Stony Brook University

Path Analysis of Multivariate Time Series fMRI Data with Subject-level Covariates

A Dissertation Presented

By

Yue Zhang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

In

Applied Mathematics and Statistics

Stony Brook University

August 2007

Stony Brook University

The Graduate School

Yue Zhang

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

Nancy Mendell
Chairperson of Defense
Professor
Department of Applied Mathematics and Statistics

Hongshik Ahn
Member
Associate Professor
Department of Applied Mathematics and Statistics

Amanda Stent
Outside Member
Assistant 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

Path Analysis of Multivariate Time Series fMRI Data with Subject-level Covariates

By

Yue Zhang

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2007

The ultimate goal of brain functional connectivity studies is to propose, test, modify, and compare certain directional brain pathways. Functional magnetic resonance imaging (fMRI) studies are routinely conducted with a group of subjects to generate relevant data for brain functional pathway analysis. Path analysis, also called structural equation modeling (SEM), is commonly regarded as the ideal statistical method for the analysis of brain functional pathways.

Three approaches are conceivable for the path analysis of multi-subject, multivariate time series data from fMRI experiments. They are: (A) summarize (e.g. average the time series data across the subjects) and then analyze, (B) analyze and then summarize, and (C) simultaneous analysis. Previously, we have developed the SEM methodology for approach (B) (Kim et al. 2006). In this thesis, we propose a path analysis framework for approach (C), the simultaneous analysis of multi-subject, multivariate time series data with subject-level covariates. Comparisons are made between the three approaches and

guidelines are provided. We also develop a two-level SEM-bootstrap method for the simultaneous resampling of the subjects and/or the individual or group average time series data. This approach would enable us to incorporate subject-level covariates into SEM analysis when other approach is not tenable.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The ultimate goal for brain functional connectivity study is to propose, test, modify and compare certain directional brain pathways. During a typical functional MRI (fMRI) experiment, each subject's functional activities of multiple brain regions of interest (ROIs) in the brain are measured longitudinally over the course of several minutes. Therefore, the multivariate time series data from the fMRI experiment is obtained for each subject. Meanwhile, usually a group of subjects participate in the same experiment in order to obtain statistical estimations for the population of interest. Thus, fMRI studies usually contain multi-subject, multivariate time series data. Furthermore, subject-level covariates such as gender, age, education, and verbal IQ (VIQ) are involved into the fMRI study. It is of interest to incorporate these covariates into the analysis to examine the relationships between functional brain pathways and subject-level covariates.

Path analysis, a special case of Structural Equation Modeling (SEM), which was designed for directional connectivity modeling, testing, and comparisons, is an ideal statistical method for studying brain pathways. Path analysis is interesting to many neuroscientists because it can quantify functional relationships between multiple brain regions in terms of unidirectional connections. In the early 1990s McIntosh introduced

SEM to neuroimaging (McIntosh and Gonzales-Lima, 1991; McIntosh et al., 1994) for modeling, testing and comparison of directional effective connectivity of the brain. SEM has quickly become popular in this field (Büchel and Friston, 1997, 1999; Honey et al., 2002).

1.1 Structural Equation Modeling (SEM)

Structural equation modeling represents the hybrid of two separate statistical traditions. The first tradition is confirmatory analysis developed in the disciplines of psychology and psychometrics. The second tradition is simultaneous equation modeling developed mainly in econometrics, but having an early history in the field of genetics. The combination of these two methodologies into a coherent analytic framework of SEM was based on the work of Jöreskog (1973), Keesling (1972), and Wiley (1973). SEM imposes a structure on the covariance/correlation matrix of the multivariate normal random variables, and sometimes on its mean vector as well. The general structural equation model as outlined by Jöreskog (1973) consists of two parts: the first is a latent variable model that is similar to the simultaneous equation model of econometrics except that all variables are latent ones, and the second part is the measurement model linking indicators to latent variables as in factor analyses. The procedure was implemented in Jöreskog and Sörbom's LISREL program. Bentler and Weeks (1980), McArdle and McDonald (1984), and others have proposed alternative representations of general structural equations. Most analysts now agree that both the new and old representations

are capable of treating the range of linear models that typically occur in practice. The "LISREL notation", which is the most widely accepted representation, will be used in this work.

Structural equation modeling is a more powerful method whose framework includes many common statistical procedures, such as multiple regression, confirmatory factor analysis, canonical correlations, and ANOVA. That is, these statistical techniques are special cases of SEM. In fact, structural equation modeling is an extension of the general linear model (GLM) of which multiple regression is a part. When there are no latent variables in the model, modeling system of structural relationships among a set of observed variables is often referred to as path analysis

With respect to using structural modeling in brain imaging, the emphasis so far has been on general SEM approaches, primarily with path analysis. It was McIntosh who introduced SEM to neuro-imaging field in a series of papers, and the methodology quickly became popular in this field (McIntosh, 1998). The first application of structural equation modeling to neuroscience was carried out by McIntosh and Gonzalez-Lima (1991), and they demonstrated how structural modeling can be used to determine the functional interrelationships between brain structures that form the auditory system. The models were discussed in the context of previous findings to demonstrate how structural modeling can not only complement, but also extract more information from auditory system 2-DG uptake mapping experiments. Büchel and Friston (1997) developed a SEM model for the interrelations among dorsal visual pathway areas under visual motion conditions while varying the attentional component of the task. Their model showed a

significant modulatory effect of prefrontal regions on a motion-selective cortical area V5 and the posterior parietal cortex. Fletcher et al. (1999) used SEM to explore time-dependent changes in inter-regional connectivity as a function of item and grammar rule learning in an experiment. They found that although there were no significant effects of item learning on the measured path strengths, rule learning was associated with a decrease in right fronto-parietal connectivity and an increase in connectivity between left and right prefrontal cortex. This previous work clearly demonstrates that path analysis has a great potential to make new progresses and developments in the field of brain imaging and brain mapping.

1.2 Current Approaches and Limitations

Although SEM is a powerful tool for the analysis of brain connectivity, the potential of structural equation modeling in this field has been explicitly recognized only recently. A milestone event was the paper by Büchel, Coull, and Friston (1999), published in *Science,* which demonstrated the general scientific acceptance of the methodology to neuroscience field. As described at the beginning of this chapter, we have a set of multi-subject multivariate time series of observations from an fMRI study. The simplest method would be to analyze the time series for each subject separately, as was done in Büchel and Friston (1997). However, they assumed that the observations are independent, which is not true for time series data since the observations are autocorrelated. In fact, no existing SEM software has the right procedure for the analysis of fMRI data because

conventional SEM assumes that the observations are independent (Bentler and Wu, 2002). An alternative method is to extract the first eigenvector of each relevant ROI, which is representative of the entire population, by employing Principal Components Analysis (PCA) (Bullmore et al., 2000). And the number of independent observations, ν, is estimated, which is smaller than the number of time series observations, *n*, due to temporal autocorrelation in fMRI time series data. In this way, the autocorrelated time series from all subjects are reduced to a single sequence of observations for each ROI, and the ensuing path analysis can be carried out using any existing general purpose SEM software when assume ν independent observations. This approach was also adopted by Honey et al. (2002) for the working memory study. However, by reducing the group data to merely the first eigenvector, one would lose a substantial amount of information and could mistakenly make a significant pathway insignificant. When we perform PCA, we assume the observations of each ROI are independent, which is not true for fMRI data. And there is as yet no general agreement about the best method for estimating ν in a given time series, nor about the best procedure to follow if ν differs markedly between regional time series. Therefore this approach is far from ideal. To follow a hierarchical data structure of fMRI data, Friston et al. (1999, 2005) employed the hierarchical linear model to analyze random effects of multi-subject fMRI studies, however, it was a univariate methodology. Mechelli et al. (2002) illustrated how differences in connectivity among subjects can be addressed explicitly using structural equation modeling by constructing a multi-subject network that comprises *k* regions of interest for each of the *m* subjects studied, resulting in a total of *k*m* nodes. Although their model allows one to

test directly for differences among subjects by comparing models that do and do not allow a particular connectivity parameter to vary over subjects, it requires a large number of observations for each subject and the model can be fairly tested only when there are few subjects, as well as few paths in the model. Kim et al. (2006) proposed a new SEM method, called unified SEM, which incorporated longitudinal as well as contemporaneous path connections into the conventional SEM, making the observations nearly independent and thus appropriate for the analysis using the existing SEM software. They analyzed fMRI data at two stages. At Stage 1, unified SEM was fitted to each subject individually, and at Stage 2, for each path, the mean of estimated path connections of all subjects ("summary statistics") was simply tested by a one-sample t-test to determine whether this path was significant. However, the two-stage unified SEM method analyzed each subject separately, and failed to consider the hierarchical structure of fMRI data.

1.3 Overview

In this work, we will present three possible approaches of analyzing multi-subject multivariate time-series fMRI data. They are: (1) summarize (e.g. average the time series data across the subjects) and then analyze, (2) analyze and then summarize (Kim et al., 2006), and (3) simultaneous analysis. The application of these three approaches will be illustrated by an analysis of the visual attention network (Driver and Mattingley, 1998; Kanwisher and Wojciulik, 2000). The study data came from an fMRI three ball-tracking

experiment with 28 normal subjects (14 females and 14 males). The appropriateness of these three methods will be demonstrated and comparisons between these three approaches will be outlined.

The path analysis is performed by the most widely used SEM software LISREL (Jöreskog and Sörbom, 1996) and RPOC CALIS in SAS.

The main goal of this work is to test and compare the path models for all subjects or different groups from the three approaches.

Chapter 2

Structural Equation Modeling (SEM) and Path Analysis

Structural equation modeling (SEM) is a very powerful multivariate statistical technique for building and testing statistical models. Structual equation modeling represents the hybrid of two separate statistical methods. The first one is confirmatory factor analysis developed in psychology and psychometrics, and the second one is simultaneous equation modeling developed in econometrics. The basic idea of SEM differs from the usual statistical approach of modeling individual observations.

In multiple regression or ANOVA (analysis of variance) the regression coefficients or parameters of the model are derived from the minimization of the sum of squared differences between the predicted and observed dependent variables for each case (Bollen 1989).

Structural equation modeling (SEM) grows out of and serves purposes similar to multiple regression, but SEM takes into account the modeling of interactions, correlated independent variables, measurement error, correlated disturbance terms, and multiple latent variables measured by multiple indicators.

The procedures of structural equation modeling emphasize covariances rather than cases, approaching the data from a different perspective. Instead of minimizing the differences between observed and predicted values of each variable, structural equation modeling estimates the parameters by minimizing the difference between the observed sample covariances and the predicted covariances by a structural or path model.

The fundamental hypothesis for structural equation modeling is that the covariance matrix of the observed variables is a function of a set of parameters. If the model is correct and if we know the values of the set of parameters, the population covariance matrix can be reproduced. The equation of the fundamental hypothesis is

$$\Sigma = \Sigma(\theta), \tag{2.1}$$

where Σ is the population covariance matrix of the observed variables, θ is a vector of free model parameters, and $\Sigma(\theta)$ is the covariance matrix written as a function of the model parameters in θ. Equation (2.1) implies that each element of the covariance matrix is a function of one or more model parameters. The relation of Σ and $\Sigma(\theta)$ is basic to an understanding of identification, estimation, and assessments of model fit (Bollen. 1989).

The parameters to be estimated in structural equation modeling are connection strengths (or path coefficients), the variances and covariances of exogenous variables, and the variances and covraiances of disturbance terms. The path coefficients are the standardized partial regression coefficients, which representing the response of the endogenous variables to a standard unit change in an exogenous variable, while the other variables in the model are held constant.

The elements of the structural model are mathematically defined in terms of a set of

simultaneous regression equations. Maximum likelihood estimation (MLE) is by far the most common method for estimation, which can generate log-likelihood and χ^2 values for each model. We will explain the estimation procedures in details in section 2.2.

The path coefficients can be estimated by LISREL (Jöreskog and Sörbom, 1996), SAS CALIS procedure, AMOS, and other SEM software.

Structural equation modeling is a more powerful alternative to multiple regression, confirmatory factor analysis, canonical correlations, ANOVA, and analysis of covariance. That is, these statistical techniques are special cases of SEM. In fact, structural equation modeling is an extension of the general linear model (GLM) of which multiple regression is a part. When there are no latent variables in the model, modeling system of structural relationships among a set of observed variables is often referred to as path analysis. Hence, path analysis is a special case of SEM, too.

Since all variables in our experiment were observed, we focus on path analysis, a type of structural equation modeling with observed variables. In this model, there are no latent variables, only directly observed or measured variables.

2.1 Structural Equation Models with Observed Variables and Implied Covariance Matrix

In the structural equation models with observed variables, we assume that all the endogenous and exogenous variables are directly observed with no measurement error.

If we assume there are p endogenous (dependent) variables and q exogenous (independent) variables, then the general representation of the structural equations with observed variables is

$$y = By + \Gamma x + \zeta \text{ ,} \tag{2.2}$$

where $y = (y_1, y_2, ..., y_p)'$ is a $p \times 1$ vector of endogenous variables, $x = (x_1, x_2, ..., x_q)'$ is a $q \times 1$ vector of exogenous variables.

B is a $m \times m$ matrix of coefficients of the y variables in the structural relationship, Γ is a $m \times n$ matrix of coefficients of the x variables in the structural relationship, and $\zeta = (\zeta_1 \zeta_2, ..., \zeta_p)'$ is a $p \times 1$ vector of errors in the equations.

The assumptions of this model are

1. $(I - B)$ is nonsingular, the expectation of ζ is zero, and ζ is uncorrelated with x.

2. The free parameter matrices to be estimated are B, Γ, $\Phi = \text{cov}(x)$, and $\Psi = \text{cov}(\zeta)$.

Equation (2.2) can be written in the reduced form as

$$y = A(\Gamma x + \zeta) \text{ ,} \tag{2.3}$$

where $A = (I - B)^{-1}$.

If we assume $\theta = \{\theta_i\}$ is the set of free parameters to be estimated in B, Γ, Φ, and Ψ, the fundamental hypothesis of the general structural equation model is Equation (2.1). Σ is the population covariance matrix of the observed variables y and x, and $\Sigma(\theta)$ is the covariance matrix written as a function of the free model parameters in θ.

$\Sigma(\theta)$ is assembled in three pieces: (1) the covariance matrix of y, (2) the covariance

matrix of x with y, and (3) the covariance matrix of x:

$$\Sigma(\theta) = \text{cov}\begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} \Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{pmatrix}. \tag{2.4}$$

We first consider $\Sigma_{yy}(\theta)$, the implied covariance structure of *y*:

$$\begin{aligned} \Sigma_{yy}(\theta) &= E(yy') \\ &= E[(I-B)^{-1}(\Gamma x + \zeta)((I-B)^{-1}(\Gamma x + \zeta))'] \\ &= E[(I-B)^{-1}(\Gamma x + \zeta)(x'\Gamma' + \zeta')(I-B)^{-1\prime}] \\ &= (I-B)^{-1}(E(\Gamma x x'\Gamma') + E(\Gamma x \zeta') + E(\zeta x'\Gamma') + E(\zeta\zeta'))(I-B)^{-1\prime} \\ &= (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B)^{-1\prime}, \end{aligned} \tag{2.5}$$

where Φ is the covariance matrix of *x*, and Ψ is the covariance matrix of ζ.

The implied covariance matrix of *y* with *x* is

$$\begin{aligned} \Sigma_{yx}(\theta) &= E(y'x) \\ &= E[((I-B)^{-1}(\Gamma x + \zeta))'x] \\ &= (I-B)^{-1}\Gamma\Phi. \end{aligned} \tag{2.6}$$

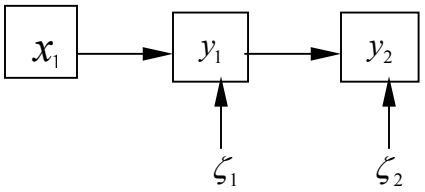
The implied covariance matrix of *x*, cov(*x*) is equal to Φ, or

$$\begin{aligned} \Sigma_{xx}(\theta) &= E(xx') \\ &= \Phi. \end{aligned} \tag{2.7}$$

After assembling Equations (2.5), (2.6) and (2.7), the implied covariance matrix of y and x is

$$\Sigma(\theta) = \begin{pmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B)^{-1\prime} & (I-B)^{-1}\Gamma\Phi \\ \Phi\Gamma'(I-B)^{-1\prime} & \Phi \end{pmatrix}. \tag{2.8}$$

Let us illustrate (2.8) with a hypothetical example that is the following structural model:



The model equations are

$$\begin{aligned} y_1 &= \gamma_{11} x_1 + \zeta_1 \\ y_2 &= \beta_{21} y_1 + \zeta_2, \end{aligned} \tag{2.9}$$

where $\text{cov}(\zeta_1, x_1), \text{cov}(\zeta_1, \zeta_2)$, and $\text{cov}(x_1, \zeta_2)$ are zero.

The parameter matrices to be estimated for this model are

$$B = \begin{pmatrix} 0 & 0 \\ \beta_{21} & 0 \end{pmatrix}, \ \Gamma = \begin{pmatrix} \gamma_{11} \\ 0 \end{pmatrix}, \ \Psi = \begin{pmatrix} \psi_{11} & 0 \\ 0 & \psi_{22} \end{pmatrix}, \text{ and } \Phi = (\phi_{11}). \tag{2.10}$$

Substituting (2.10) into (2.8) and using (2.1) leads to

$$\Sigma = \Sigma(\theta)$$

$$\begin{bmatrix} \text{var}(y_1) & & \\ \text{cov}(y_2, y_1) & \text{var}(y_2) & \\ \text{cov}(x_1, y_1) & \text{cov}(x_1, y_2) & \text{var}(x_1) \end{bmatrix} = \begin{bmatrix} \gamma_{11}^2 \phi_{11} + \psi_{11} & & \\ \beta_{11}(\gamma_{11}^2 \phi_{11} + \psi_{11}) & \beta_{21}^2(\gamma_{11}^2 \phi_{11} + \psi_{11}) + \psi_{22} & \\ \gamma_{11}\phi_{11} & \beta_{21}\gamma_{11}\phi_{11} & \phi_{11} \end{bmatrix}. \tag{2.11}$$

The left-hand side of (2.11) is the population covariance matrix of $y_1, y_2, and\ x_1$, and the right-hand side is composed of each variance and covariance in terms of the unknown parameters. The parameter estimations are derived from the relation of the covariance matrix of the observed variables to the structural parameters (Bollen 1989).

2.2 Estimation of Model Parameters

The unknown parameters of the model are (a) the variances and covariances of exogenous variables contained in Φ, (b) the variances and covariances of disturbance terms contained in Ψ, and (c) the path coefficients contained in B and Γ. The unknown parameters are estimated so that the implied covariance matrix, $\hat{\Sigma}$, is close to the sample covariance matrix S. In order to know when the estimates are as 'close' as possible, we must define 'close'. That is, we should find a discrepancy function that is to be minimized. Many different fitting functions for the task are possible.

The fitting functions defined by $F(S,\Sigma(\theta))$ are based on S, the sample covariance matrix, and $\Sigma(\theta)$, the implied covariance matrix of structural parameters. If estimates of θ are substituted in $\Sigma(\theta)$, this leads to the implied covariance matrix, $\hat{\Sigma}$. The value of the fitting function for $\hat{\theta}$ is $F(S,\hat{\Sigma})$. The fitting function $F(S,\hat{\Sigma})$ is a scalar that measures the discrepancy between the sample covariance matrix S and the implied covariance matrix $\hat{\Sigma}$ and can be characterized by the following properties

(i) $F(S,\Sigma(\theta)) \geq 0$,

(ii) $F(S,\Sigma(\theta))=0$, if and only if $\Sigma(\theta)=S$,

(iii) $F(S,\Sigma(\theta))$ is a continuous function in S and $\Sigma(\theta)$.

According to Browne (1984), minimizing fitting functions that satisfy these conditions leads to consistent estimators of θ. Three fitting functions are widely used, which are maximum likelihood (ML), unweighted least squares (ULS), and generalized least squares (GLS) functions.

The most widely used fitting function for general structural equation models is the maximum likelihood (ML) function.

The derivation of the likelihood function is illustrated as follows.

In deriving F_{ML}, the set of N independent observations are of the multinormal random variables y and x. Consider the $(p+q)\times 1$ vector of observed variables, Z, containing both y and x. Then Z follows the multivariate normal distribution with zero mean and covariance matrix Σ, i.e. $Z \sim MN(0,\Sigma)$, then its probability density function is

$$f(Z \mid \Sigma) = (2\pi)^{-\frac{(p+q)}{2}} \left|\Sigma\right|^{-\frac{1}{2}} \exp\left[\left(-\frac{1}{2}\right) Z'\Sigma^{-1}Z\right]. \tag{2.12}$$

For a random sample of N independent observations of Z, the joint probability density function is

$$f(Z_1, Z_2, \cdots, Z_N \mid \Sigma) = f(Z_1 \mid \Sigma) f(Z_2 \mid \Sigma) \cdots f(Z_N \mid \Sigma). \tag{2.13}$$

Once we observe a given sample, the likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{N} \left\{ (2\pi)^{-\frac{(p+q)}{2}} \left|\Sigma\right|^{-\frac{1}{2}} \exp\left[\left(-\frac{1}{2}\right) Z'\Sigma^{-1}Z\right] \right\} \\ &= (2\pi)^{-\frac{N(p+q)}{2}} \left|\Sigma(\theta)\right|^{-\frac{N}{2}} \exp\left[\left(-\frac{1}{2}\right) \sum_{i=1}^{N} Z_i'\Sigma^{-1}(\theta) Z_i\right]. \end{aligned} \tag{2.14}$$

Σ was substituted by $\Sigma(\theta)$ based on the covariance structure hypothesis that $\Sigma = \Sigma(\theta)$.

The log of the likelihood function is

$$\log L(\theta) = -\frac{N(p+q)}{2} \log(2\pi) - \left(\frac{N}{2}\right) \log \left|\Sigma(\theta)\right| - \left(\frac{1}{2}\right) \sum_{i=1}^{N} Z_i'\Sigma^{-1}(\theta) Z_i. \tag{2.15}$$

We can rewrite the last term of (2.15) as

$$
\begin{aligned}
-\frac{1}{2}\sum_{i=1}^{N} Z_i'\Sigma(\theta)^{-1}Z_i &= -\frac{1}{2}\sum_{i=1}^{N} tr\left[Z_i'\Sigma^{-1}(\theta)Z_i\right] \\
&= -\frac{N}{2}\sum_{i=1}^{N} tr\left[N^{-1}Z_iZ_i'\Sigma^{-1}(\theta)\right] \\
&= -\frac{N}{2} tr\left[S^*\Sigma^{-1}(\theta)\right].
\end{aligned}
\tag{2.16}
$$

where $S^* = \frac{N-1}{N}S$, and S is the sample covariance matrix of y and x.

The Equation (2.14) can be rewritten as

$$
\begin{aligned}
\log L(\theta) &= \text{constant} - \left(\frac{N}{2}\right)\log|\Sigma(\theta)| - \frac{N}{2} tr\left[S^*\Sigma^{-1}(\theta)\right] \\
&= \text{constant} - \left(\frac{N}{2}\right)\left\{\log|\Sigma(\theta)| + tr\left[S^*\Sigma^{-1}(\theta)\right]\right\}.
\end{aligned}
\tag{2.17}
$$

The next step is to maximize Equation (2.17) with respect to the parameters of the model. To maximize the log likelihood in Equation (2.17), we need to obtain the derivatives with respect to the parameters of the model, set the derivatives equal to zero and solve.

Since the constant term in (2.17) does not contain model parameters, this term will have no consequences when obtaining the derivatives and can be ignored. Second, the difference between S^* and the usual unbiased sample covariance matrix S is negligible in large samples, since $S^* = [(N-1)/N]S$. Therefore, we can rewrite Equation (2.17) as

$$
\log L(\theta) = -\left(\frac{N}{2}\right)\left\{\log|\Sigma(\theta)| + tr\left[S\Sigma^{-1}(\theta)\right]\right\}. \tag{2.18}
$$

A problem with Equation (2.18) is that it does not possess the properties of a discrepancy function as described earlier. For example, if $\Sigma(\theta) = S$ then the second term

on the right side of Equation (2.18) will be an identity matrix and the trace will be $p+q$. However, the difference between the first term and the second term will not be zero as required for a discrepancy function. To create a proper discrepancy function based on Equation (2.18), we need to add terms that do not depend on model parameters and can make the discrepancy function equal zero when $\Sigma(\theta)=S$. First we remove the term *–(N/2)*, in which case we will minimize the function instead of maximizing it. Then, we add terms that do not depend on model parameters. This gives,

$$F_{ML} = \log\left|\Sigma(\theta)\right| + tr[S\Sigma(\theta)^{-1}] - \log\left|S\right| - (p+q). \tag{2.19}$$

If the model fits perfectly, that is, $\Sigma(\theta)=S$, the first and third terms sum to zero and the second and fourth terms sum to zero and therefore Equation (2.19) in a proper fitting function as required.

If we denote $F_{ML} = F(\theta)$, a necessary condition for minimizing of $F(\theta)$ is to choose the estimates $\hat{\theta}_i$ s in that the partial derivatives of $F(\theta)$ with respect to each of $\hat{\theta}_i$ are zero. That is,

$$\frac{\partial F(\theta)}{\partial \theta_i} = 0, \quad i=1,2, \ldots\ldots,t. \tag{2.20}$$

In the general structural Equation model where $F(\theta)$ is the F_{ML}, F_{GLS} or F_{ULS} fitting functions, (2.20) results in t equations which are typically nonlinear in the parameters, and therefore explicit solutions for the parameters are not always found. In these cases minimization with numerical methods such as iterative numerical procedure are necessary to find the free or equality constrained unknown parameters in B, Γ, Φ and Ψ. The goal is to develop a sequence of values for θ such that the last vector in the sequence

minimizes F_{ML}. The three key issues in the iterative solution procedure are (1) the selection of initial or starting values $\theta^{(1)}$, (2) the rules for moving from one step in the sequence to the next (from $\theta^{(i)}$ to $\theta^{(i+1)}$), and, (3) when to stop the iteration. A numerical method for these procedures is described in Bollen, 1989. LISREL implements this method.

2.3 Statistical Inference

By using maximum likelihood estimation of the path model, one can explicitly test the hypothesis that the model fits the data. Statistical inference in the structural equation modeling is the goodness of the overall fit of the model. In other word, how significantly different are the implied and observed covariance structures, i.e. $H_0 : \Sigma = \Sigma(\theta)$. Under some conditions, $(N-1)F_{ML}$ provides chi-square estimators to test H_0. The following shows the justification for using $(N-1)F_{ML}$ as chi-square estimators based on the likelihood ratio test.

Anderson (1958) has shown that if y and x have a multivariate normal distribution, then the unbiased sample covariance matrix S has a Wishart Distribution. Bollen (1989) showed that the log of Wishart likelihood function could be written as

$$\log L(\theta) = \text{constant} - \frac{N-1}{2}\left\{\log|\Sigma(\theta)| + tr[S\Sigma(\theta)^{-1}]\right\}. \qquad (2.21)$$

Comparing (2.21) to F_{ML} (2.19), we see that the value of estimated parameters $\hat{\theta}$

that maximizes $\log L(\theta)$ will minimize F_{ML}.

And we notice that under H_0, (2.21) can be written as

$$\log L_0 = -\frac{N-1}{2}\left\{\log|\hat{\Sigma}| + tr(S\hat{\Sigma}^{-1})\right\}, \tag{2.22}$$

where $\hat{\Sigma} = \Sigma(\hat{\theta})$.

It is the log likelihood under the null hypothesis that the specified model holds in the population. The corresponding alternative hypothesis is that Σ is any positive definite matrix. Under the alternative hypothesis, if *S*, the sample covariance matrix, is the estimator of Σ, the log likelihood function $\log L_a$ attains its maximum value. Therefore, the log likelihood under the alternative hypothesis, $\log L_a$, can be written as

$$\begin{aligned} \log L_a &= -\frac{N-1}{2}\left\{\log|S| + tr(SS^{-1})\right\} \\ &= -\frac{N-1}{2}\left\{\log|S| + k\right\}, \end{aligned} \tag{2.23}$$

where $k=p+q$.

The natural logarithm of the likelihood ratio, when multiplied by –2 is distributed as chi-square variate when H_0 is true and (N-1) is large. In this case,

$$\begin{aligned} -2\log\left(\frac{L_0}{L_a}\right) &= -2\log L_0 + 2\log L_a \\ &= (N-1)\left[\log|\hat{\Sigma}| + tr(S\hat{\Sigma}^{-1})\right] - (N-1)(\log|S| + k) \\ &= (N-1)\left(\log|\hat{\Sigma}| + tr(S\hat{\Sigma}^{-1}) - \log|S| - k\right) \\ &= (N-1)F_{ML}(S,\hat{\Sigma}). \end{aligned} \tag{2.24}$$

The large sample distribution of logarithm of the likelihood ratio, when multiplied by –2 is chi-square with degrees of freedom given by the difference in the number of

nonredundant elements in Σ and the number of free parameters in the model. That is, $(N-1)F_{ML} \sim \chi^2_{df}$ with its degrees of freedom $df = \frac{1}{2}k(k+1) - t$, where the first term is the number of non redundant elements in *S* given *k* observed variables, and *t* is the number of free parameters in θ. The likelihood ratio chi-square test is used to test the null hypothesis that the population covariance matrix possesses that structure implied by the model against the alternative hypothesis that Σ is an arbitrary symmetric positive definite matrix.

The logic of significance testing here is different than that usual in testing, say, the statistical significance of the explained variance in a regression equation. In regression, the null hypothesis is set such that it runs counter to the theoretical hypothesis, and our hope is to reject the null hypothesis. In contrast, for the chi-square test of structural equation modeling the null hypothesis is that the constraints on Σ implied by the model are valid, i.e. $\Sigma = \Sigma(\theta)$. The standard of comparison is the perfect fit of $\hat{\Sigma}$ equal to S. The probability level of the calculated chi-square is the probability of obtaining a chi-square value larger than the value obtained if H_0 is correct (Bollen 1989). The higher the probability of the chi-square, the closer is the fit of H_0 to the perfect fit. Therefore, it is a type of goodness of fit, and a good fit is indicated when we cannot reject H_0 at a significant level α.

Structural equation models which provide a good account of the observed data will be associated with small χ^2-values with correspondingly large probabilities (P >0.1), low value of root mean square residual (RMR), low values for the root mean square error of

approximation (RMSEA), and high goodness of fit index (GFI) and adjusted goodness of fit index (AGFI) (GFI, AGFI > 0.95) under the null hypothesis.

On the other hand, in the context of stacked models, the χ^2 statistic difference test can be used to compare two models (e.g. data from different groups or different conditions) in the context of structural equation modeling (Grafton et al., 1994). A so-called 'null model' is constructed where the estimates of some parameters (i.e. path coefficients) are constrained to be zero or equal under two conditions. The alternative model allows these parameters to differ between two conditions, and therefore they are free to be estimated. The difference in the χ^2 goodness of fit indicator for the null model and the alternative model is calculated. This χ^2 statistic has *n* degrees of freedom, where *n* is the difference in the degrees of freedom between the null model and the alternative model. For example, if the null model constrains one parameter to be equal between groups or conditions, the resulting degrees of freedom for the χ^2 statistic would be one. If χ^2 statistic is not significant, we conclude that the constrained-equal model is the same as the unconstrained alternative model, leading to the conclusion that the model does apply across different conditions.

Chapter 3

Visual Attention fMRI Data and Unified Structural Equation Modeling

3.1 Subjects

Twenty-eight volunteers (14 females and 14 males) who had no psychiatric or neurological history participated in our study. A visual attention experiment with a "three-ball tracking" task (Figure 3.1) （Lange 1999） was conducted on a 4 T Varian MR System at the Brookhaven National Laboratory (BNL) for each subject. This study was approved by the Medical Research Center at BNL and all subjects provided verbal and written consent.

3.2 Experimental Design

A series of visual attention paradigms with variable attentional load was programmed in Matlab (Mathworks Inc., Natick, MA) and presented through MRI-compatible goggles that were mounted on the head coil. A cross hair was displayed

initially for 10 sec to focus the attention before the actual task was presented.

At the beginning of each trial, subjects first saw a message for 1.25 sec indicating whether their task would be active tracking ("TRACK") or passive viewing ("DO NOT TRACK"). Next, 10 copper-colored balls appeared at random positions on the screen, along with a central fixation cross. Subjects were asked to fixate throughout the entire trial. At the beginning of each active tracking trial, orange squares appeared for 1.5 sec around three balls that the subject was asked to track. On passive-viewing trials, the balls simply remained motionless for this 1.5 sec period. After this cue period, the balls moved in random directions. When a ball approached another ball or the edge of the screen, it changed direction to avoid collision or overlap. After 7.75 sec, the balls stopped moving and three balls, which were chosen at random with equal probability to have been a target or non-target, were highlighted for 1 sec. Subjects were asked to touch a button with their dominant hand (thumb) only if the balls were identical to those that they were tracking; their responses therefore provided an objective measure of tracking performance, with 50% being chance. The active tracking trial continued after a delay of 0.5 sec, when the balls were highlighted again for 1 sec for the next tracking session. The sequence of events was identical in the non-tracking trials, however, no balls were highlighted, and the subjects were instructed not to track the balls, but to view them passively.

Each active trial lasted for 60 sec, consisting of a total of five different active tracking modules within this period. Each passive tracking trial also lasted for 60 sec. The three-ball tracking tasks consisted of three blocks of active tracking alternated with passive tracking.

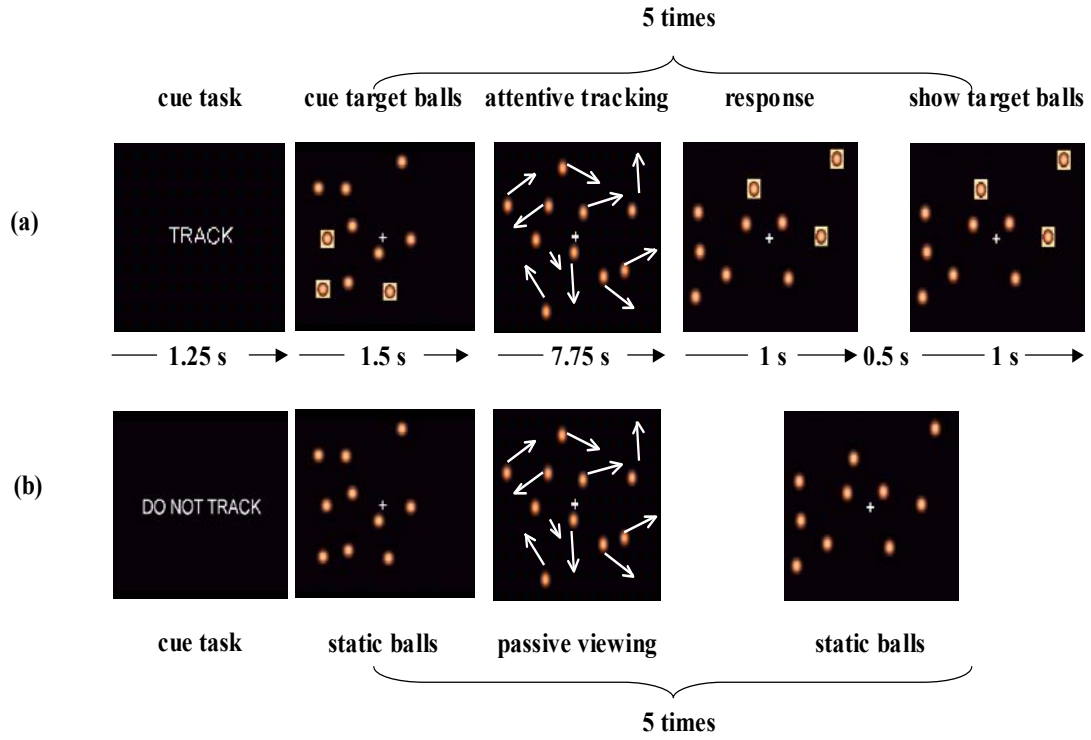


Figure 3.1: A schematic diagram of the visual stimulus used in (a) active tracking and (b) passive viewing trials. Each trial began with a text cue indicating the type of trial. This was followed by a period of static balls (1.5 s), in which the target balls were highlighted with orange squares on active trials. These highlights then disappeared and the balls moved in random directions on the screen without overlapping. After 7.75 s, the balls stopped moving and were highlighted for 1s only on active tracking trials, and subjects indicated (using a response button) whether the highlighted balls were among the balls that they had been tracking. Following this response, and after a delay of 0.5 s, the correct balls were re-highlighted for 1 s to provide feedback to the subjects on the correctness of their response.

3.3 Data Processing

Preprocessing of fMRI time series were performed in SPM99 (Statistical Parametric Mapping software, http://www.fil.ion.ucl.ac.uk/spm) and involved motion correction, spatial normalization to the Talairach frame, and spatial smoothing. (Kim et al., 2006)

The following six regions were identified, based on the consideration of the prior literature and the fact that they had strong activation during the ball-tracking task: cerebellum (CEREB), posterior parietal cortex (PPC, BA 40), anterior parietal cortex (APC, BA 7), thalamus (THAL), supplementary motor area (SMA, BA 8), and lateral prefrontal cortex (LPFC BA 6, 9, 46) (Büchel and Friston, 1997; Friston and Büchel, 2000; Chang et al., 2004). Functional MRI time series were extracted from each individual data set at coordinates corresponding to all generically activated voxels in each of these six regions as the text format.

The segments of each regional time series corresponding to presentation of the activation conditions were then extracted (Honey et al, 2002). To do this, we allowed a mean hemodynamic delay of 6 sec, i.e., two TR periods, at the beginning of each onset condition. Therefore, the segments of signal corresponding to the presentation of each of the three activation conditions without the first two time points (6 sec.) which were truncated by correction for hemodynamic delay were concatenated, resulting in T=54 time points for each subject in each region.

3.4 Unified Structural Equation Modeling

We analyze the fMRI multivariate time series data via the unified SEM approach presented by Kim et al. (Kim et al., 2006). This approach includes longitudinal as well as contemporaneous relations. Longitudinal temporal relations reflect relationships between brain regions involving different time points, and are represented in the form of a

multivariate autoregressive model (MAR). In contrast, contemporaneous relations are those relationships between brain regions at the same time points, and involve conventional SEMs.

Let $y_j(t)$ be the j^{th} variable measured at time t, $j=1,2,...,m$. The m-dimensional multivariate autoregressive process of order p (MAR(p)) with contemporaneous relations is written as:

$$\begin{aligned} y(t) &= \mathrm{B}\bullet y(t) + \Gamma(1)\bullet y(t-1) + \cdots + \Gamma(p)\bullet y(t-p) + \varepsilon(t) \\ &= \mathrm{B}\bullet y(t) + \sum_{i=1}^{p} \Gamma(i)\bullet y(t-i) + \varepsilon(t), \end{aligned} \tag{3.1}$$

where $y(t) = [y_1(t)\ y_2(t) \cdots y_m(t)]'$ is an $m \times 1$ vector of observed variables at time t, $\varepsilon(t) = [\varepsilon_1(t)\ \varepsilon_2(t) \cdots \varepsilon_m(t)]'$ is an $m \times 1$ vector of white noise with zero mean and covariance Ψ_ε, B is the parameter matrix of the contemporaneous relations, and $\Gamma(i)$, $i=1,...,p$, is a series of $m \times m$ parameter matrices representing the longitudinal relations. The diagonal elements of $\Gamma(i)$ represent the coefficients of the autoregressive process for each variable, while the off-diagonal elements represent the coefficients of the lagged relations between different variables. The parameters contained in matrices B and $\Gamma(i)$ can be free, constrained or fixed. These parameters are set by the initial path model with predefined path according to prior study.

Let $\Gamma = [\Gamma(1)\ \Gamma(2) \cdots \Gamma(p)]$ be a $m \times (m \times p)$ matrix and $x = [y(t-1)\ y(t-2) \cdots y(t-p)]'$ be a $(m \times p)$ vector. If we denote θ as the set of free parameters contained in $B, \Gamma, \Psi_\varepsilon, and\ \Phi$, the variance-covariance structure of x, the implied covariance matrix of $y(t)$ and x can be written as:

$$\Sigma(\theta) = \begin{pmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi_{\varepsilon})(I-B)^{-1\prime} & (I-B)^{-1}\Gamma\Phi \\ \Phi\Gamma'(I-B)^{-1\prime} & \Phi \end{pmatrix} \quad (3.2)$$

Therefore, the parameters of MAR-SEM can be estimated through conventional structural equation modeling procedures.

The initial path model is defined with six ROIs and seven anatomically possible directional paths for the left brain hemisphere. The posterior parietal cortex (PPC) serves as the starting region of visual attention processing in the path model, and information flows via the anterior parietal cortex (APC) to the lateral prefrontal cortex (LPFC). An attentional feedback loop starts at the supplementary motor area (SMA), with input from the LPFC, and extends through the thalamus (THAL), back to the PPC. The THAL acts a subcortical relay station, and receives additional input from the cerebellum (CEREB) (See Figure 3.2). Our model is restricted to the left hemisphere to simplify the brain network. The path model of our study incorporates the conventional contemporaneous relations as well as the longitudinal relations. According to the previous research of Kim et al. (2006), the longitudinal relations were depicted by the first-order multivariate autoregressive process (MAR(1)). Although the order of MAR for each ROI obtained from the partial autocorrelation function (PACF) analysis was not always 1, due to estimability constrains (Honey et al., 2002), the MAR of order 1 for all ROIs was chosen, which produced thirteen possible longitudinal directional paths for the left brain hemisphere. The path diagram of the unified longitudinal and contemporaneous path model is described in Figure 3.3. And Equation 3.3 is the matrix form of the unified SEM with its contemporaneous and longitudinal components.

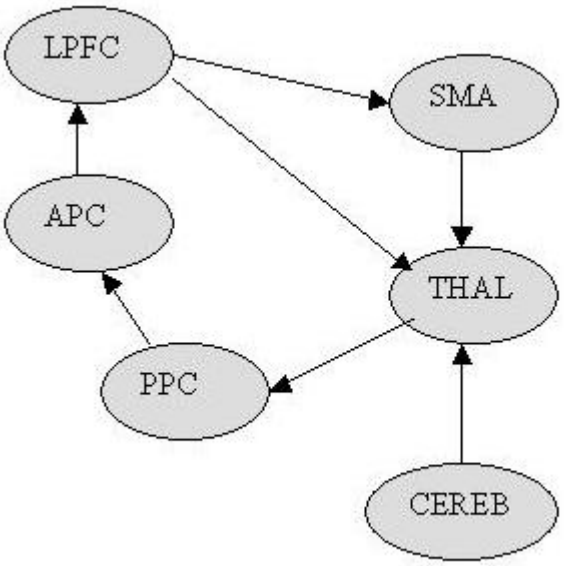


Figure 3.2: Path diagram of the theoretical contemporaneous path model with six ROIs: cerebellum (CEREB), posterior parietal cortex (PPC), anterior parietal cortex (APC), thalamus (THAL), supplementary motor area (SMA), and lateral prefrontal cortex (LPFC), and seven possible directional paths in the left hemisphere.

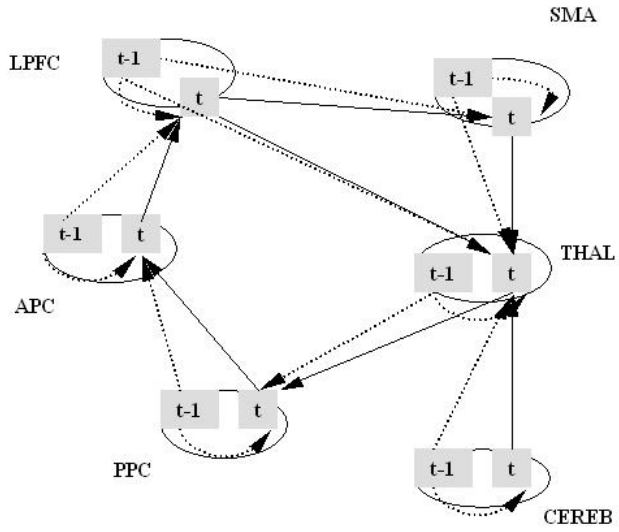


Figure 3.3: Path diagram of unified contemporaneous and longitudinal path model with six brain ROIs, seven possible directional paths (solid lines), and thirteen possible longitudinal paths (dashed lines) in the left hemisphere. The longitudinal paths are from one region at the previous time (t-1) to other regions as well as itself at the current time (t).

$$\begin{pmatrix} CEREB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{21} & 0 & 0 & 0 & \beta_{25} & \beta_{26} \\ 0 & \beta_{32} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{43} & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_{54} & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{65} & 0 \end{pmatrix} \bullet \begin{pmatrix} CEREB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix}$$
$$+ \begin{pmatrix} \gamma_{11} & 0 & 0 & 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & 0 & 0 & \gamma_{25} & \gamma_{26} \\ 0 & \gamma_{32} & \gamma_{33} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{43} & \gamma_{44} & 0 & 0 \\ 0 & 0 & 0 & \gamma_{54} & \gamma_{55} & 0 \\ 0 & 0 & 0 & 0 & \gamma_{65} & \gamma_{66} \end{pmatrix} \bullet \begin{pmatrix} CEREB(t-1) \\ THAL(t-1) \\ PPC(t-1) \\ APC(t-1) \\ LPFC(t-1) \\ SMA(t-1) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \\ \varepsilon_4(t) \\ \varepsilon_5(t) \\ \varepsilon_6(t) \end{pmatrix} \quad (3.3)$$

or

$$y(t) = B \bullet y(t) + \Gamma(1) \bullet y(t-1) + \varepsilon(t) \quad (3.4)$$

In Equation 3.4, $y(t)$ is a vector of the observed fMRI data of six ROIs at time *t*, *B* and $\Gamma(1)$ are contemporaneous and longitudinal parameter matrices, and $\varepsilon(t) = [\varepsilon_1(t)\ \varepsilon_2(t)\ \varepsilon_3(t)\ \varepsilon_4(t)\ \varepsilon_5(t)\ \varepsilon_6(t)]'$ is a vector of errors.

Equation 3.3 or 3.4 indicates that the value of one brain region at time *t* is influenced by the values of other regions at the same time and the values of the previous time *t-1* of itself and of other regions.

Chapter 4

Subject-Average Unified Structural Equation Modeling Approach and Two-Stage Multi-Subject Unified Structural Equation Modeling Approach

The typical fMRI data involves a number of subjects and each subject's functional activities of multiple brain regions of interest (ROIs) in the brain are measured longitudinally over the course of several minutes. Thus, fMRI studies usually contain multi-subject, multivariate time series data. Furthermore, it is of interest to incorporate the subject-level covariates into the analysis to examine the relationships between functional brain pathways and subject-level covariates. For example, twenty-eight subjects (14 females and 14 males) participated in our visual attention fMRI study. And the corresponding fMRI data has 54 time points for each subject in each of six regions. The subject-level covariates include age, gender, verbal IQ (VIQ), and education.

Three approaches are conceivable for the path analysis of multi-subject, multivariate time series data from fMRI experiments. They are: (1) summarize (e.g. average the time series data across the subjects) and then analyze, (2) analyze and then summarize, and (3) simultaneous analysis. In this chapter we present the first two approaches, which are integrated with the unified SEM method described in Chapter 3, to analyze the

multi-subject, multivariate time series fMRI data. And we show the application of the two approaches to our visual attention fMRI data. The appropriateness of these two approaches and comparisons between the two approaches are stated in Chapter 6. In the framework of unified SEM, the two approaches are:

(1) Approach 1: summarize (e.g. average the time series data across the subjects) and then analyze, which is the subject-average unified structural equation modeling approach, and

(2) Approach 2: analyze and then summarize, which is the two-stage multi-subject unified structural equation modeling approach.

4.1 Subject-Average Unified Structural Equation Modeling Approach

4.1.1 Method and Application

Kim (2004) proposed to take the subject-average of time-series fMRI data, producing a single sequence of observations for each ROI (which was also adopted by Hoge, 1998), and then analyze the subject-average fMRI data via SEM model. However, since the reduced average observations are time-series data, they are not independent. In order to analyze the dependent fMRI data via conventional SEM software, the unified SEM described in Chapter 3 is used to analyze subject-average multivariate fMRI data, and therefore this approach is called subject-average unified structural equation modeling

approach.

In the subject-average unified SEM approach, we take the average time series data across all the subjects (summarize first), and analyze the subject-average fMRI data with the unified structural equation modeling. The unified SEM model with its contemporaneous SEM components, its MAR(p) longitudinal components and error variances, is fitted for the subject-average fMRI data.

In our visual attention fMRI data, the unified longitudinal and contemporaneous path model (See Section 3.4) to be tested is:

$$\begin{pmatrix} CEREB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{21} & 0 & 0 & 0 & \beta_{25} & \beta_{26} \\ 0 & \beta_{32} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{43} & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_{54} & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{65} & 0 \end{pmatrix} \bullet \begin{pmatrix} CEREB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix}$$
$$+ \begin{pmatrix} \gamma_{11} & 0 & 0 & 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & 0 & 0 & \gamma_{25} & \gamma_{26} \\ 0 & \gamma_{32} & \gamma_{33} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{43} & \gamma_{44} & 0 & 0 \\ 0 & 0 & 0 & \gamma_{54} & \gamma_{55} & 0 \\ 0 & 0 & 0 & 0 & \gamma_{65} & \gamma_{66} \end{pmatrix} \bullet \begin{pmatrix} CEREB(t-1) \\ THAL(t-1) \\ PPC(t-1) \\ APC(t-1) \\ LPFC(t-1) \\ SMA(t-1) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \\ \varepsilon_4(t) \\ \varepsilon_5(t) \\ \varepsilon_6(t) \end{pmatrix}. \quad (4.1)$$

We take the average time series data across all the twenty-eight subjects, and analyze the subject-average visual attention fMRI data with the unified structural equation modeling. The unified SEM model in Equation (4.1) with its seven contemporaneous paths, its thirteen MAR(1) longitudinal paths and error variances, is fitted for the subject-average fMRI data using the SAS PROC CALIS.

The estimated longitudinal and contemporaneous path parameters with their

standard errors, t test statistics, and corresponding p values are tabulated in Table 4.1. The bold characters indicate the significant paths at the significance level of 0.05. The path model with the significant paths is displayed in Figure 4.1. The significant longitudinal path connections contain (1) paths connecting THAL to PPC, and PPC to APC, and (2) the longitudinal paths connecting each region to itself. Two contemporaneous paths, which are from THAL to PPC and from LPFC to SMA, are significant. The goodness-of-fit statistics of the model is χ^2=86.134 (d.f.=31, p value<0.01) indicating a poor model fit. The same analysis was repeated using LISREL, with identical results as those from SAS.

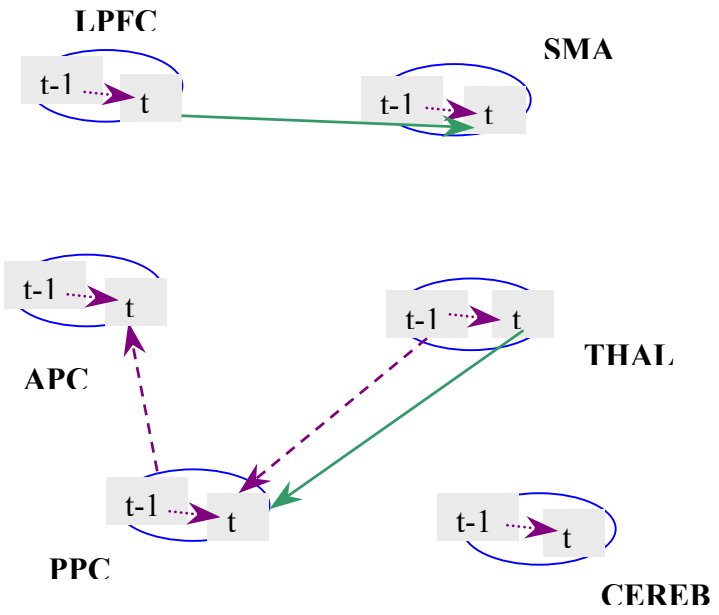


Figure 4.1 Significant path connections from the subject-average unified SEM approach. This path network contains eight significant longitudinal path connections (dashed lines) and two significant contemporaneous paths (solid lines) in the left hemisphere.

Table 4.1 Estimated longitudinal and contemporaneous path parameters from the subject-average unified SEM approach with their standard errors, t test statistics, and corresponding p values (two-sided).

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Est. (S.E.)	T-value (p-value)	Path parameters	Est. (S.E.)	T-value (p-value)
γ_{11}	**0.780 (.091)**	**8.593 (.000)**	β_{21}	-0.058 (.250)	-0.232 (.816)
γ_{21}	-0.317 (.274)	-1.158 (.247)	β_{25}	0.394 (.217)	1.815 (.070)
γ_{22}	**0.538 (.111)**	**4.842 (.000)**	β_{26}	0.270 (.182)	1.482 (.138)
γ_{25}	-0.024 (.221)	-0.110 (.912)	β_{32}	**-0.258 (.057)**	**-4.509 (.000)**
γ_{26}	0.023 (.198)	0.116 (.901)	β_{43}	-0.141 (.126)	-1.122 (.262)
γ_{32}	**0.237 (.061)**	**3.908 (.000)**	β_{54}	0.172 (.179)	0.959 (.338)
γ_{33}	**0.386 (.105)**	**3.696 (.000)**	β_{65}	**0.377 (.156)**	**2.422 (.015)**
γ_{43}	**0.246 (.110)**	**2.233 (.026)**			
γ_{44}	**0.566 (.111)**	**5.116 (.000)**			
γ_{54}	–0.284 (.178)	-1.600 (.110)			
γ_{55}	**0.582 (.113)**	**5.164 (.000)**			
γ_{65}	-0.282 (.160)	-1.757 (.079)			
γ_{66}	**0.590 (.107)**	**5.518 (.000)**			

Bold characters indicate the significant path parameters at the significance level of 0.05.

4.1.2 Gender Effect and Two-level Nonparametric Bootstrap

Because in our visual attention fMRI data, we have two gender groups, females and males, we can test the gender effect on each path coefficient, i.e., to test if the paths are significantly different between the two groups. For the subject-average unified structural

equation modeling approach, we need to take the subject average of fMRI data across all subjects in each gender group, and therefore we obtain two unified SEM models, one for females and one for males. By fitting the unified SEM model for the two groups, we have two separate estimations for each path *i*: one is from the subject-average unified SEM across female subjects, denoted as $path^{(i)}_{female}$, and one is from subject-average unified SEM across male subjects, denoted as $path^{(i)}_{male}$. The estimation of difference between females and males for each path could be computed as

$$Difference\ of\ \ path^{(i)} = path^{(i)}_{female} - path^{(i)}_{male}. \tag{4.2}$$

The standard two-sample t-test or nonparametric test is not appropriate for this data, since we only have one subject-average observation for females and males. In order to obtain the standard error of the estimated difference between two gender groups and perform hypothesis test, we propose to use a two-level nonparametric bootstrap method. Meijer et al. (1995) has illustrated implementing the two-level bootstrap for multilevel models. In this work, we implement the two-level bootstrap for SEM model with subject-level covariate, which is a new development. And we illustrate the bootstrap procedures through our visual attention fMRI data.

In the structure of fMRI data, we have two-level samples; level-one samples are the fMRI time points within a subject, and level-two samples are independent subjects. Therefore, in order to retain the fMRI data structure, we do two-level rasampling to obtain bootstrap samples.

We use the nonparametric bootstrap, whose bootstrap samples can be drawn by resampling complete cases. The bootstrap samples can be drawn in the following way.

We assume there are N_f and N_m subjects in the female group and male group, respectively. First, a sample of size N_f is drawn with replacement from the female group and a sample of size N_m is drawn with replacement from the male group. This gives a sample j^*_{fk} of female subjects and a sample j^*_{mk} of male subjects, where $fk = 1,...N_f$, $mk = 1,...N_m$. Then for each *fk* or *mk*, a nonparametric bootstrap sample of complete cases (including longitudinal and contemporaneous components) from the original subject $j = j^*_{fk}$ or $j = j^*_{mk}$ is drawn. We call this the complete cases bootstrap for both levels.

The estimator of gender group difference for each path $\hat{\mu} = path_{female} - path_{male}$ is computed from the original sample. For each bootstrap sample *b=1,...,B* (obtained in the way described above), a bootstrap estimator $\mu^{*(b)}$ is obtained in the same way the estimator $\hat{\mu}$ was obtained from the original sample, i.e., $\mu^{*(b)} = path_f^{*(b)} - path_m^{*(b)}$, where $path_f^{*(b)}$ is the bootstrap path estimator from subject-average unified SEM approach for female group and $path_m^{*(b)}$ is the bootstrap path estimator from subject-average unified SEM approach for male group. The variance of $path_f^{*(b)}$ and $path_m^{*(b)}$ are V_f^* and V_m^*, respectively. Therefore, the standard error of $\mu^{*(b)}$ is $\sqrt{V_f^* + V_m^*}$ (Efron and Tibshirani, 1993).

We have 14 females and 14 males in our visual attention fMRI data, and we perform the two-level bootstrap with *B*=1000 replicates to calculate the standard errors. The results are summarized in Table 4.2.

Three paths are significantly different between females and males from Approach 1,

which are the longitudinal and contemporaneous paths from LPFC to SMA, and the longitudinal path connecting LPFC and THAL.

Table 4.2 Gender comparisons of the estimated longitudinal and contemporaneous path parameters between two gender groups from the subject-average unified SEM approach with their bootstrap standard errors, z values, and corresponding p values (two-sided).

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Est. (S.E.)	Z-value (p-value)	Path parameters	Est. (S.E.)	Z-value (p-value)
γ_{11}	0.163 (.169)	0.965 (.334)	β_{21}	0.088 (.367)	0.241 (.809)
γ_{21}	-0.039 (.374)	-0.104 (.917)	β_{25}	0.108 (.338)	0.319 (.749)
γ_{22}	0.029 (.168)	0.173 (.863)	β_{26}	-0.056 (.354)	-0.159 (.874)
γ_{25}	**-0.879 (.336)**	**-2.620 (.009)**	β_{32}	-0.092 (.136)	-0.678 (.497)
γ_{26}	-0.115 (.341)	-0.338 (.736)	β_{43}	-0.277 (.193)	-1.432 (.152)
γ_{32}	-0.033 (.138)	-0.240 (.810)	β_{54}	-0.102 (.330)	-0.310 (.756)
γ_{33}	-0.275 (.170)	-1.617 (.106)	β_{65}	**0.440 (.210)**	**2.096 (.036)**
γ_{43}	0.097 (.173)	0.562 (.574)			
γ_{44}	-0.042 (.166)	-0.251 (.802)			
γ_{54}	–0.112 (.329)	-0.340 (.734)			
γ_{55}	0.126 (.169)	0.743 (.457)			
γ_{65}	**-0.572 (.222)**	**-2.573 (.010)**			
γ_{66}	0.149 (.173)	0.864 (.387)			

Bold characters indicate the path parameters significantly different between two gender groups at the significance level of 0.05.

4.2 Two-Stage Multi-Subject Unified Structural Equation Modeling Approach (Kim et al. 2006)

4.2.1 Methods and Application

Kim et al. (2006) developed a two-stage multi-subject unified structural equation modeling approach to analyze multi-subject, multivariate time series fMRI data with subject-level covariates. We adopt their method in this work.

In stead of taking the average of time series data across all the subject, in this approach, the unified SEM model with its contemporaneous SEM components, its MAR(p) longitudinal components and error variances, is fitted for each of the 28 subject time courses individually at the first stage. And we take the average of the estimated path parameters across all the 28 subjects for each possible path to test if the mean of the parameters for each path is significantly different from zero using one-sample t-test (two-sided) at the second stage. So it is called the two-stage multi-subject unified SEM approach.

For our visual attention fMRI data, the unified SEM model presented in Equation (4.1) with its contemporaneous SEM components of six ROIs and seven paths, its MAR(1) longitudinal components and thirteen paths, and error variances, is fitted for each of the 28 subject time course individually using the SAS PROC CALIS. Table 4.3 presents the mean values of the estimated path coefficients with their standard errors across all 28 subjects. The t-test statistics and the corresponding p values provide the significant longitudinal and contemporaneous paths at the significance level of 0.05. The path model with the significant paths is displayed in Figure 4.2. The goodness-of-fit

statistics χ^2 is large (small p value) for fitted SEM model of each subject, indicating a poor model fit.

The significant longitudinal path connections contain (1) paths connecting APC to LPFC, and LPFC to SMA, and (2) the longitudinal paths connecting each region to itself. The only significant contemporaneous path is from LPFC to SMA. The same analysis was repeated using LISREL, with identical results as those from SAS.

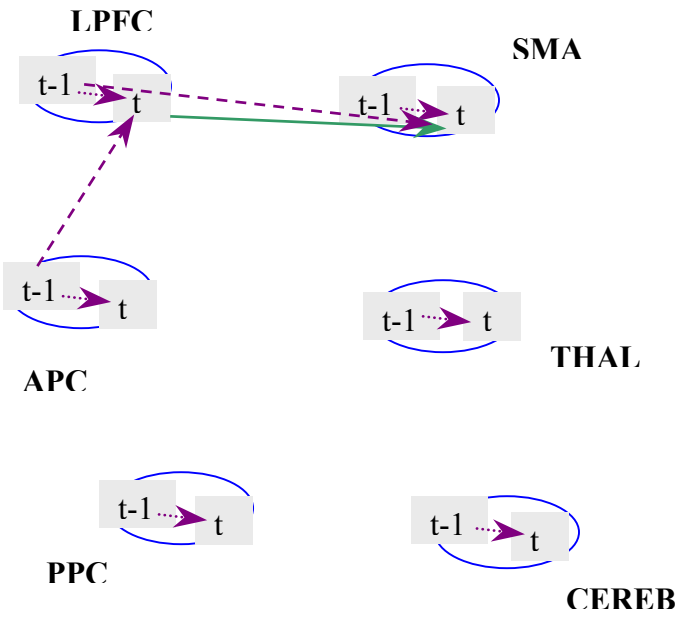


Figure 4.2 Significant path connections from the two-stage multi-subject unified SEM approach. This path network contains eight significant longitudinal path connections (dashed lines) and one significant contemporaneous path (solid line) in the left hemisphere.

Table 4.3 Mean values of the estimated longitudinal and contemporaneous path parameters across 28 subjects from the two-stage multi-subject unified SEM approach with their standard errors, t test statistics, and corresponding p values (two-sided).

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Mean (S.E.)	T-value (p-value)	Path parameters	Mean (S.E.)	T-value (p-value)
γ_{11}	**0.612 (.019)**	**31.462 (.000)**	β_{21}	0.047 (.076)	0.622 (.539)
γ_{21}	0.005 (.070)	0.074 (.942)	β_{25}	0.101 (.071)	1.424 (.166)
γ_{22}	**0.591 (.020)**	**29.391 (.000)**	β_{26}	0.030 (.087)	0.340 (.736)
γ_{25}	-0.120 (.070)	-1.718 (.097)	β_{32}	-0.034 (.026)	-1.311 (.201)
γ_{26}	0.028 (.082)	0.339 (.737)	β_{43}	-0.005 (.036)	-0.126 (.901)
γ_{32}	0.035 (.029)	1.182 (.248)	β_{54}	0.091 (.066)	1.384 (.178)
γ_{33}	**0.595 (.021)**	**28.821 (.000)**	β_{65}	**0.229 (.054)**	**4.208 (.000)**
γ_{43}	0.031 (.036)	0.872 (.391)			
γ_{44}	**0.593 (.016)**	**36.001 (.000)**			
γ_{54}	**−0.124 (.046)**	**-2.683 (.012)**			
γ_{55}	**0.615 (.019)**	**31.685 (.000)**			
γ_{65}	**-0.166 (.047)**	**-3.517 (.002)**			
γ_{66}	**0.578 (.018)**	**31.140 (.000)**			

Bold characters indicate the significant path parameters at the significance level of 0.05.

4.2.2 Gender Effect and Two-level Nonparametric Bootstrap

By using the two-stage multi-subject unified structural equation modeling approach, we can test if the paths are significantly different between two gender groups simply by two-sample t-test. In stage 2 the subject-level path coefficients obtained form the unified SEM analysis of stage 1 are merged with the subject-level covariate, gender, to examine

the impact of the gender on the brain pathways via a statistical t-test. Here we illustrate the stage 2 analysis using the fMRI visual attention data.

Two-sample t-test is performed for each path connection individually if we assume the path connections are normally distributed in each gender group. The corresponding analysis results (two-sample t-test and their p-values) are tabulated in Table 4.4. At the significance level 0.05, three out of 13 longitudinal and 7 contemporaneous paths in the model are significantly influenced by gender. The comparisons of Approach 1 and Approach 2 of detecting gender effects are shown in Chapter 6.

Table 4.4 Gender comparisons of the estimated longitudinal and contemporaneous path parameters between two gender groups from the two-stage multi-subject unified SEM approach with their standard errors, t test statistics, and corresponding p values (two-sided).

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Est. (S.E.)	T-value (p-value)	Path parameters	Est. (S.E.)	T-value (p-value)
γ_{11}	0.004 (.040)	0.096 (.925)	β_{21}	-0.029 (.154)	-0.191 (.850)
γ_{21}	0.079 (.141)	0.561 (.580)	β_{25}	0.044 (.144)	0.306 (.762)
γ_{22}	**0.082 (.038)**	**2.185 (.038)**	β_{26}	0.022 (.177)	0.124 (.903)
γ_{25}	-0.083 (.142)	-0.587 (.562)	β_{32}	-0.018 (.052)	-0.346 (.732)
γ_{26}	-0.099 (.165)	-0.599 (.554)	β_{43}	-0.120 (.071)	-1.697 (.102)
γ_{32}	-0.022 (.060)	-0.375 (.711)	β_{54}	0.102 (.132)	0.771 (.448)
γ_{33}	**-0.085 (.039)**	**-2.210 (.036)**	β_{65}	**0.304 (.093)**	**3.255 (.003)**
γ_{43}	0.063 (.071)	0.882 (.386)			
γ_{44}	0.060 (.031)	1.909 (.067)			
γ_{54}	–0.172 (.088)	-1.943 (.063)			
γ_{55}	-0.064 (.037)	-1.722 (.097)			
γ_{65}	-0.158 (.091)	-1.727 (.096)			
γ_{66}	-0.011 (.038)	-0.296 (.770)			

Bold characters indicate the path parameters significantly different between two gender groups at the significance level of 0.05.

It is worth noting that the two-sample t-test is based on the assumption that the path coefficients are normally distributed in each gender group. However, in real life, this assumption could be violated, especially when the sample size of subjects in each gender group is small. The nonparametric bootstrap method described in Section 4.1.2 could be used to estimate the standard error of the estimated difference between two gender groups for each path connection individually. We illustrate the bootstrap procedures through our visual attention fMRI data.

We use the nonparametric bootstrap, whose bootstrap samples can be drawn by resampling complete cases. The bootstrap samples can be drawn in the following way. We assume there are N_f and N_m subjects in the female group and male group, respectively. First, a sample of size N_f is drawn with replacement from the female group and a sample of size N_m is drawn with replacement from the male group. This gives a samples j^*_{fk} of female subjects and a sample j^*_{mk} of male subjects, where $fk = 1,...N_f$, $mk = 1,...N_m$. Then for each *fk* or *mk*, a nonparametric bootstrap sample of complete cases (including longitudinal and contemporaneous components) from the original subject $j = j^*_{fk}$ or $j = j^*_{mk}$ is drawn. We call this the complete cases bootstrap for both levels.

The estimator of gender group difference for each path $\hat{\mu} = \hat{\mu}_f - \hat{\mu}_m$ is computed from the original sample. For each bootstrap sample *b=1,...,B* (obtained in the way described above), a bootstrap estimator $\mu^{*(b)}$ is obtained in the same way the estimator $\hat{\mu}$ was obtained from the original sample, i.e., $\mu^{*(b)} = \mu_f^{*(b)} - \mu_m^{*(b)}$, where $\mu_f^{*(b)}$ is the

bootstrap estimator of mean for female group and $\mu_m^{*(b)}$ is the bootstrap estimator of mean for male group. The variance of $\mu_f^{*(b)}$ and $\mu_m^{*(b)}$ are V_f^* and V_m^*, respectively. Therefore, for the independent two-sample test, the standard error of $\mu^{*(b)}$ is $\sqrt{V_f^* + V_m^*}$ (Efron and Tibshirani, 1993).

Table 4.5 Gender comparisons of the estimated longitudinal and contemporaneous path parameters between two gender groups with bootstrap standard errors.

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Est. (S.E.)	Z-value (p-value)	Path parameters	Est. (S.E.)	Z-value (p-value)
γ_{11}	0.004 (.058)	0.066 (.948)	β_{21}	-0.029 (.190)	-0.155 (.877)
γ_{21}	0.079 (.184)	0.430 (.667)	β_{25}	0.044 (.170)	0.259 (.796)
γ_{22}	0.082 (.059)	1.384 (.166)	β_{26}	0.022 (.200)	0.109 (.913)
γ_{25}	-0.083 (.171)	-0.489 (.635)	β_{32}	-0.018 (.062)	-0.294 (.769)
γ_{26}	-0.099 (.190)	-0.521 (.602)	β_{43}	-0.120 (.088)	-1.357 (.175)
γ_{32}	-0.022 (.066)	-0.338 (.735)	β_{54}	0.102 (.157)	0.649 (.516)
γ_{33}	-0.085 (.057)	-1.492 (.136)	β_{65}	**0.304 (.110)**	**2.759 (.006)**
γ_{43}	0.063 (.087)	0.726 (.468)			
γ_{44}	0.060 (.056)	1.081 (.280)			
γ_{54}	–0.172 (.128)	-1.338 (.181)			
γ_{55}	-0.064 (.056)	-1.143 (.253)			
γ_{65}	-0.158 (.106)	-1.485 (.138)			
γ_{66}	-0.011 (.059)	-0.189 (.850)			

Bold characters indicate the path parameters that are significantly different between two gender groups at the significance level of 0.05.

We have 14 females and 14 males with our visual attention fMRI data, and

coefficient estimation from two-stage unified SEM model of each path within the same gender group barely follows the normal distribution, except for path γ_{43}. Hence, the t-test is appropriate for the visual attention fMRI data. However, in order to illustrate the application of two-level nonparametric bootstrap method, we perform the bootstrap with B=1000 replicates to our data and the results are listed in Table 4.5. Only the contemporaneous path from LPFC to SMA is significantly different between females and males. The nonparametric bootstrap method gives conservative conclusions due to larger standard errors.

4.2.3 Generalized Linear Model of Other Covariates

Similar to the analysis in Section 4.2.2, in Stage 2 of the two-stage multi-subject unified structural equation modeling approach, the subject-level path coefficients obtained form the unified SEM analysis of stage 1 could be merged with other subject-level covariates to examine the impact of these covariates on the brain pathways via a GLM. Here we illustrate the stage 2 analysis using the fMRI visual attention data.

We examine the impact of three other covariates---age, verbal IQ (VIQ), and education on the visual attention pathway. Three covariates of 28 subjects are fitted on a GLM for each path connection individually. The analysis results (F-statistics and corresponding p values) are tabulated in Table 4.6. Two out of 13 longitudinal and 7 contemporaneous paths are significantly influenced by age, and one path is influenced by

VIQ. No connectivity in this visual attention network, however, is significantly correlated with education.

Table 4.6. F-test statistics and the corresponding p-values of three subject-level covariates from the general linear model (GLM) analysis. Bold characters indicate the paths significantly influenced by the corresponding covariates at the significance level of 0.05.

Paths	Age	VIQ	Education
β_{21}	0.078 (.782)	1.846 (.186)	0.623 (.437)
β_{25}	0.069 (.795)	1.394 (.248)	0.005(.947)
β_{26}	3.184 (.086)	1.844 (.186)	0.539 (.470)
β_{32}	2.419 (.132)	1.566 (.222)	0.000 (.988)
β_{43}	0.056 (.815)	0.447 (.510)	1.509 (.230)
β_{54}	**6.306 (.019)**	2.739 (.135)	1.754 (.197)
β_{65}	0.006 (.938)	1.119 (.300)	0.202 (.657)
γ_{11}	0.921 (.346)	0.001(.978)	0.015 (.904)
γ_{21}	1.135 (.297)	0.731 (.400)	0.508 (.483)
γ_{22}	**4.233 (.050)**	0.012 (.913)	0.702 (.410)
γ_{25}	0.234 (.633)	**17.335 (.000)**	3.684 (.066)
γ_{26}	0.469 (.499)	0.366 (.551)	0.661 (.424)
γ_{32}	0.039 (.845)	0.003 (.968)	0.157 (.695)
γ_{33}	0.142 (.709)	1.136 (.296)	0.030 (.863)
γ_{43}	0.395 (.535)	0.689 (.414)	0.000 (.992)
γ_{44}	0.258 (.616)	2.122 (.157)	1.901 (.180)
γ_{54}	1.285 (.267)	1.572 (.221)	0.972 (.333)
γ_{55}	0.189 (.667)	0.174 (.680)	0.156 (.696)
γ_{65}	2.024(.167)	0.623 (.437)	0.005 (.945)
γ_{66}	0.016 (.900)	0.170 (.683)	0.609 (.442)

Combining the results of Section 4.2.2 and Section 4.3 from two-stage multi-subject unified structural equation modeling approach, the paths, which are significantly influenced by all the covariates---gender, age, and VIQ, are shown in Figure 4.3.

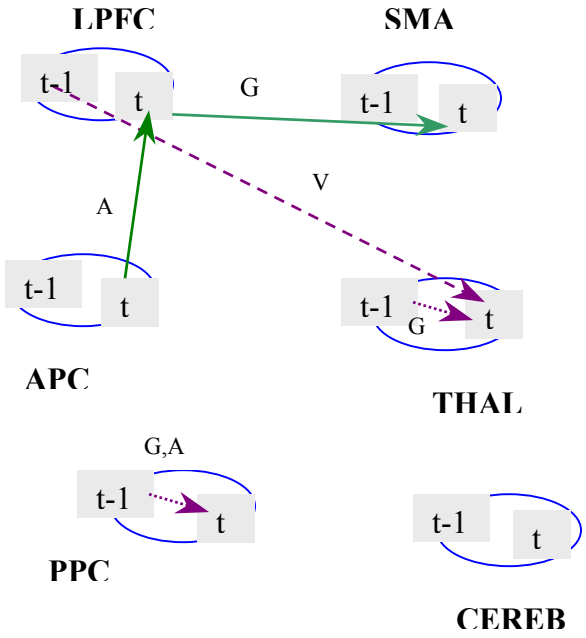


Figure 4.3 Path connections, which are significantly influenced by subject-level covariates from the two-stage multi-subject unified SEM approach. Three covariates (G, gender; A, age; and V, verbal IQ) are denoted along with the path connections. Dashed lines represent longitudinal path connections and solid lines represent contemporaneous path.

Chapter 5

Hierarchical Multi-Subject Unified Structural Equation Modeling Approaches (Simultaneous Analysis)

The two approaches described in Chapter 4 are "summarize and then analyze", and "analyze and summarize". However, we can analyze fMRI in a third way, which is "analyze and summarize simultaneously". In this chapter we present three different approaches to analyze multi-subject, multivariate time series fMRI data simultaneously. We illustrate the application of the three approaches by our visual attention fMRI data. The comparisons among the approaches of simultaneous analysis and the two approaches in Chapter 4 are shown in Chapter 6 in detail.

5.1 Single-Level Multi-Subject Unified SEM Approach

Consider the typical fMRI data, which involves a number of subjects and have multiple time points for each subject in each brain region. Instead analyzing the fMRI data by the two approaches described in Chapter 4, we can analyze all the subjects simultaneously by a single-level model. That is, we analyze a concatenated fMRI data,

which contain the time courses (observations) of all the subjects, instead of analyzing each subject individually or analyzing the average time courses across all subjects. This is a common practice in neuroscience (Penny and Holmes, 2004; Mechelli et al., 2002). However, since the concatenated fMRI data are time series data, the observations are not independent. In this work, we apply the unified SEM method to analyze the concatenated fMRI data, which is called the single-level unified SEM approach. The single-level multi-subject unified SEM approach assumes that after introducing MAR(p) longitudinal components into the fMRI data, all the fMRI observations across all the subjects were sampled independently from each other and then can be analyzed via conventional SEM model.

Our visual attention fMRI data have 54 time points (observations) for each subject and each ROI, and we have 28 subjects in this study. Therefore, the concatenated fMRI data should contain 1512 (54*28) observations of six ROIs. To apply the unified SEM approach, as a result of introducing MAR(1) longitudinal components, the multi-subject fMRI data contain 1484 (53*28) observations. The unified SEM model in Equation (4.1) with its contemporaneous SEM component of six ROIs and seven paths, its MAR(1) longitudinal components and thirteen paths and error variances, is fitted for the concatenated multi-subject fMRI data using the SAS PROC CALIS.

Table 5.1 presents the estimated longitudinal and contemporaneous path parameters with their standard errors, t test statistics, and corresponding p values. The bold characters indicate the significant paths at the significance level of 0.05. The path model with the significant paths is displayed in Figure 5.1. The significant longitudinal path

connections contain (1) paths connecting APC to LPFC, and LPFC to THAL and SMA, (2) the longitudinal paths connecting each region to itself. Five contemporaneous paths are significant, four of which form a single loop starting at THAL and connecting PPC, APC, and LPFC, and one of which is from LPFC to SMA. The goodness-of-fit statistics of the model is χ^2=122.125 (d.f.=31, p value<0.01) indicating a poor model fit. The same analysis was repeated using LISREL, with identical results as those from SAS.

Table 5.1 Estimated longitudinal and contemporaneous path parameters of the concatenating multi-subject fMRI data using the single-level unified SEM model with their standard errors, t test statistics, and corresponding p values (two-sided).

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Est. (S.E.)	T-value (p-value)	Path parameters	Est. (S.E.)	T-value (p-value)
γ_{11}	**0.609 (.020)**	**29.985 (.000)**	β_{21}	0.073 (.041)	1.781 (.075)
γ_{21}	-0.040 (.041)	-0.992 (.321)	β_{25}	**0.086 (.038)**	**2.287 (.022)**
γ_{22}	**0.630 (.020)**	**31.287 (.000)**	β_{26}	0.050 (.039)	1.293 (.196)
γ_{25}	**-0.074 (.038)**	**-1.976 (.048)**	β_{32}	**-0.040 (.016)**	**-2.528 (.012)**
γ_{26}	-0.027 (.039)	-0.689 (.491)	β_{43}	**0.040 (.020)**	**1.979 (.048)**
γ_{32}	0.024 (.016)	1.574 (.116)	β_{54}	**0.112 (.039)**	**2.882 (.004)**
γ_{33}	**0.618 (.020)**	**30.371 (.000)**	β_{65}	**0.209 (.025)**	**8.486 (.000)**
γ_{43}	0.006 (.020)	0.303 (.762)			
γ_{44}	**0.613 (.021)**	**29.661 (.000)**			
γ_{54}	**–0.133 (.039)**	**-3.448 (.000)**			
γ_{55}	**0.643 (.020)**	**32.394 (.000)**			
γ_{65}	**-0.180 (.025)**	**-7.247 (.000)**			
γ_{66}	**0.580 (.021)**	**27.354 (.000)**			

Bold characters indicate the significant path parameters at the significance level of 0.05.

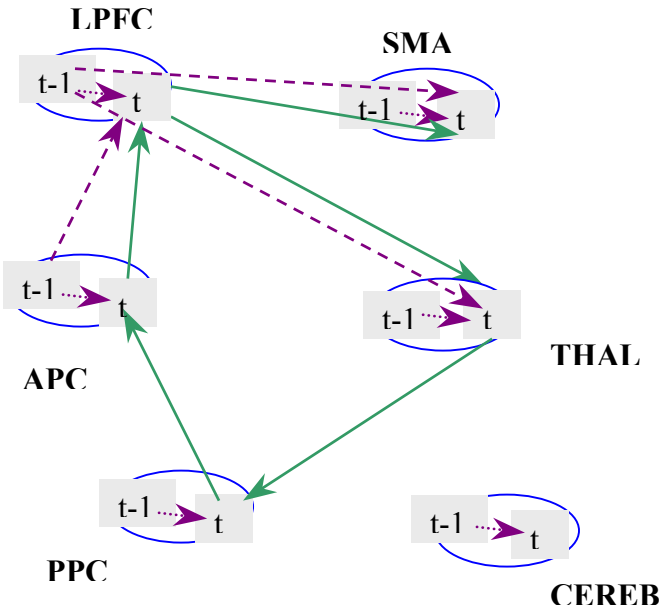


Figure 5.1 Significant path connections of the concatenating multi-subject fMRI data using the single-level unified SEM model. This path network contains nine significant longitudinal path connections (dashed lines) and five significant contemporaneous paths (solid lines) in the left hemisphere.

5.2 Hierarchical Multi-Subject Unified SEM Approach --- Multilevel Covariance Structural Analysis with Unified SEM

The single-level multi-subject unified SEM approach described in Section 5.1, which analyzes fMRI data of all the subjects simultaneously by a single-level unified SEM model, assumes that all the observations (e.g. 1484 observations in visual attention fMRI data) were sampled independently from each other. However, the typical fMRI data

have a complex hierarchical data structure, that is, the time point observations are nested within subjects. Both fMRI observations and subjects are units in the analysis and variables are measured at both levels (we have fMRI measurements for each ROI, and we have subject-level covariates, e.g., gender and age.). The sampling design of fMRI data is two-stage sampling: the population of interest consists of subpopulations, and selection takes place through the subpopulations. The time point observations are nested in subjects, so the population of time point fMRI observations consists of subpopulations of subjects that produce time point fMRI observations. In such a random two-stage sample, a random sample of the subjects is taken in the first stage, and then the secondary units (time point fMRI observations) are sampled from the selected subjects in the second stage. Therefore, the time point fMRI observations might be dependent from each other: having selected a subject increases the chances of selection of the secondary units (time point fMRI observations) from that subject. The patterns of fMRI observations from the same subject are more similar than fMRI observations from different subjects. The fMRI observations from different subjects can be independent, but fMRI observations from the same subject share common characteristics or perceptions. In other words, the two-stage sampling design could lead to dependent observations. Hence, the independent observation assumption of the approach described in Section 5.1 could be violated, and the use of single-level unified SEM model might not be valid.

In this section and the next section, we will present a hierarchical (multilevel) multi-subject unified SEM approach, which takes the hierarchical structure of fMRI data into account. It is a simultaneous analysis of fMRI data by considering both levels (subject

level and single fMRI scan level) simultaneously in the analysis. In this work, the word "hierarchical" and the word "multilevel" are exchangeable. We will show the application of this approach to our visual attention fMRI data in Section 5.4.

5.2.1. Introduction

Attempts to integrate multilevel modeling with structural equation modeling so as to provide a general methodology that can account for issues of measurement error and simultaneity as well as multistage sampling can be traced back to the late 1960s. One of the earliest attempts was by Schmidt (1969) who derived a maximum likelihood estimator for a general multilevel covariance structure model but did not attempt to introduce group level variables into the model. More recently, Longford and Muthén (1992) provided computational results for multilevel factor analysis models. Muthén and Satorra (1989) were the first to show the variety of possible special cases of multilevel covariance structure modeling. Muthén and Muthén (1998-2006) developed Mplus, a powerful software, for multilevel covariance structural analysis.

Muthén (1991, 1994) showed the decomposition of hierarchical educational data into two separate models for the within- and between-groups structures through multilevel covariance structural analysis. Multilevel covariance structural analysis assumes that only the intercepts (not path coefficients) in the structural equation modeling are varying randomly over groups. This method can be applied to general hierarchical data. However, to our knowledge no works have been done to introduce this

method into analysis of fMRI data. Since fMRI data have the similar hierarchical structure to education data where multilevel covariance structural analysis arose from, we propose to use the multilevel covariance structural analysis to analyze fMRI data. Since fMRI data are temporally correlated, in this approach, we first transform the fMRI data under the unified SEM framework to introduce longitudinal components, and then we apply multilevel covariance structural analysis to the transformed fMRI data that have both contemporaneous and longitudinal components. Following Muthén's work, if we assume that there is one population of fMRI observations that are clustered in subjects, the fMRI data are decomposed into two separate models for the within- and between-subjects structures.

The goal of multilevel covariance structure analysis is to decompose the variation in the variables (both contemporaneous and longitudinal ROIs) into variance and covariance components associated with the two levels of the hierarchical fMRI data structure and explain the variation present at each level simultaneously. For each fMRI measurement, the total score is decomposed into an individual fMRI component (i.e., the individual fMRI measurement deviation from the subject mean) and a subject component (i.e., the disaggregated subject mean). The decomposition is used to compute separate within- and between- subjects covariance matrices.

5.2.2. Intraclass Correlation

The degree of resemblance between fMRI observations belonging to the same

subject under the assumptions that the intercepts in the structural equation modeling are varying randomly over subjects can be expressed by the intraclass correlation coefficient. The term class is conventionally used and refers to the level-two units (subjects in fMRI data) in the classification system under consideration. Therefore, the intraclass correlation describes the degree of correspondence within subjects. It can be expressed as

$$\rho = \sigma_B^2 /(\sigma_B^2 + \sigma_W^2)\,, \tag{5.1}$$

where σ_B^2 is the variability between subjects and σ_W^2 is the within-subject variability for each variable (ROIs).

Therefore, ρ indicates the proportion of the total variability that can be attributed to variability between the subjects. Since intraclass correlation is calculated on the assumption that only the intercepts in the structural equation modeling are varying randomly over subjects, the intraclass correlation could be zero under two circumstances: (1) the data are independent, or (2) the assumptions are not valid, i.e., not only intercepts but also path coefficients are random. The magnitude of intraclass correlation depends on characteristics of the variable (ROIs) and the attributes of the subjects. The larger the intraclass correlation, the larger the distortion in parameter estimation that results from ignoring the similarities due to clustering from the multilevel fMRI data structure is. In the absence of between-subject variability (i.e., where the intraclass correlation is less than 0.05), however, there is little need to perform a multilevel covariance structural analysis. In such cases, a single-level analysis would provide similar parameter estimations as what multilevel covariance structural analysis would produce.

5.2.3. Multilevel Covariance Structure

If we write the fMRI data is format Y_{gi} (subscript *i* for individual fMRI observations, *i=1…N*; subscript *g* for subjects, *g=1…G*), Y_{gi} is a vector containing all variables (contemporaneous and longitudinal ROIs) for all individual fMRI observations in the subject. There are N_g individual fMRI observations in subject *g*, and $N = \sum N_g$ is the total sample size. Unlike conventional analysis, independence of observations is not assumed over all *N* observations but only over the *G* subjects.

Following the idea of Cronbach and Webb (1975), we can decompose the individual hierarchical fMRI data into a between subjects component $Y_B = \overline{Y}_g$, and a within subjects component $Y_W = Y_{gi} - \overline{Y}_g$. In other words, for each fMRI observation we replace the observed total score $Y_T = Y_{gi}$ by its components: the subject component Y_B (the disaggregated subject mean) and the individual fMRI observation Y_W (the individual fMRI observation deviation from the subject mean). These two components have the attractive property that they are orthogonal and additive (Searle, Casella & McCulloch, 1992):

$$Y_T = Y_B + Y_W . \tag{5.2}$$

This decomposition can be used to compute a between subjects covariance matrix Σ_B (the population covariance matrix of the disaggregated subject means Y_B) and a within subjects covariance matrix Σ_W (the population covariance matrix of the individual

fMRI observation deviations from the subject means Y_W). These covariance matrices are also orthogonal and additive:

$$\Sigma_T = \Sigma_B + \Sigma_W . \tag{5.3}$$

Following the same logic, we can also decompose the sample data. Suppose we have data from *N* fMRI individual observations, divided into *G* groups (subjects). If we decompose the sample data, we have for the sample covariance matrices:

$$S_T = S_B + S_W . \tag{5.4}$$

Hierarchical (multilevel) structural equation modeling assumes that the population covariance matrices Σ_B and Σ_W can be described by separate models for the between subjects and within subjects structure. To estimate the model parameters, we need maximum likelihood estimates of the population between subjects covariance matrix Σ_B and the population within subjects covariance matrix Σ_W.

We define two sample covariance matrices: the pooled within covariance matrix S_{PW} and the scaled between covariance matrix S_B.

Based on Muthén's work (Muthén 1989) we know that an unbiased estimate of the population within subjects covariance matrix Σ_W is given by the pooled within subjects covariance matrix S_{PW}, calculated in the sample by:

$$S_{PW} = \frac{\sum_{g=1}^{G}\sum_{i=1}^{N_g}(Y_{gi} - \bar{Y}_g)(Y_{gi} - \bar{Y}_g)'}{N - G}. \tag{5.5}$$

Equation (5.5) corresponds to the conventional equation for the covariance matrix of the individual deviation scores, with *N-G* in the denominator instead of the usual *N-1*.

Since the pooled within subjects covariance matrix S_{PW} is an unbiased estimated of the population within subjects covariance matrix Σ_W, we can estimate the population within subject structure directly by constructing and testing a model for S_{PW}.

The scaled between subjects covariance matrix for the disaggregated subject means S_B can be calculated in the sample by:

$$S_B = \frac{\sum_{g=1}^{G} N_g(\bar{Y}_g - \bar{Y})(\bar{Y}_g - \bar{Y})'}{G-1}. \tag{5.6}$$

However, unlike S_{PW}, S_B is not an unbiased estimate of Σ_B.

Muthén (1989, 1990) shows that in a hierarchical model, S_{PW} is the maximum likelihood estimator of Σ_W, with sample size *N*-*G*, and S_B is the maximum likelihood estimator of the composite $\Sigma_W + c\Sigma_B$ with sample size *G*:

$$E(S_{PW}) = \Sigma_W, \tag{5.7}$$

and

$$E(S_B) = \Sigma_W + c\Sigma_B. \tag{5.8}$$

For balanced fMRI data (all subjects have the same number of individual fMRI observations), *c* equal to the common subject size *n*. For unbalanced data, *c* reflects the subject size and we have:

$$c = [N^2 - \sum_{g=1}^{G} N_g^2][N(G-1)]^{-1}. \tag{5.9}$$

Equation (5.8) shows that the population counterpart of S_B is a function of both Σ_B and

Σ_W . The ML estimate of Σ_W is S_{PW} , while the ML estimate of Σ_B is (Muthén 1990)

$$c^{-1}(S_B - S_{PW}). \tag{5.10}$$

Under the assumption of multivariate normality, the maximum-likelihood (ML) estimator (Muthén 1989) minimizes the function

$$\begin{aligned} F^* = G\log\left|n^{-1}\Sigma_W + \Sigma_B\right| + (N-G)\log\left|\Sigma_W\right| + G\ tr[n^{-1}\Sigma_W + \Sigma_B)^{-1}S_B] \\ +(N-G)\ tr[\Sigma_W^{-1}S_{PW}]. \end{aligned} \tag{5.11}$$

Here *n* is the balanced subject fMRI observation size.

For unbalanced data, Muthén (1994) derived a quasi-likelihood estimator as

$$\begin{aligned} F_{muml} = G\left\{\ln\left|\Sigma_W + c\Sigma_B\right| + trace[(\Sigma_W + c\Sigma_B)^{-1}S_B] - \ln\left|S_B\right| - t\right\} \\ + (N-G)\left\{\ln\left|\Sigma_W\right| + trace[\Sigma_W^{-1}S_{PW}] - \ln[S_{PW}] - t\right\}, \end{aligned} \tag{5.12}$$

where *t* is the total number of variables.

Since in our fMRI data, there are only observed variables, we consider multilevel path analysis (Kaplan, 2000) instead of multilevel factor analysis. Following Muthén (1994) consider the within groups model written as

$$Y_{gi} = \alpha_g + B_Y Y_{gi} + \varepsilon_{gi}, \tag{5.13}$$

where α_g is a vector of intercepts which are assumed to vary over subjects, B_Y is a matrix of path coefficients relating the within subject variables (ROIs) to each other, and ε_{gi} is a disturbance term.

The specification of the model in Equation (5.13) may appear unusual, as there is no vector of *x* of exogenous variables. In fact, Equation (5.13) is referred to as an "all-y" model (Jöreskog and Sörbom, 1993). In the "all-y" specification, all variables are treated

as "endogenous" variables. Thus Y_{gi} is a p+q-dimentional vector of variables, where the first p variables are endogenous variables, while the last q variables are exogenous variables. And the first p elements of α_g intercepts for endogenous variables, while the last q elements of α_g are the means of the exogenous variables. The model as expressed in Equation (5.13) was referred to as the structural form of the "within subject" part of the model, representing the hypothesized relationships among the within-subject variables (ROIs) as they are implied by the generic unified SEM model. It is important to note that the model specified in Equation (5.13) allows one to capture variation in the intercepts and exogenous variable means, but does not allow one to capture variation in the structural relationships contained in B_Y. Indeed, it is presently not possible to model slope variation in the context of multilevel structural equation modeling. (Kaplan 2000; Bauer et al., 2006)

If we assume that the inverse of $(I - B_Y)$ exists, Equation (5.13) can be reexpressed as

$$Y_{gi} = (I - B_Y)^{-1}\alpha_g + (I - B_Y)^{-1}\varepsilon_{gi}. \tag{5.14}$$

We assume that the levels of subject level variables (contained in α_g) vary across the g subjects and that this variation can be explained by subject level variables. Thus, we can write a "between subject" model for intercepts and means as

$$\alpha_g = \alpha + B_\alpha z_g + \delta_g, \tag{5.15}$$

where assuming z_g is centered around the grand mean, α is the grand mean vector

across the *g* subjects, z_g are subject level variables, B_α is a matrix of path coefficients relating z_g to α_g, and δ_g is a vector of disturbances. Furthermore, the between subject variables z_g are allowed to follow a separate between subject simultaneous equation that can be written as

$$z_g = \tau + B_z z_g + u_g . \tag{5.16}$$

If we assume that $(I - B_z)$ is nonsingular, Equation (5.16) can be rewritten in reduced form as

$$z_g = (I - B_z)^{-1}\tau + (I - B_z)^{-1}u_g , \tag{5.17}$$

where τ is a vector of intercepts and means for the subject level equations, B_z is a matrix of coefficients relating subject level variables to each other, and u_g is a vector of disturbances for the subject level equation.

After a series of substitutions of Equations (5.17), (5.16) and (5.15) into Equation (5.14), taking into account the structural relationships within as well as between subjects, we have the final model which can be written as

$$Y_{gi} = (I - B_Y)^{-1}\alpha + \Pi\tau + \Pi u_g + (I - B_Y)^{-1}\delta_g + (I - B_Y)^{-1}\varepsilon_{gi} , \tag{5.18}$$

where $\Pi = (I - B_Y)^{-1}B_\alpha(I - B_z)^{-1}$.

We can decompose the covariance components of the model in Equation (5.18) as

$$\begin{aligned} \text{var}(Y_{gi}) = {} & \Pi\Psi_u\Pi' + (I - B_Y)^{-1}\Psi_\delta(I - B_Y)'^{-1} \\ & + (I - B_Y)^{-1}\Psi_\varepsilon(I - B_Y)'^{-1}, \end{aligned} \tag{5.19}$$

where $\Psi_u = \text{var}(u_g)$, $\Psi_\delta = \text{var}(\delta_g)$, and $\Psi_\varepsilon = \text{var}(\varepsilon_{gi})$. From here, the between subjects

and within subjects covariance matrix can be written, respectively, as

$$\Sigma_B = \Pi \Psi_u \Pi' + (I - B_Y)^{-1} \Psi_\delta (I - B_Y)'^{-1} \tag{5.20}$$

and

$$\Sigma_W = (I - B_Y)^{-1} \Psi_\varepsilon (I - B_Y)'^{-1}. \tag{5.21}$$

5.2.4. Multilevel Covariance Structural Analysis Steps

Following the suggestions of Muthén (1994), the hierarchical multi-subject unified SEM analysis of fMRI data can be preceded by four steps: single-level unified SEM analysis of S_T, estimations of between variation, estimation of within structure, and estimation of between structure.

Step 1: Single-level unified SEM analysis of S_T. We have described this analysis in Section 5.1. The analysis is incorrect when the fMRI data is multilevel due to the correlated fMRI observations. The model test of fit is usually inflated, particularly for data with large intraclass correlations, large class sizes, and highly correlated variables (ROIs). However, the test of fit might still be of practical use by giving a rough sense of fit.

Step 2: Estimation of between variations. It is wise to first check if a multilevel analysis is appropriate by testing $\Sigma_B = 0$. A simpler way, to get a rough indication of the amount of between variation is to compute the estimated intraclass correlations for each variable by Equation (5.1). In line with (5.10), σ_W^2 is estimated as s_{PW}^2 and σ_B^2 is

estimated as

$$c^{-1}(s_B^2 - s_{PW}^2) \tag{5.22}$$

These estimates may be obtained by random effects ANOVA (Viner, Brown, and Michels 1991). If all intraclass correlations are close to zero, it might not be worthwhile to go further. If all intraclass correlations are close to zero, we know that (1) the data are independent, and therefore single-level unified SEM is sufficient for path estimations, or (2) the assumptions that only the intercepts are random is not valid.

Step 3: Estimation of within structure. We fit the Σ_W structure to the pooled within matrix S_{PW} by unified SEM model. For the balanced case, this gives the same ML estimates as a multilevel structural model with Σ_B unrestricted. This analysis estimates ROI level (first-level) parameters only (it can not estimate subject-level parameters if any). The unified SEM model in this step would use a sample size *N*-*G*. Since the S_{PW} analysis is not distorted by the between covariation, it is expected to give a better model fit than the S_T analysis (Muthén 1989) and it is the preferred way to explore the first-level variation.

Step 4: Estimation of between structure. The covariance structure of Σ_B does not concern the customary first-level data but instead across-subject covariation. In the between subjects model, we can incorporate a set of subject vairables (covariates) that may impact some individual ROI (contemporaneous ROIs) scores. Additionally, variables (contemporaneous and longitudinal ROIs) in the within subjects model can also be included in the between subjects model. By minimizing the fitting function (for the

special case of balanced fMRI data, the fitting function is given by Equation (5.11)), we can obtain the maximum likelihood estimations of path parameters. Maximum likelihood estimation of such models is implemented in Mplus (Muthén and Muthén, 2004).

5.2.5. Assumptions and Limitations

The approach to multilevel SEM models only one single within subjects covariance matrix. We have assumed that the within subjects covariances are homogeneous, i.e., that all subjects have the same within subjects covariance matrix. This is not necessarily the case. The effect of violating this assumption in general hierarchical models is currently unknown (Hox, 2002). Simulation studies on the assumption of homogeneous covariance matrices of general hierarchical models show that when larger variability exists in the smaller group sizes, the between group variation is overestimated; when larger variability exists in the larger group sizes, the between group variation is underestimated.

In fMRI data, if we assume that the covariance matrices differ in different subjects, one possible solution is to divide the subjects in two or more subsets, with each subset having its own within subjects model. For example, we may assume that within subject covariances differ for male and female participants. Then we model a different within subjects model for each subsets, and a common between subjects models.

It should be noted that the multilevel SEM model described in Section 5.2.3. differs from the multilevel regression model, because it does not have random regression slopes while the latter one has. The variation and covariation on the subject level is intercept

variation. There are no cross-level and interaction effects. The interpretation of subject level path coefficients is in terms of contextual effects (Hox, 2002), which are added to the first-level (individual fMRI observations) effects. The principal difficulty is that the typical maximum likelihood estimators for multilevel SEMs allow for random intercepts but not random slopes (Bauer et al. 2006).

5.3 Hierarchical Multi-Subject Unified SEM Approach --- Random-Effects Models

As explained in the last section, the multilevel SEM model described in Section 5.2. differs from multilevel regression model, because it does not have random regression slopes while the latter one has. Friston et al. (2005) has introduced the univariate random-effects regression model into fMRI studies. In this work, we study the random-effects model under the unified SEM framework by introducing longitudinal components into fMRI data. In this section, we will start with univariate random-effects model and then move to multivariate random-effects models of fMRI data.

5.3.1 Univariate Hierarchical Model with Random-Effects

In this section, we consider a very general form of the multilevel regression model referred to as the intercepts- and slopes-as-outcomes model or random-effects model. A

number of special cases of this model can be derived (Bryk and Raudenbush, 1992). To begin, consider a simple two-level form of the model where, for example, we are interested in modeling the fMRI score of ROI THAL by ROI CEREB for the *i*th individual time point observation of the *g*th subject. Such a model can be specified as follows. Following Bryk and Raudenbush (1992) let

$$y_{ig} = \beta_{0g} + \beta_{1g} x_{ig} + r_{ig}, \tag{5.23}$$

where y_{ig} is the *i*th time point fMRI score on THAL of the *g*th subject, x_{ig} is *i*th time point fMRI score on CEREB of the *g*th subject. Assuming that x_{ig} is centered around the subject mean, the intercepts, β_{0g} can be interpreted as the unadjusted mean for subject *g* and β_{1g} is the CEREB slope for subject *g*. r_{ig} is the residual term and we assume $r_{ig} \sim N(0, \sigma^2)$.

Note that the subscript *g* on the intercept β_{0g} and slope β_{1g} in Equation (5.23) implies that the *g* subjects vary both in their levels of THAL and the relationship of CEREB to THAL. It comes from the possibility the coefficient β_{0g} and β_{1g} may depend on *g*, in other words, the effect of CEREB on THAL might be stronger for some subjects than for others and the average value of the dependent variable THAL could differ across subjects. In the hierarchical linear model, it is modeled by random intercepts and random slopes. Thus, there may be subject level variables that explain variation in the intercepts and slopes. These variables might include subject characteristics such as gender, verbal IQ, education and age. For simplicity, let W_{1g} represent the gender of subject *g* and W_{2g}

be the age of subject *g*. Under the assumption that random parameters β_{0g} and β_{1g} are normally distributed, the *g* intercepts and slopes in Equation (5.23) can be modeled as

$$\beta_{0g} = \gamma_{00} + \gamma_{01}W_{1g} + \gamma_{02}W_{2g} + u_{0g} \tag{5.24}$$

and

$$\beta_{1g} = \gamma_{10} + \gamma_{11}W_{1g} + \gamma_{12}W_{2g} + u_{1g}, \tag{5.25}$$

where

(i) γ_{00} is the mean THAL score for female (when subjects are females, W_{1g} is zero) subjects with mean age of all subjects (when W_{2g} is centered around its means),

(ii) γ_{01} is the relationship between THAL mean scores and the gender of subject holding age constant,

(iii) γ_{02} is the relationship between THAL mean scores and the age of subject holding gender constant,

(iv) γ_{10} is the average CEREB-THAL relationship for female (when subjects are female, W_{1g} is zero) subjects with mean age of all subjects (when W_{2g} is centered around its means),

(v) γ_{11} is the effect of gender on the CEREB-THAL relationship, holding age constant, and

(vi) γ_{12} is the effect of age on the CEREB-THAL relationship, holding gender constant.

And we assume residual terms $u_{0g} \sim N(0, \tau_{00})$ and $u_{1g} \sim N(0, \tau_{11})$.

In Equation (5.23), intercept β_{0g} and slope β_{1g} are random parameters instead of fixed parameters in conventional regression models.

Substituting Equation (5.24) and (5.25) into Equation (5.23), yielding the full model

$$\begin{aligned} y_{ig} &= (\gamma_{00} + \gamma_{01}W_{1g} + \gamma_{02}W_{2g} + u_{0g}) + (\gamma_{10} + \gamma_{11}W_{1g} + \gamma_{12}W_{2g} + u_{1g})x_{ig} + r_{ig} \\ &= \gamma_{00} + \gamma_{01}W_{1g} + \gamma_{02}W_{2g} + \gamma_{10}x_{ig} + \gamma_{11}W_{1g}x_{ig} + \gamma_{12}W_{2g}x_{ig} + u_{0g} + u_{1g}x_{ig} + r_{ig}, \end{aligned} \tag{5.26}$$

The last expression was rearranged so that first comes the fixed part and then the random part. The first part of (5.26), $\gamma_{00} + \gamma_{01}W_{1g} + \gamma_{02}W_{2g} + \gamma_{10}x_{ig} + \gamma_{11}W_{1g}x_{ig} + \gamma_{12}W_{2g}x_{ig}$, is called the fixed part of the model. The second part, $u_{0g} + u_{1g}x_{ig} + r_{ig}$, is called the random part.

The term $u_{1g}x_{ig}$ can be regarded as a random interaction between subject and CEREB. Model (5.26) implied that the subjects are characterized by two random effects: their intercept and their slope. We say that *X* has a random slope, or a random effect, or a random coefficient. These two subject effects will usually not be independent, but correlated. It is assumed that, for different subjects, the pairs of random effects (u_{0g}, u_{1g}) are independent and identically distributed, and they are independent of the level-one residuals r_{ig}. All r_{ig} are independent and identically distributed. The variance of the level-one residuals r_{ig} is denoted σ^2; the variances and covariance of the level-two (subject level) residuals (u_{0g}, u_{1g}) are denoted as follows:

$$\begin{aligned} &\text{var}(u_{0g}) = \tau_{00} = \tau_0^2; \\ &\text{var}(u_{1g}) = \tau_{11} = \tau_1^2; \\ &\text{cov}(u_{0g}, u_{1g}) = \tau_{01}. \end{aligned} \tag{5.27}$$

In Equation (5.26) we see that explaining the intercept β_{0g} by subject level variables W_{1g} (gender) and W_{2g} (age) leads to two main effects of W_{1g} and W_{2g}. However, explaining the coefficient β_{1g} of x (CEREB) by the subject level variables W_{1g} (gender) and W_{2g} (age) leads to two product interaction effects of x with W_{1g}, and x with W_{2g}. Such interactions between a level-one variable and a level-two (subject level) variable are called a cross-level interaction.

The preceding models can be extended by including more variables (ROIs) that have random effects, and more variables explaining these random effects. Suppose that there are Q level-one explanatory variables $X_1, X_2, \ldots, X_Q$ (ROIs) and some set of level-two (subject level) W-variables (covariates). Consider the model where all X-variables have varying slopes, and where the random intercept as well as all these slopes are explained by all W-variables. As the within subject level, the model then is regression model with Q variables,

$$\begin{aligned} y_{ig} &= \beta_{0g} + \beta_{1g}x_{1ig} + \beta_{2g}x_{2ig} + \ldots + \beta_{Qg}x_{Qig} + r_{ig} \\ &= \beta_{0g} + \sum_{q=1}^{Q} \beta_{qg}x_{qig} + r_{ig}, \end{aligned} \tag{5.28}$$

where $r_{ig} \sim N(0, \sigma^2)$. Equation (5.28) has Q+1 coefficients, any one of which could be viewed as fixed, nonrandomly varying, or random. When they are random, in the subject level model, each coefficient β_{qj} can be modeled as

$$\begin{aligned} \beta_{qg} &= \gamma_{q0} + \gamma_{q1}W_{1g} + \gamma_{q2}W_{2g} + \ldots + \gamma_{qS}W_{S_q g} + u_{qg} \\ &= \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs}W_{sg} + u_{qg}, \end{aligned} \tag{5.29}$$

for some set a subject level variables $Wsg, s = 1,\ldots,S_q$.

The subjects are now characterized by Q+1 random coefficients u_{0g} to u_{Qg}. These random coefficients are independent between subjects, but may be correlated within subjects. It is assumed that the vector $(U_{0g},\ldots,U_{Qg})$ is independent of the level-one residuals r_{ig} and that all residuals have population means 0, given the values of all explanatory variables. It is also assumed that the level-one residual r_{ig} has a normal distribution with constant variance σ^2 and that $(U_{0g},\ldots,U_{Qg})$ has a multivariate normal distribution with a constant covariance matrix. Analogous to (5.27), the variance and covariances of the subject level random effects are denoted

$$\begin{aligned} &\text{var}(u_{qg}) = \tau_{qq} = \tau_q^2 \ (q = 1,\ldots Q); \\ &\text{cov}(u_{qg}, u_{q'g}) = \tau_{qq'} \ (q, q' = 1,\ldots Q). \end{aligned} \tag{5.30}$$

If we rewrite Equation (5.28) in matrix notation, we have

$$Y_g = X_g\beta_g + r_g, \quad r_g \sim N(0,\sigma^2 I), \tag{5.31}$$

where Y_g is an n_g by 1 vector of observed values of dependent variables (each subject has n_g individual time point observations), X_g is an n_g by (Q+1) matrix of independent variables, β_g is a (Q+1) by 1 vector of unknown parameters, I is an n_g by n_g identity matrix, and r_g is an n_g by 1 vector of random errors assumed normally distributed with a mean vector of 0 and a variance-covariance matrix in which all diagonal elements are equal to σ^2 and all off-diagonal elements are 0.

At subject level, the general model for β_g in matrix notation is

$$\beta_g = W_g\gamma + u_g, \quad u_g \sim N(0,T), \tag{5.32}$$

where W_g is a (Q+1) by S matrix of subject level predictors, γ is an S by 1 vector of fixed effects, u_g is a (Q+1) by 1 vector of subject level errors or random effects, and T is a (Q+1) variance-covariance matrix. T is the residual variance-covariance matrix, indicating the dispersion of β_g about the expected value $W_g\gamma$.

Next we show the point estimations of parameters in random-effects model. Assuming X_g to be of full column rank $Q+1$, the OLS estimator of β_g from Equation (5.31) is

$$\hat{\beta}_g = (X_g^T X_g)^{-1} X_g^T Y_g . \tag{5.33}$$

The generalized least squares (GLS) estimator of γ is

$$\hat{\gamma} = (\sum W_g^T \Delta_g^{-1} W_g)^{-1} \sum W_g^T \Delta_g^{-1} \hat{\beta}_g , \tag{5.34}$$

when Δ_g is known and $\Delta_g = T + V_g = T + \sigma^2 (X_g^T X_g)^{-1}$. Given the normality assumptions of Equation (5.31) and (5.32), Equation (5.34) is also the maximum likelihood estimator for γ.

The optimal estimator of β_g is given by

$$\beta_g^* = \Lambda_g \hat{\beta}_g + (I - \Lambda_g) W_g \hat{\gamma} , \tag{5.35}$$

where

$$\Lambda_g = T(T + V_g)^{-1}, \tag{5.36}$$

and $\hat{\beta}_g$ and $\hat{\gamma}$ are given by Equation (5.33) and (5.34). β_g^* is an empirical Bayes or shrinkage estimator.

So far we have assumed that the variance and covariance components are known. Although this assumption clarifies understanding of estimation of the fixed and random effects, the variances and covariances must nearly always be estimated in practice. Next we show the methods to estimate variances and covariances.

Substituting Equation (5.32) into Equation (5.31), we have

$$Y_g = X_g W_g \gamma + X_g u_g + r_g. \tag{5.37}$$

We can rewrite the model of (5.37) in a way that allows some level-one variables to have fixed effects but not random effect:

$$Y_g = X_g W_g \gamma + Z_g u_g + r_g. \tag{5.38}$$

In Equation (5.38), Z_g is typically a subset of X_g.

Bryk and Raudenbush (1992) show that the point estimations of the unknown parameters, variances, and covariances are as follows.

We place the random errors of level-one and level-two into two vectors,

$$u = (u_1, u_2, ..., u_G)' \quad \text{and} \quad r = (r_1', r_2', ..., r_G')'. \tag{5.39}$$

And we have,

$$E\begin{bmatrix} u \\ r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad Var\begin{bmatrix} u \\ r \end{bmatrix} = \begin{bmatrix} T & 0 \\ 0 & R \end{bmatrix}, \tag{5.40}$$

where $Cov(r) = R = \sigma^2 I_N$.

Similarly, we can place the level-one vector in Equation (5.38) of all subjects into one vector, such as,

$$y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_G \end{pmatrix}, \quad \text{and} \quad Z = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_G \end{pmatrix}. \tag{5.41}$$

We define

$$V = ZTZ' + R\,. \tag{5.42}$$

Estimation is more difficult in the random effect model than in the general linear model. Not only do we have γ as in the general model, but also we have unknown parameters in T and R as well.

In many situations, the best approach is to use likelihood based methods, exploiting the assumption that *u* and *r* are normally distributed (Hartley and Rao 1967; Patterson and Thompson 1971; Harville 1977; Laird and Ware 1982; Jennrich and Schluchter 1986). A fitting function associated with maximum likelihood is constructed, and it is maximized over all unknown parameters. The corresponding log-likelihood function is as follows:

$$l(T,R) = -\frac{1}{2}\log|V| - \frac{1}{2}s'V^{-1}s - \frac{1}{2}\log(2\pi)\,. \tag{5.43}$$

where *s* is a vector containing $s_g = Y_g - X_g W_g \hat{\gamma}$, and *p* is the rank of *X*. The most multilevel softwares minimize -2 times the above functions using an Expectation-Maximum (EM) algorithm.

We can perform the hypothesis test for fixed effects. The typical null hypothesis is

$$H_0 : \gamma_{qs} = 0\,, \tag{5.44}$$

which implies that the effect of a level-2 (subject level) predictor, W_{sg}, on a particular

parameter, β_{qg}, is null. The tests for such hypotheses have the form

$$z = \hat{\gamma}_{qs} / (\hat{V}_{\hat{\gamma}_{qs}})^{1/2}, \tag{5.45}$$

where $\hat{\gamma}_{qs}$ is the maximum likelihood estimate of γ_{qs} and $(\hat{V}_{\hat{\gamma}_{qs}})^{1/2}$ is the estimated sampling variance of $\hat{\gamma}_{qs}$. Formally the *z* statistic is asymptotically normal. It will often be the case, however, that a *t* statistic with degrees of freedom equal to $G - S_q - 1$ will provide a more accurate reference distribution for testing effects of level-2 predictors. (Bryk and Raudenbush, 1992)

To test whether random variation exists, we may test a null hypothesis

$$H_0 : \tau_{qq} = 0, \tag{5.46}$$

where $\tau_{qq} = Var(\beta_{qg})$. If this hypothesis is rejected, the investigator may conclude that there is random variation in β_{qg}.

Let $\hat{V}_{qqg}$ represent the *q*th diagonal element of $\hat{V}_g = \hat{\sigma}^2 (X_g^T X_g)^{-1}$. Then, under the model

$$\beta_{qg} = \gamma_{q0} + \sum_{s=`1}^{S_q} \gamma_{qs} W_{sg}, \tag{5.47}$$

the statistic

$$\sum_g (\hat{\beta}_{qg} - \hat{\gamma}_{q0} - \sum_{s=1}^{S_q} \hat{\gamma}_{qs} W_{sg})^2 / \hat{V}_{qqg} \tag{5.48}$$

will be distributed approximately χ^2 with $G - S_q - 1$ degrees of freedom, where $\hat{\beta}_g = (X'_g X_g)^{-1} X' Y_g$. (Bryk and Raudenbush, 1992)

A second test of the hypothesis $H_0 : \tau_{qq} = 0$ is based on the estimated standard error of $\hat{\tau}_{qq}$ computed from the inverse of the second derivative matrix of the likelihood with respect to each of the variance parameters. The ratio

$$z = \hat{\tau}_{qq} / [Var(\hat{\tau}_{qq})]^{1/2} \tag{5.49}$$

is approximately normally distributed under the large sample theory of maximum likelihood estimates. However, in many cases, especially when τ_{qq} is close to zero, the normality approximation is poor. A test based on a symmetric confidence interval for τ_{qq} may be highly misleading (Bryk and Raudenbush, 1992). Hence the Chi-square test is recommended for the hypothesis test of random variation.

5.3.2 Multivariate Hierarchical Random-Effects Model

Typical fMRI data are multivariate multilevel data, however, the random-effects model described in Section 5.3.1 is a univariate multilevel model. To analyze the multivariate fMRI data appropriately, in this work we adopt the multivariate hierarchical random-effect model developed by Bauer et al. (2006) which was applied by the authors in psychology field study, and we incorporate longitudinal components into the model to make it more suitable for fMRI data study.

Due to the principal difficulty that the typical maximum likelihood estimators for multilevel SEMs allow for random intercepts but not random slopes (Bauer et al. 2006), Bauer et al. (2006) proposed to use a modified univariate hierarchical random-effects

model to analyze multivariate hierarchical data. We illustrate this approach as follows and the application of this approach to our visual attention data is shown in Section 5.4.

Suppose we have three variables, *X*, *Y*, and *M* (ROIs). The path diagram is shown in Figure 5.2.

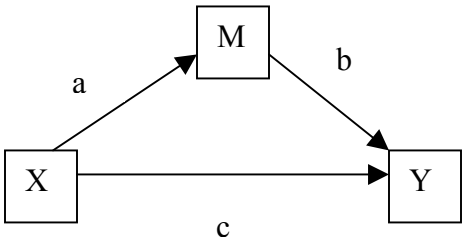


Figure 5.2 Path diagram of the example in Section 5.3.2. to illustrate Bauer's multivariate random-effects approach.

The two level-one equations following the notation in Section 5.3.1. are

$$M_{ig} = d_{Mg} + a_g X_{ig} + e_{Mig}$$

$$Y_{ig} = d_{Yg} + b_g M_{ig} + c_g X_{ig} + e_{Yig}\,. \tag{5.50}$$

The terms e_{Mig} and e_{Yig} are level-one residuals for *M* and *Y*, respectively. The other five terms are random intercepts and slopes.

In order to analyze multivariate hierarchical data simultaneously, we can formulate the model with a single level-one equation through the use of indicator variables. The basic idea is to form a new outcome variable, for instance, *Z*, by stacking *Y* and *M* for each unit *i* within *g*. This single outcome variable allows us to fit a "multivariate" model using univariate multilevel modeling software. To distinguish the two variables stacked in *Z*, we created two indicator variables, for instance, S_M and S_Y. The variable S_M is set

equal to 1 when *Z* refers to *M* and is 0 otherwise. Similarly, the variable S_Y is set equal to 1 when *Z* refers to *Y* and is 0 otherwise. We retain the variables *X* and *M* in the new data set. The purpose of rearranging the data is that we can now specify the level-one model with a single equation:

$$Z_{ig} = S_{Mig}(d_{Mg} + a_g X_{ig}) + S_{Yig}(d_{Yg} + b_g M_{ig} + c_g X_{ig}) + e_{Zig}. \quad (5.51)$$

Rewrite Equation (5.51) as

$$Z_{ig} = d_{Mg} S_{Mig} + a_g (S_{Mig} X_{ig}) + d_{Yg} S_{Yig} + b_g (S_{Yig} M_{ig}) + c_g (S_{Yig} X_{ig}) + e_{Zig}. \quad (5.52)$$

Equation (5.52) shows that we could specify a model for *Z* with no intercept but with random effects for S_M and S_Y (d_{Mg} and d_{Yg}, respectively) and with random effects for the product variables $S_M X$, $S_Y M$, and $S_Y X$ (a_g, b_g, and c_g, respectively). In addition, we must use some method to allow the residual variance $Var(e_{Zig})$ to differ depending on S_M (or, equivalently, S_Y). This represents a form of heteroscedasticity because the residual variance for *Z* is then conditional on S_M. Fortunately, most multilevel modeling software programs offer options for modeling heteroscedasticity.

5.4 Application of Hierarchical Multi-Subject Unified SEM Approach to Visual Attention fMRI Data

5.4.1 Multilevel Covariance Structural Analysis with Unified SEM

Table 5.2 shows the intraclass correlation of contemporaneous and longitudinal variables of visual attention fMRI data. And the estimated within subject covariance matrix and the estimated between subject covariance matrix are listed in Table 5.3 and 5.4, respectively. Since the calculated intraclass correlations are all small, we know that we would not go further to perform multilevel covariance structural analysis with unified SEM, because there are two possibilities: (1) the data are independent and the single-level multi-subject unified SEM approach gives appropriate path estimation, or (2) the assumptions of random intercepts but non-random path coefficients are not valid.

Nevertheless, for the purpose of this example we will continue to fit the visual attention fMRI data by within-subject unified SEM model. Table 5.5 presents the estimated longitudinal and contemporaneous path parameters with their standard errors, t test statistics, and corresponding p values. The bold characters indicate the significant paths at the significance level of 0.05. The path model with the significant paths is displayed in Figure 5.3. Without surprise, the significant longitudinal path connections are identical to those in Section 5.1, containing (1) paths connecting APC to LPFC, and LPFC to THAL and SMA, (2) the longitudinal paths connecting each region to itself. Five contemporaneous paths are significant, four of which form a single loop starting at THAL and connecting PPC, APC, LPFC, and one of which is from LPFC to SMA. The goodness-of-fit statistics of the model is χ^2 =122.331 (d.f.=31, p value<0.01) indicating a poor model fit. The same analysis was repeated using LISREL, with identical results as those from SAS.

Table 5.2 Intraclass correlations of longitudinal and contemporaneous components of the multi-subject fMRI data.

	CEREB	THAL	PPC	APC	LPFC	SMA
Intraclass correlation	0.0132	0.0068	0.0108	0.0136	0.0122	0.0131
	CEREBt	THALt	PPCt	APCt	LPFCt	SMAt
Intraclass correlation	0.0118	0.0062	0.011	0.0131	0.0116	0.0118

Table 5.3 Estimated within subject covariance matrix of the multi-subject fMRI data with longitudinal and contemporaneous components.

	CEREB	THAL	PPC	APC	LPFC	SMA	CEREBt	THALt	PPCt	APCt	LPFCt	SMAt
CEREB	0.443											
THAL	0.036	1.165										
PPC	-0.026	-0.071	0.410									
APC	-0.023	-0.050	0.064	0.254								
LPFC	0.045	0.070	-0.021	0.016	0.591							
SMA	0.044	0.041	0.028	0.030	0.092	0.487						
CEREBt	0.274	0.024	-0.014	-0.024	0.033	0.021	0.450					
THALt	0.019	0.735	-0.044	-0.057	0.067	0.025	0.036	1.167				
PPCt	-0.009	-0.070	0.255	0.051	-0.024	0.040	-0.027	-0.071	0.413			
APCt	-0.013	-0.025	0.075	0.158	-0.007	0.002	-0.026	-0.051	0.063	0.253		
LPFCt	0.027	0.030	-0.009	0.013	0.378	0.022	0.043	0.067	-0.019	0.014	0.591	
SMAt	0.053	0.030	0.008	0.041	0.064	0.276	0.043	0.045	0.027	0.027	0.087	0.481

Table 5.4 Estimated between subject covariance matrix of the multi-subject fMRI data with longitudinal and contemporaneous components.

	CEREB	THAL	PPC	APC	LPFC	SMA	CEREBt	THALt	PPCt	APCt	LPFCt	SMAt
CEREB	0.006											
THAL	-0.001	0.008										
PPC	0.001	0.000	0.004									
APC	0.000	0.001	0.000	0.003								
LPFC	-0.002	-0.001	0.001	0.000	0.007							
SMA	0.000	0.000	-0.001	-0.001	0.000	0.006						
CEREBt	-0.003	-0.001	0.001	0.000	-0.001	0.000	0.005					
THALt	0.000	-0.010	0.000	0.001	-0.001	0.001	-0.001	0.007				
PPCt	0.000	0.000	-0.002	0.000	0.001	-0.001	0.001	0.000	0.005			
APCt	0.000	0.001	0.000	-0.002	0.000	0.000	0.000	0.001	0.000	0.003		
LPFCt	-0.001	0.000	0.001	0.000	-0.003	0.001	-0.001	0.000	0.001	0.000	0.007	
SMAt	-0.001	0.000	-0.001	-0.001	0.000	-0.002	0.000	0.000	-0.001	-0.001	0.000	0.006

Table 5.5 Estimated longitudinal and contemporaneous path parameters of the multi-subject fMRI data using the multilevel within-subject unified SEM model with their standard errors, t test statistics, and corresponding p values (two-sided).

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Est. (S.E.)	T-value (p-value)	Path parameters	Est. (S.E.)	T-value (p-value)
γ_{11}	**0.607 (.020)**	**29.778 (.000)**	β_{21}	0.073 (.041)	1.769 (.077)
γ_{21}	-0.040 (.041)	-0.965 (.334)	β_{25}	**0.086 (.038)**	**2.296 (.022)**
γ_{22}	**0.629 (.020)**	**31.128 (.000)**	β_{26}	0.050 (.039)	1.297 (.195)
γ_{25}	**-0.075 (.038)**	**-1.976 (.048)**	β_{32}	**-0.039 (.016)**	**2.504 (.012)**
γ_{26}	-0.027 (.039)	-0.691 (.490)	β_{43}	**0.040 (.020)**	**1.961 (.050)**
γ_{32}	0.024 (.016)	1.564 (.118)	β_{54}	**0.111 (.039)**	**2.872 (.004)**
γ_{33}	**0.615 (.020)**	**30.106 (.000)**	β_{65}	**0.208 (.025)**	**8.439 (.000)**
γ_{43}	0.006 (.021)	0.279 (.780)			
γ_{44}	**0.612 (.021)**	**29.506 (.000)**			
γ_{54}	**−0.133 (.039)**	**-3.423 (.001)**			
γ_{55}	**0.641 (.020)**	**32.154 (.000)**			
γ_{65}	**-0.180 (.025)**	**-7.266 (.000)**			
γ_{66}	**0.578 (.021)**	**27.179 (.000)**			

Bold characters indicate the significant path parameters at the significance level of 0.05.

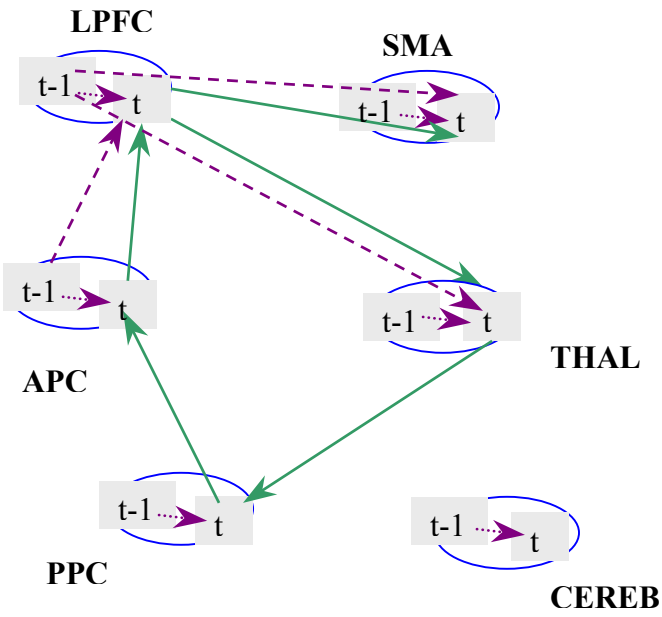


Figure 5.3 Significant path connections of the multi-subject fMRI data using the multilevel within-subject unified SEM model. This path network contains nine significant longitudinal path connections (dashed lines) and five significant contemporaneous paths (solid lines) in the left hemisphere.

5.4.2. Random-Effects Model

We test the 14 significant path found from Section 5.4.1 by random-effect model to see if the paths are random varying at the subject level. Under the framework of random-effects model, we have 14 paths and 5 intercepts to test for random effects, but we only have 28 samples at subject level. We cannot test all of them simultaneously. To implement Bauer's multivariate methods, we can break all the paths/intercepts into small systems. For instance, in Figure 5.4, we have two contemporaneous components (SMA_t

and $LPFC_t$) and two longitudinal components (SMA_{t-1} and $LPFC_{t-1}$), and one contemporaneous path (β_{65}) and three longitudinal paths γ_{55}, γ_{65}, and γ_{66}).

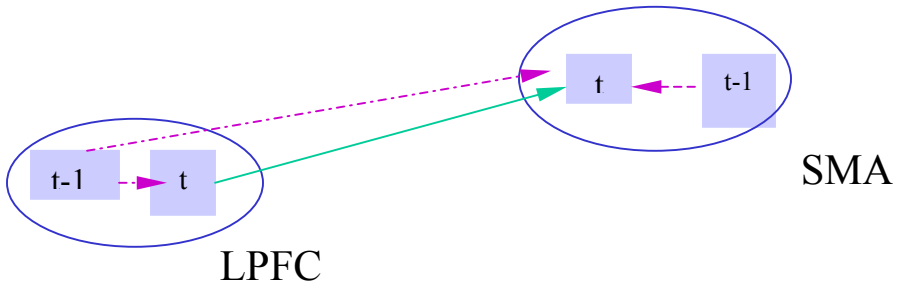


Figure 5.4 Illustration of the application of Bauer's approach to fMRI data.

By using the approach described in Section 5.3.2., we have a univariate random-effects model

$$Z_{ig} = S_{LPFCig}(d_{LPFCg} + a_g LPFCt_{ig}) + S_{SMAig}(d_{SMAg} + b_g LPFC_{ig} + c_g LPFCt_{ig} + f_g SMAt_{ig}) + e_{Zig}, \quad (5.53)$$

which has six random effects to be tested.

The multilevel random-effects model is fitted using the software HLM (Hierarchical Linear Model), and the corresponding analysis results are summarized in Table 5.6. First we test if each path is a random effect and if the estimated mean of each possible random path is significant (p value <0.05). The small p values (<0.05) of Chi-square test in Table 5.6 indicate random paths. If a path is random, we continue to test if the variation could be explained by subject-level covariates. The small p-value of t-tests of the fixed effect at the second level indicates the significant impact of subject covariates on each random

path. Five contemporaneous paths and two longitudinal paths are indeed random, and one contemporaneous path and eight longitudinal paths are significantly different from zero. However, only one random contemporaneous path and one random longitudinal path are significantly different from zero. Both of the random contemporaneous and random longitudinal paths connect LPFC to SMA. The fixed subject-effect of gender could explain the variations of these two random paths. The path model with significant paths is displayed in Figure 5.5.

Since there are significant random paths in the model, we now know that the assumptions of random intercepts only are not appropriate for our visual attention fMRI data, and therefore the multilevel covariance structural analysis described in Section 5.2 will not provide a correct model for our data.

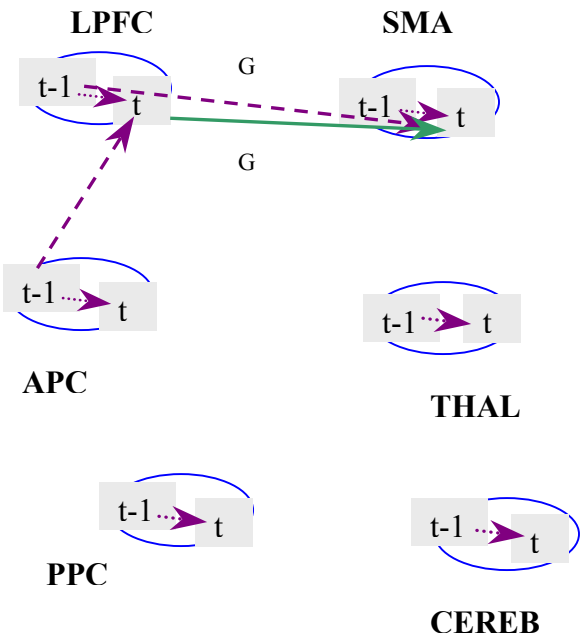


Figure 5.5 Nine significant path connections from the multilevel random-effects unified approach. Two paths are random which are significantly correlated with subject-level covariate gender. G (gender) is denoted along with the two random path connections. Dashed lines represent longitudinal path connections and solid lines represent

contemporaneous path.

Table 5.6 Longitudinal and contemporaneous paths of the multi-subject fMRI data using a multilevel random-effects unified model with their estimated means, standard errors, t-test statistics, corresponding p values, variance component, chi-square statistics, and corresponding p values. Significant and random paths with the t test statistics, and corresponding p values (two-sided) for second level fixed effects.

	Test for Random Effects				Second Level Predictor (Fixed Effects) for Random Effects		
	Est. (S.E.)	t value (p value)	Var.	Chi-square (p value)	Covariate	Intercept (p value)	Slope
β_{25}	0.097 (0.063)	1.545 (0.134)	**0.072**	**71.440 (0.000)**			
β_{32}	-0.037 (0.023)	-1.602 (0.124)	**0.084**	**52.632 (0.003)**			
β_{43}	0.002 (0.038)	0.047 (.963)	**0.025**	**83.545 (0.000)**			
β_{54}	0.107 (0.065)	1.642 (0.113)	**0.068**	**64.738 (0.000)**			
β_{65}	**0.210 (0.049)**	**4.252 (0.000)**	**0.048**	**107.175 (0.000)**	**Gender**	**0.347 (0.000)**	**-0.273 (0.003)**
γ_{11}	**0.607 (0.027)**	**22.293 (0.000)**	0.007	39.375 (0.060)			
γ_{22}	**0.629 (0.022)**	**28.383 (0.000)**	0.003	28.320 (0.398)			
γ_{25}	-0.089 (0.072)	-1.242 (0.225)	**0.104**	**93.397 (0.000)**			
γ_{33}	**0.6126 (0.022)**	**27.947 (0.000)**	0.0005	21.058 (>0.5)			
γ_{44}	**0.606 (0.022)**	**27.602 (0.000)**	0.0002	13.852 (>0.5)			
γ_{54}	**-0.134 (0.048)**	**-2.766 (0.010)**	0.022	33.335 (0.186)			
γ_{55}	**0.628 (0.023)**	**27.929 (0.000)**	0.002	20.102 (>0.5)			
γ_{65}	**-0.175 (0.048)**	**-3.638 (0.001)**	**0.044**	**92.503 (0.000)**	**Gender**	**-0.261 (0.002)**	**0.179 (0.050)**
γ_{66}	**0.585 (0.022)**	**26.783 (0.000)**	0.0002	17.495 (>0.5)			

Bold characters indicate the significant random paths, and significant fixed effects at the significance level of 0.05.

Chapter 6

Comparisons of Three Approaches

In Chapter 4 and Chapter 5, we discussed three main approaches to analyze multi-subject multivariate time series fMRI data, which are:

(1) Approach 1--- summarize (e.g. average the time series data across the subjects) and then analyze, which is the subject-average unified structural equation modeling approach,

(2) Approach 2---analyze and then summarize, which is the two-stage multi-subject unified structural equation modeling approach, and

(3) Approach 3---analyze simultaneously, which is the hierarchical multi-subject unified structural equation modeling approach.

In this chapter, we will compare these three main approaches.

6.1 Comparison between Approach 1 and Approach 2

Table 6.1 presents the significant longitudinal and contemporaneous path parameters with their estimations, standard errors, t test statistics, and corresponding p

values for Approach 1 and Approach 2. The bold characters indicate the significant paths at the significance level of 0.05. The path models with the significant paths from the two approaches are displayed in Figure 6.1. The t test statistics and the corresponding p values are calculated in the unified SEM model in Approach 1. In contrast, the t test statistics and the corresponding p values are calculated from one sample t test in Approach 2 (See Chapter 4 for details).

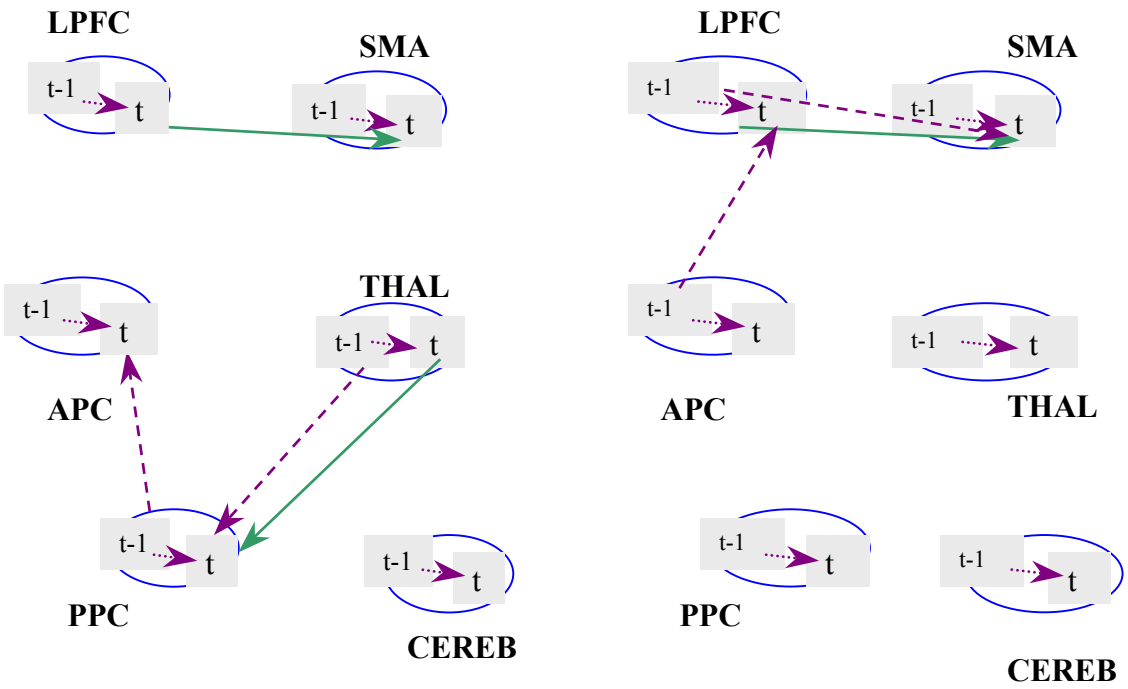


Figure 6.1 Significant path connections from the Approach 1 (left picture) and Approach 2 (right picture). Dashed lines represent longitudinal path connections solid lines represent contemporaneous path.

Table 6.1 Comparisons of longitudinal and contemporaneous path parameters from Approach 1 and Approach 2 with their estimations, standard errors, t test statistics, and corresponding p values (two-sided).

Path	Approach 1		Approach 2	
parameters	Est. (S.E.)	T-value (p-value)	Mean (S.E.)	T-value (p-value)
β_{21}	-0.058 (.250)	-0.232 (.816)	0.047 (.076)	0.622 (.539)
β_{25}	0.394 (.217)	1.815 (.070)	0.101 (.071)	1.424 (.166)
β_{26}	0.270 (.182)	1.482 (.138)	0.030 (.087)	0.340 (.736)
β_{32}	**-0.258 (.057)**	**-4.509 (.000)**	-0.034 (.026)	-1.311 (.201)
β_{43}	-0.141 (.126)	-1.122 (.262)	-0.005 (.036)	-0.126 (.901)
β_{54}	0.172 (.179)	0.959 (.338)	0.091 (.066)	1.384 (.178)
β_{65}	**0.377 (.156)**	**2.422 (.015)**	**0.229 (.054)**	**4.208 (.000)**
γ_{11}	**0.780 (.091)**	**8.593 (.000)**	**0.612 (.019)**	**31.462 (.000)**
γ_{21}	-0.317 (.274)	-1.158 (.247)	0.005 (.070)	0.074 (.942)
γ_{22}	**0.538 (.111)**	**4.842 (.000)**	**0.591 (.020)**	**29.391 (.000)**
γ_{25}	-0.024 (.221)	-0.110 (.912)	-0.120 (.070)	-1.718 (.097)
γ_{26}	0.023 (.198)	0.116 (.901)	0.028 (.082)	0.339 (.737)
γ_{32}	**0.237 (.061)**	**3.908 (.000)**	0.035 (.029)	1.182 (.248)
γ_{33}	**0.386 (.105)**	**3.696 (.000)**	**0.595 (.021)**	**28.821 (.000)**
γ_{43}	**0.246 (.110)**	**2.233 (.026)**	0.031 (.036)	0.872 (.391)
γ_{44}	**0.566 (.111)**	**5.116 (.000)**	**0.593 (.016)**	**36.001 (.000)**
γ_{54}	−0.284 (.178)	-1.600 (.110)	**−0.124 (.046)**	**-2.683 (.012)**
γ_{55}	**0.582 (.113)**	**5.164 (.000)**	**0.615 (.019)**	**31.685 (.000)**
γ_{65}	-0.282 (.160)	-1.757 (.079)	**-0.166 (.047)**	**-3.517 (.002)**
γ_{66}	**0.590 (.107)**	**5.518 (.000)**	**0.578 (.018)**	**31.140 (.000)**

Bold characters indicate the significant path parameters at the significance level of 0.05.

From Table 6.1 and Figure 6.1, we know that all of the six longitudinal paths connecting each region to itself are significant in both of the two approaches. The

contemporaneous path connecting LPFC and SMA is significant in both of approaches too. However, two longitudinal paths connecting THAL to PPC, and PPC to APC, are significant in Approach 1 but not significant in Approach 2; and two longitudinal paths, which are from APC to LPFC and from LPFC to SMA, are significant in Approach 2 but not in Approach 1. And there is one contemporaneous path, connecting THAL to PPC, is significant in Approach 1 but not in Approach 2.

Even though the two longitudinal paths, connecting APC to LPFC, and LPFC to SMA, are significant in Approach 2 but not in Approach 1 at significance level 0.05, the path from LPFC to SMA would be significant at level 0.1 (p value=0.079), and the path from APC to LPFC would be nearly significant at level 0.1 (p value=0.110). Therefore, only three path connections are different between Approach 1 and Approach 2, which are two longitudinal paths connecting THAL to PPC, and PPC to APC, and one contemporaneous path, connecting THAL to PPC.

Next we illustrate the reason why these two approaches could reach to different conclusions about the significance of some paths.

We assume we have two variables (ROIs) of fMRI data, x and y. We have m subjects and for each subject, we have N observations for each variable. To make computation simple, without losing generality, we assume the mean value of each variable for each subject is zero. Since SEM is based on covariance or correlation, and the bigger the correlation coefficient the more chance the corresponding path connection is significant, we next examine the correlation of x and y for each approach. The correlation coefficient is calculated by

$$corr(x, y) = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)\,\text{var}(y)}} = \frac{E(xy)}{\sqrt{E(x^2)E(y^2)}}. \tag{6.1}$$

For approach 1, we calculate the average of each variable across all the subjects for each fMRI time point observation. We denote the value of x for the kth fMRI time point observation of the ith subject as $x_k^{(i)}$, and the average of $x_k^{(i)}$ across all subjects as x_k^{ave}. Similarly, we have $y_k^{(i)}$ representing the value of y for the kth fMRI time point observation of the ith subject and y_k^{ave} representing the average of $y_k^{(i)}$ across all subjects.

$$x_k^{ave} = \frac{1}{m}\sum_{i=1}^{m} x_k^{(i)}, \quad \text{and} \quad y_k^{ave} = \frac{1}{m}\sum_{i=1}^{m} y_k^{(i)}. \tag{6.2}$$

Let $x^{ave} = (x_1^{ave}, ..., x_N^{ave})'$, and $y^{ave} = (y_1^{ave}, ..., y_N^{ave})'$.

Therefore,

$$E(x^{ave}y^{ave}) = \frac{1}{N}\sum_{k=1}^{N}(x_k^{ave}y_k^{ave}), \tag{6.3}$$

$$E((x^{ave})^2) = \frac{1}{N}\sum_{k=1}^{N}(x_k^{ave})^2, \tag{6.4}$$

and,

$$E((y^{ave})^2) = \frac{1}{N}\sum_{k=1}^{N}(y_k^{ave})^2. \tag{6.5}$$

By applying Equation (6.1),

$$corr(x^{ave}, y^{ave}) = \frac{E(x^{ave}y^{ave})}{\sqrt{E((x^{ave})^2)E((y^{ave})^2)}} = \frac{\sum_{k=1}^{N}(x_k^{ave}y_k^{ave})}{\sqrt{\sum_{k=1}^{N}(x_k^{ave})^2\sum_{k=1}^{N}(y_k^{ave})^2}}. \tag{6.6}$$

Substituting Equation (6.2) into Equation (6.6), we have

$$corr(x^{ave}, y^{ave}) = \frac{\sum_{k=1}^{N}(x_k^{ave} y_k^{ave})}{\sqrt{\sum_{k=1}^{N}(x_k^{ave})^2 \sum_{k=1}^{N}(y_k^{ave})^2}} = \frac{\sum_{k=1}^{N}(\frac{1}{m}\sum_{i=1}^{m} x_k^{(i)} \bullet \frac{1}{m}\sum_{i=1}^{m} y_k^{(i)})}{\sqrt{\sum_{k=1}^{N}(\frac{1}{m}\sum_{i=1}^{m} x_k^{(i)})^2 \bullet \sum_{k=1}^{N}(\frac{1}{m}\sum_{i=1}^{m} y_k^{(i)})^2}}$$
$$= \frac{\sum_{k=1}^{N}(\sum_{i=1}^{m} x_k^{(i)} \bullet \sum_{i=1}^{m} y_k^{(i)})}{\sqrt{\sum_{k=1}^{N}(\sum_{i=1}^{m} x_k^{(i)})^2 \bullet \sum_{k=1}^{N}(\sum_{i=1}^{m} y_k^{(i)})^2}}. \quad (6.7)$$

Next, we examine the correlation of *x* and *y* under Approach 2. For Approach 2, we first obtain estimated path connections through the unified SEM method for each subject individually, and then we compute the mean value of each path coefficient across all subjects (See Chapter 4 for details). For subject *i*, define $x^{(i)} = (x_1^{(i)}, ..., x_N^{(i)})'$, and $y^{(i)} = (y_1^{(i)}, ..., y_N^{(i)})'$. The correlation coefficient is

$$corr(x^{(i)}, y^{(i)}) = \frac{E(x^{(i)} y^{(i)})}{\sqrt{E((x^{(i)})^2) E((y^{(i)})^2)}} = \frac{\sum_{k=1}^{N}(x_k^{(i)} y_k^{(i)})}{\sqrt{\sum_{k=1}^{N}(x_k^{(i)})^2 \sum_{k=1}^{N}(y_k^{(i)})^2}}. \quad (6.8)$$

Since for SEM the bigger the correlation coefficient the more chance to be significant the corresponding path connection, we examine the average value of correlation coefficients of $x^{(i)}$ and $y^{(i)}$ for all subjects in order to obtain a rough idea of the significance of correlation across all subjects. The average of correlation coefficients of all subjects is given by

$$\frac{1}{m}\sum_{i=1}^{m} corr(x^{(i)}, y^{(i)}) = \frac{1}{m}\sum_{i=1}^{m} \frac{\sum_{k=1}^{N}(x_k^{(i)} y_k^{(i)})}{\sqrt{\sum_{k=1}^{N}(x_k^{(i)})^2 \sum_{k=1}^{N}(y_k^{(i)})^2}}. \quad (6.9)$$

Comparing Equation (6.7), which is the correlation from Approach 1, and Equation (6.9), which is from Approach 2, the two equations are not equivalent to each other. We now illustrate it by a simple example.

Let $x^{(1)} = (1, 1, 1, 1, 1, 1, 1, 1)'$, $y^{(1)} = (1, 2, 3, 4, 5, 6, 7, 8)'$, $x^{(2)} = (1, 2, 3, 4, 5, 6, 7, 8)'$, and $y^{(2)} = (1, 1, 1, 1, 1, 1, 1, 1)'$. The average of $x^{(1)}$ and $x^{(2)}$ is $x^{ave} = (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5)'$ and the average of $y^{(1)}$ and $y^{(2)}$ is $y^{ave} = (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5)'$. In this example, $x^{(1)}$ and $y^{(1)}$ are not correlated, $x^{(2)}$ and $y^{(2)}$ are not correlated either. However, the subject averages, x^{ave} and y^{ave}, are perfectly correlated. Although it is a hypothetical example, it shows that the correlation of subject average could be far from the pattern of individual subject correlation. The correlation of subject average could be weaker or stronger than the pattern of individual subject correlation.

Back to our fMRI data, we now examine the contemporaneous path connecting THAL to PPC, which is significant from Approach 1 but not significant from Approach 2. The correlation coefficient of the subject-average THAL and PPC from Approach 1 is –0.4309. The correlation coefficient of THAL and PPC, however, varies over subjects. For instance, the correlation of THAL and PPC is -0.0175 for subject 5, 0.0014 for subject 11, -0.1280 for subject 15, and 0.1344 for subject 23. However, the correlation coefficient of the subject-average THAL and PPC averaging subject 5 and subject 11 is 0.2766, which is far bigger than the individual correlation coefficient alone; the correlation coefficient of the subject-average THAL and PPC averaging subject 5 and subject 15 is –0.2732, which is stronger than the individual correlations, and the correlation coefficient of the

subject-average THAL and PPC averaging subject 5 and subject 23 is 0.0179, which is weaker than the correlation for subject 23. Therefore, by taking subject average for THAL and PPC, the correlation as well as covariance pattern could be distorted. Meanwhile, we found that this path, THAL to PPC, is significant for only 5 subjects (three females and two males) out of 28 subjects. And without surprise, this path is not significant by Approach 2. We also found that the gender effects on path connections by using Approach 1 (from Table 4.2) are different from those (from Table 4.4) obtained by Approach 2. Only one path is in common for both approaches. This could also be explained by the fact that taking subject average would distort the true correlation structures and hence produce unreliable path estimations.

By investigating the correlation of subject average from Approach 1, and the correlation of each individual subject from Approach 2, we believe that the conclusion obtained by Approach 1 could be misleading. When a particular path connection is significant (or equivalently, nonsignificant) across most of the subjects, such as the longitudinal paths connecting each region to itself, which are significant for every subject, Approach 1 and Approach 2 would reach to the same conclusion. However, when the path connection is significant only for some subjects, such as the contemporaneous path connecting THAL to PPC as we described above, the correlation of subject average for this path could be far different from the true pattern, and therefore Approach 1 could lead to wrong conclusions.

Although Approach 1 is conceivable, simple to implement, and some people do use this approach in their research, we proved that it is not appropriate under some

circumstances, and hence, not recommended for fMRI data analysis.

6.2 Comparison between Approach 2 and Approach 3

In Chapter 5 we present three approaches under the framework of Approach 3, which are:

(i) Approach 3-1: Single-level multi-subject unified SEM approach,

(ii) Approach 3-2: Multilevel covariance structural analysis with unified SEM, and

(iii) Approach 3-3: Multilevel random-effects approach.

As we discussed in Chapter 5, Approach 3-2 is superior to Approach 3-1, since Approach 3-1 is a special case of Approach 3-2 when all the observations are independent. Therefore, we only consider the comparisons among Approach 2, Approach 3-2 and Approach 3-3. First we compare the results from Approach 2 with those from Approach 3-2 of our visual attention study.

Table 6.2 lists the significance of longitudinal and contemporaneous path parameters with their estimations, standard errors, t test statistics, and corresponding p values for Approach 2 and Approach 3-2. The bold characters indicate the significant paths at the significance level of 0.05. The path models with the significant paths from the two approaches are displayed in Figure 6.2. The t test statistics and the corresponding p values are calculated in the unified SEM model for Approach 3-2. In contrast, the t test statistics and the corresponding p values are calculated from one sample t test for Approach 2 (See Chapter 4 and Chapter 5 for details).

From Table 6.2 and Figure 6.2, we know that all the significant paths from Approach 2 are significant in Approach 3-2. The paths include the six longitudinal paths connecting each region to itself, the contemporaneous path connecting LPFC and SMA, and two longitudinal paths connecting APC to LPFC, and LPFC to SMA. Four contemporaneous paths are significant only in Approach 3-2, which form a single loop starting at THAL and connecting PPC, APC, and LPFC. One longitudinal path from LPFC to THAL is significant in Approach 3-2 only.

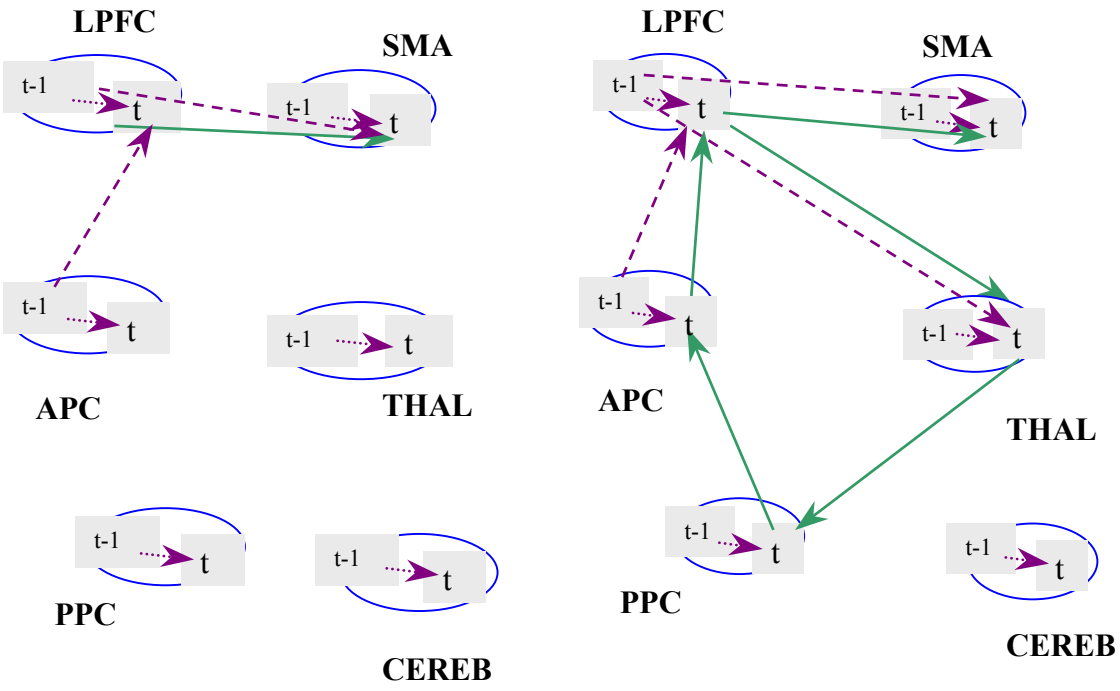


Figure 6.2 Significant path connections from Approach 2 (left picture) and Approach 3-2 (right picture). Dashed lines represent longitudinal path connections solid lines represent contemporaneous path.

Table 6.2 Comparisons of longitudinal and contemporaneous path parameters from Approach 2 and Approach 3-2 with their estimations, standard errors, t test statistics, and corresponding p values (two-sided).

Path	Approach 2		Approach 3-2	
parameters	Mean (S.E.)	T-value (p-value)	Est. (S.E.)	T-value (p-value)
β_{21}	0.047 (.076)	0.622 (.539)	0.073 (.041)	1.781 (.075)
β_{25}	0.101 (.071)	1.424 (.166)	**0.086 (.038)**	**2.287 (.022)**
β_{26}	0.030 (.087)	0.340 (.736)	0.050 (.039)	1.293 (.196)
β_{32}	-0.034 (.026)	-1.311 (.201)	**-0.040 (.016)**	**-2.528 (.012)**
β_{43}	-0.005 (.036)	-0.126 (.901)	**0.040 (.020)**	**1.979 (.048)**
β_{54}	0.091 (.066)	1.384 (.178)	**0.112 (.039)**	**2.882 (.004)**
β_{65}	**0.229 (.054)**	**4.208 (.000)**	**0.209 (.025)**	**8.486 (.000)**
γ_{11}	**0.612 (.019)**	**31.462 (.000)**	**0.609 (.020)**	**29.985 (.000)**
γ_{21}	0.005 (.070)	0.074 (.942)	-0.040 (.041)	-0.992 (.321)
γ_{22}	**0.591 (.020)**	**29.391 (.000)**	**0.630 (.020)**	**31.287 (.000)**
γ_{25}	-0.120 (.070)	-1.718 (.097)	**-0.074 (.038)**	**-1.976 (.048)**
γ_{26}	0.028 (.082)	0.339 (.737)	-0.027 (.039)	-0.689 (.491)
γ_{32}	0.035 (.029)	1.182 (.248)	0.024 (.016)	1.574 (.116)
γ_{33}	**0.595 (.021)**	**28.821 (.000)**	**0.618 (.020)**	**30.371 (.000)**
γ_{43}	0.031 (.036)	0.872 (.391)	0.006 (.020)	0.303 (.762)
γ_{44}	**0.593 (.016)**	**36.001 (.000)**	**0.613 (.021)**	**29.661 (.000)**
γ_{54}	**−0.124 (.046)**	**-2.683 (.012)**	**−0.133 (.039)**	**-3.448 (.000)**
γ_{55}	**0.615 (.019)**	**31.685 (.000)**	**0.643 (.020)**	**32.394 (.000)**
γ_{65}	**-0.166 (.047)**	**-3.517 (.002)**	**-0.180 (.025)**	**-7.247 (.000)**
γ_{66}	**0.578 (.018)**	**31.140 (.000)**	**0.580 (.021)**	**27.354 (.000)**

Bold characters indicate the significant path parameters at the significance level of 0.05.

Since Approach 3-2 incorporates all the subjects simultaneously (See Chapter 5 for details), it has much larger sample size and hence smaller standard error for each

estimated path coefficient, leading to more significant path connections. Therefore, some paths are not significant in Approach 2 but significant in Approach 3-2; and those paths that are significant in Approach 2 are still significant in Approach 3-2.

Approach 3-2 is built on the assumption that only intercept of SEM model is random over subjects, however, from the results of Chapter 5 we know this assumption is not appropriate in our visual attention data. Therefore, the significant path connections found by Approach 3-2 may not be truly significant.

Next we compare Approach 2 with Approach 3-3 based on the results of our visual attention study.

Due to the lack of sufficient sample size of subjects, to apply Approach 3-3, we only test the significant paths detected via Approach 3-2. Table 6.3 lists the significance of longitudinal and contemporaneous path parameters with their estimations, standard errors, t test statistics, and corresponding p values for Approach 2 and Approach 3-3. The bold characters indicate the significant paths at the significance level of 0.05. The significant paths from Approach 3-3 and the significant paths from Approach 2 are displayed in Figure 6.3. These two approaches give identical conclusions.

For Approach 2, the path estimations for each subject are obtained at Stage 1, and the subject mean of path estimations for each path is tested by one sample t-test to see whether it is significantly different from zero at Stage 2. Meanwhile, the path estimations are merged with subject-level covariates to examine the impact of subject-level covariates on the pathways via a GLM. In contrast, the paths tested by Approach 3-3 are

Table 6.3 Comparisons of longitudinal and contemporaneous path parameters from Approach 2 and Approach 3-3 with their estimations, standard errors, t test statistics, and corresponding p values (two-sided).

	Approach 2		Approach 3-3	
path	Mean (S.E.)	t-value (p-value)	Est. (S.E.)	t-value (p-value)
β_{25}	0.101 (.071)	1.424 (.166)	0.097 (0.063)	1.545 (0.134)
β_{32}	-0.034 (.026)	-1.311 (.201)	-0.037 (0.023)	-1.602 (0.124)
β_{43}	-0.005 (.036)	-0.126 (.901)	0.002 (0.038)	0.047 (.963)
β_{54}	0.091 (.066)	1.384 (.178)	0.107 (0.065)	1.642 (0.113)
β_{65}	**0.229 (.054)**	**4.208 (.000)**	**0.210 (0.049)**	**4.252 (0.000)**
γ_{11}	**0.612 (.019)**	**31.462 (.000)**	**0.607 (0.027)**	**22.293 (0.000)**
γ_{22}	**0.591 (.020)**	**29.391 (.000)**	**0.629 (0.022)**	**28.383 (0.000)**
γ_{25}	-0.120 (.070)	-1.718 (.097)	-0.089 (0.072)	-1.242 (0.225)
γ_{33}	**0.595 (.021)**	**28.821 (.000)**	**0.6126 (0.022)**	**27.947 (0.000)**
γ_{44}	**0.593 (.016)**	**36.001 (.000)**	**0.606 (0.022)**	**27.602 (0.000)**
γ_{54}	**−0.124 (.046)**	**-2.683 (.012)**	**-0.134 (0.048)**	**-2.766 (0.010)**
γ_{55}	**0.615 (.019)**	**31.685 (.000)**	**0.628 (0.023)**	**27.929 (0.000)**
γ_{65}	**-0.166 (.047)**	**-3.517 (.002)**	**-0.175 (0.048)**	**-3.638 (0.001)**
γ_{66}	**0.578 (.018)**	**31.140 (.000)**	**0.585 (0.022)**	**26.783 (0.000)**

random effects, which are assumed to vary across subjects. Two significant paths, which are longitudinal and contemporaneous paths from LPFC to SMA are random paths detected by Approach 3-3, and their variations could be explained by a subject-level covariate, gender. The contemporaneous path from LPFC to SMA is also significantly correlated with gender from Approach 2.

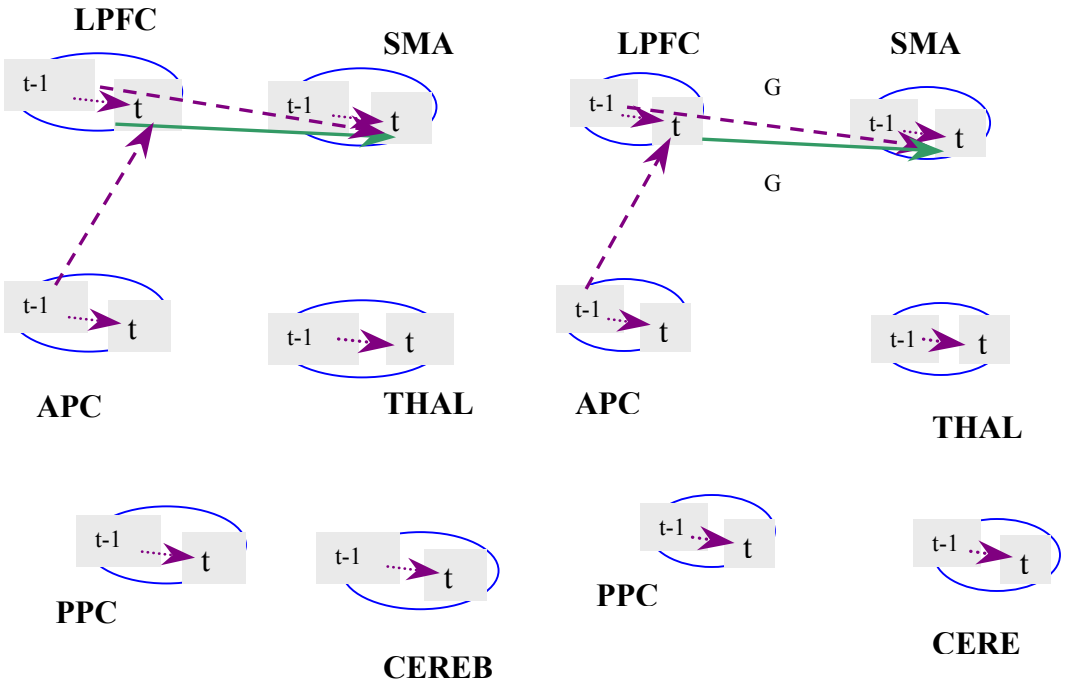


Figure 6.3 Significant path connections from Approach 2 (left picture) and Approach 3-3 (right picture). Dashed lines represent longitudinal path connections solid lines represent contemporaneous path.

Recall the univariate two-level random-effects model of Equation (5.31) and Equation (5.32) presented in Chapter 5:

$$Y_g = X_g\beta_g + r_g, \quad r_g \sim N(0,\sigma^2 I), \tag{6.10}$$

$$\beta_g = W_g\gamma + u_g, \quad u_g \sim N(0,T). \tag{6.11}$$

The generalized least squares (GLS) estimator of γ is

$$\hat{\gamma} = (\sum W_g^T \Delta_g^{-1} W_g)^{-1} \sum W_g^T \Delta_g^{-1} \hat{\beta}_g, \tag{6.12}$$

when Δ_g is known and

$$\Delta_g = \text{var}(\hat{\beta}_g) = T + V_g = T + \sigma^2 (X_g^T X_g)^{-1}. \tag{6.13}$$

The optimal estimator of β_g is given by

$$\beta_g^* = \Lambda_g \hat{\beta}_g + (I - \Lambda_g) W_g \hat{\gamma}\,, \tag{6.14}$$

where

$$\Lambda_g = T(T + V_g)^{-1}, \tag{6.15}$$

and

$$\hat{\beta}_g = (X_g^T X_g)^{-1} X_g^T Y_g \tag{6.16}$$

is the OLS estimator of β_g from Equation (6.10).

$\hat{\beta}_g$ is also the estimator of β_g at Stage 1 of Approach 2 for univariate case (multiple regression). When the data are perfectly balanced such that each subject have the same number of observations, the same values of the predictor matrix *X*, and the same set of subject-level predictors for each component of β_g, $V_g = \sigma^2 (X_g^T X_g)^{-1}$ would be the same for each subject, and each $\hat{\beta}_g$ would have the same dispersion Δ. In this case, the unique, minimum-variance, unbiased estimator of γ would be the OLS regression estimator of Equation (6.11) given $\hat{\beta}_g$ (Bryk and Raudenbush 1992). Therefore,

$$W_g \hat{\gamma} = \hat{\beta}_g\,, \tag{6.17}$$

and

$$\beta_g^* = \Lambda_g \hat{\beta}_g + (I - \Lambda_g) W_g \hat{\gamma} = \Lambda_g \hat{\beta}_g + (I - \Lambda_g) \hat{\beta}_g = \hat{\beta}_g\,. \tag{6.18}$$

Thus for the univariate model, we establish the equivalence of β_g^*, the estimator of β_g obtained by Approach 3-3, and $\hat{\beta}_g$, the estimator of β_g by Approach 2 under

assumptions of perfectly balanced data. However, the assumption that each subject have the same values of the predictor matrix X is too strict, which is not realistic for fMRI data. Interestingly, even though the assumptions of perfectly balanced data are not satisfied in our visual attention data, we still reach to the same conclusion by Approach 2 and Approach 3-3. Frison et al. (2005) also demonstrated the robustness of Approach 2 in univariate models.

At the end of this chapter, we summarize the advantages and disadvantages of Approach 2,Approach 3-2, and Approach 3-3.

The advantages of Approach 2 are as follows

(1) It is easy to implement,

(2) It can be used to analyze each subject individually, obtaining subject specific unified SEM model,

(3) We can examine the impact of subject-level covariates for any path via a GLM, and

(4) Although it is not a complete random-effects SEM model, its robustness has been shown by studies.

The disadvantages of Approach 2 are:

(1) The paths of the SEM model are not completely random over subjects, and

(2) When the observations are independent, it will give fewer significant paths due to the smaller sample size at Stage 1.

The advantages of Approach 3-2 are:

(1) It is a multilevel SEM model taking the hierarchical data structure into account,

(2) Both the within-subjects SEM model and the between-subjects SEM model can be fitted, and

(3) It can be used for latent variables.

The disadvantages of Approach 3-2 are:

(1) Although Approach 3-2 is a SEM approach, it assumes homogeneous within subject covariance structures for all subjects and it cannot analyze random path connections, and

(2) The between-subjects covariance matrix is required to be positive definite for computation of likelihood function by software, however, the estimator of between-subjects covariance is frequently not positive definite (Muthén 1994).

The advantages of Approach 3-3 are:

(1) It incorporates the hierarchical data structure and it takes the proper error variance structures into consideration, and

(2) It assumes that the paths are random variables, varying across subjects, and it can detect the subject covariates that explain the variations of random paths.

The disadvantages of Approach 3-3 are:

(1) Approach 3-3 can only examine the covariates effect for random paths. It cannot detect the covariate impact for non-random paths, and

(2) Due to the difficulties of developing typical maximum likelihood estimations of multilevel structural equation modeling with random intercepts and random slopes, Approach 3-3 is not a SEM approach.

In summary, Approach 3-3 and Approach 3-2 take the hierarchical structure of

fMRI data into consideration. Approach 3-3 can be used to analyze random paths while Approach 3-2 cannot. However, Approach 3-2 and Approach 2 are SEM models, while Approach 3-3 is not. Approach 3-3 can detect effects of subject-level covariates on brain pathways only when the corresponding paths are random effects. In contrast, Approach 2 can test any path with subject covariates via a GLM. Even though Approach 2 is not a completely random-effects model, it is robust for analysis of hierarchical fMRI data. When the main purpose it to test the subject covariates, Approach 2 is better than Approach 3-3. Approach 2 is a good alternative against Approach 3, and a simulation study is suggested to further investigate the robustness of Approach 2. We believe, by combining the results of Approach 2 and Approach 3 we would have a better understanding of fMRI data.

Chapter 7

Conclusion and Discussion

We have presented three conceivable methodological approaches for the analysis of multi-subject multivariate time series fMRI data. They are: (1) summarize (e.g. average the time series data across the subjects) and then analyze, (2) analyze and then summarize, and (3) simultaneous analysis.

As described in Chapter 1, structural equation modeling assumes the interactions are instantaneous in the sense that structural equation models are not time-series models. This is in contradistinction to analyze the human brain connectivity network from neurophysiological times series such as fMRI data. We adopt the unified SEM method, which introduces longitudinal MAR(p) components into SEM model, and we incorporate this method into the three main approaches.

Approach 1 is subject-average unified SEM approach, by averaging time series fMRI data across subjects and analyzing the reduced average data via unified SEM. Approach 2 is two-stage multi-subject unified SEM approach, which analyze each subject individually via unified SEM in Stage 1 and analyze the mean of path connections across subjects in Stage 2. We have demonstrated the difference between Approach 1 and Approach 2, pointing out that the correlation or covariance structure of original data

could be possibly changed by Approach 1. Therefore, under some circumstances, Approach 1 can provide misleading results.

For Approach 1 and Approach 2, we developed a two-level bootstrap method to examine the influence of subject-level covariates on the brain pathways by resampling the subjects and within subject fMRI observations simultaneously.

There are three approaches under the framework of Approach 3, which are (1) Approach 3-1: single-level multi-subject unified SEM approach, (2) Approach 3-2: multilevel covariance structural analysis with unified SEM, and (3) Approach 3-3: multilevel random-effects approach.

Approach 3-1 is a special case of Approach 3-2 when all the observations are independent. Approach 3-2 arises from multilevel SEM method, which takes the hierarchical data structure into account. It assumes homogeneous within subject covariance structures for all subjects and it can analyze random intercepts in SEM. Both within-subjects and between-subjects SEM models are fitted simultaneously. However, when there exists random paths, Approach 3-2 gives unreliable results.

Approach 3-3 is a complete random-effects model, however, it is not a SEM method in nature. Meanwhile, Approach 2 and Approach 3-2 are SEM methods, but they are not complete random-effects models. Fortunately, Approach 2 may be a robust alternative method to random-effects model even though it is equivalent to Approach 3-3 only when some strict assumptions are satisfied.

In this work, we have presented the applications of those approaches to test and compare path models from visual attention fMRI study. We used the initially suggested

path model for visual attention network with six ROIs in the left-brain hemisphere which were identified by previous literatures and experts in fMRI field. Since the goodness-of-fit of each SEM model is poor for the visual attention fMRI data, other related ROIs may exist and are not considered in the study. The overall model fit would be improved when a better-designed fMRI experiment is conducted. The sample size is small for each individual SEM model. We expect the model fit will be improved by taking more fMRI time points for each subject.

The three approaches are compared based on the results of their applications to this visual attention fMRI study. And the general outlines are provided.

Although numerous methods have been developed for fMRI data analysis, many problems are still not solved. We suggest a simulation study to further investigate the robustness of Approach 2, a good alternative against Approach 3-3. And a maximum likelihood estimator of multilevel SEM with random intercepts and random paths is badly in need of being developed. When this difficulty is conquered, we can analyze multi-subject multivariate time series fMRI data more adequately.

REFERENCES

[1] Anderson, T.W. (1958). An introduction to multivariate statistical analysis. New York: Wiley.

[2] Bauer, D.J., Preacher, K.J. & Gil, K.M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychological Methods*, (2006) Vol. 11, No. 2, 142-163.

[3] Bentler, P. M. (1995). EQS structural equations program manual. Encino, CA: Multivariate Software.

[4] Bentler, P. M., & Wu, E. J. C. (2002). EQS 6 for windows user's guide. Encino, CA: Multivariate Software.

[5] Bentler, P.M. and Weeks, D.G. (1980). Linear structural equations with latent variables. *Psychometrika*, 45, 289-308.

[6] Bentler, P.M. and Wu, E.J.C. (1993). EQS/Windows user's guide, Version 4. Cork, Ireland: BMDP Statistical Software.

[7] Bentler, P.M. (1992). EQS structural equation program manual. Cork, Ireland: BMDP Statistical Software.

[8] Bollen K A. (1989). Structural equations with latent variables. New York: John Wiley.

[9] Browne,M.W. (1984). Asymptotic distribution free methods in analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37:62-83.

[10] Bryk, A.S., & Raudenbush, S.W. (1992). Hierarchical linear models: applications and data analysis methods. Sage Publications Newbury Park, N. J.

[11] Bullmore, E.T., Rabe-Hesketh, S., Morris, R.G., Williams, S.C.R., Gregory, L., Gray, J.A., and Brammer, M.J., (1996a). Functional magnetic resonance image analysis of a large-scale neurocognitive network. *NeuroImage* 4:16-33.

[12] Bullmore, E.T., Vrammer, M.J., Williams, S.C.R., (1996b). Statistical methods of estimation and inference for functional MR image analysis. *Magn Reson Med* 35:261-277.

[13] Bullmore, E.T., Horwitz, B., Honey, G., Brammer, M., Williams, S., and Sharma, T. (2000). How good is good enough in path analysis of fMRI data? *NeuroImage* 11:289-301.

[14] Büchel, C., Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modeling and fMRI. *Cerebral Cortex Dec. V7*: 768-778;1047-3211.

[15] Büchel, C., Coull, J. T., & Friston, K. J. (1999). The predictive value of changes in effective connectivity for human learning. Science, 283, 1538-1541.

[16] Van Buuren, S. (1997). Fitting ARMA time series by structural equation models. *Psychometrika,* Vol 62, 215-236.

[17] Chang, L., Tomasi, D., Yakupov, R., Lozar, C., Arnold, S., Caparelli, E.C., Ernst, T. (2004): Adaptation of the attention network in human immunodeficiency virus brain injury. *Annals of Neurology* 56:259-272.

[18] Cronbach, L.J., & Webb, N. (1975). Between-class and within-class effects in a reported aptitude X treatment interaction: reanalysis of a study by G. L. Anderson. *Journal of Educational Psychology*, (1975) 67, 6, 717-24.

[19] Driver J., Mattingley, JB., (1998). Parietal neglect and visual awareness. *Nat*

Neurosci. May;1(1):17-22. review.

[20] Efron, B., Tibshirani, R.J. (1993). An introduction to the bootstrap. Chapman & Hall/CRC.

[21] Fletcher, P., Büchel, C., Josephs, O., Friston, K., Dolan, R. (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex* Mar 1999 9:168-178; 1047-3211.

[22] Friston, K.J., Stephan, K.E., Lund, T.E., Morcom, A., Kiebel, S. (2005) Mixed-effects and fMRI studies. *NeuroImage* 24 (2005): 244-252.

[23] Friston, K.J., Phillips, J., Chawla, D., Buchel, C. (1999) Revealing interactions among brain systems with nonlinear PCA. *Human Brain Mapping* 1999;8(2-3):92-7.

[24] Friston, K.J., Holmes, A.P., Price, C. J., Büchel, C., and Worsley, K.J. (1999). Multisubject fMRI studies and conjunction analyses. *NeuroImage* 10 (1999), 385-396.

[25] Friston, K.J., Buchel C., (2000). Attentional modulation of effective connectivity from V2 tot V5/MT in humans. *Proc.Natl.Acad.Sci. USA* 97,7591-7596.

[26] Friston, K.J., Mechelli, A., Turner, R., Price, C.J., (2000). Nonlinear responses in fMRI: the Balloon model, Volterra kernels and other hemodynamics. *NeuroImage* 12,466-477.

[27] Friston, K.J., 2002. Bayesian estimation of dynamical systems: an application to fMRI. *NeuroImage* 16,513-530.

[28] Friston, K.J., Harrison, L, and Penny, W., 2003. Dynamic causal modeling. *NeuroImage* 19 (2003) 1273-1302.

[29] Grafton S.T., Sutton J., Couldwell W., Lew M., Waters C. (1994). Network analysis of motor system connectivity in Parkinsonis disease: modulation of thalamocortical

interactions after pallidotomy. *Hum Brain Mapping* 2:45-55.

[30] Hartley, H.O., & Rao, J.N.K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, (Jun., 1967) Vol. 54, No. 1/2, 93-108.

[31] Hoge, R.D., (1998). Magnetic resonance imaging cerebral oxygen consumption and perfusion. Ph.D. dissertation. Department of Biomedical Engineering. McGill University.

[32] Holmes, A.P., Fristion, K.J., (1998). Generalisability, random effects and population inference. *NeuroImage* 7, S754.

[33] Honey, G.D., Fu, C.H.Y., Kim, J., Brammer, M.J. (2002). Effects of verbal working memory load on corticocortical connectivity modeled by path analysis of functional magnetic resonance imaging data. *NeuroImage* 17:573-582.

[34] Horwitz, B., Friston, K.J., and Taylor, J.G. (2000). Neural modeling and functional brain imaging: An overview. *Neural Netw*. 13:829-846.

[35] Hox, J. (2002). Multilevel analysis: techniques and applications. Lawrence Erlbaum Associates.

[36] Jennings, J. M., McIntosh, A. R., & Kapur, S. (1998). Mapping neural interactivity onto regional activity: an analysis of semantic processing and response mode interactions. *Neuroimage*, 7, 244-254.

[37] Jennrich, R.I., & Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, (1986), Vol. 42, No. 4, 805-820.

[38] Jöreskog, K.G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger and O. D. Duncan, eds., Structural Equation Models in the Social Sciences. New York: Academic Press, pp. 85-112.

[39] Jöreskog, K.G., and Sörbom, D. (1996). LISREL 8. User's reference guide.

Scientific Software International, Chicago.

[40] Kim, J., Zhu, W., Chang, L., Bentler, P., Ernst, T. (2006). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Human Brain Mapping*, 2006, Vol. 28, No. 2, 85-93.

[41] Kim, J. (2004). Path Analysis of the visual attention network for functional MRI data. Ph.D. dissertation. Department of Applied Math and Statistics: Stony Brook University.

[42] Kanwisher N., Wojciulik E. (2000). Visual attention: insights from brain imaging. *Nat Rev Neurosci*. Nov;1(2):91-100. Review.

[43] Kaplan, D. (2000). Structural equation modeling, foundations and extensions. California, Sage.

[44] Keesling, J. W. (1972). Maximum likelihood approaches to causal analysis. Ph.D. dissertation. Department of Education: University of Chicago.

[45] Laird, N.M., & Ware, J.H., (1982). Random-effects models for longitudinal data. *Biometrics*, (1982) Vol. 38, No. 4, 963-974.

[46] Lange, N. (1999). Statistical procedures for functional MRI. In C. T. W. Moonen, P. A. Bandettini, & G. K. Aguirre (Eds.), Functional MRI (pp. 301-335). New York: Springer Medicine.

[47] Longford, N.T., & Muthén, B. (1992). Factor analysis of clustered observations. *Psychometrika*, 57, 581-597.

[48] McArdle, J. J., and McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, 27: 234-251.

[49] McIntosh, A. R., Gonzalez-Lima, F. (1991). Structural modeling of functional neural pathways mapped with 2-deoxyglucose: effects of acoustic startle habituation on the auditory system. *Brain Research*, 547, 295-302.

[50] McIntosh, A.R., and Gonzalez-Lima, F., (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Hum.Brain Mapp*.2:2-22.

[51] McIntosh, A.R., Grady, C.L., Haxby, J.V., Ungerleider, L.G., and Howeitz, B., (1994). Network analysis of cortical visual pathways mapped with PET. *J.Neurosci*. 14:655-666.

[52] McIntosh, A. R. (1998). Understanding neural interactions in learning and memory using functional neuroimaging. *Ann. NY Acad. Sci*. 855: 556–571.

[53] Mechelli, A., Penny, W.D., Price, C.J., Gitelman, D.R., and Friston, K.J. (2002). Effective connectivity and intersubject variability: using a multisubject network to test differences and commonalities. *NeuroImage* 17 (2002), 1459-1469.

[54] Meijer, E., Van der Leeden, R., & Busing F.M.T.A. (1995), Implementing the bootstrap for multilevel models. *Multilevel Modeling Newsletter,* **7**, 2, June 1995.

[55] Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Presidential address to the Psychometric Society (July, 1989). *Psychometrika*, 54, 557-585.

[56] Muthén, B., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. Invited paper for the conference, "Multilevel Analysis of Educational Data," Princeton, NJ, April 1987. In D. R. Bock (Ed.), Multilevel Analysis of Educational Data (pp. 87-99). San Diego, CA: Academic Press.

[57] Muthén, B. (1990). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society meeting in Princeton, NJ, June 1990. UCLA Statistics Series 62.

[58] Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement.* 1991, 28, 4, pp. 338-354.

[59] Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox, & I. Kreft (Eds.), Multilevel Modeling, a special issue of Sociological Methods & Research, 22, 376-398.

[60] Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P.V. Marsden (Ed.), Sociological Methodology (pp. 267-316). Washington, DC: American Sociological Association.

[61] Muthén, L., & Muthén, B. (1998-2004). Mplus user's guide. Third Edition. Los Angeles, CA.

[62] Patterson, H.D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* (1971) 58(3):545-554.

[63] Penny, W., & Holmes, A.P., (2004). Random-effects analysis. *Human Brain Function*. Elsevier, San Diego, pp. 843-850.

[64] Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., & Congdon, R.T. (2004). HLM 6: Hierarchical linear and nonlinear modeling. Scientific software international.

[65] Schmidt, W.H. (1969). Covariance structure analysis of the multivariate random effects model. Unpublished doctoral dissertation, University of Chicago.

[66] Searle, S. R., Casella, G., & McCulloch, C. E. (1992), Variance components, New York: Wiley.

[67] Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger and O. D. Duncan, eds., *Structural Equation Models in the Social Sciences*. New York: Academic Press, pp. 69-83.