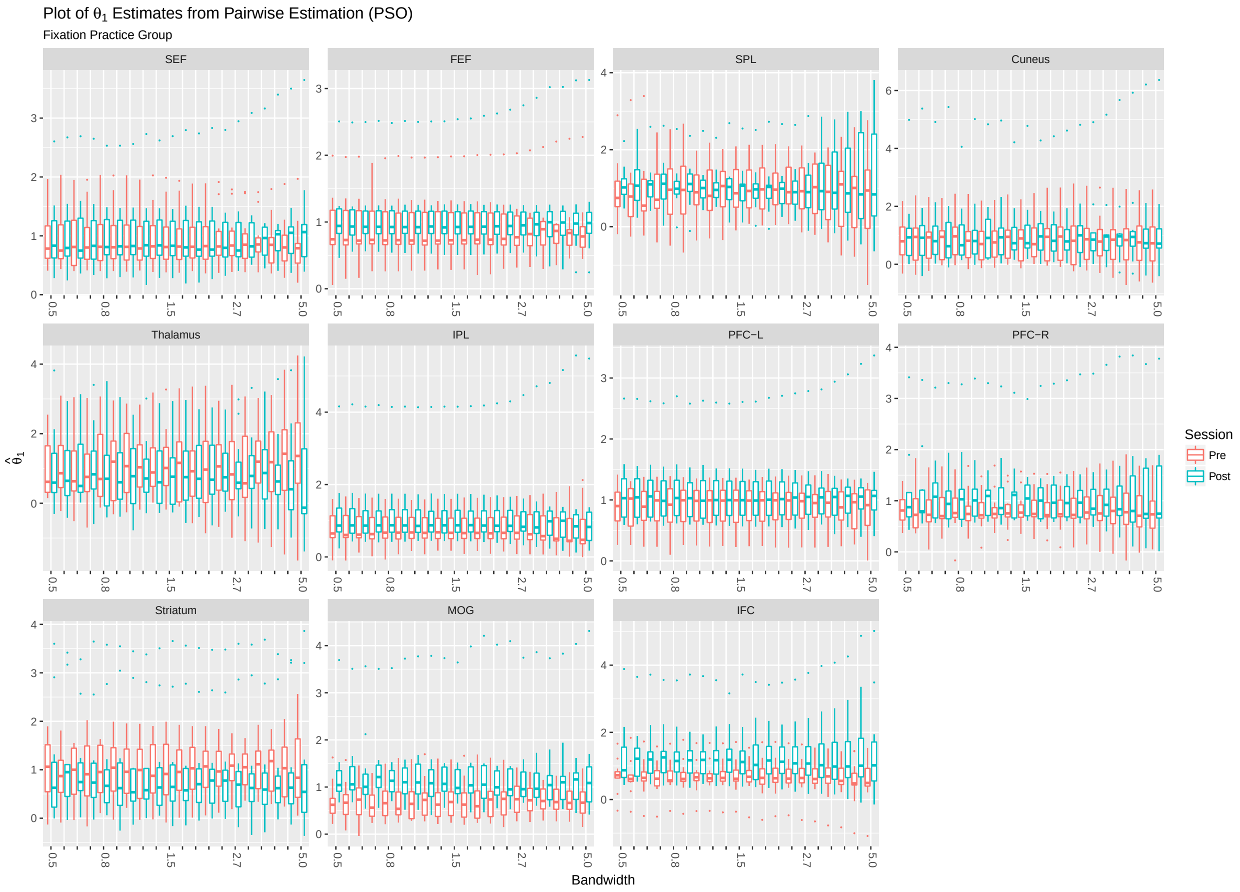


Figure A.18: Boxplots of pairwise estimation of attenuation for the fixation practice group using the Particle Swarm Optimization method.

A.3 THREE GROUP ESTIMATION

A.3.1 SIMULATION

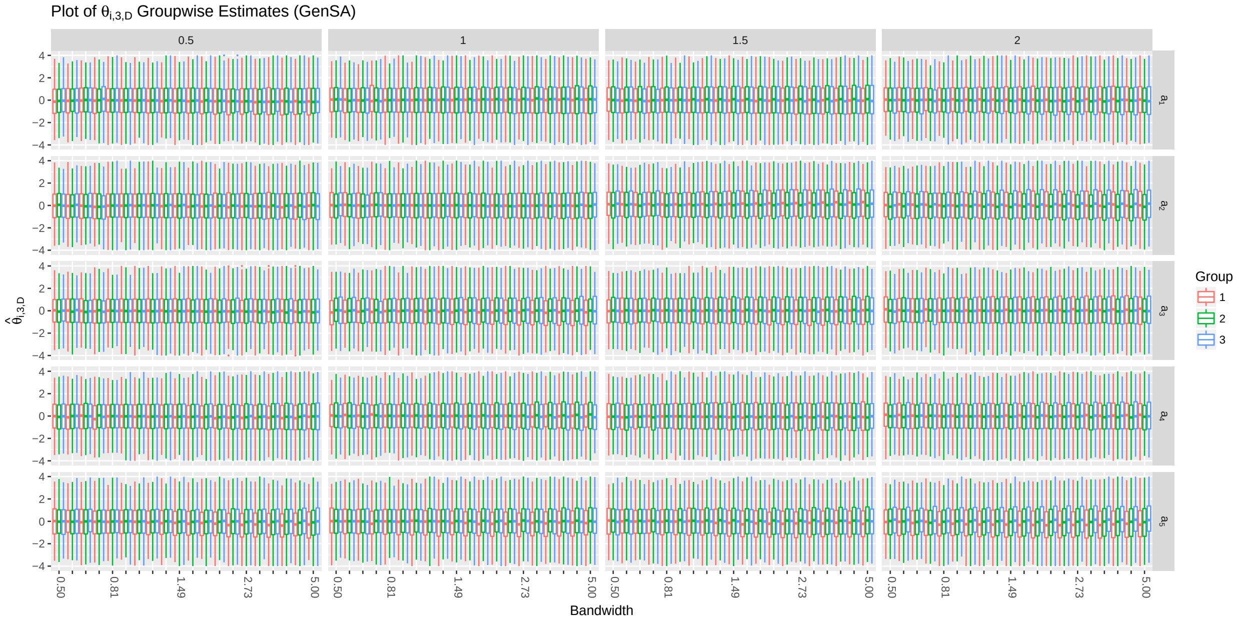


Figure A.19: Boxplots of $\theta_{i,3,D}$ Estimates from Generalized Simulated Annealing and three group estimation from Section 5.2.1.

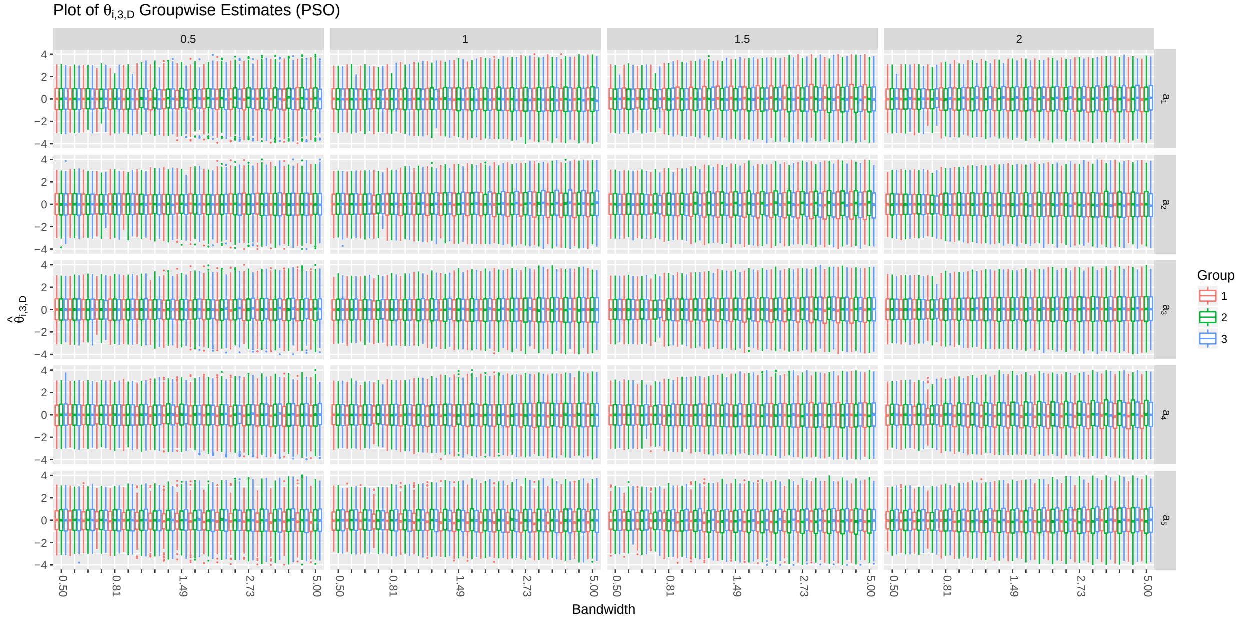


Figure A.20: Boxplots of $\theta_{i,3,D}$ Estimates from Particle Swarm Optimization and three group estimation from Section 5.2.1.

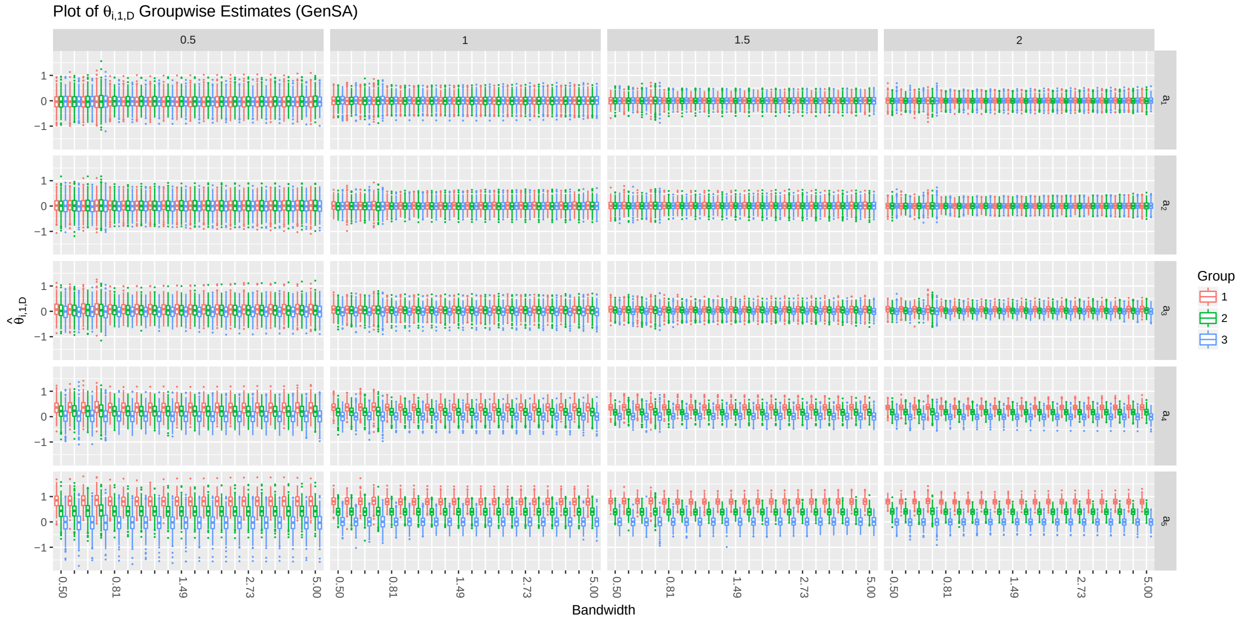


Figure A.21: Boxplots of $\theta_{i,1,D}$ Estimates from Generalized Simulated Annealing and three group estimation from Section 5.2.1.

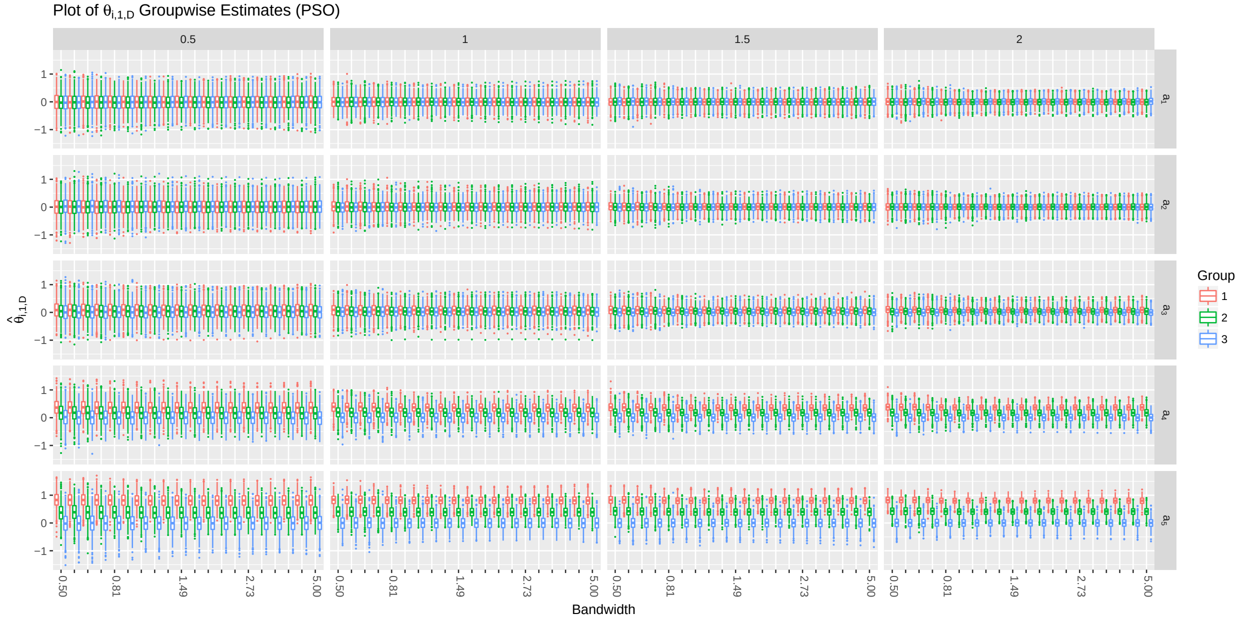


Figure A.22: Boxplots of $\theta_{i,1,D}$ Estimates from Particle Swarm Optimization and three group estimation from Section 5.2.1.

A.3.2 Real Data

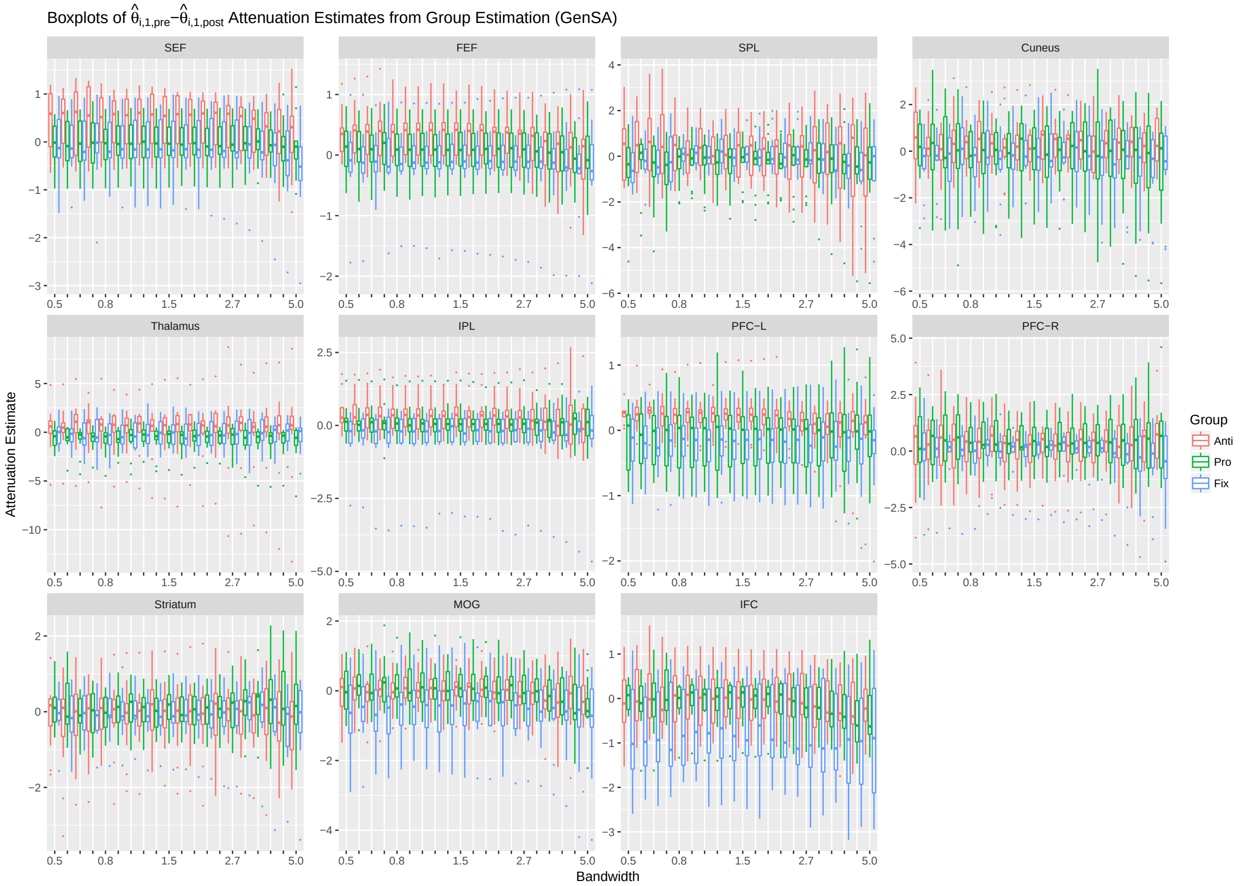


Figure A.23: Boxplots of the three group estimated attenuations using the Generalized Simulated Annealing method.

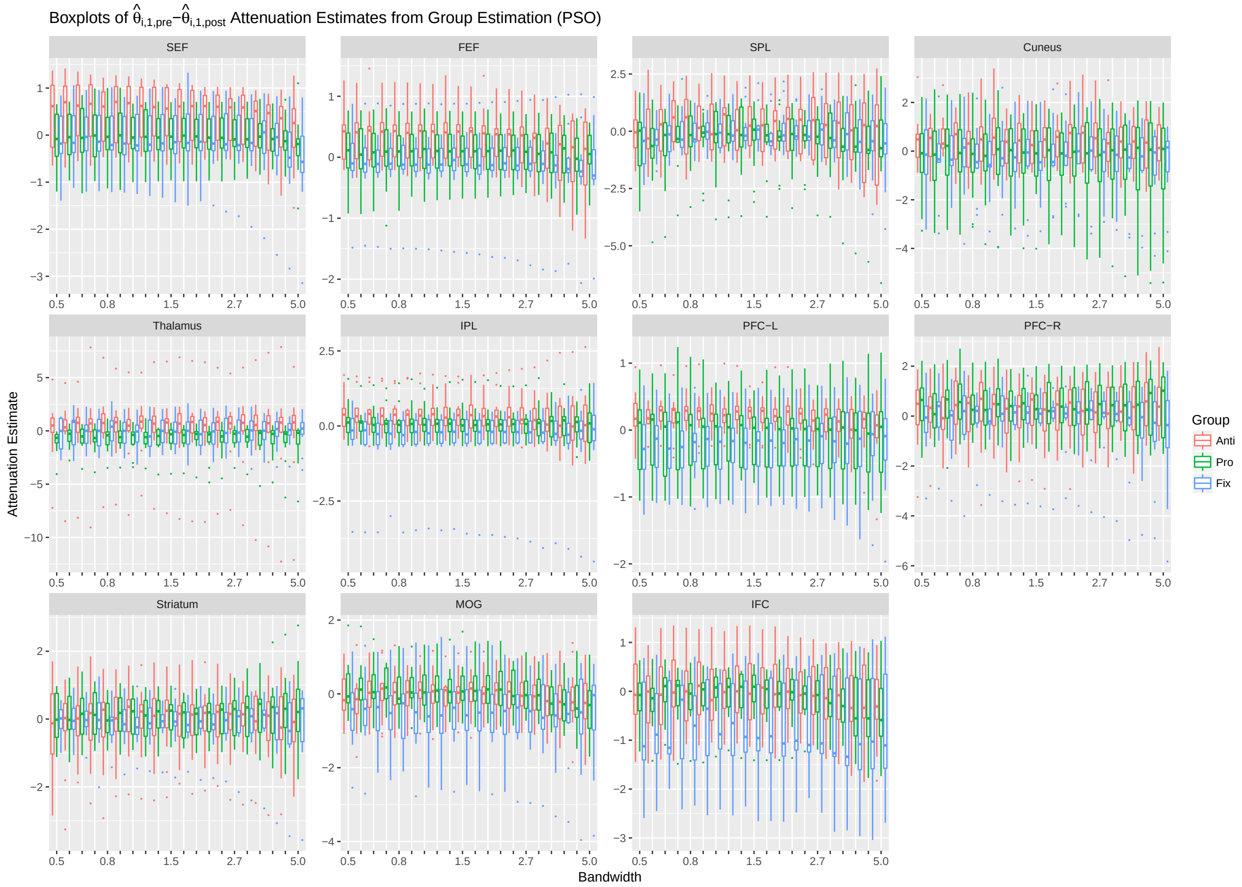


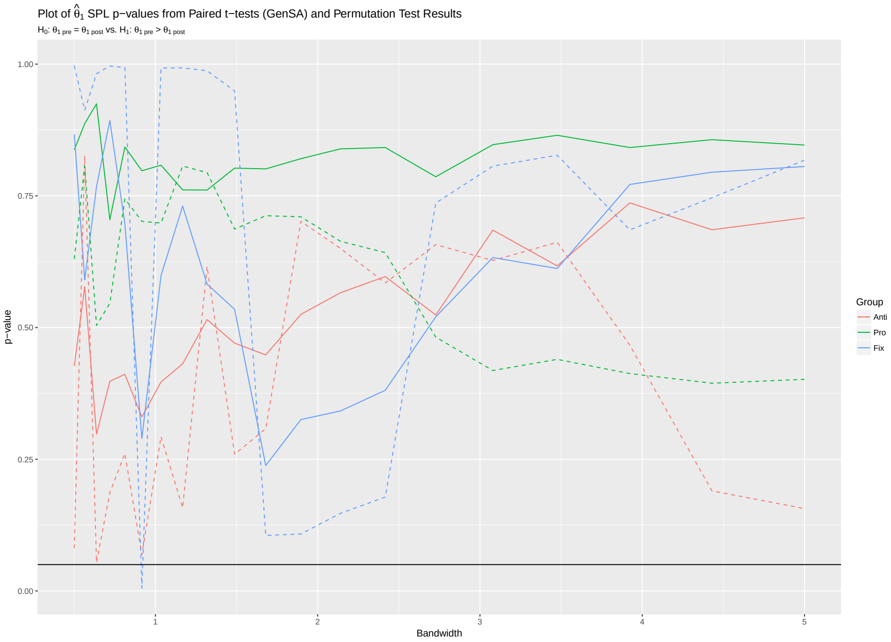
Figure A.24: Boxplots of the three group estimated attenuations using the Particle Swarm Optimization method.

A.4 PERMUTATION TESTS FOR PAIRWISE RESULTS

In this section, we present the results of permutation tests for the pairwise estimations. We take a set of results for a given bandwidth, ROI and practice group. We then shuffle the pre-practice estimates and then perform pairwise t-tests. The solid lines are the original p-values from original analysis and the dashed lines are the results of the permutation tests, or the proportion of permuted test statistics greater than the original test statistic. This process is repeated 10,000 times for each bandwidth, ROI and practice group.

Figure A.25: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the SEF region.

Figure A.26: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the FEF region.



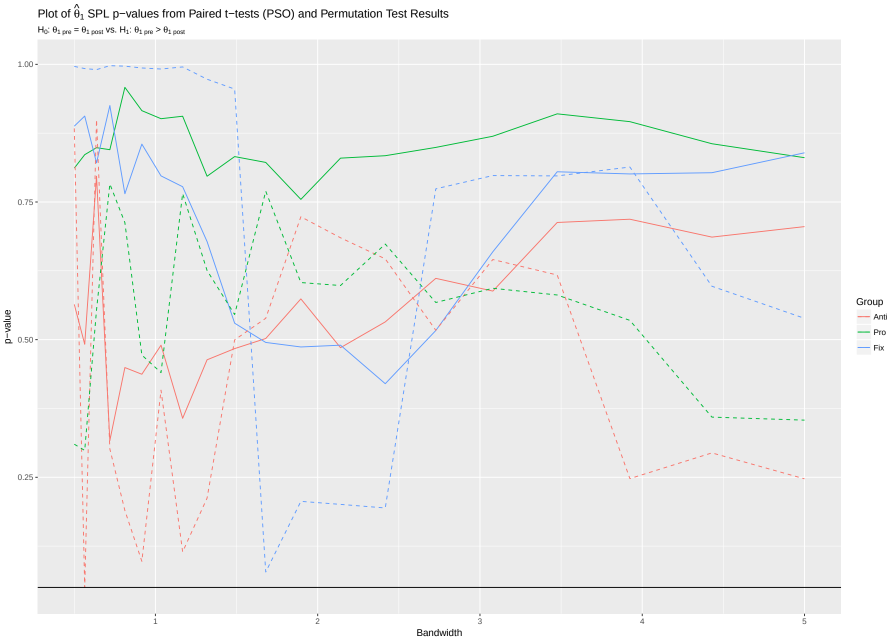


Figure A.27: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the SPL region.

Figure A.28: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the Cuneus region.

Figure A.29: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the Thalamus region.

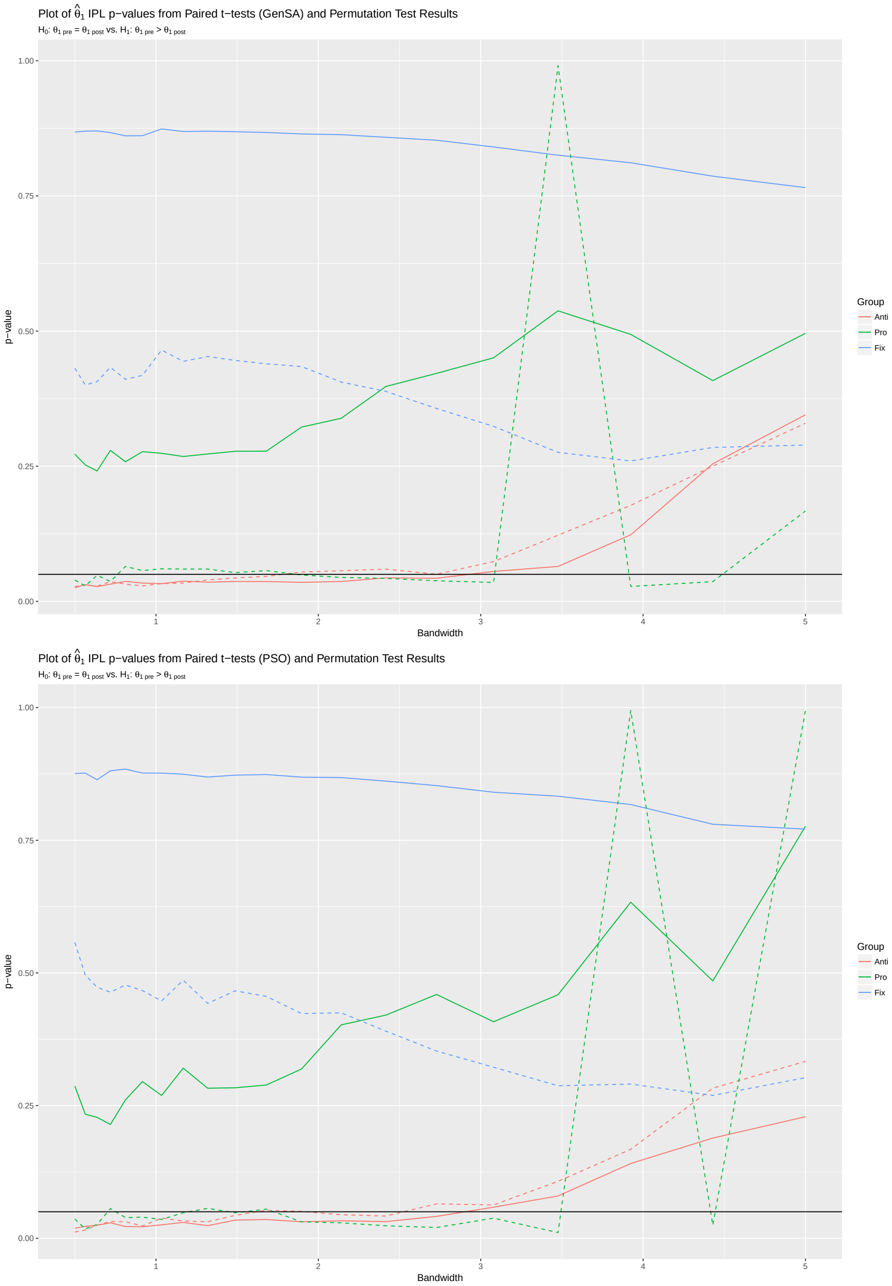
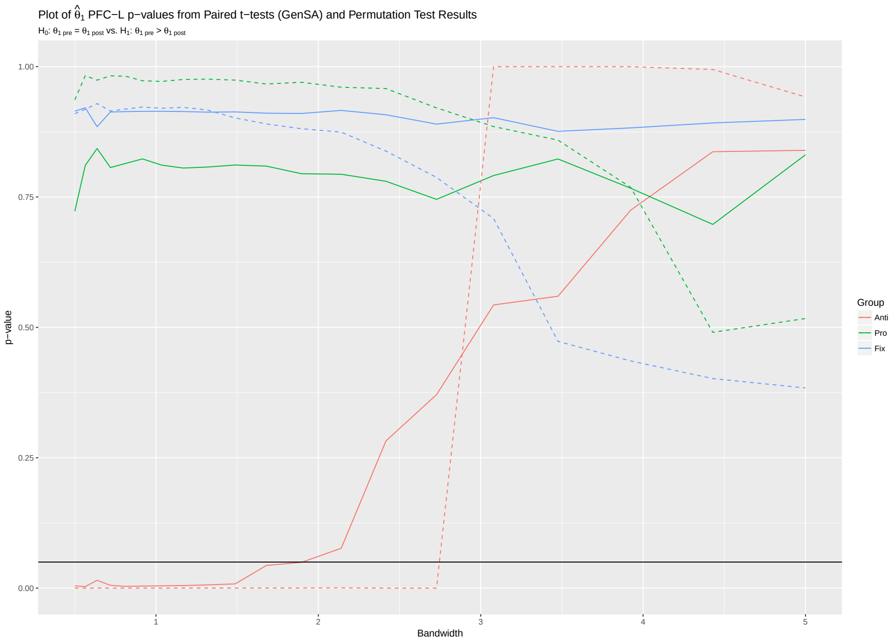


Figure A.30: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the IPL region.



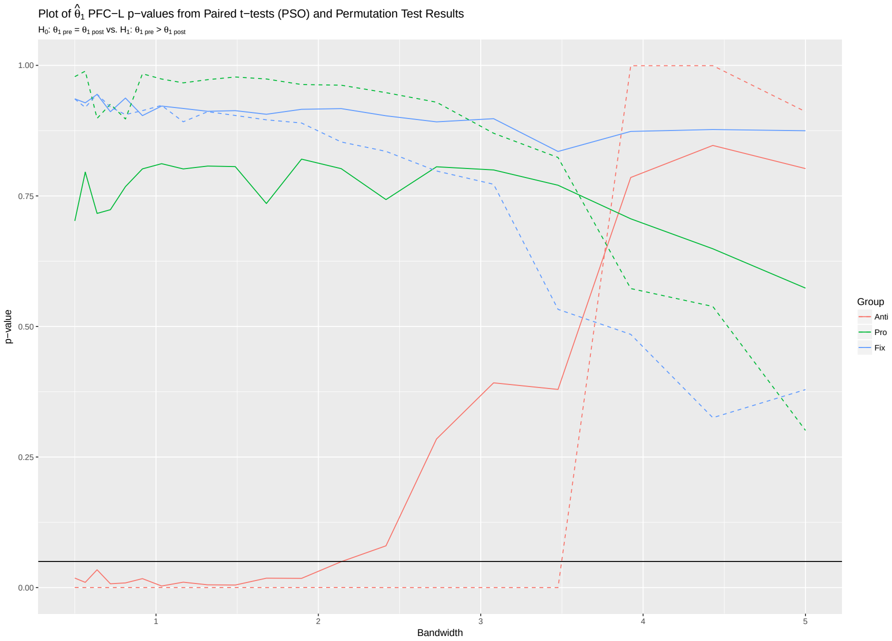
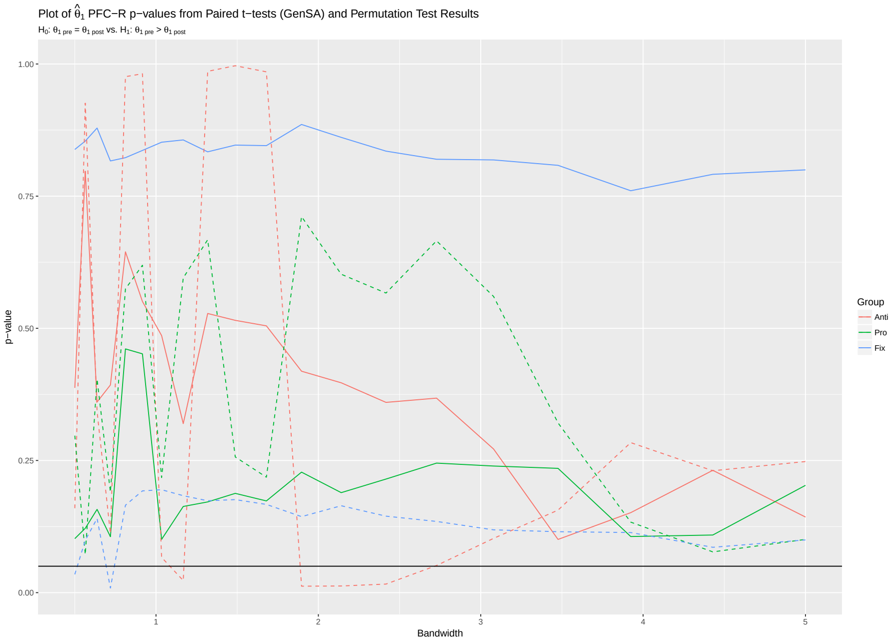


Figure A.31: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the PFC-L region.



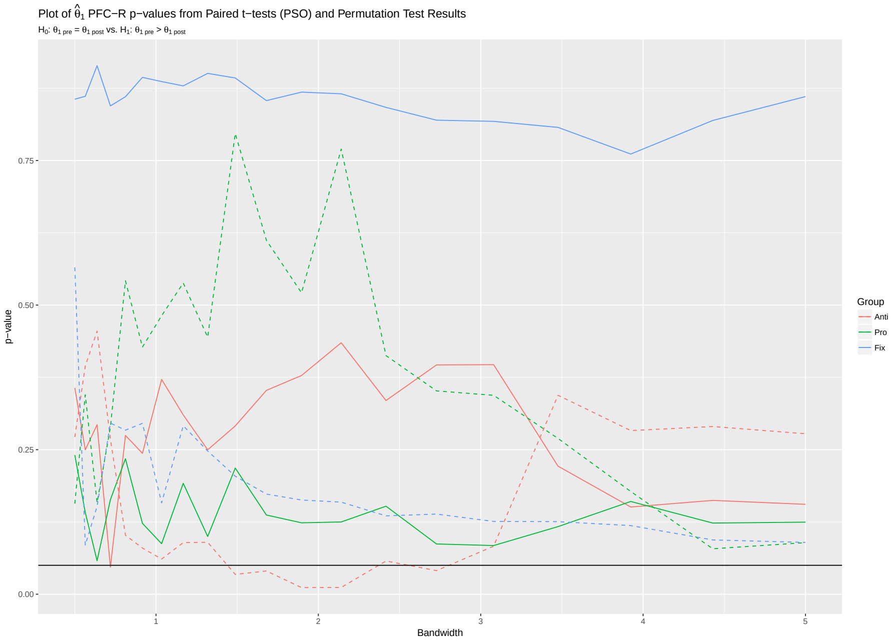


Figure A.32: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the PFC-R region.

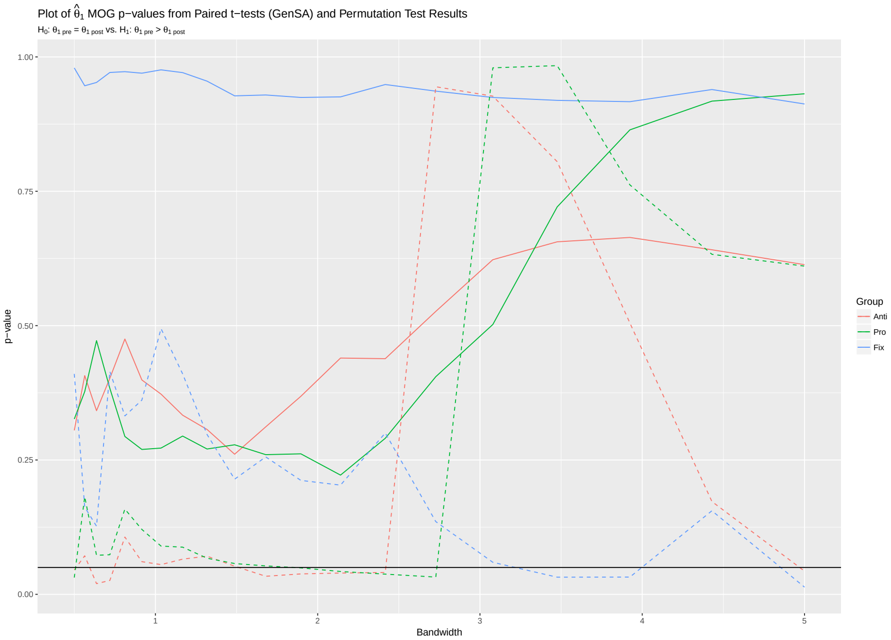
Plot of $\hat{\theta}_1$ Striatum p−values from Paired t−tests (GenSA) and Permutation Test Results

H_0: $\theta_{1\ pre} = \theta_{1\ post}$ vs. H_1: $\theta_{1\ pre} > \theta_{1\ post}$

Plot of $\hat{\theta}_1$ Striatum p−values from Paired t−tests (PSO) and Permutation Test Results

H_0: $\theta_{1\ pre} = \theta_{1\ post}$ vs. H_1: $\theta_{1\ pre} > \theta_{1\ post}$

Figure A.33: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the Striatum region.



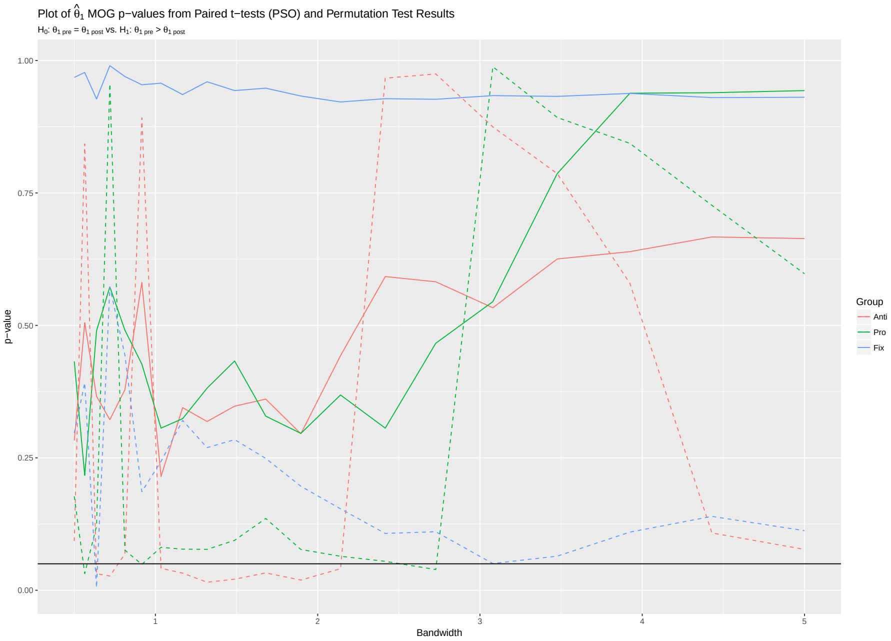


Figure A.34: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the MOG region.

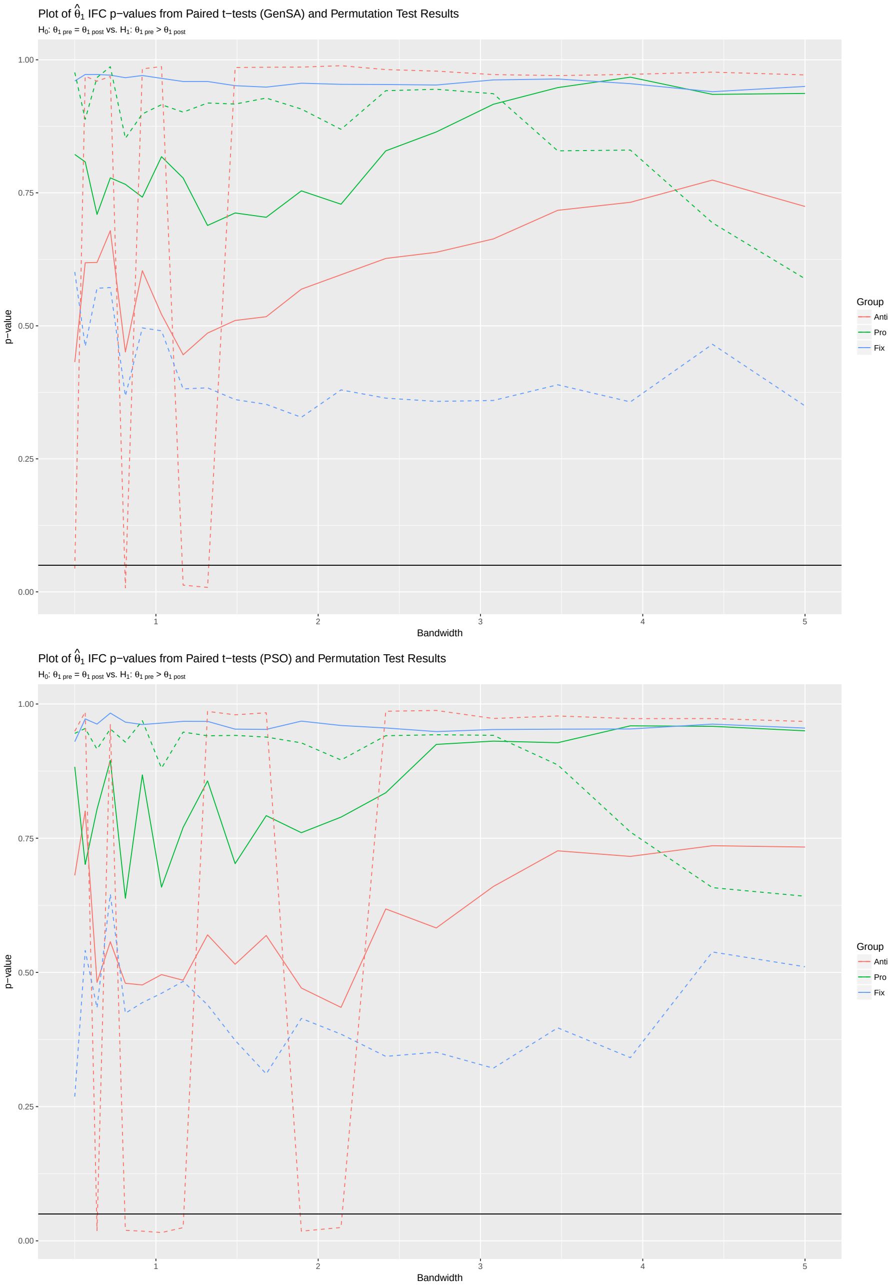


Figure A.35: These are plots of t-test p-values (solid lines) and permutation test results (dashed lines) for the IFC region.