Stony Brook University

Partial Correlation Analysis in Functional Brain Imaging Studies

A Dissertation Presented

by

Kith Pradhan

to

The Graduate School

in partial fulfillment of the

degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

STONY BROOK UNIVERSITY

December 2009

Stony Brook University

The Graduate School

Kith Pradhan

We, the Dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Wei Zhu – Dissertation Advisor
Professor, Department of Applied Mathematics and Statistics,
Stony Brook University

Nancy Mendell – Chairperson of Defense
Professor, Department of Applied Mathematics and Statistics,
Stony Brook University

Stephen Finch
Professor, Department of Applied Mathematics and Statistics,
Stony Brook University

Klaus Mueller
Associate Professor, Department of Computer Science,
Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Partial Correlation Analysis in Functional Brain Imaging Studies

by Kith Pradhan

Doctor of Philosophy

in

Applied Mathematics and Statistics

STONY BROOK UNIVERSITY

2009

There is a great deal of knowledge about the anatomical organization of the human brain but much less is known about the functional interactions among the brains components and how these interactions are altered under abnormal conditions such as disease, drugs or alcohol. Confirmatory multivariate analysis methods, especially structural equation modeling (SEM), are ideal for the verification of hypothesized brain functional pathways. However data driven methods, especially the modern partial correlation network method, would be more suitable for network discovery when little is known about the underlying pathways. In this thesis, we develop novel correlation network analysis methods that will enable us to examine the influence of factors such as disease or drugs, as well as the effect of potential confounders such as age and gender, on the underlying brain functional pathways. Two approaches are presented with the resulting methods applied to both simulated and real functional brain imaging positron emission tomography (PET) data from the Brookhaven National Laboratory.

Contents

List of Figures

List of Tables

Abbreviations

SEM	**S**tructural **E**quation **M**odelling
LRT	**L**ikelihood **R**atio **T**est
PET	**P**ositron **E**mission **T**omography
CT	**C**omputed **T**omography
BLUS	**B**est **L**inear **U**nbiased **S**caled
MRI	**M**agnetic **R**esonance **I**maging
CCA	**C**ross **C**orrelational **A**nalysis
OLS	**O**rdinary **L**east **S**quares
PCNA	**P**artial **C**orrelation **N**etwork **A**nalysis
MVN	**M**ulti**V**ariate **N**ormal
FDG	**F**luoro**D**eoxy**G**lucose
VS	**V**entral **S**triatum
THAL	**Thal**amus
INS	**Ins**ula
PUT	**Put**amen
MFC	**M**otor **F**rontal **C**ortex
CER	**Cer**ebellum
AMYG	**Amyg**dala
OFC	**O**rbito **F**rontal **C**ortex
CG	**C**ingulate **G**yrus

Acknowledgements

My thanks go out to the wonderful professors at Stony Brook University, my collaborators at Brookhaven National Lab, and my advisor Prof. Wei Zhu.

Chapter 1

Path Analysis in Neuroimaging

1.1 Introduction

Path analysis in neuroimaging is an attempt to discover relationships between the various anatomical/functional regions present in the brain. In the past, obtaining information of such pathways was only possible through in-vivo studies on animal subjects, or the physical observation of rare medical incidents where damage to one brain region leads to the loss of functionality associated with another (Novack et al., 2001). Such studies continue today, but are conducted very infrequently on human subjects, for obvious reasons. Fortunately, scientific breakthroughs in medical imaging, originating with the invention of MRI (Lauterbur, 1973), have allowed us to quantitatively measure neurological attributes of live subjects thus opening doors to a myriad of novel methodologies.

1.2 Background

In recent years there have been many mathematical frameworks developed to aid our understanding of neurological pathways, the foremost being structural equation modeling (SEM) (McLntosh and Gonzalez-Lima, 1994). In this methodology, an a priori model is declared that describes the expected directional pathways between anatomical regions, usually formulated with the aid of experienced neurologists. This theoretical a priori model implies a specific covariance structure which is validated or rejected by fitting it with the observed data. If a path in the starting model turns out to be invalid, a new model can be incrementally formed correcting for the mis-formed path and thus, better models can be found (Bullmore, 2000; Stein et al., 2007). But an incremental methodology for model improvement has its own problems. As the number of studied regions grows, so does the number of possible pathway combinations; and there are an

exponential number of combinations (regions, pathways) to test before arriving at the final model. The biggest difficulty stems from SEM's requirement of a starting model, founded in biology, which can be difficult to construct without consulting someone experienced in the field.

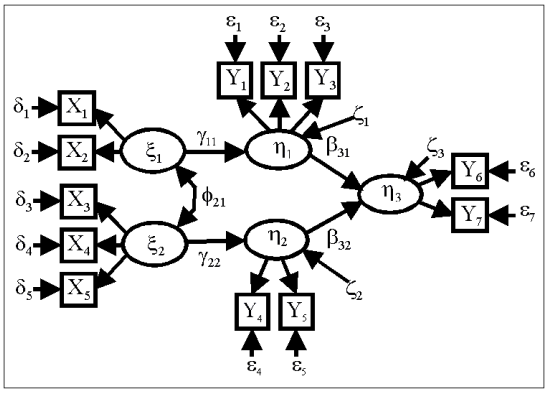


FIGURE 1.1: Structural equation modelling requires a starting model describing the assumed relationships between endogenous(η's) and exogenous(ξ's) variables as well as the distributions assumed of their error sources. The SEM procedure estimates the coefficients β's and γ's of the proposed model and can be compared to other models through a goodness of fit measurement.

The main challenge in the use of SEM in brain pathway analysis is the construction of the a priori model. SEM can validate a proposed pathway model, but it has no way to discover one itself unless the number of observations far exceed the number of parameters in a saturated SEM model. Fortunately, the field of unsupervised pathway analysis has seen rapid development in the recent years, and though relatively little has been applied to functional brain images there have been notable breakthroughs. It is suggested in (Marrelec et al., 2007) to use partial correlation pathways as a starting model for SEM. In his paper, he finds the partial correlation between all region pairs as well as their confidence intervals based on a resampling methodology. If the partial correlation between regions is significantly different from zero, then a link is added between pathways in the starting SEM model. In his experiment, he found that the model based on partial correlation corresponds well to the a priori model based on biology.

There are many methods of pathway analysis that have been applied to data other than functional imaging, such as genomic microarrays. The amount of literature available on this subject is huge, but most techniques focus on datasets with a very large number of variables and few observations. Bayesian networking (Friedman et al., 2000) involves using conditional probabilities to determine dependence across the variable pairings. Displayed graphically, with brain regions as nodes and dependencies as edges, this network can give us an idea of the true neurological patterns in the brain. It is not a method

to be recommended for networks containing large numbers of regions as finding the best model becomes extremely difficult and has been proven to be NP complete (Chickering, 1996).

Relevance networks are another method that has been used to determine relationships among a large collection of variables. Variable/pathway relationships are represented graphically through nodes/edges whose relationships are calculated as a specific functional metric. Pathways have been determined through use of correlations (Butte and Kohane, 1999), and partial correlations (Fuente et al., 2004) with promising results on gene datasets. Partial correlation is of more use to us because it shows the relationship between variables controlling for the others, but if the number of variables is larger than the number of observations, it becomes impossible to compute exactly. In these cases, the partial correlation matrix must be estimated instead as is done in (Schafer and Strimmer, 2005a) and (Peng et al., 2009).

Another popular method in determining pathways in the brain is Cross Correlational Analysis (Welsh et al., 2001). Here we start with the measurement of a single anatomical region, then we compute its correlation with all other points in the brain space. The results of such an analysis are a 3D map showing the strength of the relationship between the chosen region and the rest of the brain, see figure 1.2. The benefit of such a procedure is that it depicts an intuitive, visual display of the correlations with respect to the rest of the brain, and is very useful when it is known beforehand which regions are of true importance. Although marginal correlations are used as default, there is also capability to support partial and canonical correlations as well. As a downside, only one region's correlations may be observed at a time and navigating a dataset can become unwieldy as the number of regions increases.

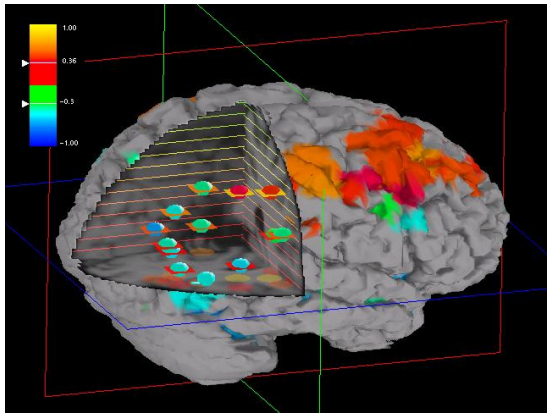


FIGURE 1.2: Brainminer: A visualization suite developed at Stony Brook to deal with the many aspects of Cross Correlation Analysis. This screenshot shows the correlations of a selected brain region to the remaining regions simultaneously in 3D format.

In (Cao and Worsley, 1999) an attempt is made to find significant pathways by computing the correlation between every possible pair of voxels in the brain space then culling those that fall below a threshold based on random field theory. As with CCA, marginal correlations do not yield information on whether the significant relationships are mediated by other regions but partial correlations cannot be substituted as they can not be calculated at voxel level.

1.3 Goal

My research focuses on using partial correlation as a measurement of the true pathway functionality between brain regions. Partial correlation shows us how well one region can be modeled by another excluding the linear contribution from a set of control regions. This measurement is useful in pathway analysis because, so often, neurological pathways are affected by neighboring regions. Partial correlation can be used for checking independence as well (Baba et al., 2004), a diagnostic that can be of great use in the formulation of new neurological models.

The novel methods proposed in this thesis deal with the addition of a descriptive factor into the partial correlation model indicating the inclusion/exclusion status of a specific observation vector to a group. The major question we address is whether this group status affects the strength of the brain pathways under a specific experimental condition. In our experimental case, we have two sets of brain images, one from normal healthy control subjects, the other from cocaine addicted subjects, both taken after subjects have been administered a dose of the psychostimulant methylphenidate. We wish to determine how addiction affects the neurological paths involved in the brain's reward mechanism.

When dealing with marginal and partial correlations, there are many methods to check if one network is equivalent to another or if a network is equal to a specific form (Brien et al., 1984); but most procedures focus only on testing the network as a whole and not on discovering which specific regions pairs are significantly different. Instead, we focus on where the individual links are distinct by looking at each region pair separately.

This approach stems from the need to validate specific pathways controlling for known region dependencies. In practice we already assume that two networks are different, it is finding the exact locations where they are different that interests us. Unfortunately, testing every region pair introduces all the problems that come with multiple comparisons. If each test has a slight chance for error, the chance of at least one test erring is more than slight. For this reason, we want tests that have as high a power as possible

so that the inevitable introduction of multiple test corrections does not extinguish all traces of significance.

Fortunately, we show through simulation that our novel methods do indeed have a higher power than the traditional methods for testing the equality of two partial correlation coefficients.

1.4 Outline of the Thesis

The remaining chapters discuss the current approaches related to comparing partial correlations and introduce two new methods based on a regression analysis framework and a likelihood ratio testing framework.

Chapter 2 describes two widely accepted solutions to the problem of determining equality of two partial correlation coefficients.

Chapter 3 introduces our first new approach to this problem. A test based on a two-level regression model.

Chapter 4 presents our second approach involving a likelihood ratio test.

Chapter 5 enumerates the numerical simulations, the real experimental data to which we applied the two new and two traditional methods, and the results that follow.

Chapter 6 details an extension to the two-level regression model that can incorporate continuous as well as binary covariate information.

Chapter 7 is a summary of what the thesis has accomplished as well as topics available for further research.

Chapter 2

Existing Methods for Comparing Partial Correlations

2.1 Overview

The distribution of the sample correlation coefficient has been known for many years (Fisher, 1915), and tables of computed values have been published as early as 1938 (David, 1938). More recently, the distribution has been formulated symbolically in Mathematica (Barabesi and Greco, 2002) and exact tests have been found for a single marginal correlation coefficient (Goria, 1980; Chang et al., 2008). Similarly, tests for partial correlation are known because the partial correlation coefficient follows the distribution of a marginal correlation coefficient with fewer degrees of freedom (Fisher, 1924).

However, the test comparing two partial correlation coefficients to each other is not known in the exact form. The accepted solution, described in many advanced (Kendall and Stuart, 1973) and elementary (Dowdy et al., 2004) statistics textbooks, is to use the Fisher transform on the two coefficients, and compare them with a standard z-test. For marginal correlation, when the sample sizes are the same, this test has been shown to be equivalent to the one based on the likelihood ratio test (Brandner, 1933), but even so, the exact distribution of the test statistic is not known. Although it is very easy to calculate, a major drawback to the Fisher method is its reliance on the large number theorem and is thus only recommended on large sample sizes (Hotelling, 1953).

Another accepted solution is to use bootstrap resampling to find the confidence intervals of the partial correlation coefficients in each group and analyze their overlapping regions. Unfortunately, bootstrap procedures typically require large samples and have theoretical

assumptions that are hard to verify under all but the most simple conditions. What's more, accurate results often require ten's of thousands of resamples and requires ample computing resources.

2.2 Fisher's Transformation

There is no exact test that can determine the equality of two partial correlation coefficients, but a very accurate approximation involving Fisher's transformation is known for moderate to large sample sizes. First noted in 1915 (Fisher, 1915), Fisher's transformation has been the most widely used method of testing the significance of correlations and partial correlations. Throughout the years there have been many attempts to understand this transformation and why it provides such an accurate approximation. In (Winterbottom, 1979) the transformation is rediscovered as a normalizing and variance stabilizing operation. More recently (Bond and Richardson, 2004), the transformation is interpreted from a geometric standpoint.

A very important relationship between the marginal correlations and the partial correlations controlling for k variables is described in (Fisher, 1924) which states that the distribution of the partial correlation coefficient is exactly that of a marginal correlation coefficient having k fewer degrees of freedom. This relationship allows us to make the comparison of partial correlations by using a modified version of the test that compares marginal correlations.

It was proven in (Anderson, 1984) that when the original data comes from bivariate normal distribution, the asymptotic variance of the fisher transformed correlation coefficient depends on the sample size N but is independent to the actual population correlation coefficient ρ. However, it is shown in (Hawkins, 1989), that no such independence holds between the asymptotic variance of the transformed coefficient and ρ when the data comes from distributions other than bivariate normal. This dependence plays a role in the acceptance/rejection rate of the test based on the Fisher transformation, as evidenced in our simulations.

2.2.1 Hypothesis Test

We are interested in determining the equality of two independent partial correlation coefficients through the use of the Fisher Transformation. The relationship between the partial correlation and the marginal correlation allows us to approximate the distribution of the sample partial correlation coefficient as that of a correlation coefficient

with reduced degree of freedom. The test involves transforming each coefficient into an approximate normal variable, then compares the transformed variables with a simple z-test.

The test comparing two population partial correlations ρ_1 and ρ_2 based on two independent samples controlling for k other variables, where the sample partial correlations are r_1 and r_2 with sample sizes n_1 and n_2 respectively is

$$Z_0 = \frac{\frac{1}{2}(ln(\frac{1+r_1}{1-r_1}) - ln(\frac{1+r_2}{1-r_2}))}{\sqrt{\frac{1}{n_1-k-3} + \frac{1}{n_2-k-3}}} \sim N(0,1) \tag{2.1}$$

under the null hypothesis $H_0 : \rho_1 = \rho_2$.

2.2.2 Drawbacks

A major problem of this method is that the statistical accuracy depends on a large sample. The normalizing properties of the transformation exist only at an asymptotic level and experiments involving small samples may introduce significant errors (Hotelling, 1953).

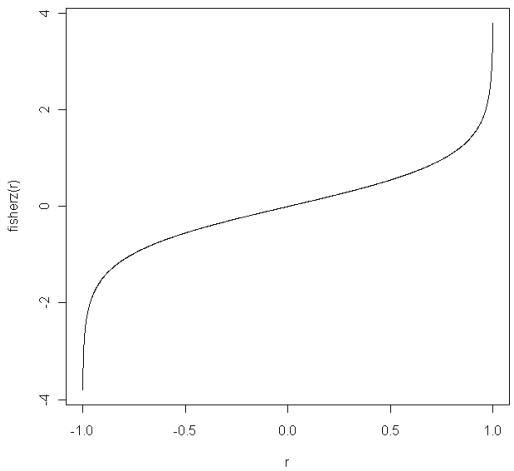


FIGURE 2.1: A graph of Fisher's transformation of the correlation coefficient. The transformed values of the stronger correlations slope faster than the weaker correlations.

In Chapter 5, it is shown that the test based on the Fisher transformation yields the most conservative results of all the methods. This seems to stem from the fact that partial correlation of high absolute value are mapped to a very wide range of z-values, see figure 2.1. The transformation takes an almost linear curve in the range of $-.75$ to .75, but slopes sharply at the higher values. Because the transformed values are further

apart when both partial correlations have very high absolute value than when they both have low absolute values, it makes intuitive sense that the acceptance rate of the Fisher based z-test is more dependent on the strength of partial correlation coefficient than tests not relying on the Fisher transformation. This phenomenon is noted in (Hawkins, 1989) in data originating from distributions other than bivariate normal. Further evidence of the bias in acceptance rate determined via Fisher z-test is given when we present our simulation results in Chapter 5.

2.3 Bootstrapping

A widely used technique to infer qualities of an estimate is the bootstrap resampling method (Efron and Tibshirani, 1994). In its most basic form, a confidence interval is created for the parameter of interest under the null distribution through resampling, and if the parameter estimate taken from the original sample is found to lie outside the constructed boundary, the null hypothesis is rejected.

This simple technique works well for parameters that follow a symmetric distribution, like the mean of a normal distribution, but correlation does not have a symmetric distribution. To illustrate this point, imagine a random sample taken from a bivariate distribution with a very high correlation parameter. Because the correlation is so high it is more likely that the observed correlation of the random sample would lie below the true parameter than above it. This is demonstrated in figure 2.2, through a Monte-Carlo simulation.

There are a number of ways to compute the bootstrap confidence intervals to deal with non-symmetrical parameters of this sort (Carpenter and Bithell, 2000). Our problem is best dealt with through the use of the bias corrected accelerated method, also known as BCa (Efron, 1987), that more accurately represents the confidence interval of a non-symmetric parameter. In BCa, two terms are introduced in the formulation of the confidence interval addressing the correction of the estimate's bias and the acceleration of the estimate's variance to its asymptotic value.

A simple way to tell if two groups' partial correlations are equal to each other is to generate two bootstrap confidence intervals and see if they overlap. But computing the confidence intervals in this manner in order to test the equality of the two parameters yields a very conservative result. If two 95% confidence intervals do not overlap, then it can be said that the values are different at a significance level of at least 5%. But the same can not be said of the converse, i.e. if the confidence intervals do overlap, there is still a chance that the two values are in fact different.

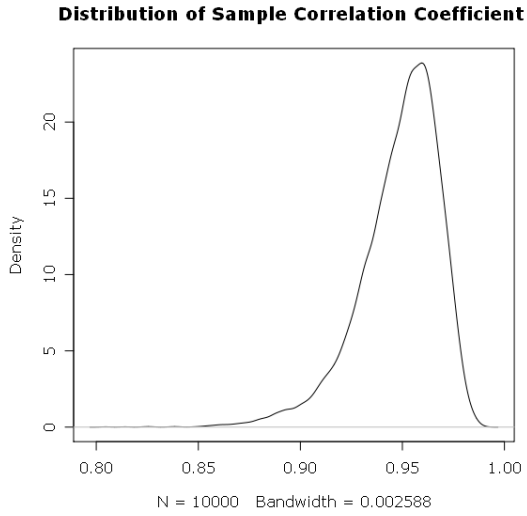


FIGURE 2.2: A sample of size 30 was taken from a bivariate distribution with means 0, variances 1, and correlation coefficient 0.95 from which the sample correlation coefficient was computed. This was repeated 10000 times and the density of the sample correlation coefficients is estimated as shown (Sheather and Jones, 1991). This skewness is present for all non-zero values of the correlation coefficient.

The conservative nature of the test comparing two confidence intervals in this manner has been examined in (Payton et al., 2003). Through a simulation study involving standard normal variables, they found the test rejecting the null when two 83% confidence intervals fail to overlap to have a true significance level of 5% . We could apply this same methodology to our own by computing the bootstrap confidence intervals of partial correlation for each group to see if the intervals overlap.

Instead, we found it easier to determine the equality of partial correlations by bootstrapping the differences directly. In each resampling step, the partial correlation is computed for both groups, and the differences noted. In the end, the partial correlation difference of the original sample is compared to the list of the N computed bootstrapped differences to see if it lies outside the accepted boundary of values. Examining the parametric estimates in this manner gives us less information than the dual confidence interval approach, as only the differences are investigated, but it may improve the statistical accuracy because there are fewer parameters under consideration. In the end, we decided on a test that rejects the null hypothesis of equality in partial correlations if the bias corrected accelerated bootstrapping method produced a confidence interval that does not include 0.

2.3.1 Hypothesis Test

The bootstrapping algorithm is described below.

Given observation matrices X and Y of the first and second group we want to see if the partial correlation between two columns controlling for the remaining columns is the same between the groups.

- repeat N times
  - draw n_1 resamples with replacement from matrix X, then compute sample partial correlation ρ_x.
  - draw n_2 resamples with replacement from matrix Y, and compute sample partial correlation ρ_y
  - compute the difference of the two partial correlations ρ_d.
- sort the N values of ρ_d taking a lowest and highest quantiles, depending on the significance level desired, to serve as the confidence interval.
- if the confidence interval does not contain 0 then we reject the null hypothesis that the two partial correlations are the same between the two groups.

2.3.2 Drawbacks

Bootstrap's great boon is that there is no need to make the assumption of which distributions the data originates from. But, by disregarding all parametric qualities that may be present in the data, it is hard to gauge the accuracy of the bootstrap method from an entirely mathematical approach. The assumptions required to meet bootstrap's asymptotic properties are hard to check when dealing with anything but trivial examples. Additionally, because dealing with small samples can lead to overly optimistic bounds on bootstrapped confidence intervals (Schenker, 1985), this method is usually considered viable only for large starting samples.

2.4 New Methods

The following chapters describe two novel ways to determine the equality of two partial correlation coefficients. Like the currently accepted methods, they are not exact, but they will be shown to perform on par with the traditional methods, and in some cases, result in greater accuracy, even on smaller sample sizes.

Chapter 3

Two Level Regression

3.1 Overview

This chapter deals with determining the equality of two partial correlation coefficients through use of a two-level regression modeling approach. We make use of the fact that the partial correlation between two variables is equal to the correlation of the variables' residuals after being regressed on the controlling variables. By further decomposing the model relating the two residuals to account for group status, we determine the significance of the influence that group status plays on the partial correlation between the variables under consideration.

The statistical methods employed in this chapter operate under the same assumptions required for ordinary least squares (OLS) regression, namely:

1. There exists a true linear relationship between the variables being regressed and the variables being regressed upon, such that the expectation of the residual variable is zero.
2. The observations of the variable being regressed are independent to each other.
3. The variables of interest have a constant variance over their ranges (homoscedasticity).

Assumptions 1 and 2 are typical of regression analysis and their validity is usually no cause of concern. The first can be verified with a scatterplot of the two variables of interest, and the second is implicit in the random nature of clinical designs. As for assumption 3, the homoscedasticity of the residuals is not effected by the two-level design introduced in the next section. If the starting observations have satisfied the

variance conditions, then their residuals will also satisfy the conditions. If on the other hand, the starting observations do not satisfy the conditions, then the usual procedures of data transformation, using robust estimators for standard error or using weighted least squares regression can be applied.

3.2 Two Level Regression

Our observations come from one of two groups $G = \{0, 1\}$, and we are interested in the effect this group factor has on the partial correlation between some pair of variables given the other controlling measurements. For ease of notation, let us say we are interested in the partial correlation between variables (Y_1, Y_2) controlling for measures $(Y_3, \ldots, Y_p)$.

The first step, using ordinary least squares, is to obtain the residuals of the variables of interest regressed on the controlling measures

$$\hat{Y}_1 = \hat{\beta}_0 + \sum_{i=3}^{p} \hat{\beta}_i Y_i$$

$$\hat{Y}_2 = \hat{\gamma}_0 + \sum_{j=3}^{p} \hat{\gamma}_j Y_j$$

$$R_1 = Y_1 - \hat{Y}_1 \tag{3.1}$$

$$R_2 = Y_2 - \hat{Y}_2, \tag{3.2}$$

where the β_i's and γ_j's are the OLS estimates of regressing Y_1 on the controlling measurements, and Y_2 on the controlling measurements, respectively.

In order to analyze the partial correlations between the main variables (Y_1, Y_2), we will be working with measures of correlation between the residuals (R_1, R_2). We draw on the fact that the significance of the test for an independent variable's coefficient being equal to zero in ordinary least squares regression gives the same significance as the test for the correlation being zero between that independent variable and the dependent variable.

In the next stage, we regress residual R_1 onto residual R_2 in the first level of our model

$$R_1 = b_0 + b_1 R_2 + \epsilon, \tag{3.3}$$

then further decompose the coefficient b_1 in the second level of our model.

$$b_1 = a_0 + a_1 G, \tag{3.4}$$

where G is the group status of the observation.

Putting these together we get our two-level model

$$\begin{aligned} R_1 &= b_0 + b_1 R_2 + \epsilon \\ R_1 &= b_0 + (a_0 + a_1 G) R_2 + \epsilon \\ R_1 &= b_0 + a_0 R_2 + a_1 G R_2 + \epsilon. \end{aligned} \tag{3.5}$$

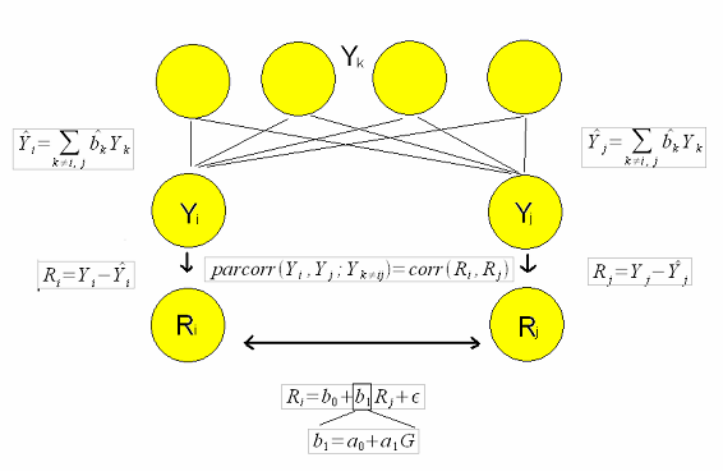


FIGURE 3.1: An overview of two-level regression model. Residuals are obtained by regressing the k controlling measurements onto variables Y_i and Y_j. We then decompose the model relating the correlation of residuals R_i and R_j to the partial correlation between the original variables Y_i and Y_j to account for the effect of group status. The b_1 coefficient in the model $R_i = b_0 + b_1 R_j + \epsilon$ shows the significance the correlation between R_i and R_j. Similarly, the a_1 coefficient in $b_1 = a_0 + a_1 G$ tells us the significance group status, G, has in the same relationship.

The coefficient a_1 in 3.5 shows us the significance of group status on the linear relationship between residuals R_1 and R_2, i.e. the correlation between R_1 and R_2, thus showing us the significance of the partial correlation between variables Y_1 and Y_2. The problem here is that the residuals have a slight dependence on each other and further analysis requires one additional step.

Using traditional notation, let us rewrite our its model

$$Y_r = R_1, \qquad X_r = \begin{bmatrix} | & | & | \\ \mathbf{1} & R_2 & R_2 G \\ | & | & | \end{bmatrix}, \qquad B_r = \begin{bmatrix} b_0 \\ a_0 \\ a_1 \end{bmatrix},$$

$$Y_r = X_r B_r + \epsilon \tag{3.6}$$

The covariance structure of residuals Y_r

$$cov(Y_r) = cov(\epsilon) = (I_n - H)\sigma^2 = \Sigma_r\sigma^2 \tag{3.7}$$

$$H = X_{3..p}(X_{3..p}^T X_{3..p})^{-1} X_{3..p}^T, \tag{3.8}$$

has non zero entries in the off-diagonals and therefore fails assumption 2 required for ordinary least squares regression (Draper and Smith, 1998).

3.2.1 Dependent Residuals

We want to perform OLS on the residuals in order to determine the significance of the a_1 coefficient, but the residuals are not independent and therefore the mathematical basis for the significance tests is degraded. Although assumption 2 is not satisfied, in practice, the correlation amongst residuals is very small (Anscombe and Tukey, 1963) and we may be able to use ordinary least squares methods to perform our regression with only slight inaccuracy. Disregarding the dependence of residuals is generally accepted for graphical methods and *by eye* approximations, but the accuracy of a significance test requires more rigorous assumptions to hold.

To deal with the dependency problem we try to apply the usual technique of multiplying the residuals by a matrix that will decorrelate the covariance structure 3.6, as described in (Seber, 1977).

$$\Sigma_r^{-\frac{1}{2}} Y_r = \Sigma_r^{-\frac{1}{2}}(X_r B_r + \epsilon) \tag{3.9}$$

$$Y_r^* = X_r^* B_r + \epsilon^*, \tag{3.10}$$

where the $cov(\epsilon) = \Sigma_r$.

Supposing we could find the proper matrix $\Sigma_r^{-\frac{1}{2}}$, the covariance structure of transformed residuals Y_r^* would be

$$cov(Y_r^*) = cov(X_r^* B_r + \epsilon^*) \tag{3.11}$$

$$= cov(X_r^* B_r) + cov(\epsilon^*) + cov(X_r^* B_r, \epsilon^*) \tag{3.12}$$

$$= cov(\epsilon^*) \tag{3.13}$$

$$= cov(\Sigma_r^{-\frac{1}{2}}\epsilon) \tag{3.14}$$

$$= \Sigma_r^{-\frac{1}{2}}\Sigma_r\sigma^2{\Sigma_r^{-\frac{1}{2}}}^T \tag{3.15}$$

$$= I_n\sigma^2 \tag{3.16}$$

and the residuals would be decorrelated with one another.

Unfortunately, this routine leads to a dead end. The covariance matrix $\Sigma_r = (I_n - H)\sigma$ is idempotent, so any number of applications of the proposed transformation to the covariance matrix will yield the same covariance matrix.

There are a number of ways to deal with this type of dependency issue. The problem with using the original n residuals is that statistical tests based on them will ultimately be making use of redundant information. Recursive residuals (Brown et al., 1975), and recovered residuals (Jensen and Ramirez, 1999) are two popular methods of extracting the $n - k$ residuals that contain only the non redundant information.

Recursive residuals, introduced in (Plackett, 1950) are constructed from a starting model of k variables, and built sequentially in such a way that the trailing residuals in the sequence are not in the computation of the earlier ones. Construction in this way makes the residuals independent from each other, but the order of the observations will have an effect on the results.

Recovered residuals make use of of spectral decomposition methods to recover the $n-k$ residuals that can fully describe the information of the n residuals. We make use of the Best Linear Unbiased Scaled (BLUS) residuals, introduced in (Theil, 1965), that accomplishes this in a way that minimizes the expected squared length of the transformed residuals and thus achieves certain optimally properties (Grossman and Styan, 1972), most importantly, the unbiasedness of the residual estimates.

Theil's procedure gives us a transformation matrix A that, under the assumption that the data comes from a normal distribution, transforms the residuals into normally distributed and independent variables. Under less restrictive assumptions where the original data is not assumed normally distributed, the BLUS residuals are still uncorrelated to each other.

The matrix A is sized $(n - k) \times n$, and therefore reduces the number of residuals after transformation. This reduction eliminates redundancy allowing us to make inferences on the now uncorrelated residuals.

Using BLUS residuals, our new model becomes

$$A^T Y_r = A^T X_r B_r + A^T \epsilon \tag{3.17}$$

$$Y_r^* = X_r^* B_r + \epsilon^* \tag{3.18}$$

$$cov(Y_r^*) = cov(Y_r^*) = I_n \sigma^2, \tag{3.19}$$

where A is Theil's transformation matrix.

3.2.2 Swapping Residual Roles

The above procedure gives us a measure of the group factor's effect on partial correlation between variables (Y_1, Y_2), but it is not the end of the analysis. An important question arises: what happens to the significance level of group factor's coefficient if we switch the order of the regression around, i.e. we regress R_2 onto R_1?

$$\begin{aligned} R_2 &= c_0 + c_1 R_1 + \xi \\ R_2 &= c_0 + d_0 + d_1 G R_1 + \xi \\ R_2 &= c_0 + d_0 R_1 + d_1 G R_1 + \xi. \end{aligned} \tag{3.20}$$

The answer: we get a different p-value.

Without the group factor included, switching the dependant/independent variables has no effect on the significance of the regression coefficient. But with the group factor included, the problem essentially becomes a multiple regression model on the residuals with an interaction term included, but only one main effect present. In the first procedure 3.21, because R_1 is the dependent variable, the coefficient a_1 tells of the residual R_2's effect on R_1, and thus is a measure of linearity from Y_1 to Y_2. Similarly, the second procedure 3.22, d_1 gives us a measure of linearity from Y_2 to Y_1. In practice, switching the order on the residuals gives two different, and possibly conflicting measurements for the effect that group status plays on the partial correlation between variables Y_1 and Y_2. We are left with two different measurements but no clear method to choose one over the other.

$$R_1 = b_0 + a_0 R_2 + a_1 G R_2 + \epsilon \tag{3.21}$$

$$R_2 = c_0 + d_0 R_1 + d_1 G R_1 + \xi \tag{3.22}$$

As a measure of partial or marginal correlation, coefficients in a regression equation have no causal direction. We cannot determine causality from the significance of the correlation coefficient. But, there is an inherent assumption in the dependence relation between residuals when we model them as 3.21 or 3.22. In the first model, residual R_1 is considered dependant on R_2 and in the second model, it is R_2 that is considered dependant on R_1.

If we have directional information on the causal pathways between the variables beforehand, then it is clear which of the equations, 3.21 or 3.22 to choose. For example, if we

know that x causes y, then we would choose to model the residuals of y as the dependent variable and residuals of x as the independent. To choose roles in this way, we would require a starting model, like SEM, to specify the causal dependences.

For the simulation and experimental data of chapter 5, we are not privy to any information on causal structure and one equation is as logical as the other. In the results section of the next chapter, we do not make use of causality information and simply report the average p-value of the two models.

The silver lining to the two-level regression method is that this drawback is indeed also its advantage if we consider the partial correlation testing as a data exploration/hypothesis generating step, before an SEM confirmatory analysis. See section 7.1 for more on this topic.

3.3 Drawbacks

Since this method relies on ordinary least squares regression, all the drawbacks and pitfalls present in OLS are present in our method. The biggest problem is what to do when there exists a non-linear relationship between the variables of interest. In this case, OLS can not give us an accurate picture, and since partial correlation is a measure of linearity, a true measurement relating the two variables will not be found.

Chapter 4

Likelihood Ratio Test

4.1 Overview

In this chapter I describe our LRT for determining whether the partial correlation of two variables (Y_1, Y_2) are the same between groups. For clarity, the LRT procedure is explained first by stepping through a simpler problem, then through the proposed problem of comparing two groups.

For the first section we start with the assumption that our dataset consists of n iid samples taken from the multivariate normal distribution of the form

$$(y_1, y_2) \overset{iid}{\sim} MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{4.1}$$

and the controlling factors $(y_3, \ldots, y_4)$ are known constants from which we can linearly model (y_1, y_2).

We are interested in the partial correlation between y_1 and y_2 given the values of the $y_3, \ldots, y_p$ which can be calculated by performing a linear regression of the main variables on the conditional variables and taking the marginal correlation of the resulting residuals. The following sections detail our approach to inferring the equality of two partial correlations through the likelihood ratio test approach.

4.2 LRT Method

4.2.1 An Easier Problem

Before describing our method for determining whether the partial correlation between two groups is equal let us start with a simpler example: determining whether the partial correlation in a single sample is equal to 0. The solution to this example is already known in exact form (Kendall and Stuart, 1973). We solve it here with a LRT to illustrate the method.

Like the previous approach, our method starts with a regression of the variables of interest onto the variables we wish to control for. Let $Y = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ be our regression model, where our two variables of interest are represented by $Y = (\boldsymbol{y_1}, \boldsymbol{y_2})$, our controlling variables by $X = (\boldsymbol{y_3}, \ldots, \boldsymbol{y_p})$, and $\boldsymbol{\beta}$ is the $(p-2) \times 2 = \alpha \times 2$ matrix of regression coefficients, with $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, I_n \otimes \Sigma)$. Least squares estimates of $\boldsymbol{\beta}$ are calculated as $X(X^TX)^{-1}X^TY = HY$, and the residuals $R = (\boldsymbol{r_1}, \boldsymbol{r_2}) = Y - \hat{Y}$ follow a singular multivariate normal distribution formulated by Khatri (1968).

$$(\boldsymbol{r_1}, \boldsymbol{r_2}) \sim N_{n\times 2}^{n-\alpha,2}((\mathbf{0}, \mathbf{0}), (I_n - H) \otimes \Sigma). \tag{4.2}$$

By definition, the partial correlation between two variables given a set of controlling variables is equal to the marginal correlation of the residuals of the variables' residuals when regressed on the controlling variables (Kendall and Stuart, 1973). In the same sense, we make an inference on the equality of partial correlation between groups by analyzing the correlation of the residuals via LRT.

Because the residuals have dependence on each other, the usual method of constructing the likelihood function from the product of independent pdf's does not work here.

$$L(\theta; \boldsymbol{r}) \neq \prod_{i=1}^{n} f_i(\boldsymbol{r}; \theta) \tag{4.3}$$

However, this fact does not affect us since the joint distribution of the residuals is known (Diaz-Garcia et al., 1997) and can be used directly as the likelihood function.

$$\begin{aligned} L(\theta; \boldsymbol{r}) &= f_R(\boldsymbol{r}; \Theta, \Sigma) \\ &= \frac{1}{(2\pi)^{(n-\alpha)p/2}(\prod_{i=1}^{p} \lambda_i^{(n-\alpha)/2})(\prod_{j=1}^{(n-\alpha)} \delta_j^{p/2})} etr\left\{-\frac{1}{2}\Theta^{-}\boldsymbol{r}\Sigma^{-}\boldsymbol{r}^T\right\}, \end{aligned} \tag{4.4}$$

where $\Theta = I_n - H$, λ_i are the eigenvalues of Θ, δ_j are the eigenvalues of Σ and A^- stands for the pseudoinverse of A.

Looking at equation (4.2) we can see the covariance structure of the residuals is composed of two parts: $(I_n - H)$ deals with the dependence across the n observed residual vectors and Σ deals with the covariance of the residual variables themselves. It is Σ that details the partial correlation between variables (y_1, y_2). If the off diagonal entries in the residual's Σ are 0, it tells us that the correlation between the residuals is 0, which in turn tells us that the partial correlation between (y_1, y_2) is 0.

Knowing the likelihood distribution, we make use of the LRT to determine whether the partial correlation is equal to 0 [1].

$$\lambda = \frac{L_0}{L_1} = \frac{f_R(\boldsymbol{r}; \Theta, \hat{\Sigma}_0)}{f_R(\boldsymbol{r}; \Theta, \hat{\Sigma}_1)}, \tag{4.5}$$

where $\hat{\Sigma}_0 = \begin{bmatrix} \hat{\sigma}^2_{1,H_0} & 0 \\ 0 & \hat{\sigma}^2_{2,H_0} \end{bmatrix}$ is the matrix maximizing the numerator(the likelihood function under the null hypothesis), and $\hat{\Sigma}_1 = \begin{bmatrix} \hat{\sigma}^2_{1,H_1} & \hat{\sigma}^2_{12,H_1} \\ \hat{\sigma}^2_{12,H_1} & \hat{\sigma}^2_{2,H_1} \end{bmatrix}$ is the matrix maximizing the denominator(likelihood under alternative hypothesis).

From here we use the χ^2 approximation for the LRT to determine the p-value of the test. The statistical significance of the test depends on the goodness of fit of the likelihood ratio under the null hypothesis to that of a χ^2 distribution. In the next chapter, we show evidence that these conditions are indeed met via simulation study.

4.2.1.1 Regularity Conditions

The procedure for calculating p-value makes use of the χ^2 approximation to the LRT, and thus requires all the regularity conditions necessary in problems of this nature. Because we do not follow the typical method of finding the likelihood function we can not blindly accept the regularity assumptions, but it will be shown that all the conditions necessary for the use of the χ^2 approximation can be met.

The asymptotic properties that allow our use of the χ^2 approximation to the LRT requires certain regularity conditions. The main condition is that the Fisher information, which is a measure of the sharpness of the support curve at the location of the maximum likelihood estimator, is finite and is defined everywhere. This is due to the fact that the χ^2 approximation makes use of the Taylor expansion with derivatives up to the 3rd

[1] These functions can be maximized numerically using various methods freely available in statistical computing software packages. For our simulations, we use the simplex algorithm proposed by Nelder and Mead in 1965 (Nelder and Mead, 1965)

degree. For the Taylor expansion to be accurate it is imperative that the support curve centered around the maximum likelihood estimators be smooth with no breaks or sharp points. In this case, it is enough to check that the Fisher information exists at all values of the parameters being estimated.

It can be shown that the Fisher information matrix of the singular multivariate normal distribution exists for all values of parameter Σ. The pdf for the singular MVN distribution, given by 4.4, can be simplified to the following form in terms of the variables of interest Σ

$$\begin{aligned} f_R(\boldsymbol{r};\Theta,\Sigma) &= k \cdot etr\left\{A \times \Sigma^{-}\right\} \\ &= k \cdot etr\left\{ \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} \times \begin{bmatrix} a & c \\ c & b \end{bmatrix}^{-} \right\} \\ &= k \cdot e^{\left\{ \frac{ak_{22}}{ab-c^2} - \frac{ck_{21}}{ab-c^2} - \frac{ck_{12}}{ab-c^2} + \frac{bk_{11}}{ab-c^2} \right\}} \end{aligned} \tag{4.6}$$

From this simplified form 4.6, it is easy to see that the Fisher information requirement $\frac{\delta}{\delta\theta} lnf(\boldsymbol{r};\theta) < \infty$, for $\theta = \{a, b, c\}$ is satisfied for all values of k and for all values of a, b, and c, as long as $ab \neq c^2$, i.e. as long as $\sigma_1\sigma_2 \neq \sigma_{12}$ or equivalently, as long as $-1 < \rho < 1$.

4.2.2 Back to our original problem...

Now that we've dealt with the case of one group, expanding the test to infer the equality of two groups is easy. As before, let's assume we have two sets of observations: X from the first group and Y from the second.

$$\begin{aligned} (x_1, x_2) &\overset{iid}{\sim} MVN(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_x}) \\ (y_1, y_2) &\overset{iid}{\sim} MVN(\boldsymbol{\mu_y}, \boldsymbol{\Sigma_y}) \end{aligned}$$

with values of the controlling variables as $(x_3, \ldots, x_p)$ and $(y_3, \ldots, y_p)$. In this case, we regress variables (x_1, x_2) on $(x_3, \ldots, x_p)$ to get residuals

$$\boldsymbol{r_x} \sim N_{n_x \times 2}^{(n_x-\alpha),2}((\mathbf{0},\mathbf{0}), (I_{n_x} - H_x) \otimes \Sigma_x), \tag{4.7}$$

and (y_1, y_2) on $(y_3, \ldots, y_p)$ to get residuals

$$\boldsymbol{r_y} \sim N_{n_y \times 2}^{(n_y-\alpha),2}((\mathbf{0},\mathbf{0}), (I_{n_y} - H_y) \otimes \Sigma_y). \tag{4.8}$$

This time we are interested in the relationship between the off diagonal entries of the matrices Σ_x and Σ_y. The correlation between residuals can be compared by observing the part of the covariance structure pertaining to the residuals.

Because of the relationship between the marginal correlation of the residuals and the partial correlation of the original vectors, when the correlation values off diagonal entries are equal in Σ_x and Σ_y the partial correlations of $(x_1, x_2|x_3, \ldots, x_p)$ and $(y_1, y_2|y_3, \ldots, y_p)$ will be also be equal. By maximizing the likelihood functions under the null where the correlations are the same between groups, and under the alternative, where the correlations are allowed to vary, we can determine the significance of the test through the likelihood ratio test.

Using the same procedures as the simpler case, we compute the likelihood ratio, using the fact that the observations from Y and X are independent, and $f(X, Y) = f(X)f(Y)$ the likelihood ratio becomes:

$$\lambda = \frac{L_0}{L_1} = \frac{f_X(\boldsymbol{r_x}; \Theta_x, \hat{\Sigma}_{0x}) f_Y(\boldsymbol{r_y}; \Theta_y, \hat{\Sigma}_{0y})}{f_X(\boldsymbol{r_x}; \Theta_x, \hat{\Sigma}_{1x}) f_Y(\boldsymbol{r_y}; \Theta_y, \hat{\Sigma}_{1y})}, \tag{4.9}$$

where the numerator is maximized over 5 parameters: a, b, c, d, ρ.

$$\hat{\Sigma}_{0x} = \begin{bmatrix} a & ab \cdot \rho \\ ab \cdot \rho & b \end{bmatrix}, \qquad \hat{\Sigma}_{0y} = \begin{bmatrix} c & cd \cdot \rho \\ cd \cdot \rho & d \end{bmatrix} \tag{4.10}$$

and the Denominator is maximized over 6 parameters: a, b, c, d, ρ_1, ρ_2.

$$\hat{\Sigma}_{1x} = \begin{bmatrix} a & ab \cdot \rho_1 \\ ab \cdot \rho_1 & b \end{bmatrix}, \qquad \hat{\Sigma}_{1y} = \begin{bmatrix} c & cd \cdot \rho_2 \\ cd \cdot \rho_2 & d \end{bmatrix}. \tag{4.11}$$

From this we calculate a p-value based on the χ^2 approximation to the likelihood ratio test.

4.3 Drawbacks

The performance of this method relies on the ability to approximate the likelihood ratio test with a χ^2 distribution, a relationship that holds true only under asymptotic conditions. For large samples this approximation is adequate, but as the sample size gets smaller, so to does the accuracy of the χ^2 approximation.

In the next chapter we compare the LRT under the null to a χ^2 distribution through simulation and find an adequate fit. But comparison in this manner is not possible with

the single observation sample of our real world data. If there is reason to believe the LRT does not conform to the assumptions required for the significance testing, there are still other ways to proceed.

One such way is to perform bootstrapping on the LRT as proposed in McLachlan (1987). In this paper, the data are resampled and the LRT is computed each time to construct a null distribution to the LRT. The original sample's LRT is compared to the range of the constructed LRT distribution to determine the significance of the test. Rotnitzky et al. (2000) pointed out that special asymptotic distributions for the LRT can be formulated to deal with sampling distributions having a singular Fisher information matrices.

The other drawback of this method is its reliance on numerical methods to maximize the likelihood function under the null and alternative hypothesis. If the numerical algorithm selects a local maximum instead of the global maximum, there is a possibility that our LRT and its χ^2 approximation becomes invalid. One such way this can happen is if a local maximum is found under the alternative conditions and the global maximum is found under the null in such a way as to make the likelihood ratio greater than 1. Because the conditions under the null are assumed a subset of the conditions under the alternative, it is never possible for the null likelihood to be greater than the alternative likelihood, but this may happen in practice due to the numerical nature of determining our maximizing estimates.

Chapter 5

Simulation and Experiment

5.1 Overview

In this chapter, we apply the previously described methods for determining the differences in partial correlations to a series of simulated datasets and analyze their results.

We then apply all four methods to the experimental brain image dataset consisting of 16 normal controls, and 25 cocaine addicted subjects.

5.2 Simulation Diagnostics

The most direct approach to test the new methods from the previous chapters is to generate a set of data and compare their performance relative to the established/existing methods. Since the approach based on the Fisher transformation is the most widely used test for determining whether two partial correlations are equal to each other, we will be using it as a baseline to gauge the other two methods.

With any hypothesis test there is a chance of getting an incorrect outcome. This can occur in one of two ways: a type 2 error where the test mistakenly accepts the null hypothesis when the condition stated in the alternative hypothesis is true, and the more severe type 1 error where the test rejects the null hypothesis in the case that the null is true. We will be viewing the new tests in terms of the type 1 and type 2 errors and comparing them to the test based on Fisher's approximation to see how they behave.

5.2.1 Simulated Power Curve (1 - Type 2 Error)

If a test accepts the null hypothesis when it should have rejected it, it has committed a type 2 error. When we know the distribution of the test statistic, we can calculate the power of the test given the specified parameter under the alternative hypothesis. For example the Fisher transformed statistic follows an approximate normal distribution, as does the test based on two level regression, while the likelihood ratio test follows an approximate χ^2 distribution and the power curves can be constructed for these approximations. However, the exact distributions to these statistics are either unknown, or very complicated, and calculating the power curve can be a very drawn out procedure.

Instead, we focus on a Monte-Carlo approach that can give us the same information but does not require the knowledge of the statistic's exact distribution. Basically, we generate many random samples where we know the alternative hypothesis should be accepted and see how often the test under consideration accepts/rejects the null. If we bin the results together by alternate hypothesis parameter we can construct a power curve by noting the percentage of tests that have correctly rejected the null.

The power curve simulation is performed as follows:

- Repeat N times
  - Generate data for the two groups. We will assume that data from the first group comes from a multivariate normal distribution with a mean vector $\boldsymbol{\mu_x}$ and covariance matrix Σ_x. Similarly, the second group's data comes from a multivariate normal of mean $\boldsymbol{\mu_y}$ and covariance matrix Σ_y.
    1. Generate a random positive definite matrix Σ_x of size $k \times k$ that will be used as the covariance matrix of group 1 (Joe, 2006).
    2. Duplicate Σ_x into Σ_y changing the values so as to make the partial correlation between the first two elements zero. This is done by transforming covariance matrix Σ_x into its corresponding partial correlation matrix through the inverse operation (Schafer and Strimmer, 2005b), setting the elements $\rho_{12} = \rho_{21} = 0$, then transforming the matrix back into its original form.
    3. Generate X, n_1 samples from a multivariate normal distribution of mean $\mathbf{0}$ and covariance Σ_x (Genz, 1992) and Y, n_2 multivariate samples with mean $\mathbf{0}$ and covariance Σ_y.
  - Perform the two group comparison testing $H_0 : \rho_x = \rho_y$, where ρ_x is partial correlation of the first two columns of X controlling for the remaining columns

of X, and ρ_y is defined similarly for Y. For each test, report the p-val of the test, as well as the actual difference in partial correlations.

* Using the Fisher transformation procedures described in chapter 2.
* Using the two-level regression described in chapter 3.
* Using the likelihood ratio test described in chapter 4.

Once we have a suitable number of data points, we can plot a simulated power curve by showing the percentage of tests that correctly reject the null hypothesis against the binned values of the true partial correlation difference between groups. This plot shows us how likely the methods under consideration will reject the null given a true difference in partial correlation. For any given significance level α, we expect to see a low rejection percentage when the true difference in partial correlation is zero, and a higher rejection percentage as the difference grows. Because we only generate a few thousand data points, the binned results are somewhat noisy. To help us make sense of the data we smooth the simulated power curve via LOWESS (Cleveland, 1979).

5.2.2 Simulated Type 1 Error Curve

When a hypothesis test incorrectly rejects the null it has committed a type 1 error. As before, the exact error curve can be constructed only if the exact distribution of the statistic is known. Instead we utilize the simulated type 1 error curve, constructed it in a similar manner to the type 2 curve described above, the only difference is that this time the covariance matrices are equal ($\Sigma_x = \Sigma_y$) thereby giving us a set in which the null hypothesis is always true.

- Repeat N times
  - Generate data for the two groups. We will assume that data from the first group comes from a multivariate normal distribution with a mean vector $\boldsymbol{\mu_x}$ and covariance matrix Σ_x. Similarly, the second group's data comes from a multivariate normal of mean $\boldsymbol{\mu_y}$ and covariance matrix Σ_y.
    1. Generate a random positive definite matrix Σ_x of size $k \times k$ that will be used as the covariance matrix of group 1.
    2. Duplicate Σ_x into Σ_y.
    3. Generate X, n_1 samples from a multivariate normal distribution of mean $\mathbf{0}$ and covariance Σ_x (cite) and Y, n_2 multivariate samples with mean $\mathbf{0}$ and covariance Σ_y.

- Perform the two group comparison testing $H_0 : \rho_x = \rho_y$, where ρ_x is partial correlation of the first two columns of X controlling for the remaining columns of X, and ρ_y is defined similarly for Y. For each test, report the p-val of the test, as well as the true partial correlations.
  - Using the Fisher transformation procedures described in chapter 2.
  - Using the two-level regression described in chapter 3.
  - Using the likelihood ratio test described in chapter 4.

This time we plot the percentage of rejected tests to the true partial correlation which is in fact the same between groups. We expect the percentage of rejected tests to be constant and equal to the significance level α regardless of the true value of partial correlation.

5.3 Simulation 1

For our first simulation we choose a relatively large sample size of $n_1 = n_2 = 30$ and $k = 6$, in order to verify that the assumptions required for each method are met. Since we are working with a large sample size, we can expect the methods based on bootstrapping and Fisher's approximation to be accurate. Since the data is coming from a multivariate normal distribution, the assumptions required for the two-level regression approach are also satisfied.

Under the null hypothesis, the distribution of the LRT statistic should be distributed as a χ^2 with 1 degree of freedom. To validate the LRT procedure, we construct the distribution of the LRT under the null hypothesis condition by generating a large number of datasets, and compare it to a random sample from a χ^2 with the Kolmogorov-Smirnov test (Conover, 1971). The resulting statistic $D = 0.0128$ and p-value of 0.3857 shows this condition to be satisfied. A visual display of the sample density (Sheather and Jones, 1991) of the LRT performed on 10k sets, compared to 10k samples taken from a χ^2 distribution with 1 degree of freedom can be seen in figure 5.1.

The simulated power curve in figure 5.2 shows us the percentage of tests rejecting the null hypothesis given the true difference in partial correlation between groups. A perfect test would would reject the null for every non zero difference and would display a straight line at the value of 1.0 for its power *curve*. Since a test like this does not exist, a more reasonable curve would have its lowest points around the area where the partial correlation difference is zero and would rise to 1.0 as the the true difference expands. Such a curve is exhibited in figure 5.2. In this simulation all methods display reasonable

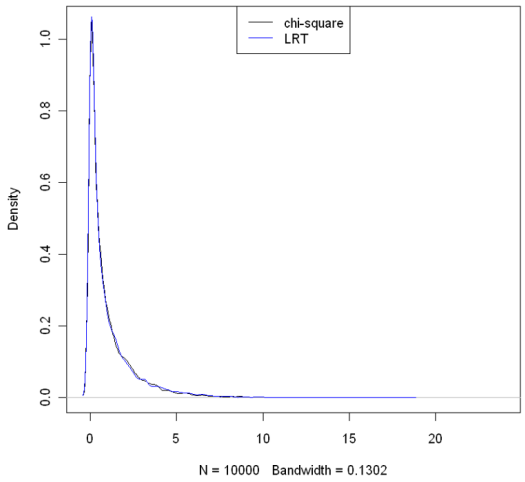


FIGURE 5.1: Simulation 1($n1 = n2 = 30$, $k = 6$): This figure shows the estimated density of the 10k independent observations LRT under the conditions stated in the null hypothesis, along with the density of 10k χ^2 variables. The procedure for estimating the density of the empirical data distribution is accomplished through through the application of an optimized smoothing kernel to the histogram.

curves, with the LRT test coming out on top with the highest power, and the Fisher transform method with the lowest for all differences in partial correlation.

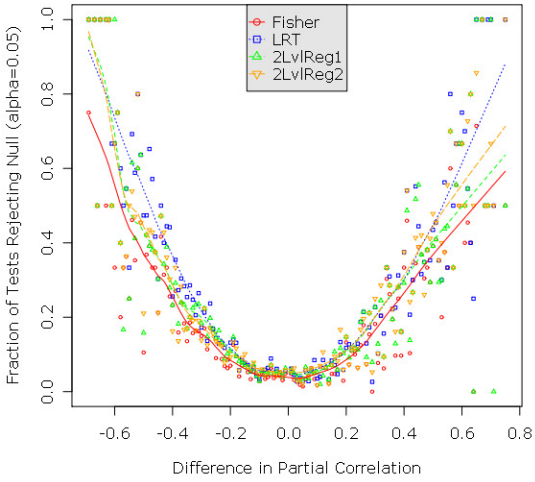


FIGURE 5.2: Simulation 1($n1 = n2 = 30$, $k = 6$): Simulated power curve based on 10k samples and a significance level $\alpha = 0.05$. The results are as expected, with a larger true differences between groups yielding higher rejection rates across all the methods.

The simulated type 1 error curve, see figure 5.3, shows similar results for all nearly all the methods, as expected. A constant error for all values of the partial correlation shows that there is no bias in the rejection of the test based on the true partial correlation value, which is the case for the two new methods proposed in this thesis. The Fisher method however, shows it is more likely to accept the null hypothesis when the true partial correlation takes on extreme values close to 1 and -1. This, coupled with the

lower power demonstrated in 5.2 makes for a more conservative test when using the Fisher method.

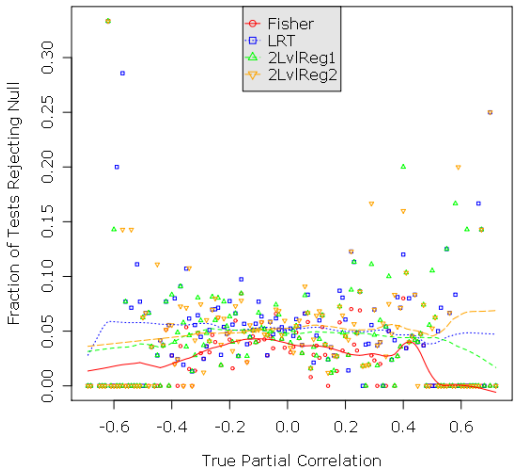


FIGURE 5.3: Simulation 1($n1 = n2 = 30$, $k = 6$): Simulated type 1 error curve based on 10k samples, $\alpha = 0.05$. The tests show nearly constant rejection levels for all values of partial correlation. Only the method based on Fisher's transform shows any variation based on range.

5.4 Simulation 2

In this simulation we use a smaller sample size $n_1 = n_2 = 20$, and fewer regions $k = 4$, in order to see how the methods perform on a dataset that does not meet all the large sample requirements suggested for the bootstrap and Fisher's transformation methods. Real world datasets, especially those from functional brain imaging studies, rarely contain as many samples as a researcher would like. The cost for a single PET scan can range in the thousands of dollars, which restricts the number of subjects that can be recruited into the study.

As before, it is necessary for the LRT to follow a χ^2 distribution with one degree of freedom for the test to have any accuracy. And again, the data shows this to be so, with a Kolmogorov-Smirnov distance of $D = 0.0145$, and corresponding p-value of 0.2439 showing the similarity between repeated samples of the LRT under the null conditions and random samples from a χ^2 distribution. The plots of the estimated density of the LRT under the null hypothesis is shown along with 10000 random points from the χ^2 distribution with one degree of freedom.

As before, the simulated type 1 and type 2 error curves are shown for this dataset in figures 5.5 and 5.6. The plots show results similar to the first simulation, despite the smaller sample size. It is important to note that under the null hypothesis, the method

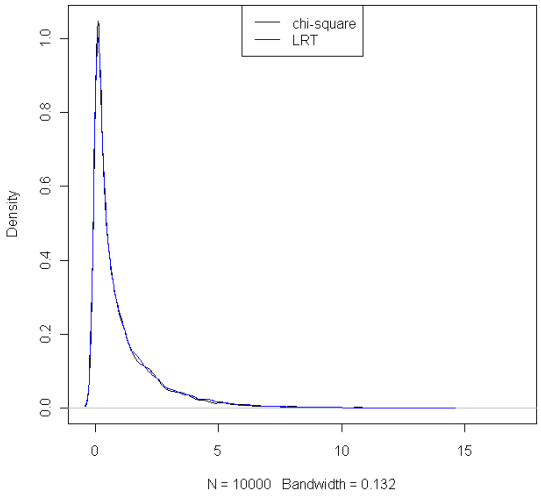


FIGURE 5.4: Simulation 2($n1 = n2 = 20$, $k = 4$): This figure shows the estimated density of the 10k independent observations LRT under the conditions stated in the null hypothesis, along with the density of 10k χ^2 variables. The procedure for estimating the density of the empirical data distribution is accomplished through through the application of an optimized smoothing kernel to the histogram.

based on the Fisher transformation rejects the null even less than under the conditions of simulation 1. With a smaller sample size and fewer controlling regions the test rejection rate is even more dependant to the actual partial correlation value.

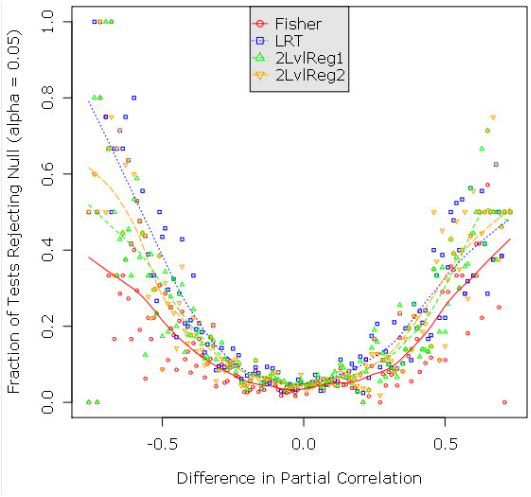


FIGURE 5.5: Simulation 2($n1 = n2 = 20$, $k = 4$): Simulated power curve based on 10k samples and a significance level $\alpha = 0.05$. The results are as expected, with larger true differences between groups yielding higher rejection rates across all the methods.

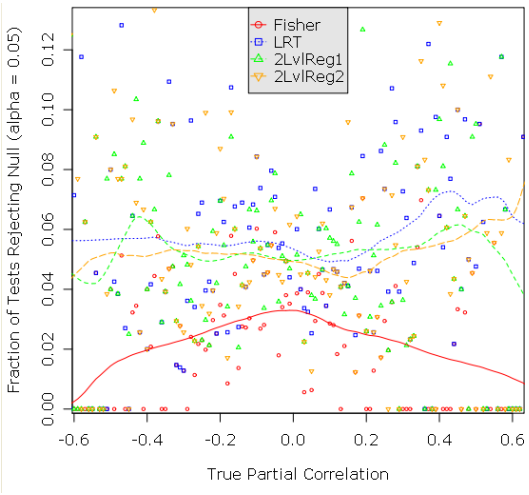


FIGURE 5.6: Simulation 2($n1 = n2 = 20$, $k = 4$): Simulated type 1 error curve based on 10k samples

5.5 Simulation 3

The purpose of this simulation is to see how a dataset of the same size as our experiment performs. Like our real world dataset, we have $n_1 = 16$ corresponding to the 16 normal controls, $n_2 = 25$ corresponding to the 25 cocaine subjects, and $k = 9$ corresponding to the measurements taken from the 9 brain regions. Reasonable curves at this sample size assures that we can apply these methods to our real world dataset.

The first step is to make sure the LRT follows the χ^2 distribution under the null hypothesis. A Kolmogorov-Smirnov test gives a D stat of 0.0132 and p-value of 0.3483. The plot can be seen in figure 5.7.

The type 2 error curves shown in figure 5.9 display patterns resembling those from the the previous simulations except that they are much lower. Increasing the number of controlling variables has a detrimental effect on the power of the all tests we compared in this study. The type 1 error curves in figure 5.8 show a possible problem with the LRT. Despite the fact that the rejection significance level was set at $\alpha = 0.05$ for all the tests in the simulation, the rejection rate of the LRT is much higher, falling around 0.08. This coincides with the higher power present in this simulation leading us to the conclusion that the LRT is more likely than the other methods to reject the null hypothesis regardless of whether or not it is true. Still, an observed type 1 error of 0.08 does not invalidate the method entirely.

The three simulations tell us much about the testing procedures we will apply to our real world dataset. For one, they show that method based on Fisher's transformation is very

conservative, a property we welcome when the null hypothesis is true, but it causes difficulty achieving significance when there really is a difference in the partial correlations. The LRT method has the highest power and has a type 1 error curve consistent with the proposed simulation significance level at lower sample sizes and controlling measures, but the addition of controlling measurements increases the likelihood of rejecting the null hypothesis despite the true state of the parameters. Of the three methods, it is the two-level regression procedure that gives the most well rounded results. In all three simulations the type 1 error curve falls in its expected places at 0.05 at all ranges of partial correlation and in all three simulations it has a higher power than the Fisher method.

When applied to entire networks, tests of this nature must account for multiple comparisons by increasing the threshold required for significance. Because of this need, it may be beneficial to use our new methods, especially the two-level regression, which achieved higher power curve than the Fisher z-test, while keeping the type 1 error rate consistent at all levels of true partial correlation.

Using the results of the simulations as a guide, we should expect the Fisher method to give us the least number of significant paths, and the LRT method to give the most. But, as we have learned through the simulations, we should be wary in our acceptance all of the paths based on the LRT. The results based on the two-level regression are the most likely to show the true nature of the partial correlations between groups.

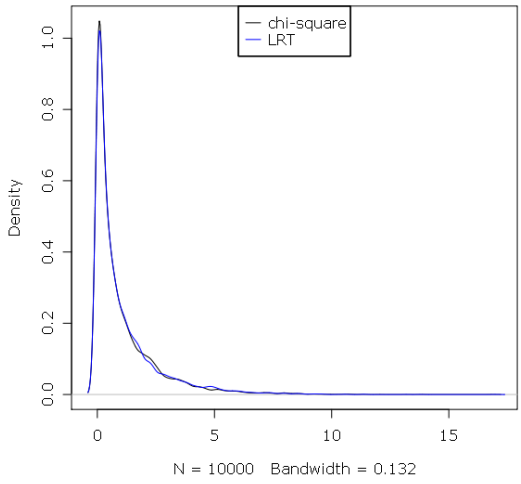


FIGURE 5.7: Simulation 3 ($n1 = 16$, $n2 = 25$, $k = 9$): This figure shows the estimated density of the 10k independent observations LRT under the conditions stated in the null hypothesis, along with the density of 10k χ^2 variables. The procedure for estimating the density of the empirical data distribution is accomplished through through the application of an optimized smoothing kernel to the histogram.

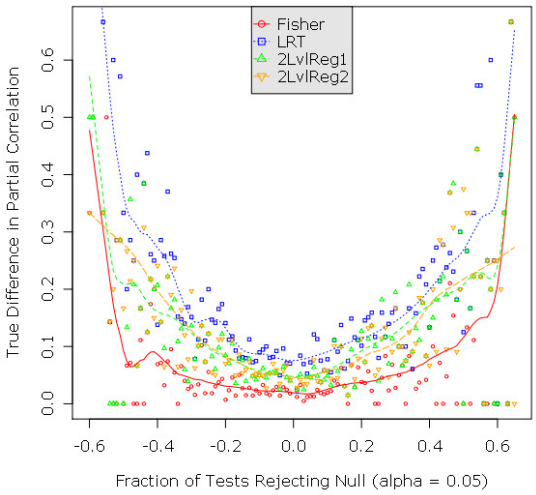


FIGURE 5.8: Simulation 3 ($n1 = 16$, $n2 = 25$, $k = 9$): Simulated power curve based on 10k samples and a significance level $\alpha = 0.05$. The results are as expected, with larger true differences between groups yielding higher rejection rates across all the methods.

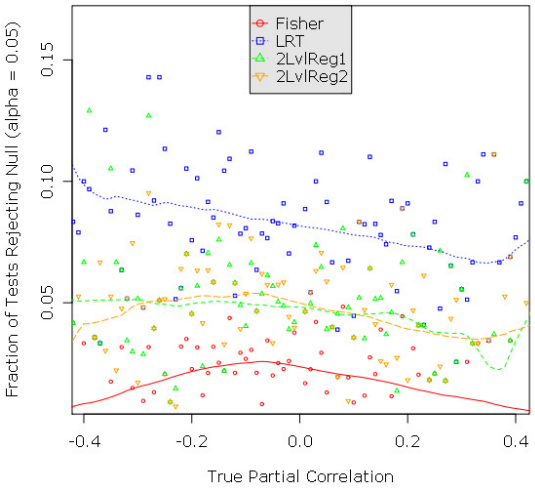


FIGURE 5.9: Simulation 3 ($n1 = 16$, $n2 = 25$, $k = 9$): Simulated type 1 error curve based on 10k samples

5.6 Experimental Study

Our data consists of PET scans taken from 16 normal control subjects and 25 cocaine addicted subjects. We will attempt to find differences in the neurological pathways between the normal control group and the cocaine addicted group under the condition that the subjects have been administered a dose of methylphenidate, a drug similar to cocaine.

This dataset has been analyzed before with traditional methods (Volkow et al., 1997), (Volkow et al., 2006), (Volkow et al., 2005), (Volkow et al., 2008) but an analysis based

on partial correlations has not been approached yet.

5.6.1 Data Acquisition and Preprocessing

Nuclear imaging is the process of extracting a 3D representation of an object based on the amount of radiation it emits. There are a number methods to accomplish this, including computed tomography (CT) which is essentially a 3D x-ray, magnetic resonance imaging (MRI) which measures the speed at which atoms return to their resting state after an initial alignment due to magnetism, and positron emission tomography (PET) which measures the amount of radiation given off by an object. In our experiment, we are working with images taken from a PET machine.

When we take PET scans using FDG, we are effectively measuring brain activity. If a human subject was put in a PET scanner as is, we would get a meaningless picture. Our bodies give off small quantities of radiation all the time but of a type that is unmeasurable, or insignificant to the topic we are studying. For this reason, the subjects of our study have been given a dose of FDG (Phelps et al., 1979), a radioactive chemical that binds to glucose. The idea is that when certain regions of the brain are more active, more blood will be flowing to them, blood containing measurable amounts of FDG bound glucose. The PET image gives us a 3D measurement of how active the various parts of the brain were during the time of the scan.

A direct comparison of PET intensity based on a fixed coordinate position would yield meaningless results. Not everyone has the same sized brain so the measurement at a fixed coordinate could yield the intensity of the amygdala in one patient and the intensity of the eye socket in another. Moreover, other than bolting the head down, there is no physical way to make sure the subjects' heads are aligned in exactly the same way. Therefore, a number of procedures must be performed on the 3D scans to make them directly comparable to each other.

The first procedure is known as spatial normalization which corrects the problem of differently shaped brains. To start off, a template is generated either from an amalgamation of scans or by selecting a single scan at random, to represent a coordinate system to which all the other scans will be transformed to. Then, for each image, the affine transformation is found to minimizes the difference between the transformed image and the template, see figure 5.10. The application of this affine transformation warps the image under consideration into the space of the template, thereby making the intensity at any given point comparable across the dataset.

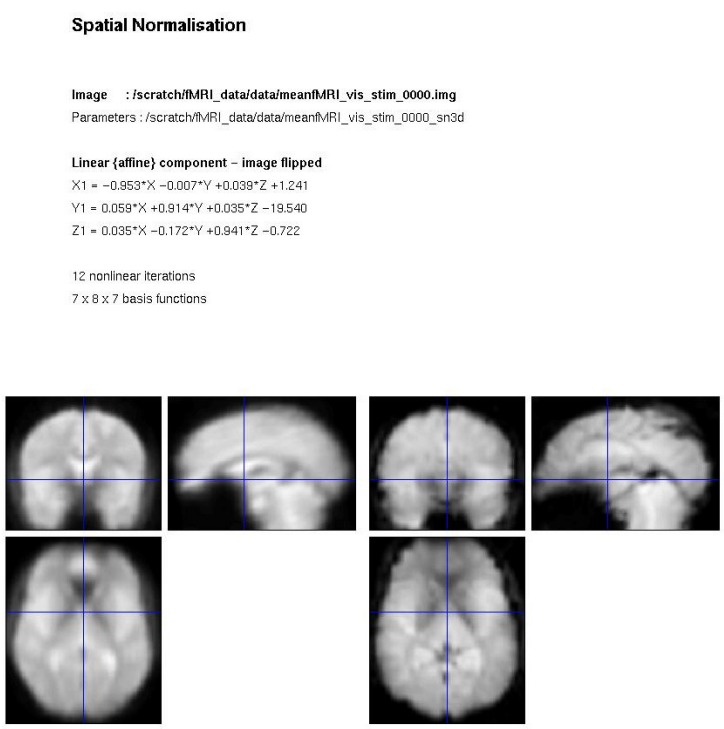


FIGURE 5.10: The normalization procedure as implemented through SPM software (Ashburner and Friston, 1999). A 3×4 affine transformation matrix is found that minimizes a metric describing the distance of the warped brain scan to the chosen brain template.

The second procedure partitions the 3D image into a set of well documented regions and extracts those that are of interest to the current study. To each scan, we apply a mask that assigns each location to an anatomical region. This automatic method has been utilized in various studies, and has been shown to produce accurate and reproducible results compared to the labor intensive manual approach (Scheinin et al., 2008). We use the original mask based on Talairach's work (Lancaster et al., 2000), but the procedure could be done using any atlas, even a probabilistic map as studied in (Lee et al., 2004).

The third procedure is to normalize the dataset across subjects by intensity. The process of acquiring images is a very lengthy one, and can take months to complete an entire study. Throughout the study, it is impossible to control all parameters so that they are exactly the same on each subject. On large studies, like the one we are analyzing, data comes from multiple scanners, is taken on different days, and by different machine operators. What's more, people's brain differ by volume. A camera's reading of higher intensity could be caused by the actual differences between groups, or by any number of these unavoidable parameters. We try to control for this effect by utilizing an intensity normalizing procedure. Dividing each scan by its total intensity equalizes these discrepancies somewhat.

FIGURE 5.11: The autoROI software package is used to extract regions of interest automatically from our dataset. The figure shows a 2D slice in the transverse plane of a spatially normalized PET image overlayed with various anatomical regions. Our experiment consists of the nine regions associated with the brain's reward mechanism, but new anatomical sets can be easily substituted into the analysis work flow.

The last procedure involves finding single values to best represent the regions found in the previous step. Each region is a collection of positions in the brain space, see figure 5.11. The easiest to to represent the intensity of a region is to use the mean or median of the region. This method is easily implemented, but we have found that taking a small neighborhood around the maximum of the region gives results that better match those of a trained doctor performing the region selection manually (Ma et al., 2004). Despite the normalization step, there are always slight differences in the brain configuration between patients. In the manual procedure the doctor selects a small region based on visual inspection and hand draws a selection region. When we base the region selection around the maximum value instead of the whole region, we hope to reduce errors due to slight differences in spatial distortion between subjects.

What we end up with is a set of numbers representing the key areas of the brain involved in the reward pathway for each patient. These areas have been found by our collaborator to be directly responsible for the reward mechanism in human subjects (Volkow et al., 2002), and consist of the following regions: Ventral Striatum (VS), Thalamus (THAL), Insula (INS), Putamen (PUT), Motor Frontal Cortex (MFC), Cerebellum (CER), Amygdala (AMYG), Orbito Frontal Cortex (OFC), Cingulate Gyrus (CG).

5.6.2 Normality Check

Some of the procedures we use operate under the assumption that the data follows a multivariate normal distribution. We checked this assumption through the use of a parametric bootstrap energy test (Szkely and Rizzo, 2005). The resulting p-value of 0.3353 shows clear evidence that the assumption is valid. Additionally, scatterplots of all region pairings, see figure 5.12, show visually that this assumption is reasonable for the bivariate case.

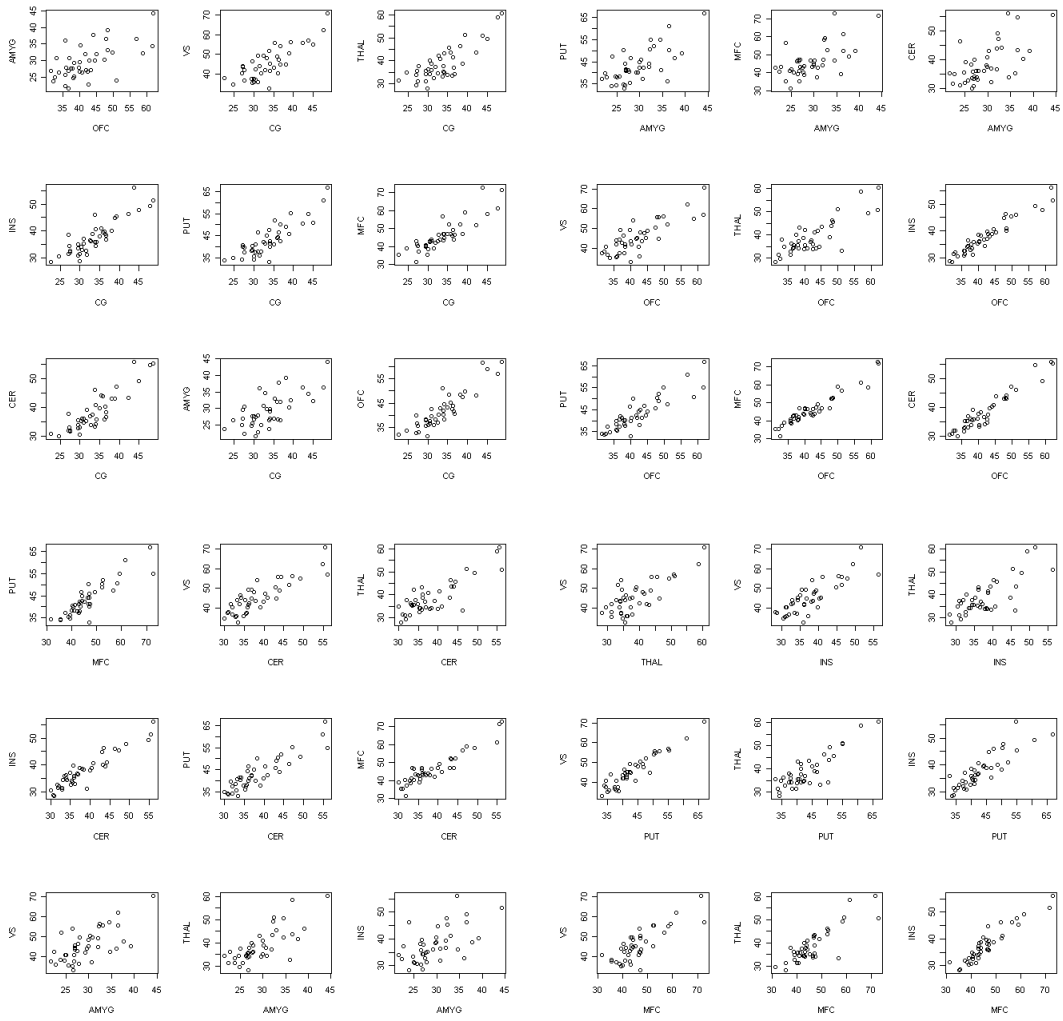


FIGURE 5.12: Scatterplots of all region pairings show that it is reasonable to assume that some form of linear relationship exists.

5.6.3 Procedure

After all the preprocessing steps, the data ends up in matrix form, see figure 5.1, to which we apply the four methods of the previous chapters: bootstrapping, Fisher's transformation, two-level regression, and LRT.

(a) Normal Control

.	Region 1	Region 2	Region ...	Region k
nml 1	x_{11}	x_{12}	...	x_{1k}
nml 2	x_{21}	x_{22}	...	x_{2k}
...	...	...	...	...
nml n_1	$x_{n_1 1}$	$x_{n_1 2}$	...	$x_{n_1 k}$

(b) Cocaine addicted

.	Region 1	Region 2	Region ...	Region k
coc 1	y_{11}	y_{12}	...	y_{1k}
coc 2	y_{21}	y_{22}	...	y_{2k}
...	...	...	...	...
coc n_2	$y_{n_2 1}$	$y_{n_2 2}$	...	$y_{n_2 k}$

TABLE 5.1: The data matrices obtained after preprocessing the PET image scans

Because we are applying this test to all region pairs, obtaining an accurate p-value requires us to deal with the problem of multiple comparisons. Let's say the probability of a single test to incorrectly reject the null hypothesis by chance is $\alpha = 0.05$. If there are many tests taking place, the probability of at least one of the tests rejecting the null by chance is much greater than α. The simplest, but most restrictive, solution is to use Bonferroni correction and divide the individual test significance levels by the number of tests. It was Boole who first noted that the probability of any set of events is less than or equal to the sum of the probability of each individual event, i.e. $P(A \cup B \cup C) \leq P(A)+P(B)+P(C)$. Bonferroni later applied this fact to the multiple testing problem by dividing each individual significance by the number of tests being carried out, there bye setting an upper limit on the probability of any one of the tests incorrectly rejecting the null by chance. Since we have 9 regions in the current study, there are $\binom{9}{2} = 36$ possible pairings, so we would have to adjust each individual significance level to $\alpha/36 = 0.0014$ to have a total groupwise significance level of 0.05. Bonferronis method sets an upper bound, on the groupwise significance level, but it is very conservative, so much so that on the current dataset, none of the four methods under consideration find any significant differences in partial correlation between groups. The use of multiple test correction in partial correlation network analysis (PCNA) is in general, considered unnecessary as PCNA is considered an exploratory analysis vehicle. The resulting network must be verified through confirmatory analysis, preferably based on an independent data set, through structural equation modeling SEM analysis (Peng et al., 2009).

However, correcting for multiple comparisons plays a less important role in the comparison of the various tests to each other which is what we want to do. Since there are the same number of tests being applied to the network for each of the methods, a direct comparison of p-values can be done to see how the methods perform against each other.

5.6.4 Results

To begin, lets look at the partial correlation matrices of the two groups separately, see figure 5.2. In this format it is hard to determine which of the region pairs have different partial correlations between the groups. It is easier to make sense of the data by visualizing it, see figure 5.13. Here, we represent each region by a node, and the partial correlations as lines connecting the nodes, with weaker partial correlations showing up with a greater transparency. In the diagram, the blue lines represent the partial correlations of the normal control group, and the red lines are from the cocaine addicted subjects. From the picture alone, it seems that the cocaine addicted group has higher partial correlation in the OFC, INS, PUT, VS and CER regions, while the normal controls exhibit stronger partial correlation in the OFC, MFC, INS, PUT and VS regions. Of course, this is just a quick run through by eye.

(a) Normal Control

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	1.000	0.379	0.155	0.918	0.484	0.744	-0.164	-0.783	-0.360
THAL	0.379	1.000	-0.146	-0.259	-0.378	-0.156	0.147	0.459	0.340
INS	0.155	-0.146	1.000	-0.213	0.367	-0.411	0.269	0.439	0.090
PUT	0.918	-0.259	-0.213	1.000	-0.345	-0.709	0.220	0.756	0.388
MFC	0.484	-0.378	0.367	-0.345	1.000	-0.274	-0.225	0.510	0.386
CER	0.744	-0.156	-0.411	-0.709	-0.274	1.000	0.120	0.810	0.377
AMYG	-0.164	0.147	0.269	0.220	-0.225	0.120	1.000	-0.173	0.361
OFC	-0.783	0.459	0.439	0.756	0.510	0.810	-0.173	1.000	-0.278
CG	-0.360	0.340	0.090	0.388	0.386	0.377	0.361	-0.278	1.000

(b) Cocaine Addicted

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	1.000	-0.446	0.204	0.748	0.443	0.448	-0.068	-0.461	-0.088
THAL	-0.446	1.000	-0.126	0.505	0.368	0.248	0.400	-0.152	0.151
INS	0.204	-0.126	1.000	0.079	0.097	0.258	0.313	0.285	0.373
PUT	0.748	0.505	0.079	1.000	-0.295	-0.315	-0.072	0.504	0.037
MFC	0.443	0.368	0.097	-0.295	1.000	-0.323	-0.308	0.506	0.206
CER	0.448	0.248	0.258	-0.315	-0.323	1.000	-0.025	0.498	0.005
AMYG	-0.068	0.400	0.313	-0.072	-0.308	-0.025	1.000	0.109	0.103
OFC	-0.461	-0.152	0.285	0.504	0.506	0.498	0.109	1.000	-0.173
CG	-0.088	0.151	0.373	0.037	0.206	0.005	0.103	-0.173	1.000

TABLE 5.2: Partial correlation matrices of the two groups

The statistical analysis takes place by applying each of the methods from the previous chapters to our dataset region by region. Each region pair was analyzed with the four methods and figures 5.3, 5.4, 5.5 and 5.6 show their resulting p-values. As with the partial correlation matrices, it is easier to make sense of the data through visualization, see figure 5.14. Here, p-values are represented by lines connecting the regions to which

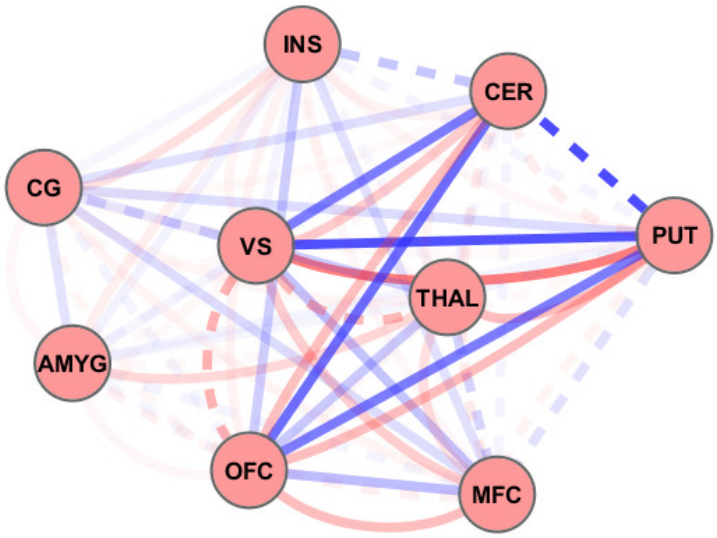


FIGURE 5.13: A visual display of the partial correlation matrices. The nine brain regions are shown as nodes, and partial correlation between regions as edges connecting the nodes. The partial correlation strength is signified by the opacity of the line connecting regions, with higher absolute values appearing more opaque. Positive partial correlations are shown in a solid line, and negative partial correlations in a dotted line. Partial correlations of the normal control subjects are shown in blue and the cocaine addicted subjects in red.

the tests were applied, each test in a different color. Although all test pairings are displayed, with more significant p-values shown with more opaque lines, for the purpose of reducing clutter, less opacity was given to the p-values below 0.15. In the next section, we report all region pairings that show an uncorrected significance level below 0.10.

Looking at the results of all four methods, we come to the conclusion that the Fisher method, which is by far the most popular way to compare two partial correlation coefficients, had the lowest rate of finding significance in group effect. As in the simulation studies, the Fisher method found the fewest significant pathways. The other widely used technique of comparing partial correlations, the bootstrap, gives use a result matrix very similar to the that of the two-level regression approach. Except for one region pair (insula, cerebellum), bootstrap came to the same conclusions as two-level regression at significance level $\alpha = 0.10$.

Moving to the next method, the LRT picks up all the region pairings that bootstrap found and more. In particular, it seems to have an easier time distinguishing partial correlations that are both very high, or both very low, as in the case of region pairs (ventral striatum, cerebellum) and (putamen, cerebellum). Recall simulations 5.3, 5.6, 5.9 to note the difficulty Fisher's method had in this task. Unfortunately, the LRT failed to converge properly on some of the tests, including the (insula, cerebellum) pair that two-level regression determined to have significantly different partial correlations.

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	.	0.046	0.912	0.705	0.926	0.516	0.824	0.467	0.549
THAL	.	.	0.968	0.075	0.083	0.356	0.570	0.153	0.671
INS	.	.	.	0.503	0.556	0.120	0.921	0.724	0.535
PUT	.	.	.	.	0.909	0.376	0.506	0.570	0.440
MFC	.	.	.	.	.	0.914	0.861	0.992	0.685
CER	.	.	.	.	.	.	0.758	0.486	0.401
AMYG	.	.	.	.	.	.	.	0.520	0.573
OFC	.	.	.	.	.	.	.	.	0.812
CG	.	.	.	.	.	.	.	.	.

TABLE 5.3: Two sided p-values comparing partial correlations between groups based on the bootstrap method with 5k resamples.

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	.	0.087	0.919	0.724	0.933	0.540	0.842	0.505	0.573
THAL	.	.	0.967	0.114	0.123	0.402	0.600	0.205	0.696
INS	.	.	.	0.545	0.577	0.166	0.928	0.750	0.559
PUT	.	.	.	.	0.917	0.415	0.545	0.602	0.468
MFC	.	.	.	.	.	0.918	0.863	0.994	0.709
CER	.	.	.	.	.	.	0.763	0.518	0.442
AMYG	.	.	.	.	.	.	.	0.560	0.594
OFC	.	.	.	.	.	.	.	.	0.828
CG	.	.	.	.	.	.	.	.	.

TABLE 5.4: Two sided p-values comparing partial correlations between groups using the traditional Fisher approximation.

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	.	0.020	0.687	0.530	0.809	0.431	0.634	0.402	0.366
THAL	.	.	0.952	0.045	0.067	0.191	0.419	0.139	0.655
INS	.	.	.	0.229	0.655	0.063	0.587	0.522	0.430
PUT	.	.	.	.	0.967	0.207	0.398	0.364	0.302
MFC	.	.	.	.	.	0.643	0.966	0.851	0.490
CER	.	.	.	.	.	.	0.475	0.491	0.487
AMYG	.	.	.	.	.	.	.	0.595	0.477
OFC	.	.	.	.	.	.	.	.	0.778
CG	.	.	.	.	.	.	.	.	.

TABLE 5.5: Two sided p-values comparing partial correlations between groups using the two-level regression approach.

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	.	0.033	0.846	1.000	0.920	0.030	0.711	0.172	0.703
THAL	.	.	0.713	0.062	0.056	0.202	0.772	0.091	0.357
INS	.	.	.	1.000	0.326	1.000	0.226	1.000	0.294
PUT	.	.	.	.	0.870	0.055	0.174	0.264	0.461
MFC	.	.	.	.	.	0.325	0.608	0.974	0.574
CER	.	.	.	.	.	.	0.189	1.000	0.575
AMYG	.	.	.	.	.	.	.	0.169	0.854
OFC	.	.	.	.	.	.	.	.	0.906
CG	.	.	.	.	.	.	.	.	.

TABLE 5.6: Two sided p-values comparing partial correlations between groups using the likelihood ratio test.

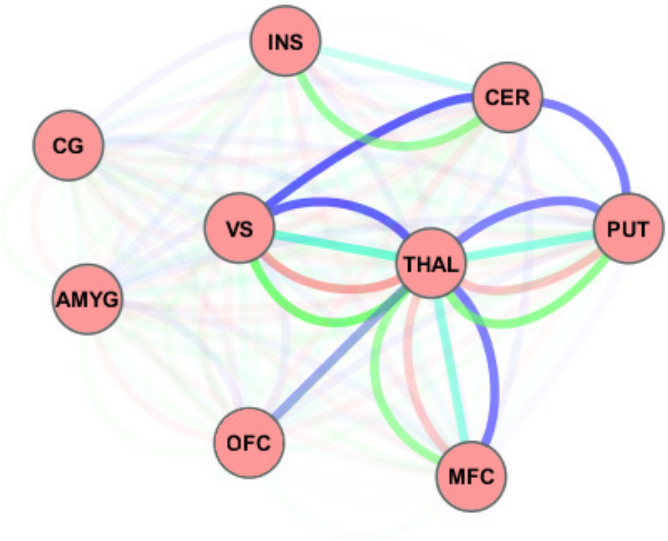
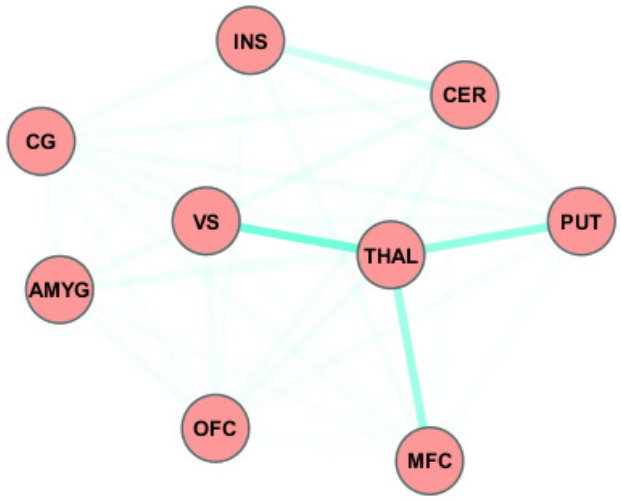


FIGURE 5.14: The result matrices of the four methods are displayed visually. The brain regions are shown as the nodes in the graph and group's effects on the partial correlation are shown as the edges. The results each of the four methods are displayed in a different color with Fisher's in red, bootstrapping in cyan, two-level regression in green and LRT in blue. The significance of the tests are shown through edge opacity, with lower p-values appearing more opaque than a higher p-values.

5.6.4.1 Bootstrapping

This method gives us a network consisting of four main regions: VS, Thal, Put, and MFC. Figure 5.15 shows us the VS, Put, and MFC are all connected to the Thal hub node. For this experiment, we used a bootstrap sample of 5000 to determine the simulated p-values.



(a) graphical display

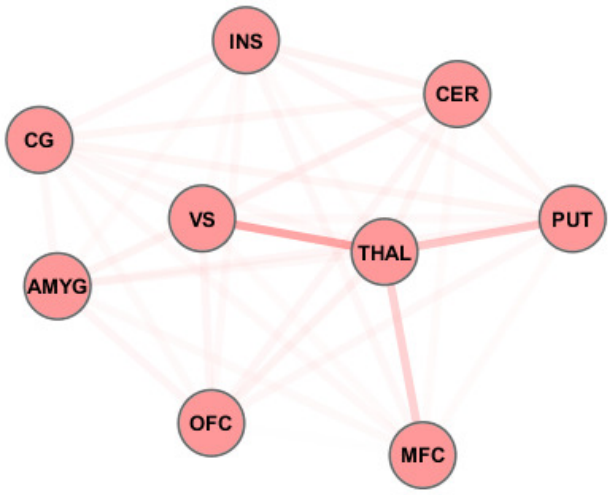
Region 1	Region 2	Sample PC (nml, coc)	p-value
VS	Thal	(0.379, -0.446)	0.046
Thal	Put	(-0.259, 0.505)	0.075
Thal	MFC	(-0.378, 0.368)	0.083

(b) tabular display

FIGURE 5.15: Bootstrap method: region pairs with an uncorrected p-value < 0.10

5.6.4.2 Fisher's Approximation

As expected from our simulation results, the traditional method based on Fisher's transformation is the most conservative of all methods. There is only one region pair that meets the uncorrected significance level of 0.10. Although they are not significant, the weak pathways found with this method follow a similar pattern to those of the bootstrap and two-level regression.



(a) graphical display

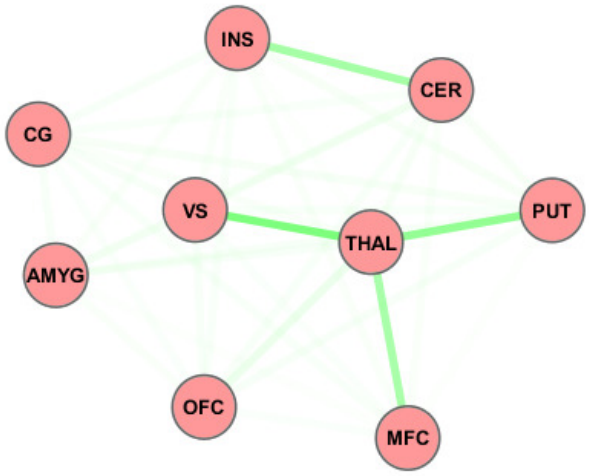
Region 1	Region 2	Sample PC (nml, coc)	p-value
VS	Thal	(0.379, -0.446)	0.087

(b) tabular display

FIGURE 5.16: Fisher method: region pairs with an uncorrected p-value < 0.10

5.6.4.3 Two Level Regression

Two-level regression yields a network similar to the bootstrapping approach. However, this method deems the (insula, cerebellum) region pair to have a significantly different partial correlation between groups. The sample partial correlations of this region pair are -0.411 for the normal controls and 0.258 for the cocaine addicted subjects.



(a) graphical display

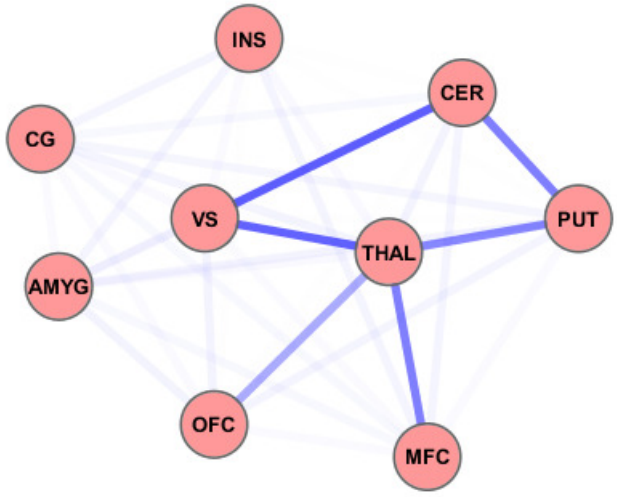
Region 1	Region 2	Sample PC (nml, coc)	p-value
VS	Thal	(0.379, -0.446)	0.020
Thal	Put	(-0.259, 0.505)	0.045
Thal	MFC	(-0.378, 0.368)	0.067
Ins	Cer	(-0.411, 0.258)	0.063

(b) tabular display

FIGURE 5.17: Two Level Regression method: region pairs with an uncorrected p-value < 0.10

5.6.4.4 LRT

The likelihood ratio test gives us a network with six main regions, and contains pairings that none of the other methods pick up on, most notably between the ventral striatum and cerebellum which had a sample partial correlation of 0.744 in the normal controls, but only 0.448 in the cocaine group. Unfortunately, it also failed to pick up on the (insula, cerebellum) pair that the two-level regression found. Because this is a numerical method, sometimes it is not possible to find a valid estimates that maximize the likelihood function, and thus it fails to obtain any result at all. This was the case with the (insula, cerebellum) pair as well as a few other pairings.



(a) graphical display

Region 1	Region 2	Sample PC (nml, coc)	p-value
VS	Thal	(0.379, -0.446)	0.033
VS	Cer	(0.744, 0.448)	0.030
Thal	Put	(-0.259, 0.505)	0.062
Thal	MFC	(-0.378, 0.368)	0.056
Thal	OFC	(0.459, -0.152)	0.092
Put	Cer	(-0.709, -0.315)	0.055

(b) tabular display

FIGURE 5.18: LRT method: region pairs with an uncorrected p-value < 0.10

Chapter 6

Dealing with Continuous Covariates

6.1 Overview

Real world data often contains information that does not directly interest us, but may affect our inference into the variables we do wish to examine. For example, in the clinical study from from Chapter 5 we are interested in the ways substance abuse changes the neurological pathways involved in the brain's reward mechanism, but these changes may be affected by factors beyond addiction status, like age and education level. In this chapter, we determine how additional covariate information can influence the pathways under consideration.

6.2 Extended Two-Level Regression

The two-level regression method introduced in Chapter 3 can be readily extended to measure how additional factors affect the partial correlation between regions. This is accomplished by inserting the covariate information into the residual equations before performing the regression procedure.

Recall equation 3.3 that describes the residuals of two regions of interest controlled for the k remaining regions. In our current problem, we have information from two covariates, C_1 and C_2, which are not required to be simple binary factors. We want to model the effect of covariate C_2 on the partial correlation while controlling for the effect of C_1.

By reformulating the b_1 coefficient to account for the effects of both covariates, i.e.,

$$b_1 = a_0 + a_1 C_1 + a_2 C_2 \tag{6.1}$$

we achieve this goal. Resubstituting b_1 back into the equation relating residuals, we end up with the following model:

$$\begin{aligned} R_1 &= b_0 + b_1 R_2 + \epsilon \\ R_1 &= b_0 + (a_0 + a_1 C_1 + a_2 C_2) R_2 + \epsilon \\ R_1 &= b_0 + a_0 R_2 + a_1 C_1 R_2 + a_2 C_2 R_2 + \epsilon. \end{aligned} \tag{6.2}$$

Incorporating the new covariates simultaneously gives us a model relating the partial correlations of regions y_1 and y_2 to covariate C_2 while controlling for the covariate C_1, and similarly, it relates covariate C_1 to partial correlation while controlling for C_2. In a multiple regression model the coefficient of an independent variable is proportional to the partial correlation of that variable and the dependent variable controlling for the other independent variables of the model. For example, in the model $y = b_0 + b_1 x_1 + b_2 x_2 + \epsilon$, the coefficient b_1 has the same significance level as $\mathrm{PCorr}(y, x_1 | x_2)$. In principle, using model 6.2 gives us the significance of the factor of interest, C_1, controlled for the effects of the extraneous covariate C_2.

Continuing as usual with the regression procedure described in Chapter 3, we find the significance of covariate C_1 coefficient and thus the effect C_1 has on partial correlation while controlling for covariate C_2.

6.3 Application to Experimental Data

Our experimental data from Chapter 5 includes the additional information of subject age, and education level, see figure 6.1. In addition to studying how cocaine addiction affects the neurological pathways under the proposed experimental conditions, it may also be interesting to examine how age or education influence these pathways. Younger subjects might display a different brain network when dealing with reward and expectation than older subjects, a discrepancy that unless accounted for, could distort our findings. The latent factors involved in education, such as discipline and intelligence could play a role similar to age, and end up skewing the pattern addiction lays in the brain's network.

We can examine these relationships through our extended two-level regression model by including the information of these covariates in our regression model.

.	Region 1	Region 2	Region ...	Region k	Group	Age	Education
nml 1	x_{11}	x_{12}	...	x_{1k}	0	A_{x1}	E_{x1}
nml 2	x_{21}	x_{22}	...	x_{2k}	0	A_{x2}	E_{x2}
...	...	...	...	...	0	A_{x3}	E_{x3}
nml n_1	$x_{n_1 1}$	$x_{n_1 2}$	...	$x_{n_1 k}$	0	A_{xn_1}	E_{xn_1}
coc 1	y_{11}	y_{12}	...	y_{1k}	1	A_{y1}	E_{y1}
coc 2	y_{21}	y_{22}	...	y_{2k}	1	A_{y2}	E_{y2}
...	...	...	...	...	1	A_{y3}	E_{y3}
coc n_2	$y_{n_2 1}$	$y_{n_2 2}$	...	$y_{n_2 k}$	1	A_{yn_2}	E_{yn_2}

FIGURE 6.1: The data matrix of the preprocessed PET image scans along with subject covariate information. Group status G is a binary variable with $G = 0$ for the normal controls, and $G = 1$ for the cocaine addicted subjects. Age A is measured in years, and education E is measured in the number of years spent in school/college.

6.3.1 Age and Addiction Status

First, let us look at the model with both age and addiction status included. The notation follows the same style as in Chapter 5, with G depicting the group status of each subject with 0 being normal control and 1 being cocaine addicted. A is a continuous variable depicting the subject's age in years.

$$\begin{aligned} R_1 &= b_0 + b_1 R_2 + \epsilon \\ R_1 &= b_0 + (a_0 + a_1 A + a_2 G) R_2 + \epsilon \\ R_1 &= b_0 + a_0 R_2 + a_1 A R_2 + a_2 G R_2 + \epsilon. \end{aligned} \tag{6.3}$$

The coefficients of equation 6.3 tell us how each factor affects the partial correlation between regions of interest. The a_1 coefficient tells us how strongly age influences the relationship between residuals while controlling for group, as the a_2 coefficient tells us the how group affects the relationship controlling for age. The a_0 coefficient shows the strength of the total partial correlation using observations from both groups while controlling for both factors, age and group.

It should be noted that the mean ages of the subjects are significantly different between groups, with the cocaine addicted subjects being slightly older. A Student's t-test on the means gave a p-value of 0.03281, see table 6.1.

Group	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
nml	27.60	30.28	34.00	36.23	41.85	49.20
cocaine	32.40	38.50	40.50	40.49	43.30	46.00

TABLE 6.1: A summary of ages between groups. There seems to be a significant difference in the means of the two groups.

It seems the discrepancy in ages between the two groups does make a difference in the significance levels of the partial correlation network. The p-values of the a coefficients of our two-level model are shown in table 6.2 and they are displayed visually side by side with the two-level results from Chapter 5 in figure 6.2. The visual display shows us that age seems to affect the partial correlation network pattern, particularly in the (amygdala, Motor Frontal Cortex) pair with a p-value of 0.041, and the Motor Front Cortex, Cerebellum) pair with a p-value of 0.069. It also demonstrates that the group effect on the network pattern is degraded when information of the subject's ages are included in the model. Notice how the (ventral striatum, thalamus) and the (insula, cerebellum) links are no longer significant in the model containing both the group indicator and the age covariate.

(a) Two sided p-values of the a_0 coefficient

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	0.000	0.291	0.115	0.086	0.717	0.211	0.650	0.084	0.151
THAL	0.000	0.000	0.919	0.546	0.315	0.692	0.363	0.930	0.461
INS	0.000	0.000	0.000	0.077	0.354	0.116	0.751	0.440	0.306
PUT	0.000	0.000	0.000	0.000	0.731	0.289	0.096	0.174	0.063
MFC	0.000	0.000	0.000	0.000	0.000	0.112	0.028	0.563	0.766
CER	0.000	0.000	0.000	0.000	0.000	0.000	0.347	0.293	0.138
AMYG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.482	0.623
OFC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.065
CG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

(b) Two sided p-values of the a_1 coefficient

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	0.000	0.194	0.144	0.691	0.512	0.637	0.764	0.335	0.228
THAL	0.000	0.000	0.873	0.624	0.213	0.649	0.409	0.822	0.308
INS	0.000	0.000	0.000	0.110	0.548	0.212	0.820	0.780	0.305
PUT	0.000	0.000	0.000	0.000	0.595	0.680	0.152	0.492	0.112
MFC	0.000	0.000	0.000	0.000	0.000	0.069	0.041	0.334	0.891
CER	0.000	0.000	0.000	0.000	0.000	0.000	0.258	0.482	0.212
AMYG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.367	0.423
OFC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.084
CG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

(c) Two sided p-values of the a_2 coefficient

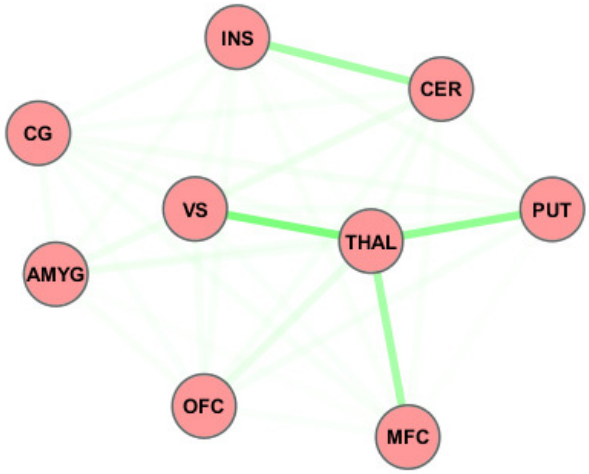
.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	0.000	0.006	0.850	0.503	0.812	0.567	0.697	0.758	0.899
THAL	0.000	0.000	0.936	0.085	0.034	0.167	0.334	0.153	0.554
INS	0.000	0.000	0.000	0.518	0.720	0.202	0.637	0.573	0.265
PUT	0.000	0.000	0.000	0.000	0.889	0.325	0.675	0.650	0.910
MFC	0.000	0.000	0.000	0.000	0.000	0.909	0.522	0.818	0.499
CER	0.000	0.000	0.000	0.000	0.000	0.000	0.309	0.436	0.803
AMYG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.423	0.351
OFC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.566
CG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 6.2: Following the same procedure as Chapter 5, we applied the two-level regression model $R_1 = b_0 + a_0 R_2 + a_1 A R_2 + a_2 G R_2 + \epsilon$ to all region pairs. Because we are trying to find the significance that addiction plays in the partial correlation network of the brain's reward mechanism, it is the a_2 matrix that is of most interest to us which shows the significance of group controlling for age.

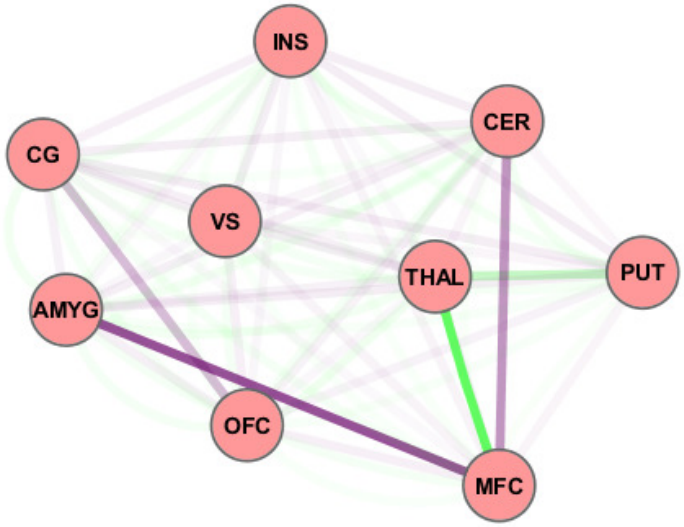
6.3.2 Education and Addiction Status

Let us look at the model including both education level and addiction status. As before, group status is indicated by G, and education level, which is measured by the number of years the subject has spent in school, by E.

$$\begin{aligned} R_1 &= b_0 + b_1 R_2 + \epsilon \\ R_1 &= b_0 + (a_0 + a_1 E + a_2 G + \ldots) R_2 + \epsilon \\ R_1 &= b_0 + a_0 R_2 + a_1 E R_2 + a_2 G R_2 + \ldots + \epsilon. \end{aligned} \tag{6.4}$$



(a) two-level regression, only group covariate. $R_1 = b_0 + a_0 R_2 + a_1 G R_2 + \epsilon$



(b) two-level regression, group and age covariates. $R_1 = b_0 + a_0 R_2 + a_1 A R_2 + a_2 G R_2 + \epsilon$

FIGURE 6.2: This visual display shows two interpretations of group's affect on the partial correlation network. The top figure shows the significance levels of the group coefficient in the model consisting of group status alone, $R_1 = b_0 + a_0 R_2 + a_1 G R_2 + \epsilon$. The bottom figure shows the significance levels of the age and group coefficients in the model containing both, $R_1 = b_0 + a_0 R_2 + a_1 A R_2 + a_2 G R_2 + \epsilon$. In both figures, group significance is shown in green, age significance in dark purple, with the more significant pathways shown with more opacity.

This time, the covariate factor is not significantly different in the two groups. A Student's t-test on the means gave a p-value of 0.1034, see figure 6.3.

Using model 6.4, we applied the two-level regression method to each regional pair. The matrices of coefficient significance levels are shown in tables 6.4 and the significance of

Group	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
nml	7.00	12.75	14.50	13.88	16.00	18.00
cocaine	8.00	11.00	12.00	12.52	14.00	16.00

TABLE 6.3: A summary of education level between groups. Education levels are valued as the number of years spent in school.

the two coefficients relating education and addiction status to partial correlation are displayed visually in 6.3. As before, the current model is compared to the model of Chapter 5 to see how education plays a role in partial correlation network. This time, the a_2 coefficient gives us a picture relating group effect to partial correlation controlling for the influences of education level.

The network describing group's affect on partial correlation controlling for education is very similar to the network that excludes education level information, see figure 6.3. The four significant paths obtained through the original model are essentially the same as the new model including education levels. Although education level does seem to have some effect on the partial correlation between the Insula and Thalamus, this relationship does not effect the significant pathways caused by group status. Remember that there was no significant difference in education levels between the two groups, so similarities between the two models are to be expected.

(a) Two sided p-values of the a_0 coefficient

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	0.000	0.914	0.243	0.239	0.606	0.161	0.312	0.205	0.928
THAL	0.000	0.000	0.056	0.785	0.242	0.756	0.431	0.506	0.767
INS	0.000	0.000	0.000	0.183	0.800	0.725	0.724	0.839	0.343
PUT	0.000	0.000	0.000	0.000	0.538	0.314	0.734	0.117	0.820
MFC	0.000	0.000	0.000	0.000	0.000	0.704	0.276	0.773	0.111
CER	0.000	0.000	0.000	0.000	0.000	0.000	0.504	0.482	0.668
AMYG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.204	0.225
OFC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.552
CG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

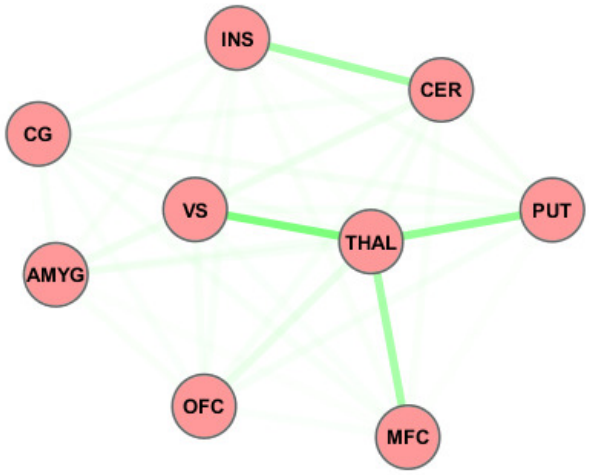
(b) Two sided p-values of the a_1 coefficient

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	0.000	0.794	0.309	0.089	0.620	0.583	0.217	0.521	0.718
THAL	0.000	0.000	0.023	0.787	0.333	0.696	0.351	0.300	0.749
INS	0.000	0.000	0.000	0.252	0.536	0.439	0.894	0.425	0.325
PUT	0.000	0.000	0.000	0.000	0.417	0.754	0.522	0.478	0.589
MFC	0.000	0.000	0.000	0.000	0.000	0.685	0.359	0.907	0.187
CER	0.000	0.000	0.000	0.000	0.000	0.000	0.376	0.225	0.483
AMYG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.130	0.338
OFC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.445
CG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

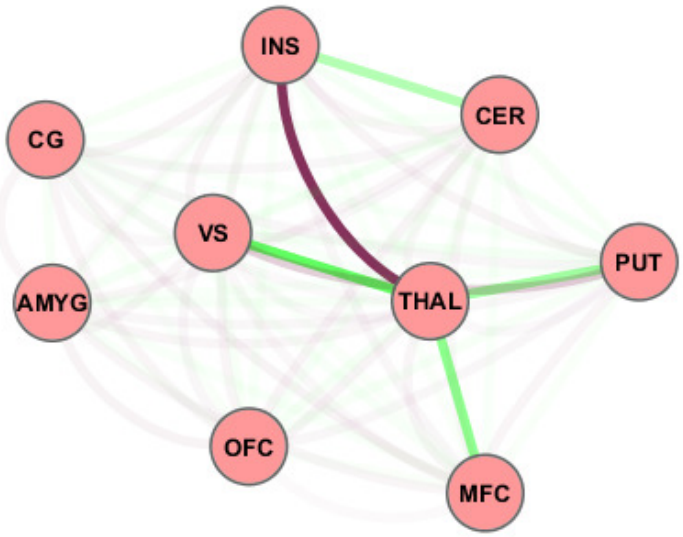
(c) Two sided p-values of the a_2 coefficient

.	VS	THAL	INS	PUT	MFC	CER	AMYG	OFC	CG
VS	0.000	0.026	0.620	0.125	0.792	0.389	0.869	0.305	0.543
THAL	0.000	0.000	0.677	0.059	0.050	0.213	0.246	0.141	0.654
INS	0.000	0.000	0.000	0.170	0.735	0.070	0.583	0.533	0.216
PUT	0.000	0.000	0.000	0.000	0.753	0.214	0.575	0.257	0.544
MFC	0.000	0.000	0.000	0.000	0.000	0.638	0.672	0.858	0.340
CER	0.000	0.000	0.000	0.000	0.000	0.000	0.635	0.421	0.692
AMYG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.717	0.281
OFC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.792
CG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 6.4: Following the same procedure as above, we applied the two-level regression model $R_1 = b_0 + a_0 R_2 + a_1 ER_2 + a_2 GR_2 + \epsilon$ to all region pairs. Because we are trying to find the significance that addiction plays in the partial correlation network of the brain's reward mechanism, it is the a_2 matrix that is of most interest to us which shows the significance of group effect controlling for education level.



(a) two-level regression, group covariate alone. $R_1 = b_0 + a_0 R_2 + a_1 G R_2 + \epsilon$



(b) two-level regression, group and education covariates. $R_1 = b_0 + a_0 R_2 + a_1 E R_2 + a_2 G R_2 + \epsilon$

FIGURE 6.3: This visual display shows two interpretations of group's affect on the partial correlation network. The top figure shows the significance levels of the group coefficient in the model consisting of group status alone, $R_1 = b_0 + a_0 R_2 + a_1 G R_2 + \epsilon$. The bottom figure shows the significance levels of the education and group coefficients in the model containing both, $R_1 = b_0 + a_0 R_2 + a_1 E R_2 + a_2 G R_2 + \epsilon$. In both figures, group significance is shown in green, education significance in dark red, with more significant pathways shown with more opacity.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

This thesis dealt mainly with the problem of determining whether the partial correlation between two groups are significantly different. Currently there are no exact methods to determine this, but there are widely accepted approximate methods, such as the procedures based on bootstrapping and Fisher's transformation. We introduced two new approximate methods, the first a two-level regression and the second a LRT test, that approach this problem from an entirely novel direction.

In numerical simulations of varying sample sizes and numbers of controlling variables, the new methods are, in general, found to be more powerful and at the same time less conservative than the traditional methods. In all simulations, the new methods have a higher rejection rate under the alternative hypothesis, but also a lower acceptance rate under the null. Of the three methods compared, the one based on two-level regression has the most consistent rejection rate with the significance level $\alpha = 0.05$ set in the simulations.

We applied the four methods to an experimental PET dataset taken from a study of addiction's role in the brain's reward mechanism and constructed a series of networks describing the effect that group status has on the partial correlations between brain regions. All four networks showed a similar pattern involving the regions VS, THAL, PUT, and MFC, but varied in path significance.

Finally, we extended the two-level regression approach to help determine a continuous covariates role in the strength of the partial correlation between regions, as well as the inclusion of extraneous information which may impact the significance of our inference. We applied this methodology to our experimental data using two separate extended

models containing information on age and education level. Only age demonstrated a notable effect on the partial correlations between regions.

7.2 Future Work

The groupwise analysis of partial correlation networks is a rich field with many avenues open to future research. One topic that merits further investigation is the construction of an SEM starting model that takes into account the bidirectional pathways generated from the role reversed equations of our two-level regression model. Another topic involves dealing with nonlinear relationships and how they associate with partial correlation. A third topic that could be pursued is the extension of the LRT to allow for dependent observations between groups. Lastly, the current implementation of our LRT utilizes a numerical algorithm that is not guaranteed to provide a valid statistic; new optimization procedures can be attempted to ensure that true global maximisations are found.

7.2.1 Starting Model for SEM

As we mentioned earlier, a major issue in using structural equation modelling is the necessity of a starting model. The two-level approach has the most potential here because of the bidirectional pathways that can be obtained through the use equations, 3.21 and 3.22, that utilize both independent and dependant roles of the variables. In the first equation, one region is modelled as a dependent variable to the other, and when the roles are reversed we get a second model showing the same region as the independent variable. These two measures together can be used to construct a bidirectional pathway diagram to be fed directly into the SEM model with the dependency of the variable under consideration determining the direction of the arrow.

Not only are we getting bidirectional pathways, but we are getting the pathways that have been determined to be most affected by the covariate factors under analysis. For example, if we wanted to study how cocaine addiction might influence a neurological network through the use of an SEM framework, we would need a starting model relating the regions of interest, as well as some insight on how group status might affect this starting model. By using two-level regression before hand, we have a starting point from which we can base our SEM analysis, with bidirectional arrows showing how likely group status is to affect the relationship between regions.

Figure 7.1 illustrates the joint network exploratory and confirmatory analysis paradigm using the covariate PCNA and the covariate SEM. First we start with the PCNA to

arrive at a data-driven non-directional pathway (a → b). This pathway is subsequently split into a directional pathway (b → c) formulating the SEM hypothesis (i.e. the hypothesized pathway), which, after the SEM analysis, will yield the confirmed directional pathways (c → d). To avoid over-fitting, our data will be randomly split into two - the training data set for PCNA exploratory analysis (a,b), and the test data set for SEM confirmatory analysis (c,d).

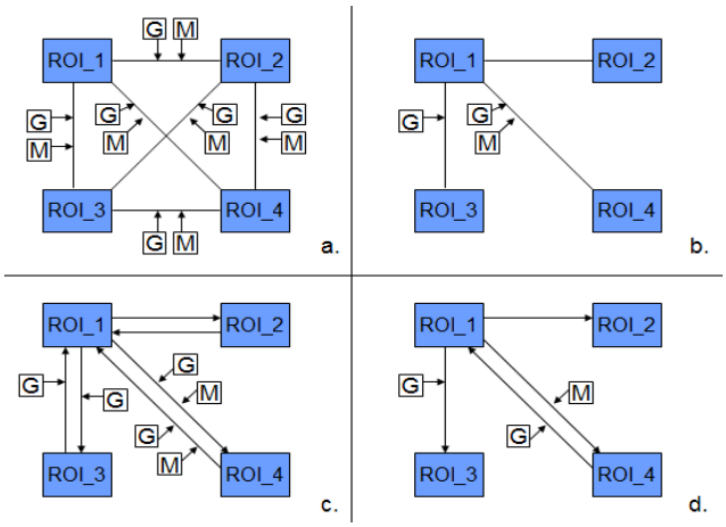
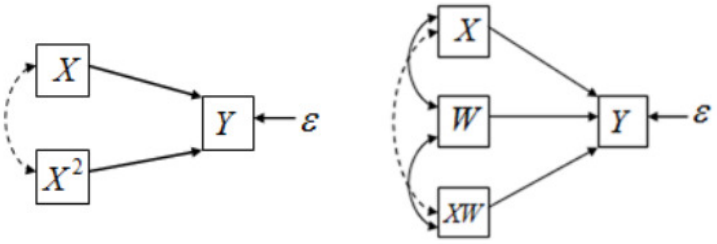


FIGURE 7.1: Identification and confirmation of novel pathways using covariate PCNA/SEM. (a.) Initial unrestricted covariate PCNA model with two covariates group (G) and gender (M). (b.) Final covariate PCNA model derived through the training data. (c.) Directional covariate SEM hypothesis based on the covariate PCNA results. (d.) Final covariate SEM model confirmed by the testing data.

In a recent paper (Marrelec et al., 2007), a similar (while much simpler, with no involvement of any factor nor covariate) methodology was carried out with the partial correlations as the input into a SEM model and subsequently obtained results supporting their expected biological model. Our method contrasts their's in that instead of a single measure of partial correlation mapping to a single non directional link in the SEM model, we would use both measures of partial correlation, giving us directional paths and further reducing the need for a field expert in the initial construction of the pathway design.

We also considered the incorporation of nonlinear and interactive terms in PCNA and SEM because the brain is considered by many to have nonlinear and non-additive functional relationships. Multiple regression analysis offers a straightforward way to represent curvilinear and interactive effects of observed variables. This method can be used directly in PCNA/SEM analysis as illustrated in Figure 7.2. Note that the unanalyzed association is represented in this model with a dashed line instead of a solid one. This represents the possibility that a power/interaction term can be created so that it has

reduced or even zero correlations with its component variables. Higher-order effects such as a cubic term or a linear × quadratic interaction can also be added to these models similarly.



(a) Inclusion of non-linear effect in PC-NA/SEM.

(b) Inclusion of interaction terms in PC-NA/SEM

FIGURE 7.2:

7.2.2 Accounting for Nonlinear Relationships

In accordance to the definition of partial correlation, we have dealt with the linear relationship between two variables controlling for a linear contribution of other variables. But what if the relationships under consideration have a nonlinear form?

Our two-level method would benefit from the study of using different regression techniques to determine both the residuals as well as the coefficients of the factors included in the model. To start with, the residuals could be computed using various nonlinear regression methods and compared. This however would necessitate the coining of new terminology, because partial correlation is defined as the linear relationship between variables controlling for a linear contribution from controlling measurements.

Nonlinear regression could be used to determine the significance of each of the model's coefficients as well. But this too, would mean we are no longer dealing with partial correlations as traditionally defined.

7.2.3 Extending the LRT to Allow for Dependent Observations

In the experimental framework examined in this thesis, we studied the relationship between two separate groups of subjects. In future studies however, we may be presented with two sets of observations from the same group of subjects. For example, we might have access to brain scans taken from the same subject under different conditions or at different time frames. In this case, the simple likelihood function describing the joint

distribution of both groups, see equation 4.9, is no longer valid, because the observations from both groups are no longer independent to each other.

If we are to analyze datasets of this sort using the LRT framework developed in this thesis, we will need a new joint likelihood function that accurately takes into consideration the dependent nature that may be present in the observations.

7.2.4 Utilizing Different Numerical Algorithms

If the LRT method of chapter 4 is to be pursued, we need to make sure the optimization algorithm finds estimates pertaining to global rather than local maximums. Besides substituting a different optimization algorithm, one thing we can try is to perform the optimization many times with different starting points each time and make sure they all come to the same maximum. If the various starting points yield different maxima, we can take the one yielding the highest value. This simple procedure has no convergence criteria and does not guarantee a global maximum, it does however provide an ever monotonically increasing estimate for the maximum.

Bibliography

Anderson, T. W. (1984). *An Introduction to Multivariate Analysis*. New York: John Wiley, 2nd edition.

Anscombe, F. J. and Tukey, J. W. (1963). The examination and analysis of residuals. *Technometrics*, 5(2):141–160. ArticleType: primary_article / Full publication date: May, 1963 / Copyright 1963 American Statistical Association and American Society for Quality.

Ashburner, J. and Friston, K. J. (1999). Spatial normalization. In *Brain Warping*, pages 27—44. Academic Press.

Baba, K., Shibata, R., and Sibuya, M. (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics*, 46(4):657–664.

Barabesi, L. and Greco, L. (2002). A note on the exact computation of the student t, snedecor f and sample correlation coefficient distribution functions. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 51(1):105–110. ArticleType: primary_article / Full publication date: 2002 / Copyright 2002 Royal Statistical Society.

Bond, C. F. and Richardson, K. (2004). Seeing the FisherZ-transformation. *psychometrika*, 69(2):291–303.

Brandner, F. A. (1933). A test of the significance of the difference of the correlation coefficients in normal bivariate samples. *Biometrika*, 25(1/2):102–109.

Brien, C. J., Venables, W. N., James, A. T., and Mayo, O. (1984). An analysis of correlation matrices: equal correlations. *Biometrika*, 71(3):545–554.

Brown, R., Durbin, J., and Evans, J. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2):192, 149.

Bullmore, E. (2000). How good is good enough in path analysis of fMRI data? *Neuroimage*, 11(4):289–301.

Butte, A. J. and Kohane, I. S. (1999). Unsupervised knowledge discovery in medical databases using relevance networks. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 711–715. PMID: 10566452.

Cao, J. and Worsley, K. (1999). The geometry of correlation fields with an application to functional connectivity of the brain. *The Annals of Applied Probability*, 9(4):1021–1057.

Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in Medicine*, 19(9):1141–1164. PMID: 10797513.

Chang, C., Lin, J., and Pal, N. (2008). Exact test critical values for correlation tseting with application. *WSEAS Transactions on Mathematics*, 7(6):363–381.

Chickering, D. M. (1996). Learning bayesian networks is NP-Complete. *Learning From Data: Artificial Intelligence and Statistics V*, pages 121—130.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.

Conover, W. J. (1971). Practical nonparametric statistics. pages 295–301. New York: John Wiley & Sons.

David, F. N. (1938). *Tables of the correlation coefficient*. The" Biometrika" office.

Diaz-Garcia, J. A., Jaimez, R. G., and Mardia, K. V. (1997). Wishart and Pseudo-Wishart distributions and some applications to shape theory. *journal of multivariate analysis*, 63:7387.

Dowdy, S. M., Wearden, S., and Chilko, D. M. (2004). *Statistics for research*. Wiley-Interscience.

Draper, N. and Smith, H. (1998). *Applied Regression Analysis*. Wiley-Interscience.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185.

Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1 edition.

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

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

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000). Using bayesian networks to analyze expression data. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 7(3-4):601–620. PMID: 11108481.

Fuente, A. D. L., Bing, N., Hoeschele, I., and Mendes, P. (2004). *Discovery of meaningful associations in genomic data using partial correlation coefficients*, volume 20. Oxford Univ Press.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1:141—150.

Goria, M. N. (1980). On testing the correlation coefficient of a bivariate normal distribution. *Metrika*, 27(1):189–194.

Grossman, S. I. and Styan, G. P. H. (1972). Optimality properties of theil's BLUS residuals. *Journal of the American Statistical Association*, 67(339):672–673.

Hawkins, D. L. (1989). Using u statistics to derive the asymptotic distribution of fisher's z statistic. *The American Statistician*, 43(4):235–237.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B (Methodological)*, 15(2):193–232.

Jensen, D. R. and Ramirez, D. E. (1999). Recovered errors and normal diagnostics in regression. *Metrika*, 49(2):107–119.

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97(10):2177–2189.

Kendall, M. and Stuart, A. (1973). In *The Advanced Theory of Statistics*, volume 2. 3rd edition.

Khatri, C. G. (1968). Some results for the singular normal multivariate regression models. *Sankhya: The Indian Journal of Statistics, Series A*, pages 267–280.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S. A., and Fox, P. T. (2000). Automated talairach atlas labels for functional brain mapping. *Human Brain Mapping*, 10(3):120–131. PMID: 10912591.

Lauterbur, P. C. (1973). Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191.

Lee, J. S., Lee, D. S., Kim, Y. K., Kim, J., Lee, H. Y., Lee, S. K., Chung, J., and Lee, M. C. (2004). Probabilistic map of blood flow distribution in the brain from the internal carotid artery. *NeuroImage*, 23(4):1422–1431. PMID: 15589106.

Ma, Y., Volkow, N., Zhu, W., Rao, M., and Pradhan, K. (2004). Automated region of interest (ROI) analysis for PET studies. In *Proceeding of the Organization for Human Brain Mapping 10th annual Meeting*, volume Methodology Section, Budapest. CD-ROM.

Marrelec, G., Horwitz, B., Kim, J., Plgrini-Issac, M., Benali, H., and Doyon, J. (2007). Using partial correlation to enhance structural equation modeling of functional MRI data. *Magnetic Resonance Imaging*, 25(8):1181–1189. PMID: 17475433.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test stastistic for the number of components in a normal mixture. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3):318–324.

McLntosh, A. R. and Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping*, 2(1-2):2–22.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

Novack, T. A., Bush, B. A., Meythaler, J. M., and Canupp, K. (2001). Outcome after traumatic brain injury: Pathway analysis of contributions from premorbid, injury severity, and recovery variables. *Archives of Physical Medicine and Rehabilitation*, 82(3):300–305.

Payton, M., Greenstone, M., and Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *J Insect Sci*, 3(23).

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression model. *Journal of the American Statistical Association*, 104:735–746.

Phelps, M. E., Huang, S. C., Hoffman, E. J., Selin, C., Sokoloff, L., and Kuhl, D. E. (1979). Tomographic measurement of local cerebral glucose metabolic rate in humans with (F-18)2-fluoro-2-deoxy-D-glucose: validation of method. *Annals of Neurology*, 6(5):371–388.

Plackett, R. L. (1950). Some theorems in least squares. *Biometrika*, 37(1/2):149–157.

Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000). Likelihood-Based inference with singular information matrix. *Bernoulli Society for Mathematical Statistics and Probability*, 6(2):243–284.

Schafer, J. and Strimmer, K. (2005a). An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6):754–764.

Schafer, J. and Strimmer, K. (2005b). A shrinkage approach to Large-Scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1).

Scheinin, N., Aalto, S., Wilson, I., Kemppainen, N., Nagren, K., Kailajarvi, M., Scheinin, M., and Rinne, J. (2008). P2-073: Reproducibility of automated ROI analysis of PET amyloid ligand [11C]PIB uptake. *Alzheimer's and Dementia*, 4(4):T389–T389.

Schenker, N. (1985). Qualms about bootstrap confidence intervals. *Journal of the American Statistical Association*, 80(390):360–361.

Seber, G. A. F. G. A. F. (1977). *Linear regression analysis / G.A.F. Seber.* Wiley series in probability and mathematical statistics. Wiley, New York :. Includes index. Bibliography: p. 433-457.

Sheather, S. J. and Jones, M. C. (1991). A reliable Data-Based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):683–690.

Stein, J. L., Wiedholz, L. M., Bassett, D. S., Weinberger, D. R., Zink, C. F., Mattay, V. S., and Meyer-Lindenberg, A. (2007). A validated network of effective amygdala connectivity. *NeuroImage*, 36(3):736–745.

Szkely, G. J. and Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1):58–80.

Theil, H. (1965). The analysis of disturbances in regression analysis. *Journal of the American Statistical Association*, 60(312):1067–1079.

Volkow, N. D., Fowler, J. S., Wang, G., and Goldstein, R. Z. (2002). Role of dopamine, the frontal cortex and memory circuits in drug addiction: insight from imaging studies. *Neurobiology of Learning and Memory*, 78(3):610–624. PMID: 12559839.

Volkow, N. D., Wang, G., Fowler, J. S., Logan, J., Gatley, S. J., Hitzemann, R., Chen, A. D., Dewey, S. L., and Pappas, N. (1997). Decreased striatal dopaminergic responsiveness in detoxified cocaine-dependent subjects. *Nature*, 386(6627):830–833.

Volkow, N. D., Wang, G., Ma, Y., Fowler, J. S., Wong, C., Ding, Y., Hitzemann, R., Swanson, J. M., and Kalivas, P. (2005). Activation of orbital and medial prefrontal cortex by methylphenidate in Cocaine-Addicted subjects but not in controls: Relevance to addiction. *J. Neurosci.*, 25(15):3932–3939.

Volkow, N. D., Wang, G., Telang, F., Fowler, J. S., Logan, J., Childress, A., Jayne, M., Ma, Y., and Wong, C. (2006). Cocaine cues and dopamine in dorsal striatum: Mechanism of craving in cocaine addiction. *J. Neurosci.*, 26(24):6583–6588.

Volkow, N. D., Wang, G., Telang, F., Fowler, J. S., Logan, J., Childress, A., Jayne, M., Ma, Y., and Wong, C. (2008). Dopamine increases in striatum do not elicit craving in cocaine abusers unless they are coupled with cocaine cues. *NeuroImage*, 39(3):1266–1273.

Welsh, T., Mueller, K., Zhu, W., Volkow, N., and Meade, J. (2001). Graphical strategies to convey functional relationships in the human brain: a case study. In *Proceedings of the conference on Visualization '01*, pages 481–484, San Diego, California. IEEE Computer Society.

Winterbottom, A. (1979). A note on the derivation of fisher's transformation of the correlation coefficient. *The American Statistician*, 33(3):142–143.