Internship Report

Covariance prediction for causal discovery in brain connectivity

Author:
Sophie Tascedda

Supervisor:
Christof Seiler (DKE)

Research Internship Project DSDM

Maastricht University
Department of Data Science and Knowledge Engineering
October 25, 2022

Contents

1 Introduction

Cognitive neuroscience aims to describe and understand the neural underpinnings of cognitive functions. The objective of this field of studies is to characterize brain activity and cognitive functions, and to relate one to the other. While probabilistic models describe the dependence structure between observed variables, causal models go one step further: they predict, for example, how cognitive functions are affected by external interventions that perturb neuronal activity [1].
Causal inference concerns the design and analysis for evaluating the effects of a treatment. A mainstream statistical framework for causal inference is the potential outcomes framework, under which each unit has a set of potential outcomes corresponding to all possible treatment levels [2].

Cognition and behavior emerge from brain network interactions, thus investigating causal interactions should be central to the study of brain function. A good starting point for characterizing statistical associations among neural time series is Functional Connectivity (FC), which refers to statistical dependence between time series of electro-physiological activity and (de)oxygenated blood levels in distinct regions of the brain [3], [4]. Currently, only the subset of FC methods called Effective Connectivity (EC) is specifically designed to infer causality.

So far, there are few applications of causal discovery in the cognitive neuroscience field. Some of the methods already experimented are Granger causality (among the most popular approaches for the analysis of connectivity between time-evolving processes) and constraint-based methods (methods that rely on two assumptions that connect properties of the causal graph with conditional independence statements in the induced distribution), but they all have their issues and limitations, as discussed by Weichwald et al.[3].

The aim of this research is to combine the potential outcomes framework from social sciences [5] and the causal discovery framework [6] to improve the statistical efficiency of currently applied neuroimaging techniques. From a statistical viewpoint, functional connectivity is the problem of estimating covariance matrices, precision matrices, or correlation matrices from timeseries data. These matrices encode the level of connectivity between any two brain regions. The timeseries are derived from resting-state fMRI (rfMRI) by averaging individual voxels over parcels in the gray matter [7]. Building on Seiler C and Holmes S [7], this research investigates two simple estimators for the covariance matrices of the missing condition in the potential outcomes framework.

2 Materials and methods

2.1 Data

Building on the work of Seiler C and Holmes S [7], the research uses data from the WU-Minn HCP 1200 Subjects Data Release, focusing on the functional-resting fMRI (rfMRI) data of 820 subjects. The images were acquired in four runs of approximately 15 min each. Acquisition ranged over 13 periods $(Q_{01}, Q_{02}, \ldots, Q_{13})$. Following the same preprocessing pipeline, I separated the subjects into short sleepers ($\leq$ 6 h) or conventional sleepers (7–9 h) as defined by the National Sleep Foundation [8]. This divides subjects into 489 conventional and 241 short sleepers.

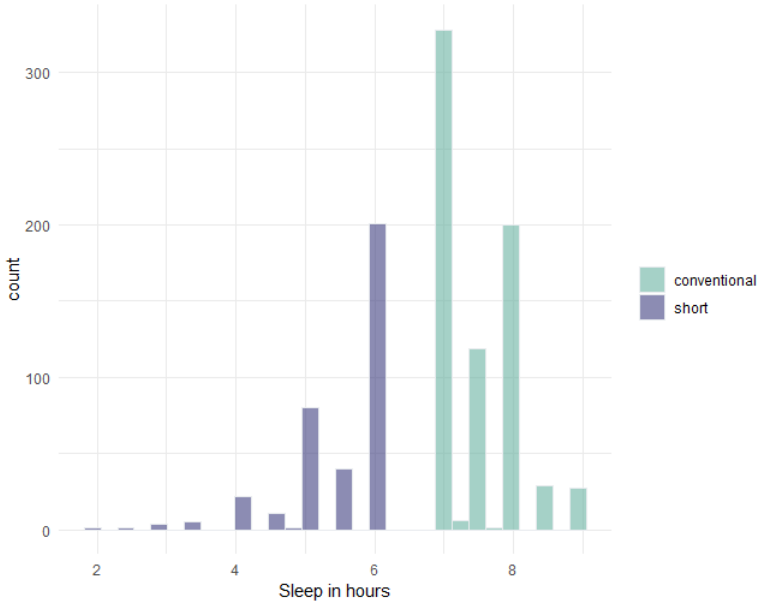

Figure 1: Distribution of sleepers

Preprocessed time series obtained from rfMRI were used. The preprocessing pipeline includes temporal [9] and spatial preprocessing [10], followed by group-PCA denoising [11] and mapping of spatial-group ICA components to each subject to define parcels [10]. Of the range of ICA components offered by HCP, we chose to keep 15 to allow for easier comparison of brain regions between conditions.

2.2 Average based covariance prediction

This section introduces the analysis performed with two different estimator choices for the missing covariance matrices.

2.2.1 The setting

The data after preprocessing are p-dimensional multivariate vectors $y_1 \dots y_N$. Those are centered to guarantee mean-centered observations and each timeseries is subsampled at n<N timepoints to remove temporal dependencies between observations [7]. From the set of subject information provided, one covariate is selected to work with the simplest possible scenario. Covariate X is *average amount of sleep in hours*, on the base of which the subjects were divided in short and conventional sleepers. Consequently, the setting is the following:

- p = 15 brain regions
- 241 short sleepers
- 489 conventional sleepers
- for each subgroup, Y is a list of 4800 observations for p brain regions (so 4800x15 matrices)
- for each subgroup, X is the covariate vector

2.2.2 The potential outcomes framework

Causality is linked to an action (manipulation, treatment, intervention), applied to a unit at a selected point in time. From this perspective, a causal statement presumes that, although a unit was (at a particular point in time) subject to a specific action, the same unit could have been exposed to an alternative action (at the same timepoint). Given a unit and a set of actions, we associate each action-unit pair with a potential outcome. The reason for calling them "potential" outcomes is because only one of them will be concretely realized and so possibly observed: the potential outcome for to the action actually taken. The causal effect of one action or treatment relative to another involves the comparison of these potential outcomes: realized, and the others not realized and therefore not observable.
In this scenario, there's one realized potential outcome and one not realized potential outcome. More formally, the defined setting is the following:

- *Treatment "by nature":* short sleeper
- *Control:* conventional sleeper

With reference to the Neyman–Rubin causal model [12], we want to identify the missing elements of the following table in order to assess the causal effect $Y_t - Y_c$:

	Y_t	Y_c
short	$*_{t,1}$	?
...	$*_{t,i}$	?
conventional	?	$*_{c,s_1+1}$
...	?	$*_{c,j}$

With the aim of surveying the potential presence of causal effect, we use covariance matrices to represent and compare the connectivity between different anatomical regions. Consequently, the * in the table are, for each subject, its sample covariance matrix. This is the realized potential outcome. For the missing potential outcome a prediction of the covariance matrix is needed, so what we need for estimating the causal effect is:

- For short sleepers:
  - Y_t: sample covariance matrices
  - Y_c: predicted covariance matrices
- For conventional sleepers:
  - Y_t: predicted covariance matrices
  - Y_c: sample covariance matrices

Defining treatment as X=1 and control as X=0, the average causal effect is

$$ACE = E(Y_i|X_i = 1) - E(Y_i|X_i = 0)$$

2.2.3 Average covariance

The first and most simple estimator selected for the covariance forecasting is the average covariance. For each subject in each condition, we use the average covariance matrix computed from all the sample covariance matrices of the opposite condition to simulate the non realized outcome. Formally, the sample covariance matrix for each subject is:

$$\begin{bmatrix} Var(x_1) & ... & Cov(x_n, x_1) \\ ... & ... & ... \\ Cov(x_n, x_1) & ... & Var(x_n) \end{bmatrix}$$

where each off diagonal element is a salmple covariance:

$$cov(x, y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n - 1}$$

and each main diagonal element is a sample variance:

$$var(x) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})}{n - 1}$$

This results in s_1=241 sample covariance matrices for short sleepers and s_2=489 sample covariance matrices for conventional sleepers. The average covariance matrix is computed as:

$$\frac{1}{s}\sum_{i=1}^{s} Cov(Y_i)$$

where $s = s1$ for average covariance matrix from short sleepers and $s = s2$ for average covariance matrix from conventional sleepers. Since the average of a list of covariance matrices is one covariance matrix, all predicted elements under treatment will have the same matrix and all predicted elements under control will have the same matrix, but those are different between the two conditions. The previous table now has the following structure:

	Y_t	Y_c
short	$*_{t,1}$	$\hat{*}_c$
...	$*_{t,i}$	$\hat{*}_c$
conventional	$\hat{*}_t$	$*_{c,s_1+1}$
...	$\hat{*}_t$	$*_{c,j}$

2.2.4 Knn average covariance

The second estimator for the predicted covariance matrices is the average covariance matrix computed from the k nearest neighbours from the opposite condition, based on values of the chosen covariate. Consequently, this prediction is still average based but differently from the precedent it involves also the predictor. Recalling that the covariate is *average amount of sleep in hours* and that for short sleepers this is inferior or equal to 6 hours, and between 7 and 9 hours for conventional sleepers, for every subject from a condition we find the k subjects from the opposite condition that have the closest covariate value and use their sample covariance matrices for the averaging. This will lead again to having the same predicted covariance matrix for all subjects under the same condition because independently from the covariate value of the current subject, the nearest neighbors will always be the same for how the separation of the two conditions is defined. More explicitly, the nearest neighbors for a short sleeper will always be the k subjects from conventional sleepers with the lowest X value, and for a conventional sleeper they will always be the k highest X values from the short sleepers. The chosen number of nearest neighbours is $k = 5$.

2.3 Hypothesis testing

For every combination of brain regions, we performed two sided paired t-test between treatment and control covariance values, for all subjects. The hypothesis tested is the following:

- H(0): The mean covariance value between 2 brain regions across all subjects is the same under treatment and under control

- H(1): The mean covariance value between 2 brain regions across all subjects is different under treatment and under control

This translates in:

- H(0): $E(Y_{ij}|X = 1) - E(Y_{ij}|X = 0) = 0$
- H(1): $E(Y_{ij}|X = 1) - E(Y_{ij}|X = 0) \neq 0$

where Y_{ij} is related to i-th and j-th brain regions, i≠j (we have a test for every combination of different regions). The sample and predicted covariance matrices are all $15 * 15$ symmetric matrices. We want to test difference in the covariance value of distinct brain regions, so we can exploit the symmetry of the matrices and keep either the lower or upper triangular matrix. With the dimensionality of the setting, this leads to 105 tests.

2.3.1 Benjamini - Hochberg adjustment procedure

When dealing with multiple hypothesis testing, it's important to remember that a low *false discovery rate* does not imply a low number of *false discoveries* (where false discoveries are incorrect rejections of the null, also called *type I errors*). False discovery (FDR) correction methods limit the expected value of the fraction of rejected tests that come from the null hypothesis [13]. Here, the chosen method is the Benjamini-Hochberg (BH) adjustment procedure [14]. Given a set of p-values and α level set at 5%, the procedure is the following:

1. sort the p-values in ascending order and rank them accordingly
2. for each individual p-value compute it's BH critical value, using the formula $\frac{i}{m}Q$, where:
   - i = rank of the p-value
   - m = total number of tests
   - Q = false discovery rate (a set percentage)
3. comparing the two lists, find the largest p-value also smaller than it's critical value.

By rejecting all the p-values smaller than the new threshold, the FDR is guaranteed to be under the chosen percentage.

2.4 Validation

The validity analysis was performed on simulated data, building up from the simplest scenario up to the dimensionality of the original data.

2.4.1 Sampling

The simulated data is obtained sampling from a standard normal distribution:

$$\mathcal{N}(0,1) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{1}(x^2)}$$

Firstly, we sample 100 observations from the standard normal distribution twice and perform unpaired two sided ttest on the values. This is repeated $m = 10000$ times to check for uniform distribution of the obtained p-values.

The successive step is sampling twice from a multivariate normal distribution with zero mean and $2*2$ identity covariance matrix. We sample 100 observations for 2 variables, each time for 10 subjects. For each subject the sample covariance matrix is computed and the same matrix element is selected from each subject under both conditions for the unpaired t-test. Again, this is repeated numerous times for investigating the distribution of p-values.

Finally, the same analysis is carried out on simulated data with the original dimensionality.

This first part of the validation process is carried out to guarantee the data is generated correctly.

2.4.2 Validating the analysis pipeline

The further step in the validity analysis is testing on simulated data the two prediction methods. This is done by sampling 100 observations for 2 variables from multivariate normal distribution with zero mean and $2 * 2$ identity covariance matrix and running the complete analysis on this data $m = 10000$ times.

3 Results

In this section the results of the analysis and validation are presented.

3.1 Average covariance prediction

When predicting the covariance matrices for the missing potential outcome using the average of all subjects from the opposite condition, we obtain a strongly unbalanced distribution of the 105 p-values, as can be seen in 2
With alpha level is set at 0.05, the tests leads to:

- 76 rejections of the null hypothesis
- 29 not significant p-values

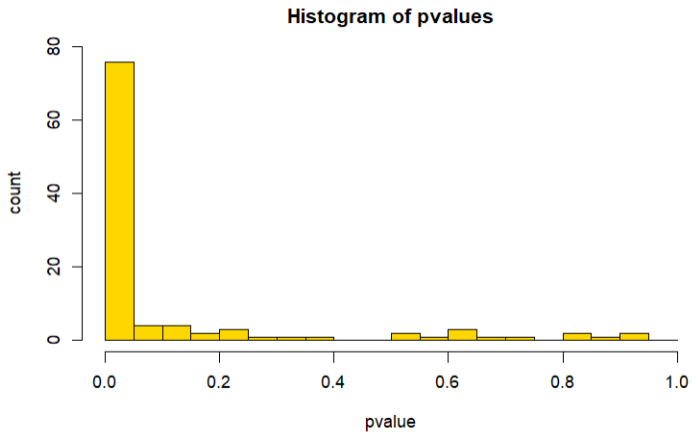

Figure 2: Distribution of p-values, using average covariance prediction

This translates in 76 pairs of regions having values with a significant difference in the mean, and 29 pairs with a t-test p-value not significant to reject the null hypothesis of zero mean difference. The visualization in 3 allows to better appreciate which pairs of regions show a significant pvalue versus which don't. The blue tiles represent values below the threshold of 0.05 so the pairs for which

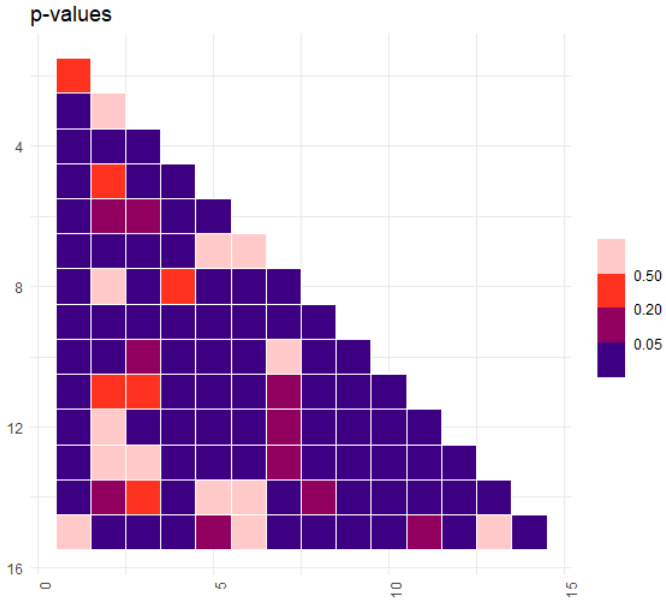

Figure 3: Tiles visualization of t-test p-values

we would reject the null hypothesis.

BH adjustment
The BH adjustment procedure lowers the cutoff point at $p = 0.02777567$ keeping a number of 76 rejections.

3.2 Knn average covariance prediction

When the prediction of the covariance matrices for the missing potential outcome is performed by averaging over the covariance matrices of the k nearest neighbours from the opposite condition based on the covariate value, we get a very similar result that is even more unbalanced 4 with now a total of:

- 95 rejections of the null hypothesis
- 10 not significant p-values

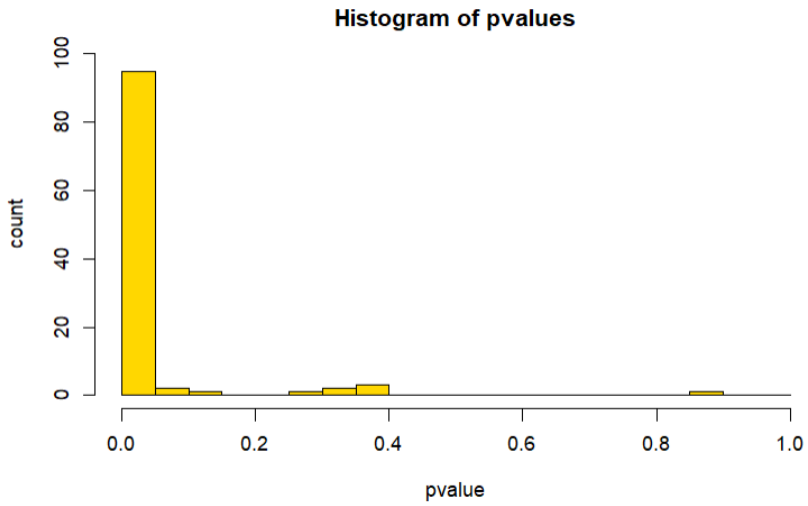

Figure 4: Distribution of p-values, using knn average covariance prediction

This translates in 95 pairs of regions for which we would reject the null hypothesis that treatment and control have a zero mean difference. In 5, we visualize a tile representation of those p-values, where we can appreciate which pairs of regions now give a significant p-value compared to the previous prediction.

BH adjustment
Again, to guarantee a FDR lower than 5%, the BH procedure keeps the same number of rejections, lowering the cutoff to 0.02824.

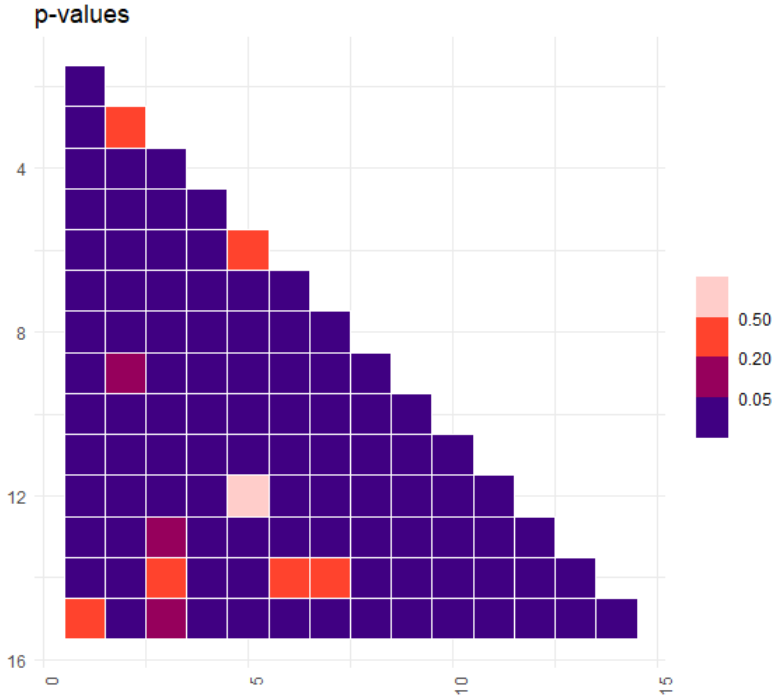

Figure 5: Tiles visualization of t-test p-values

3.3 Validation

3.3.1 Validating the sampling process

The first three experiments of the validation task are performed with the purpose of guaranteeing the correct generation process of the simulated data. We are sampling from a standard normal distribution, and for both conditions we're sampling from the same distribution, so we expect the p-values of an unpaired t-test to follow a uniform distribution. Figure 6 shows the distribution of the p-values for the simple univariate sampling and for the sampling from multivariate normal distribution for 2 and for 15 variables. As we can observe, the p-values are indeed distributed uniformly, confirming we are sampling correctly for the purpose of the esperiment.

3.3.2 Validating the estimators

Biased vs unbiased estimator
For both chosen covariance estimators, it was important to test the importance of using sample averages versus population averages. Because of the particular results given by the hypothesis testing, it was fundamental to understand if averaging over $s - 1$ instead of s has a role in the outcome of the analysis. For

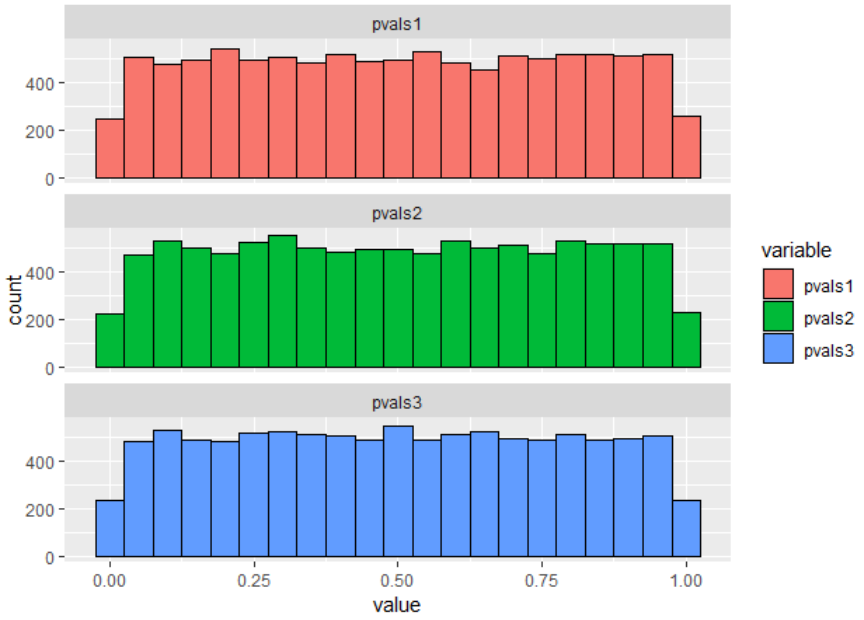

Figure 6: Distribution of t-test p-values on data sampled from (a) univariate normal distribution (b) multivariate normal with 2 variables (c) multivariate normal with 15 variables.

this, the validity analysis is performed:

- Averaging over s. The predicted covariance matrices are obtained as:

$$\hat{Cov} = \frac{1}{s}\sum_{i=1}^{s} Cov(Y)_i$$

- Averaging over $s-1$. The predicted covariance matrices are obtained as:

$$\hat{Cov} = \frac{1}{s-1}\sum_{i=1}^{s} Cov(Y)_i$$

where $Cov(Y)_i$ is the i_{th} sample covariance matrix of the subgroup, $s = s1$ for short sleepers and $s = s2$ for conventional sleepers.

Average covariance prediction
When performing the complete prediction and hypothesis testing on the simulated data, the distribution of the p-values changes compared to the distribution obtained when testing on the generated values or on the values of the sample covariance matrices. In Fig.7 we can observe the distribution of the p-values

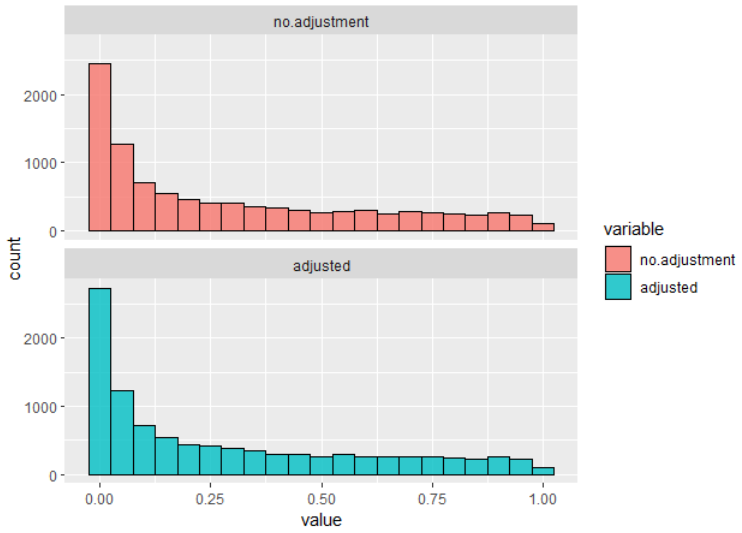

Figure 7: Distribution of t-test p-values after average covariance prediction on simulated data, averaging over (a) s subjects (b) s-1 subjects.

obtained when averaging over s (top plot) and over $s-1$ (bottom plot).
The two histograms seem to still have an overall almost uniform distribution, with exception for very low p-values. If we have a look at the numbers, we have:

- 3248 p-values$<$ 0.05 in the top histogram
- 3484 p-values$<$ 0.05 in the bottom histogram

We notice that averaging over $s-1$ actually leads to even more rejections than if we average over s.

BH adjustment

Here we're performing a lot more tests compared to the real data, and we can see the actually expected results from the BH procedure. The new cutoff that guarantees an FDR lower than 5% is 0.0083 resulting in 1675 rejections (versus the 3248 ones found before). Similarly, when averaging over $s-1$ the new cutoff is at 0.01 with 2008 rejections (versus the previous 3484).

Knn average covariance prediction
Using a k nearest neighbours average approach leads to very similar results. As can be observed in Fig.8 the p-values have the same distribution shape as the previous experiment, but with an even higher number of p-values under the 0.05 α level. In numbers, we now have:

- 4778 p-values$<$ 0.05 in the top histogram

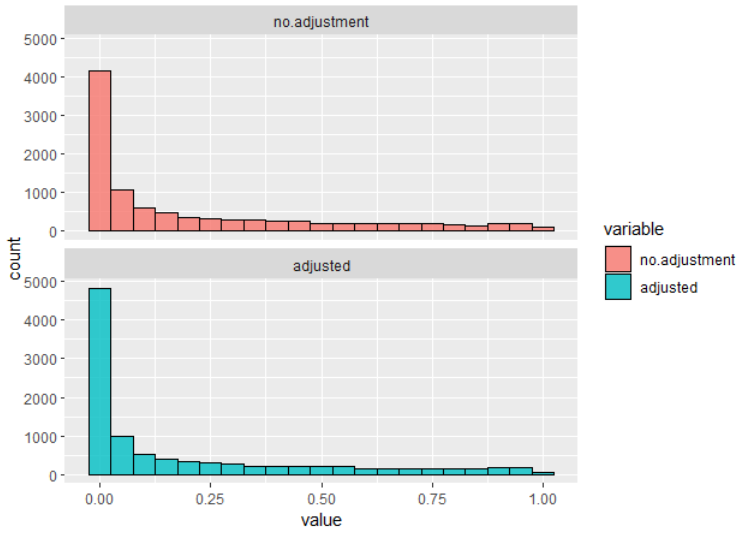

Figure 8: Distribution of t-test p-values after knn average covariance prediction on simulated data, averaging over (a) s subjects (b) s-1 subjects.

- 5454 p-values<0.05 in the bottom histogram

Again, averaging over $s-1$ gives a higher number of p values under the α level compared to averaging over s.

BH adjustment
Here the adjustment is a little less evident than before. In the first case, we get a reduction of the cutoff to 0.0197 with 3956 rejections (versus the precedent 4778) while in the second the cutoff is lowered to 0.0241 with 4813 (versus the 5454 before).

4 Discussion

This research investigates two potential estimators for predicting covariance matrices with the purpose of combining the potential outcomes framework and causal discovery in the context of neuroimaging and exploring brain connectivity. The chosen estimators are the average covariance matrix and a knn average covariance matrix computed from the opposite condition. This prediction is necessary to apply the potential outcomes framework and estimate the average causal effect because only one of the potential outcomes is truly realised. For every subject there is one sample and one predicted $15 * 15$ covariance matrix and the t-test is performed between every couple of distinct brain regions. This coincides with taking the same element from all the lower (higher) triangular covariance matrices for each subject, from Y_t and Y_c, and feeding those to the paired t-test. Both estimators lead to a very high number of p-values that are $\leq \alpha$, 76 for the first and 95 for the second. Benjamini-Hochberg adjustment doesn't alter the number of rejections in neither case.

Constructing a histogram from the p-values is a natural graphical method method for visualizing the body of a large number of hypothesis tests. When all nulls are true, the histogram results to be flat, showing thus a uniform distribution [13]. This is what we observe in Fig.6 and corresponds to what we expect to see since the data was generated from a standard normal distribution specifically to behave like this. If we observe the distributions in Fig.7 and Fig.8, there is something different going on. We get a high number of p-values close to zero, and specifically under 0.05, but the rest of the distribution is close to uniform. This can be labelled as an *anti-conservative* behavior of the p-values. Because we're performing a higher number of tests, the Benjamini Hochberg procedure reduces more impactfully the number of rejections.

No clear causal relationship can be extrapolated from the analysis but the concept is worth to be investigated more deeply. A good starting point is to use more accurate estimators for the covariance. Specifically, it would be interesting to take into consideration estimators for the covariance matrices that don't result in the same prediction for every subject under the same condition. In Appendix 6 I briefly discuss a prediction formula derived in the first stages of the research from the work of Zhao, Yi, et al. [15], that hasn't been included in the final estimators because to better understand the anomalies in the results the decision was to consider the most simple estimators. The use of the algorithms in the cap package [16] allows an estimation for a rotation vector γ and regression vector and parameter β_1 and β_0 (respectively). This would provide all the needed elements to use the derived prediction formula (4) in 6 to estimate the covariance matrices for the missing potential outcome. In my opinion this method is very promising. It would give an interesting insight on how and if having a prediction method that doesn't give the same estimation for every subject under the same condition impacts the performance of the analysis.

5 Reproducibility and Supplementary Material

The entire analysis workflow can be found on the dedicated Git repository:

- `https://github.com/SophieTax/CovPred_ACE`

Data preparation and analysis can be found in:

- Covariance prediction.Rmd

and the validity analysis in:

- Testing.Rmd

The HCP data can be found at:

- `https://www.humanconnectome.org/data/`

References

[1] Sebastian Weichwald and Jonas Peters. Causality in Cognitive Neuroscience: Concepts, Challenges, and Distributional Robustness. *Journal of Cognitive Neuroscience*, 33(2):226–247, 02 2021.

[2] Peng Ding and Fan Li. Causal inference: A missing data perspective. *Statistical Science*, 33(2):214–237, 2018.

[3] Parinaz Babaeeghazvini, Laura M Rueda-Delgado, Jolien Gooijers, Stephan P Swinnen, and Andreas Daffertshofer. Brain structural and functional connectivity: A review of combined works of diffusion magnetic resonance imaging and electro-encephalography. *Frontiers in human neuroscience*, page 585, 2021.

[4] Andrew T Reid, Drew B Headley, Ravi D Mill, Ruben Sanchez-Romero, Lucina Q Uddin, Daniele Marinazzo, Daniel J Lurie, Pedro A Valdés-Sosa, Stephen José Hanson, Bharat B Biswal, et al. Advancing functional connectivity research from association to causation. *Nature neuroscience*, 22(11):1751–1760, 2019.

[5] Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

[6] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

[7] Christof Seiler and Susan Holmes. Multivariate heteroscedasticity models for functional brain connectivity. *Frontiers in neuroscience*, 11:696, 2017.

[8] Max Hirshkowitz, Kaitlyn Whiton, Steven M Albert, Cathy Alessi, Oliviero Bruni, Lydia DonCarlos, Nancy Hazen, John Herman, Eliot S Katz, Leila Kheirandish-Gozal, et al. National sleep foundation's sleep time duration recommendations: methodology and results summary. *Sleep health*, 1(1):40–43, 2015.

[9] Stephen M Smith, Christian F Beckmann, Jesper Andersson, Edward J Auerbach, Janine Bijsterbosch, Gwenaëlle Douaud, Eugene Duff, David A Feinberg, Ludovica Griffanti, Michael P Harms, et al. Resting-state fmri in the human connectome project. *Neuroimage*, 80:144–168, 2013.

[10] Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

[11] SM Smith. Hyva [r with umlaut] inen a., varoquaux g., miller kl, beckmann cf (2014). *Group-PCA for very large fMRI datasets. Neuroimage*, 101:738–749.

[12] Jasjeet S Sekhon. The neyman-rubin model of causal inference and estimation via matching methods. *The Oxford handbook of political methodology*, 2:1–32, 2008.

[13] Patrick Breheny, Arnold Stromberg, and Joshua Lambert. p-value histograms: Inference and diagnostics. *High-Throughput*, 7(3), 2018.

[14] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[15] Yi Zhao, Bingkai Wang, Stewart H Mostofsky, Brian S Caffo, and Xi Luo. Covariate assisted principal regression for covariance matrix outcomes. *Biostatistics*, 22(3):629–645, 2021.

[16] Stewart Mostofsky Brian Caffo Xi Luo Yi Zhao, Bingkai Wang. cap: Covariate assisted principal (cap) regression for covariance matrix outcomes. `https://CRAN.R-project.org/package=cap`, 2018.

6 Appendix

6.1 Covariate assisted principal regression

A very interesting reflection point was provided by the work of Zhao, Yi, et al. [15]. The aim of the paper is to link n sample covariance matrices Σ_i to the corresponding covariance vectors $\mathcal{X}_i$ for $i = 1, ..., n$. Under the assumptions that the response variable is multidimensional and that is must be positive (semi-) definite, the idea is to borrow concepts from *Principal Component Analysis* (PCA) applied to covariance matrices. Specifically, what they exploit is to identify a covariate-dependent rotation vector γ such that:

$$\gamma^T \Sigma_i \Gamma = \textit{Response in a generalized linear model of } \mathcal{X}_i$$

They propose a *CAP (Covariance Assisted Principal) regression model* which:

- integrates the mentioned PCA principle with generalized linear modeling of multiple covariates
- similarly to PCA, aims to identify linear projections of covariance matrices while being computationally feasible for large data
- differently from PCA, targets specifically those projections that are associated to covariates. This enables to study changes in covariance matrices that are associated with subject specific factors.

6.1.1 The CAP regression model

For every subject/unit $i \in \{1, ...n\}$ let $y_{it} \in \mathbb{R}^p$ $(t = 1, ..., T_i$ be iid random samples from a multivariate normal distribution with:

- zero mean
- covariance matrix Σ_i, where the latter may depend on explanatory variables $X_i \in \mathbb{R}^{q-i}$

where:

- $T_i = \#$ of observations from subject i
- $n = \#$ of subjects
- $q = \#$ of predictors (including the intercept)

In the specific application of the paper, which coincides with the scenario investigated in this research, $\overline{y}_{it}$ is a sample of fMRI measurements of p brain regions and $\overline{X}_i$ is a vector of covariates, both collected for subject i. Assume the existence of a vector $\underline{\gamma} \in \mathbb{R}^p$ such that:

$$z_{it} = \underline{\gamma}^T \underline{y}_{iT}$$

satisfies the following *heteroscedasticity model*:

$$\log\{Var(z_{it})\} = \log(\gamma^T \Sigma_i \gamma) = \beta_0 + \underline{x}_i^T \underline{\beta}_1 \tag{1}$$

where β_0 and β_1 are model coefficients:

- $\beta_0 \in \mathbb{R}$
- $\beta_1 \in \mathbb{R}^{q-1}$

The model (1) generalizes a variance regression model in which Σ_i is scalar.
The paper then proposes 3 algorithms to, under the considered model, estimate the parameters maximizing the likelihood function of the data in the projection space.

6.1.2 A prediction formula for covariance matrices

By manipulating model (1) we can derive a prediction formula for the covariance matrices. Specifically, we can eliminate the logarithm using the exponential operator and obtain the equivalent form:

$$\gamma^T \Sigma_i \gamma = \exp(\beta_0 + x_i^T \beta_1) \tag{2}$$

In the setting of this research, Σ_i is a $15 * 15$ matrix and consequently γ has dimensionality $15 * 1$ and γ^T $1 * 15$.
On each side of equation (2) we left multiply (matrix multiplication) by γ and right multiply by γ^T, obtaining:

$$\gamma\gamma^T \Sigma_i \gamma\gamma^T = \gamma \exp(\beta_0 + x_i^T \beta_i)\gamma^T \tag{3}$$

Both sides of equation (3) have dimensionality $15*15$. Left and right multiplying both sides of the equation by the inverse of $\gamma\gamma^T$ allows to isolate the covariance matrix Σ_i and obtain the following prediction formula for covariance matrices:

$$\hat{\Sigma}_i = [\hat{\gamma}\hat{\gamma}^T]^{-1}\hat{\gamma} \exp(\hat{\beta}_0 + x_i^T \hat{\beta}_1)\hat{\gamma}^T[\hat{\gamma}\hat{\gamma}^T]^{-1} \tag{4}$$

In (4), $\hat{\Sigma}_i$ is the predicted covariance matrix for subject i. Vectors $\hat{\gamma}$ and $\hat{\beta}_1$ and parameter $\hat{\beta}_0$ are estimated by one of the 3 algorithms proposed by Zhao, Yi, et al. that can be found in the dedicated R package on CRAN [16].

In a setting as the one investigated in this paper, the data includes all the needed elements to exploit the cap package [16] to estimate the missing parameters and use the equation in (4) to predict the covariance matrices for the missing potential outcomes in a covariate driven way.