

# Reproducing fMRI Data Analysis on Brain Connectivity

Li, Jie  
Jay4869

Li, Zeyu  
lizeyuyuz

Yun, Chuan  
ay2456

Zhang, Qiangyuan  
amandazhang

December 14, 2015

## Abstract

In this research paper, we analyze the fMRI data from the study done by Repov and Barch exploring the relationship between schizophrenia and brain connectivity [2, 3]. Instead of reproducing ANOVA analysis of within and between brain network connectivity, we focus on analyzing fMRI scans of one subject using generalized linear model (GLM). Before data analysis, we smoothed the data spatially and corrected noise in the fMRI data. On the preprocessed data, we performed student t tests to identify the voxels responding to designed stimulus. Having localized the activated voxels, we closely examined the behavior of these voxels and validated our assumptions through various statistical tests.

## 1 Introduction

The paper of which our research is based is titled “Working memory related brain network connectivity in individuals with schizophrenia and their siblings” [2, 3]. Schizophrenia is a chronic, severe, and disabling brain disorder. Previous studies have shown that changes in the function of a single brain region, or even a brain system, cannot explain the functional impairments seen in this illness [1, 2]. However, Repovs and Barch attempts to show that individuals with schizophrenia have reduced connectivity within and between neural networks, which could be a forward step to understanding schizophrenia.

Repov and Barch systematically examined changes in functional connectivity across rest and different task states in order to make inferences on characteristics of schizophrenia. They assessed connectivity using blood oxygen level dependent (BOLD) time series acquired using fMRI. They found four types of participants, individuals with schizophrenia, the siblings of individuals with schizophrenia, healthy controls and the siblings of healthy controls. They designed three working memory loads of an N-back task and designated four regions of interest (ROIs). The four brain networks are: (1) Dorsal fronto-parietal network (FP), (2) Cingulo-opercular network (CO), (3) Cerebellar network (CER) and (4) “Default mode” network (DMN). The objective of the study is to examine the altered functional connectivity within and between these four brain networks when the participant perform a designated N-back task. With ANOVA analysis, it was found that individuals with schizophrenia and their siblings showed consistent reductions in connectivity between both the FP and CO networks with the CER network.

We narrowed our focus of analysis on a single schizophrenic subject. The fMRI data in the study shows the changes in blood flows in the brain. Analyzing this data can help us understand how schizophrenic brains respond during task performances. We detailed in the following the steps we took to analyze the data set. We attempted both spatial and temporal smoothing on the voxels in order to reduce noise variance. Having preprocessed the data, we fit a multiple linear regression model using two different sets of experiment conditions. Having obtained coefficient estimates, we performed a Student t-test on each voxel to examine whether the particular voxel has show significant activity subject to appropriate thresholds. Performing t-tests relying on a set of assumptions that we need to validate. We performed the Shapiro-Wilk test to check the normality assumption. In addition, we performed a heteroskedasticity test (White test) that checked the constant variance and linear model assumption. We also performed ANOVA tests that compared different sub models. We performed cross validation (CV) to select the optimal model for our data. And lastly, we compared 0-back and 2-back tests through t-test.

## 2 Data

The fMRI data used in the published paper is available on the website OpenFMRI.org. There are many data files grouped by subjects, but we will be focusing on only one schizophrenic subject—subject 001, who is male, Caucasian and Schizophrenic. The raw data has four dimensions ( $91 \times 109 \times 91 \times 137$ ), consisting of three dimensional voxels that are moving across time. We removed the first four outlying data along the time dimension, which leaves us with 133 images.

### 2.1 Mean Voxels Mask

We plotted the sample images of the brain from three different perspectives, which gives us a general sense of our data. Since the raw data is noisy, we averaged our data by time and plotted a histogram to check noise variance. The dataset contains a lot of signal noise at the outskirts of the brain. The signals are correlated across several voxels, because nearby brain locations usually have similar response to the task. However, we assume that the noise in the data is mostly independent from one voxel to the next. Therefore, we reduce the noise by smoothing in space using averaging across the independent noise in the voxels. Based on the histogram, we decided to set the threshold to 8000 that leaves out most useless information. Then we extracted the data points larger than 8000 such that noise variances are removed. Moreover, we apply the Gaussian filtering to smooth by 2 standard deviations in all three spatial dimensions.

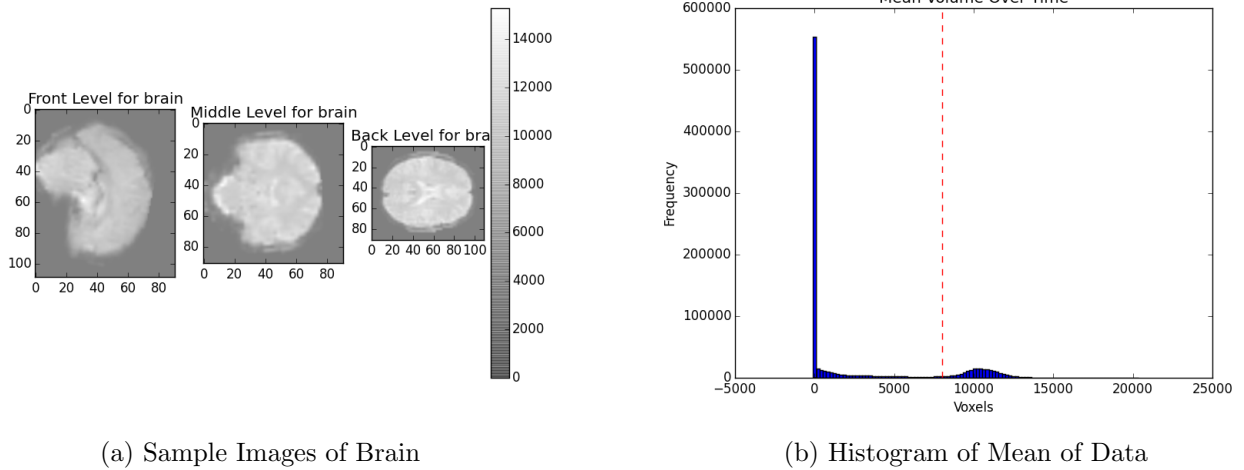


Figure 1: Data Preprocessing

### 2.2 Bonferroni Correction

The Bonferroni correction is a multiple comparison correction method used when several dependent or independent statistical tests are being performed simultaneously. In order to avoid spurious detections, the alpha value would have to be lowered to account for the number of comparisons being performed.

Suppose that the model is true so that  $t_1, \dots, t_n$  have the  $t(n - p - 2)$  distribution. They are not necessarily independent, and we would expect about  $n\alpha$  number of  $t_1, \dots, t_n$  to be larger in absolute value than than  $\alpha/2$  critical value of the  $t(n - p - 2)$ . Therefore, we would be tagging  $n\alpha$  of the subjects as outliers even when there are no outliers in the data.

To deal with this problem, one takes a value of  $\alpha$  much smaller than 0.05 while testing whether the  $i$ th observation is an outlier or not. A particularly conservative value is  $\alpha = 0.05/n$ . In this case, one can show that the probability that at least one subject is tagged as an outlier when in fact there are none is almost 0.05.

### 3 Convolution and Design Matrix

We need a good design matrix, containing data on the explanatory variables, to be used in our linear models to explain as much variation in the observed fMRI data as possible. We studied the design of the N-back experiment closely. fMRI scans were acquired while participants perform specified memory tasks. They are to respond to each letter shown as to whether it was the same as a pre-specified letter (0-back), the same as the immediately preceding letter (1-back), or the same as the letter shown two trials previously (2-back). For our subject, there were three BOLD runs, each consisting of two blocks of 0-back, 1-back, or 2-back working memory task.

We examined the seven study conditions closely. The conditions are unusual from what we are familiar with in that they are consisted of fractions of durations and negative amplitudes. The first condition consists of the start-cues, the second condition consists of the task targets (amplitudes are all one), the third condition consists of the task non-target (the positive amplitude indicates correct identification of the target and the negative amplitude indicates the non-response to the non-target), the fourth condition consists of end-cues, and the fifth condition consists of durations of the two task blocks. There is no information on what the sixth file mean, and the seventh condition consists of the erroneous responses. All figures about convolution response are in the Appendix.

Since we were only familiar with convolution for block design, we initially used only the condition one, four and five, which give us the start time, the duration, and end time of the two task blocks. Due to the delay of signal after the onset of neural activity, we make use of the hemodynamic response function (HRF). The idea is to model the signal response as the convolution of the stimulus function with the HRF. Convoluting with stimulus timing allows us to get idealized response as an explanatory variable (a regressor included in design matrix). We bear in mind the limitations of this design matrix due to its omission of trials within each of two task blocks.

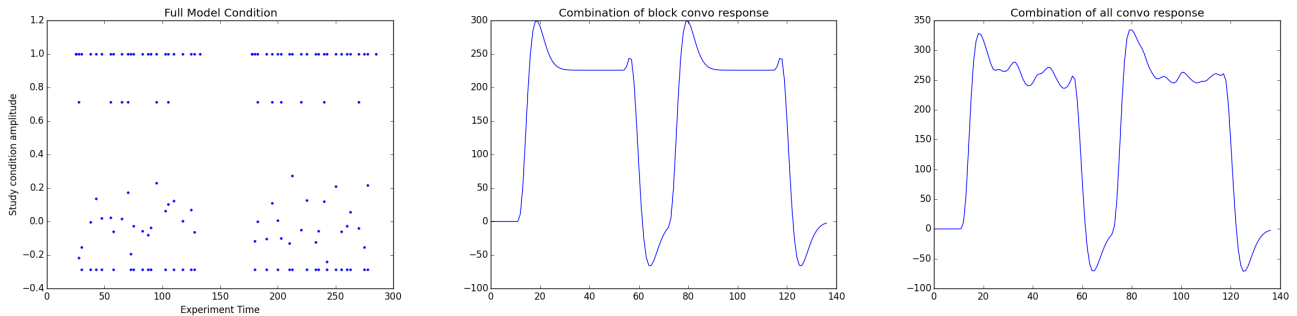


Figure 2: Combination of Block Design and Mixed Design

We are going to focus on the block design and mixed design in order to research on the behaviors of individual voxels and the relationship between blood flows and task stimulus.

### 4 Linear modeling

The basic idea of linear models is to find the relationship between a response variable and one or more explanatory variables (regressors). For our purpose, we attempt to define a relationship between the BOLD signals and the working memory tasks. Mathematically, our model is as follows:

$$y = X\beta + \epsilon, X = \{1, x_1, x_2, \dots, x_k\}.$$

To perform hypothesis tests using this model, we assume linearity of the relationship, and  $E[\epsilon|X] = 0, Var[\epsilon|X] = \Omega$ . We solve for linear regression using generalized least square:

$$\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y.$$

After finding estimates for the  $\beta$ s, we perform Student t-test for each  $\beta_i$ , where  $H_0 : \beta_i = 0, i = \{1, 2, \dots, k\}$ .

## 4.1 Block Design

We are interested in the simple block experiment in this research, so we first fitted a block design experiment to ordinary least square (OLS) regression in order to define the basic relationship between BOLD signals and task stimulus. Here, we assumed that the errors follow independent and identically normal distribution and have a constant variance.

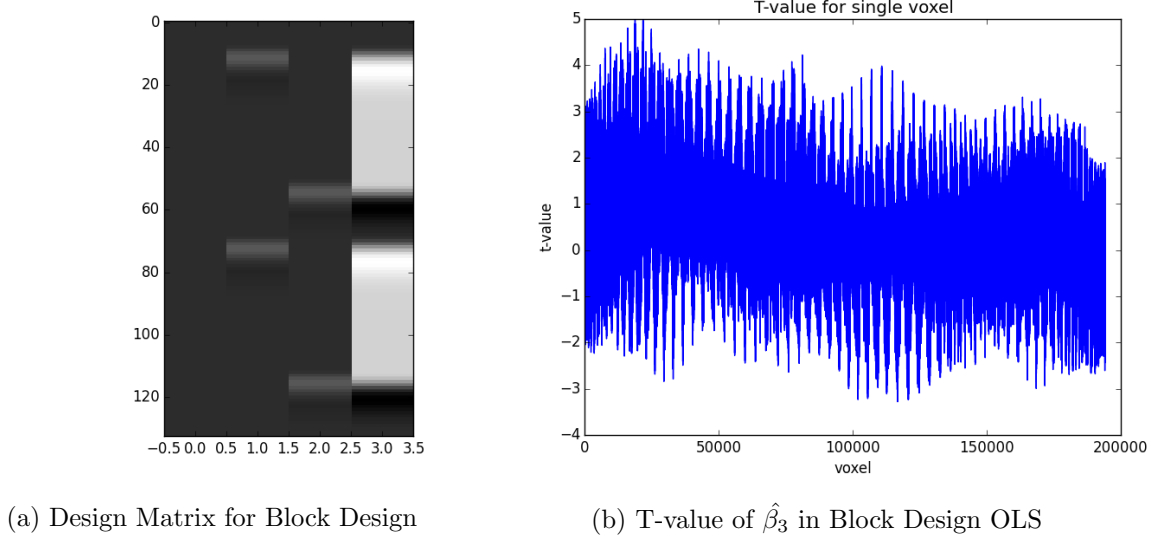


Figure 3: Block Design Model

Since  $\hat{\beta}_3$  in 2 dimensions is not easy to recognize which voxels are significantly active, we would like to reshape it to 3 dimensions. Then, we plotted the  $\hat{\beta}_3$  and P-value of front, middle and back perspectives respectively. Figure 4 is for the middle perspective. The plots of front and back perspectives are in Appendix.

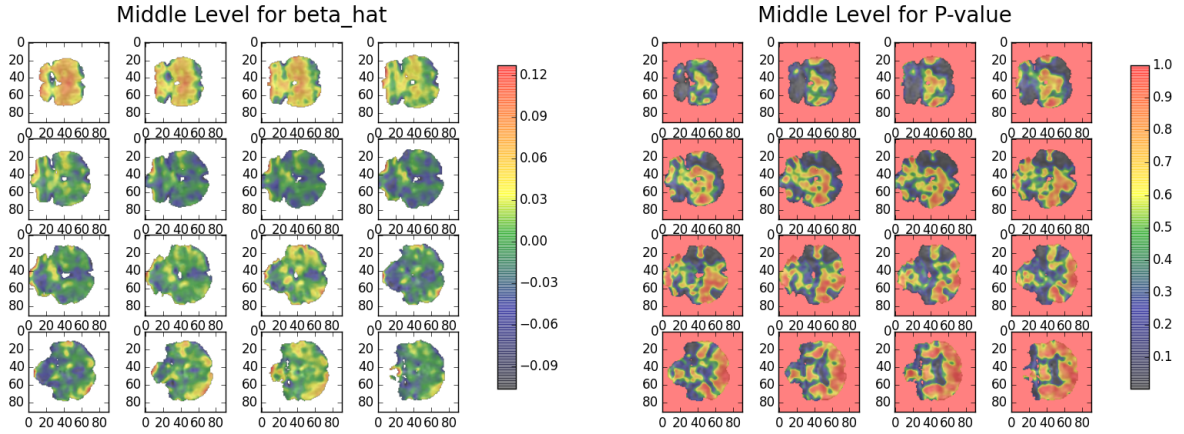


Figure 4: Middle Perspective of Brain in Block Design

In the  $\hat{\beta}$  map, the warmer color is showing  $\hat{\beta}$  is bigger, which means there is higher positive relationship between BOLD signals and task stimulus. However, if the color is more yellow, it means  $\hat{\beta}$  is close to 0 and there is little relationship. Therefore, we conclude that the voxels are highly active during the experiment test located at the center of front and back brain. On the other hand, in the middle brain, majority of voxel are showing the negative relationship between BOLD signals and task stimulus.

In the p-value map, the dark region is showing higher confidence on the  $\hat{\beta}$  values. From the color of the region, we can judge whether or not these voxels are related to our experiment test. We could not

find many regions are highly significantly. If the voxels are on the edge of the brain, most of them did not provide any information in the blood flow, which makes sense.

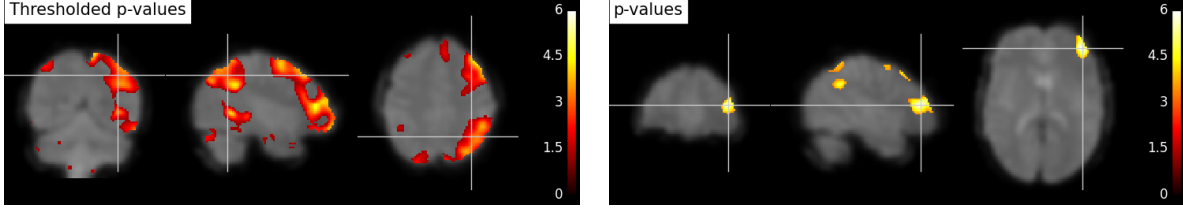


Figure 5: Highly Significant Voxels in the Block Design

In the left image, we set the threshold to 0.05, so find the red region contains highly significant active voxels. In the right image, as Bonferroni correction suggests, we decide to set the threshold to  $0.05/133$ , then we get a few very significant regions. However, most of them are located at the edge of brain, which does not make sense in the common knowledge. Therefore, we did not firmly believe this results, so in the next section, we are going to check heteroscedasticity test.

## 4.2 Heteroscedasticity Test

In the paper, it introduces that there is non-constant variance of blood measurement in the dataset, which means it breaks the assumption of ordinary least squares. If the variance of the Y is not constant, then the error variance will not be constant because it assumes there is linear relationship between true Y and X. The most common form of such heteroscedasticity in Y is that the variance of Y may increase as the mean of Y increases, for data with positive X and Y. We would like to apply heteroscedasticity test, called White test in order to statistically test on variance.

1. Set the null hypothesis: no heteroscedasticity;  $H_a$ : there is heteroscedasticity
2. Estimate by OLS, save squared residuals in  $e^2$
3. Regress  $e^2$  on all variables, their squares and all possible non-redundant cross-products
4. Form the auxiliary regression:  $nR^2$   $X^2(p)$
5. Reject if  $nR^2$  is too large

Intuition: The auxiliary model can be seen as trying to model the variance of the error term. If the  $R^2$  of this auxiliary regression were high, then we could explain the behavior of the squared residuals, providing evidence that they are not constant.

In the figure, we can see only few points under the red line, which voxels are been rejected the null hypothesis test.

The result we got is that there are total 194287 voxels, but only 3868 voxels have the significant P value which means reject our null hypothesis. Because only 2% voxels have the constant variance, we judge the most voxels keep non-constant variance in the dataset. Therefore, we could not apply OLS regression, however, a weighted least squares linear regression may be the preferred method of dealing with non-constant variance of error.

## 4.3 Mixed Design Model

Previously we used block design that has too long of duration, and therefore unable to explain much of the variation in the BOLD signals. We revised convolution of stimulus using a finer and higher resolution on time. With trials within task blocks convolved and included in the design matrix, we expect the result from linear regression have higher explanatory power. However, we need to be cautious because the task block conditions (conditions 1, 4, and 5) are likely to have colinearity problem with the trial conditions (conditions 2 and 3).

Figure 7 shows the new design matrix has dimensions  $133 \times 7$ , where the first 6 columns are the convolved neural responses and the last column is the constant.

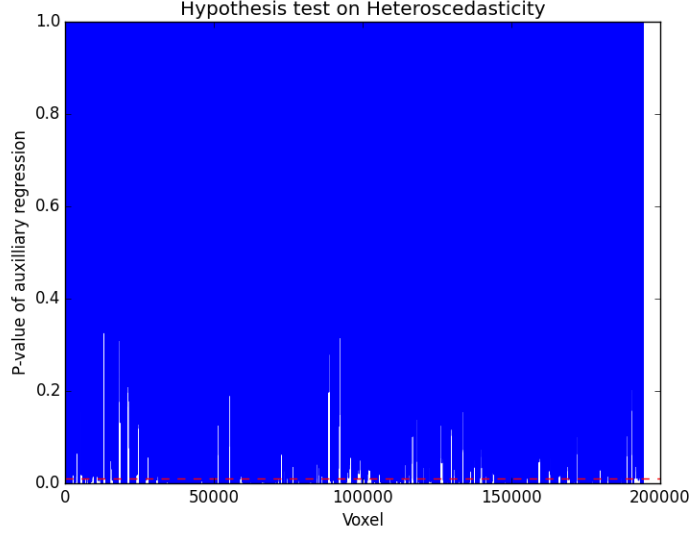


Figure 6: P value for Heteroscedasticity Test in Block Design

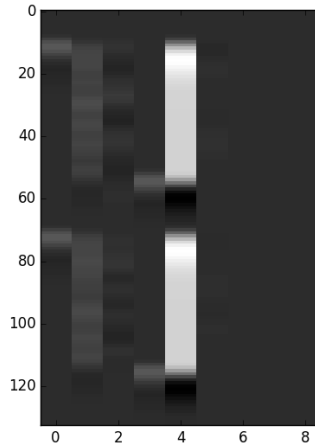


Figure 7: Design Matrix of Mixed Design

#### 4.3.1 Mixed Model Linear Regression Results

With the  $133 \times 7$  mixed model design matrix, we obtained 6 beta estimates for each voxel using GLM. We perform t-test on each voxel, where the null hypothesis claims that  $\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 = 0$ . Figure 8 shows the statistical parametric maps. Each significant voxel is color-coded according to the size of its beta value, t value and p value. Since we are making multiple comparisons, we use the Bonferroni correction for setting threshold. Using  $0.05/133$  as the threshold, we found a number of voxels near the center of the brain to be significant. The mean MRSS across all voxels using all 6 study conditions is 2041.804.

#### 4.3.2 Temporal Smoothing

To reduce noise and improve signal in the fMRI data, we attempted smoothing of the fMRI data. We experimented with smoothing temporally using discrete cosine transformations. We added a set of discrete cosine transformation basis to the design matrix to filter out the low frequency noise present in the BOLD signal. This result in a larger design matrix of dimensions  $133 \times 14$ , which included the discrete cosine transformation basis.

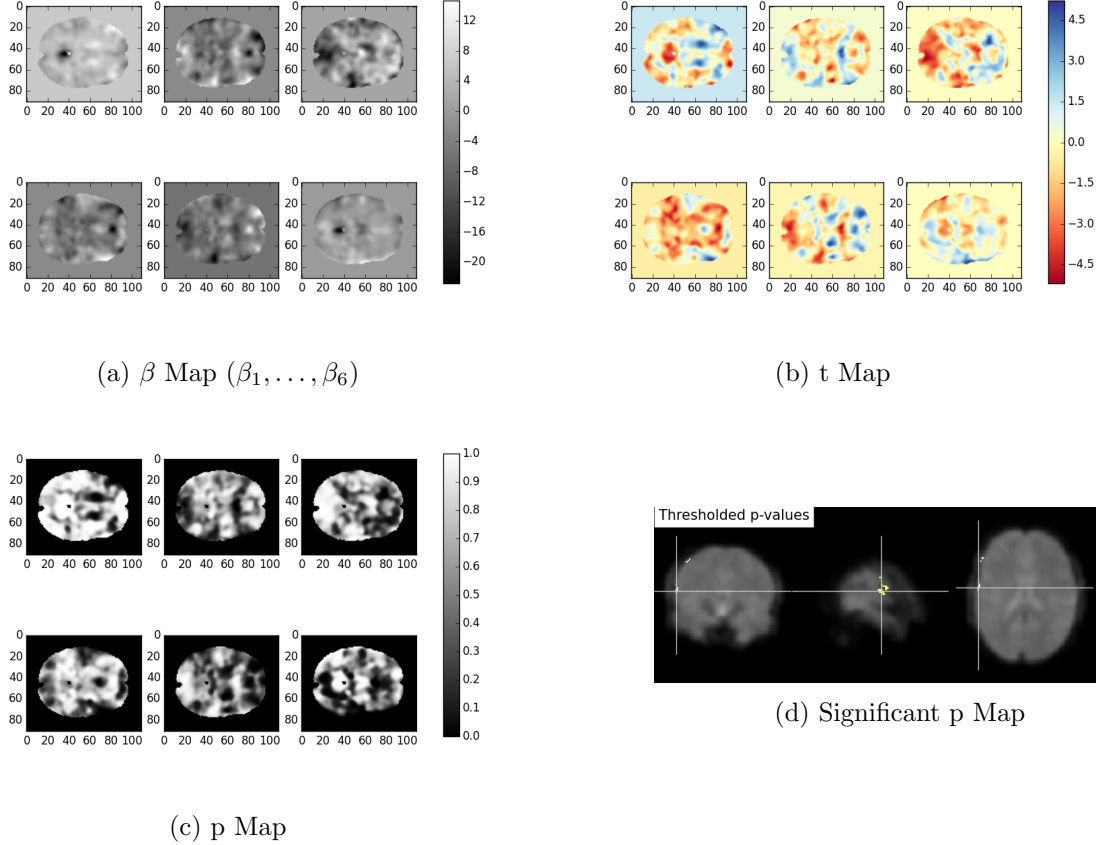


Figure 8: Mixed Model Parametric Maps

As the block model and the mixed model, we used GLM to obtain 6 betas of interest. We perform t-test on each voxel, where the null hypothesis claims that  $\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 = 0$ . The mean MRSS across all voxels using all 6 study conditions is 2033.76. The mean MRSS is smaller than the mixed model, but we cannot conclude that this model is better yet. There are no voxels that are significant at the Bonferroni corrected threshold. We raised our conservative threshold to about 0.05/110, and there are two voxels that are significant.

## 5 Diagnostic

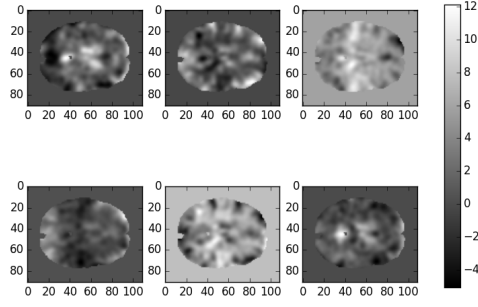
### 5.1 Normality Test

In our analysis part, we used linear models to fit fMRI data. There are several assumptions on linear modeling, one of which is normality assumption on the errors term. Normality of the errors can be checked by checking normality of the residuals. The usual way is to draw a Q-Q plot and check whether it forms a straight line. However in this paper, since there are around 200,000 voxels, it would be hard to plot all of them and check for normality. Thus we use the Shapiro-Wilk test in the paper to test for normality.

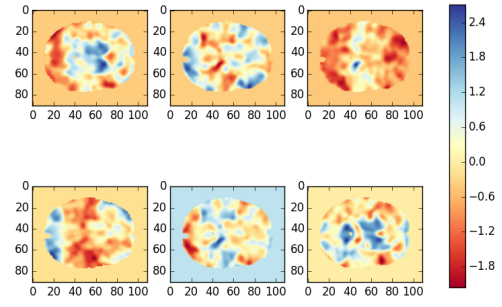
In Shapiro-Wilk test, a small p-value indicates non-linearity, and thereby rejecting the null hypothesis that the errors are distributed normally. We write a function to compute p-value of each voxel and plot those p-values in three brain perspectives. There are three linear models used in the paper. The 6 plots (Figure 10) below are p-value plots: the three plots to the left are p-values for three different linear models using raw data. The other three plots are p-values for three different linear models using smoothed data.

The bright areas in the plots means errors in that area don't follow normal distribution. We can see that data before smoothed seems to have a lot of areas that don't follow normal distribution. After

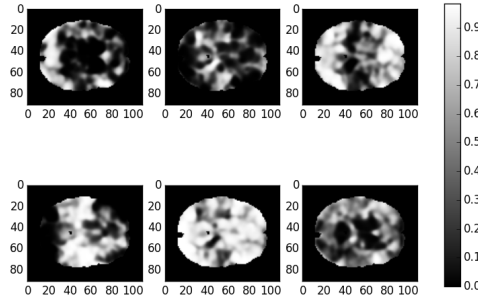




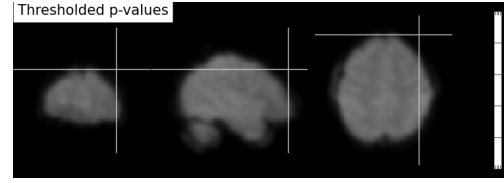
(a)  $\beta$  Map ( $\beta_1, \dots, \beta_6$ )



(b) t Map



(c) p Map



(d) Significant p Map

Figure 9: DCT Model Parametric Maps

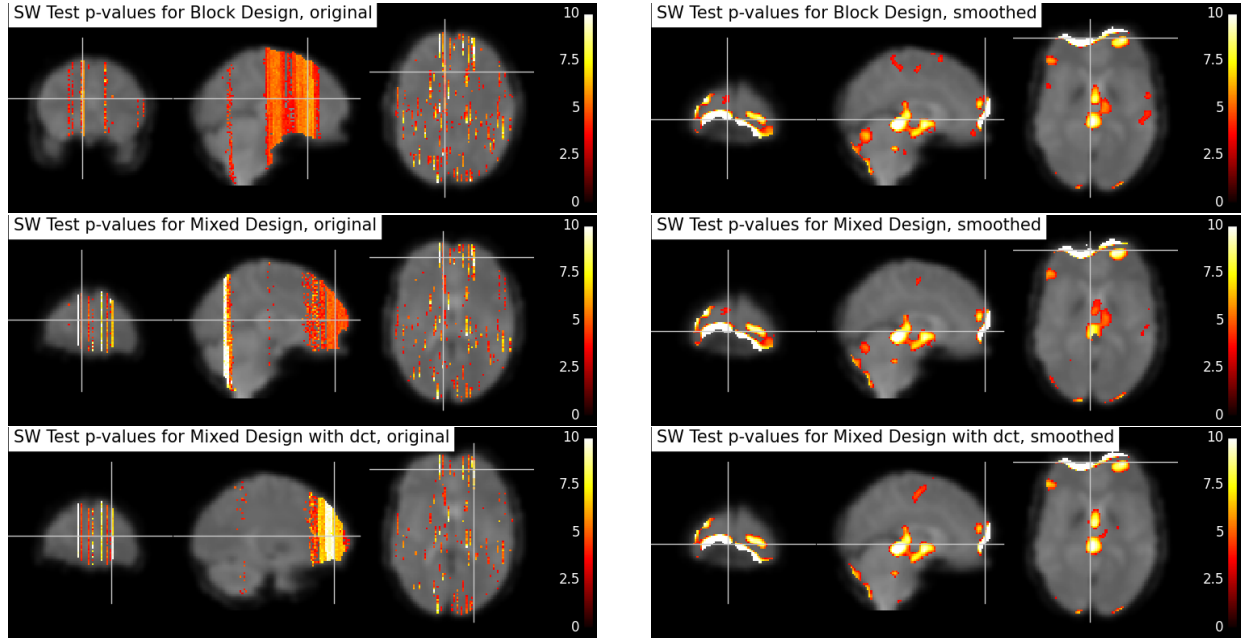


Figure 10: SW Test for Three Linear Models

smoothing, the plots look much better. They show that only some of the areas are without normality. This makes sense because there should be significant points for anatomy.



## 5.2 ANOVA Test

First of all, ideally, since the block design is highly correlated to trial design, we might get the collinearity issue in our model, we would like to check that and keep only either one design that would be good enough for our generalized linear regression model. To address this question, we apply ANOVA test to figure which design will be the best candidate. In the null hypothesis, we would like to test  $\widehat{\beta}_{block} = 0$  or  $\widehat{\beta}_{trial} = 0$ , and we would prefer the model that has the lowest residual sum of square (RSS). The RSSs we obtained from the ANOVA test are 253183.7, 263744.3, 260204.09 for mixed design, model without block design, and model without trial design, respectively. Since the mixed design model has the lowest RSS, we decide to choose the mixed model to do further analysis.

## 5.3 Cross Validation

Since we have already tested the mixed design model with ANOVA, we would like to apply a 5-folds cross validation (CV) to select a model from the mixed models with and without discrete cosine transformation (DCT), and we are going to select a model base on cross validation score, calculated as the mean of squared errors (MSE). Mean of squared errors (mixed design no dct): 71461.0961108, Mean of squared errors (mixed design with dct): 191332664.585. Therefore, the final model we select is the mixed design model without DCT because its MSE is the lower one.

## 5.4 Comparing 2-back and 0-back

Finally, in the experiment design, we know the 2-back test asks for two letters back, while 0-back asks for a specified letter. We are interested in find the voxels that responded differently when performing these two tasks. We fitted our final design models to both 0-back and 2-back task, and obtained  $\hat{\beta}$  from each model. We set the null hypothesis test for t-test to be  $\widehat{\beta}_{0-back} - \widehat{\beta}_{2-back} = 0$ . We reject the null hypothesis test if the p value is lower than 0.05/133 from Bonferroni correction. We found only 108 voxels that rejected the null hypothesis, meaning that these voxels in the brain responded differently when performing 0-back and 2-back tasks.

## 6 Obstacles

Due to our limited knowledge on brain anatomy, we are unable to specify which part of the brain are tested to be responding to the Nback tasks. The biggest challenge is figuring out which direction our research would take and which analysis we have learned to apply to our fMRI data. We also spent a lot of time on understanding the design of the Nback task. The task design is not the usual block design we have seen from class, and the data provides incomplete information on condition files. In addition, since we did not learn how to analyze multiple subjects who have different brain shapes, we focused on a single subject. It took us sometime to become familiar with Github workflow, and writing linear models on Python was also challenging. However, as a group we worked together and helped each other whenever there's a question. We eventually narrowed down our research to linear modeling, and we were able to meet as a group regularly to discuss our project and review each other's codes.

# Appendix

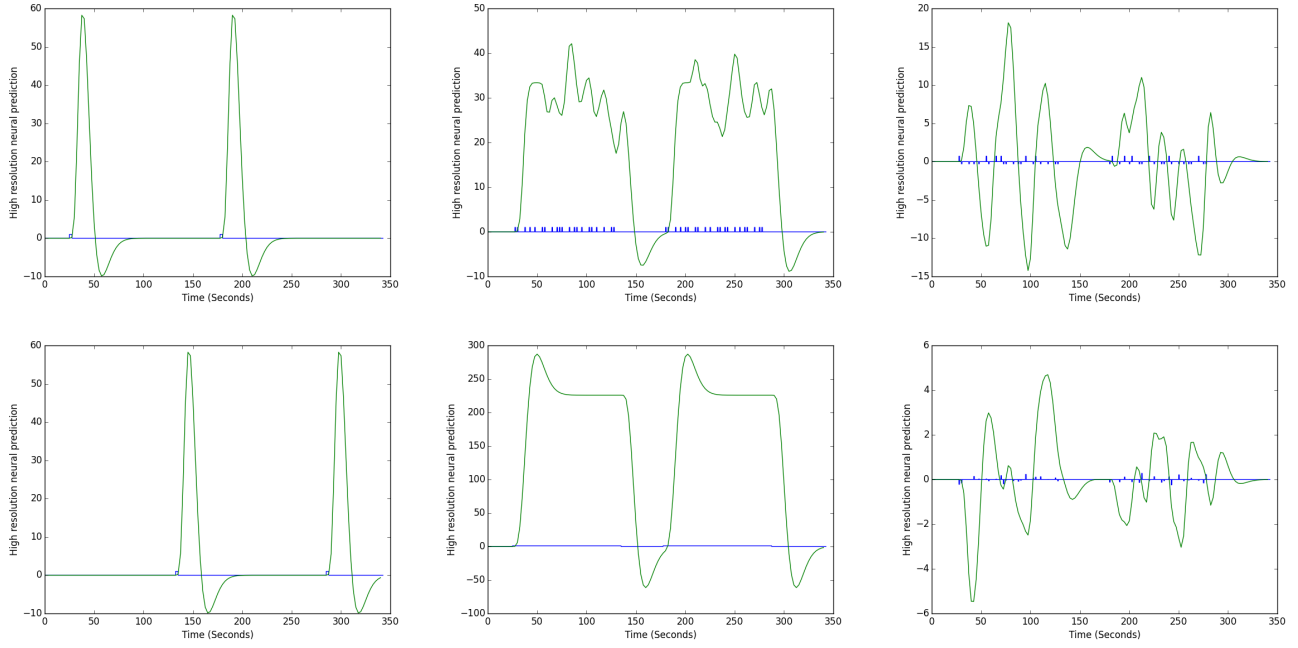


Figure 11: Convolution Response from 6 Conditions

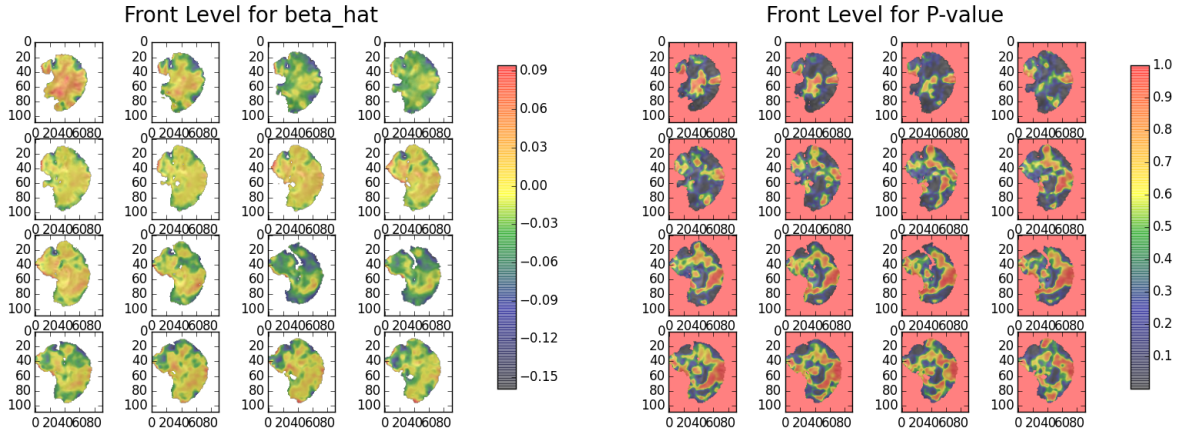


Figure 12: Front Perspective of Brain in Block Design

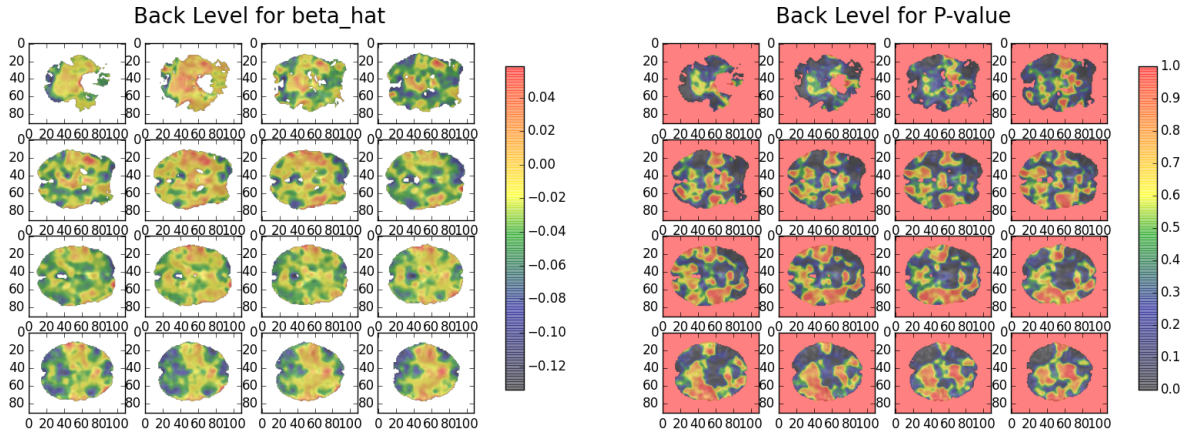


Figure 13: Back Perspective of Brain in Block Design

## References

- [1] T. T. KIRCHER AND R. THIENEL, *Functional brain imaging of symptoms and cognition in schizophrenia*, Progress in brain research, 150 (2005), pp. 299–604.
- [2] G. REPOVS AND D. M.BARCH, *Brain network connectivity in individuals with schizophrenia and their siblings*, Biological Psychiatry, 69 (2011), pp. 967–973.
- [3] ———, *Working memory related brain network connectivity in individuals with schizophrenia and their siblings*, Frontiers in Human Neuroscience, 6 (May 2012), pp. 1–15.