Stony Brook University

Structural Equation Modeling with Time Series Data

A Dissertation Presented

By

Tianyi 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

May 2010

Stony Brook University

The Graduate School

Tianyi 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

Stephen Finch
Chairperson of Defense
Professor, Department of Applied Mathematics and Statistics

Haiping Xing
Member
Assistant Professor, Department of Applied Mathematics and Statistics

Ellen Li
Outside Member
Professor, Department of Medicine

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Structural Equation Modeling with Time Series Data

By

Tianyi Zhang

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2010

Structural Equation Modeling (SEM), also commonly referred to as path analysis in the absence of latent variables, is a powerful multivariate analysis approach to explore and to confirm causal relationships. It imposes a structure on the covariance matrix and the imposed structure is subsequently validated by the data. In recent years, SEM has been extended to analyze autoregressive moving average (ARMA) time series data assuming time-constant path coefficients. The mechanism of ARMA-based SEM makes it the ideal procedure for the analysis of directional brain functional pathways based on the multi-subject, multivariate time series data generated through the functional magnetic resonance imaging (*f*MRI) studies. However the time-constant path coefficient assumption is unrealistic and overly restrictive. In this work, based on converting the overall SEM to the sectional SEM approach for vector ARMA(p, q) time series, we extend the ARMA-based SEM to allow time-varying coefficients (TVC). The statistical inference framework based on the maximum likelihood method is derived and the advantage of the novel TVC SEM approach is demonstrated through simulation studies. In addition, we also applied the new method to examine the brain visual-attention pathway based on an *f*MRI experiment conducted at the Brookhaven National Laboratory. Other than brain functional pathways studies, the TVC SEM method can be readily applied to analyze other longitudinal data such as the financial time series.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

Structural Equation Modeling (SEM) is a powerful multivariate analysis approach to confirm causal relationships. It is also commonly referred to as path analysis, especially when no latent variable is present. It imposes a structure on the covariance matrix and the imposed structure is subsequently validated by the data. In recent years, SEM tools have been employed to analyze autoregressive moving average (ARMA) time series data. The mechanism of ARMA-based SEM makes it the ideal procedure for the analysis of directional brain functional pathways based on longitudinal brain image data from the functional magnetic resonance imaging (*f*MRI) experiments.

Our research group (Kim, 2004; Kim et al., 2007; Zhang, 2007) have developed a unified SEM framework combining the contemporaneous brain functional pathways (instantaneous directed relations among brain regions) and the longitudinal brain functional pathways (between the same brain region and among different brain regions at adjacent time periods) represented by a vector autoregressive time series process of order p, VAR(p). In this thesis, we aim to extend our unified SEM paradigm in two directions: (1) enabling time varying (path) coefficient (TVC) for the contemporaneous pathways because intuitively, the relations among different brain regions could/should evolve as the subject's learning of a certain task (visual attention or mathematical computation) evolves while lying inside the fMRI scanner; (2) extend the longitudinal models from Vector AR(p) to the more versatile Vector ARMA(p,q) processes. In the process, we also discuss several other relevant issues including the analysis of a single versus multiple (-subject) time series data, the different approaches to the analysis of multiple (-subject) time series data, and partial correlation network analysis (PCNA) for data-driven pathway discovery – to complement the structural equation modeling (SEM) which is a hypothesis driven confirmatory data analysis approach.

1.1 Structural Equation Modeling

Structural equation modeling (SEM) is a statistical technique for testing and estimating causal relationships using a combination of statistical data and qualitative causal assumptions. It is a versatile multivariate statistical method including many common statistical procedures such as multivariate regression and confirmatory factor analysis etc. as special cases.

Early work of SEM can be traced to the geneticist Sewall Wright around 1921 under the name of "path analysis". Wright's accomplishments include pioneering the estimation of supply and demand equations with a treatment of identification and developing estimations for covariances of variables {Goldberger, 1972 #328}. In the 1970s, Jöreskog (1973), Keesling (1972), and Wiley (1973) combined factor analysis with

econometric simultaneous equation models, and thus greatly enriched and improved the methodology of SEM (Bollen, 1989; Bollen, 1993; Jöreskog, 1996; Loehlin, 2004). This tool is now available in several commercial software packages including LISREL, EQS, AMOS, and SAS (Bentler, 1995; Bentler, 2002; Jöreskog, 1996)

SEM is also suitable for models with latent variables (Bock and Bargmann, 1966; Bollen, 1989) – a scenario that cannot be easily handled by many traditional multivariate statistical techniques. In this thesis, however, we will focus on SEM with only observed variables (Bentler, 1983; Browne, 1982; Browne, 1984), and make it part of our future work to further extend the novel TVC SEM methods developed here to time series data with latent variables.

1.2 ARMA-Based SEM

The use of Structural Equation Modeling for autoregressive (AR), moving average (MA) and autoregressive-moving average (ARMA) data has been studied and discussed increasingly in the past ten years albeit the majority of the work concerns only with the longitudinal pathways, and not the joint longitudinal and contemporaneous pathways (Kim et al., 2007; Zhang 2007).

There are two ways in which SEM has been used to fit ARMA models, depending on whether we have only a single time series or multiple time series data from say, N independent subjects. A, what we call, "overall" SEM approach, has been developed for the multiple time series data where the number of samples, N, is much larger than the total number of time points, T. Du Toit and Browne (2007) presented the most comprehensive model specifications and the corresponding parameter estimations and statistical inference with an example analyzing the South African school children tests data. Other relevant early work includes Jöreskog (1971) and du Toit and Browne(2001).

The second approach was introduced to analyze ARMA data via SEM when there is only one time series, which is the standard problem in time series modeling. A, what we refer to as the "sectional" SEM approach, has been developed with its properties in parameter estimation and statistical inference thoroughly studied in both theory and simulation for the univariate problems during the past decade (Hamaker et al, 2002; Molenaar, 1999; Nesselroade et al., 2002; vanBuuren, 1997).

In this work, we introduce several principles for translating the overall SEM to the sectional SEM approach for ARMA (p, q) time series. Simulation studies demonstrated the advantage of the sectional SEM approach in terms of better computational efficiencies and convergence properties. Furthermore, we extend the SEM analysis of longitudinal and contemporaneous pathways to allow time-varying contemporaneous path coefficients (TVC). We also discuss the transformation from the TVC model to a time constant (path) coefficient (TCC) model, as subsequently, the transformation from an over TVC SEM model to a sectional TVC SEM, and then to a sectional TCC SEM

model. The novel TVC SEM model is subsequently applied to analyze a brain visual-attention pathway model based on an *f*MRI experiment conducted at the Brookhaven National Laboratory.

1.3 SEM in fMRI Data Analysis

The ultimate goal of brain functional connectivity studies is to propose, test, modify, and compare certain directional brain pathways. In the early 1990s McIntosh introduced SEM to neuroimaging to analyze effective brain connectivity based on data from positron emission tomography (PET), which is not longitudinal (Mcintosh and Gonzalezlima, 1991; McIntosh, 1994; Mcintosh et al., 1994). Subsequently SEM has quickly gaine dground in this field (Bullmore et al., 1996; Buchel and Friston, 1997; Jennings et al., 1998; Buchel et al., 1999; Fletcher et al., 1999; Bullmore et al., 2000; Honey et al., 2002; Grafton, 2004) especially with the advent of the less invasive, less expensive, and less restrictive functional magnetic resonance imaging (*f*MRI) modalities.

Nowadays the *f*MRI experiments are routinely conducted, usually with a group of subjects, to generate relevant data for brain functional pathway analysis (Gusnard and Raichle, 2001; Kim et al., 2000; Kwong et al., 1992; Malonek and Grinvald, 1996; Raichle and Mintun, 2006). During a typical functional *f*MRI experiment, each subject's functional activity in the brain is measured continuously at the rate of many images per second, over the course of several minutes to half an hour. The original brain imaging data measured at the voxel level are usually transformed to the brain regions of interest (ROIs) level for better noise control and more meaningful interpretations. Thus, one obtains a multivariate (multiple ROIs) time series per subject from the *f*MRI experiment (Bandettini et al., 1993; Bonvento et al., 2000; Rogowska and Wolf, 1992).

Conventional SEM procedure assumes independent observations and thus can not be applied directly to analyze auto correlated time series data. In the recent years, our group had successfully developed a unified SEM approach for analyzing multivariate time series data by combining longitudinal pathways represented by a multivariate autoregressive (MAR) model, and contemporaneous pathways represented by a conventional SEM (Kim, 2004; Kim et al., 2007; Zhang, 2007). A draw-back of our approach was that we had assumed time-constant path coefficients (TCC) for the contemporaneous pathways. In this work, we aim to extend our unified SEM approach to include time-varying path coefficients (TVC).

The rest of this dissertation is organized as follows. A general review of Structural Equation Modeling will be given in Chapter 2, including the matrix form notations, point estimates and statistical inference. Chapter 3 will review different approaches to analyze ARMA-based SEM. Chapter 4 will define, more rigorously, the "overall" and the "sectional" SEM and propose the rules of rewriting the overall SEM into the much

simpler sectional form. ARMA-based SEM with time-varying coefficients will be studied in Chapter 5. The TVC SEM will be applied to re-analyze the *f*MRI pathway, with results from different modeling approaches compared in Chapter 6. We also introduce there, an exploratory data analysis tool, partial correlation network analysis (PCNA), to refine and to correct the initial hypothesis of brain connections (based on field knowledge). The last chapter will conclude our findings and discuss possible future works.

Chapter 2 Structural Equation Modeling

The parameters θ to be estimated in structural equation modeling are connection strengths (or path coefficients), the variances and covariances of exogenous (independent) variables, and the variances and covariances of disturbance terms. The path coefficients are standardized partial regression coefficients, which representing the response of the endogenous (dependent) variables to a standard unit change in an exogenous variable, while the other variables in the model are held constant. The primary interests in SEM analysis lie in the estimation and test of the related path coefficients.

The variables 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. We will explain the estimation procedures in Section 2.2.

There are several common SEM software, including LISREL (Jöreskog, 1996), SAS CALIS and TCALIS procedures, EQS (Bentler, 2002), AMOS (Arbuckle, 1995), and Mx (Neale et al, 1999). Our group is in the process of compiling a joint partial correlation network analysis (PCNA) and SEM visual-analytic tool, entitled "BrainMiner" in collaboration with researchers at the Brookhaven National Laboratory, SBU Computer Science Department and the National Institute on Drug Abuse (NIDA). This software package will include the novel development in PCNA and SEM we have made over the past 10 years. Besides pathway analysis based on brain functional images, we have also started to explore biological pathway analysis based on molecular data in a systems biological study framework.

2.1 Model Description and Notations

In the structural equation models, we assume there are m endogenous (dependent) variables, denoted by a vector $\boldsymbol{y}$, and n exogenous (independent) variables, denoted by a vector $\boldsymbol{x}$. If we ignore the measurement errors (it is the common situation in practice), the model can be written as

$$y = \mathrm{B}y + \Gamma x + \zeta \tag{2.1}$$

Here $\mathbf{B}$ is an $m \times m$ matrix of coefficients for endogenous variables, while $\mathbf{\Gamma}$ is an $m \times n$ coefficient matrix for exogenous variables. The vector $\boldsymbol{\zeta}$ is the independent error terms. In addition, we denote $\mathbf{\Phi}$ as the covariance matrix of $\boldsymbol{x}$, and $\mathbf{\Psi}$ as the covariance matrix of $\boldsymbol{\zeta}$. These terms are summarized in Table 2.1.

Table 2.1 Notations and Definitions in Structural Equation Modeling

Symbol	Dimension	Definition
$\boldsymbol{y}$	$m\times 1$	endogenous variables
$\boldsymbol{x}$	$n\times 1$	exogenous variables
$\boldsymbol{\zeta}$	$m\times 1$	errors
$\mathbf{B}$	$m\times m$	coefficient matrix for endogenous variables
$\mathbf{\Gamma}$	$m\times n$	coefficient matrix for exogenous variables
$\mathbf{\Phi}$	$n\times n$	covariance matrix of $\boldsymbol{x}$
$\mathbf{\Psi}$	$m\times m$	covariance matrix of $\boldsymbol{\zeta}$

The model assumptions are further summarized as follows:

(1) The error terms have mean 0. ($\boldsymbol{x}$ and $\boldsymbol{y}$ are also centered in most cases, so we have $E(\boldsymbol{x})=E(\boldsymbol{y})=E(\boldsymbol{\zeta})=0$.)

(2) $\boldsymbol{\zeta}$ is uncorrelated to $\boldsymbol{x}$.

(3) $\mathbf{I}-\mathbf{B}$ is nonsingular.

(4) The free parameter matrices to be estimated are $\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

2.2 Point Estimation and Statistical Inference of Model Parameters

Let $\theta=\{\theta_i\}$ be the set of free parameters to be estimated in $\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and $\mathbf{\Psi}$. Further, let $\Sigma(\theta)$ be the covariance matrix of the observed variables $\boldsymbol{y}$ and $\boldsymbol{x}$ written as a function of the free model parameters θ. $\Sigma(\theta)$ is assembled in three pieces: (1) the covariance matrix of $\boldsymbol{y}$, (2) the covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{y}$, and (3) the covariance matrix of $\boldsymbol{x}$:

$$\Sigma(\theta)=\operatorname{cov}\begin{pmatrix} y \\ x \end{pmatrix}=\begin{pmatrix} \Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{pmatrix} \tag{2.2}$$

Notice Equation (2.1) can be rewritten as

$$y=(\mathrm{I}-\mathrm{B})^{-1}(\Gamma x+\zeta) \tag{2.3}$$

Hence,

$$\begin{aligned} \Sigma_{yy}(\theta) &= E(yy') \\ &= E[(\mathrm{I}-\mathrm{B})^{-1}(\Gamma x+\zeta)((\mathrm{I}-\mathrm{B})^{-1}(\Gamma x+\zeta))'] \\ &= E[(\mathrm{I}-\mathrm{B})^{-1}(\Gamma x+\zeta)(x'\Gamma'+\zeta')(\mathrm{I}-\mathrm{B})^{-1'}] \\ &= (\mathrm{I}-\mathrm{B})^{-1}(E(\Gamma xx'\Gamma')+E(\Gamma x\zeta')+E(\zeta x'\Gamma')+E(\zeta\zeta'))(\mathrm{I}-\mathrm{B})^{-1'} \\ &= (\mathrm{I}-\mathrm{B})^{-1}(\Gamma\Phi\Gamma'+\Psi)(\mathrm{I}-\mathrm{B})^{-1'} \end{aligned} \tag{2.4}$$

in which we make use of $E(x\zeta') = E(\zeta x') = 0$.

Similarly, we can derive

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

$$\Sigma_{xx}(\theta) = E(xx') = \Phi \tag{2.6}$$

Based on Equations (2.4), (2.5) and (2.6), the implied covariance matrix of ***y*** and ***x*** is

$$\Sigma(\theta) = \begin{pmatrix} (\mathrm{I}-\mathrm{B})^{-1}(\Gamma\Phi\Gamma'+\Psi)(\mathrm{I}-\mathrm{B})^{-1'} & (\mathrm{I}-\mathrm{B})^{-1}\Gamma\Phi \\ \Phi\Gamma'(\mathrm{I}-\mathrm{B})^{-1'} & \Phi \end{pmatrix} \tag{2.7}$$

We should find an estimated Σ "close" to the sample covariance matrix *S*. The fitting function $F(S, \Sigma(\theta))$ is defined to quantify the discrepancy between *S* and $\Sigma(\theta)$. The measurements of the discrepancy usually satisfy the following properties:

(i) $F(S, \Sigma(\theta))\geq 0$;
(ii) $F(S, \Sigma(\theta))=0$, if and only if $S=\Sigma(\theta)$;
(iii) $F(S, \Sigma(\theta))$ is a continuous function in *S* and $\Sigma(\theta)$.

Apparently, different $F(S, \Sigma(\theta))$ mean different estimation approaches. However, if the conditions listed above are satisfied, fitting functions will lead to consistent estimators of θ (Browne, 1984). Generally speaking, three fitting functions are widely used: they 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. Suppose the vector ***z*,** concatenating the vectors ***y*** and ***x***, with dimension $(m+n) \times 1$, follows the multivariate normal distribution, and we have *N* independent observations on ***z***. The derivation of the likelihood function is illustrated as follows.

z~MN(0, Σ), so its probability density function (*pdf*) is

$$f(z \mid \Sigma) = (2\pi)^{-\frac{m+n}{2}} \mid \Sigma \mid^{-\frac{1}{2}} \exp\left[\left(-\frac{1}{2}z'\Sigma^{-1}z\right)\right] \tag{2.8}$$

Let $z_1, z_2, \ldots, z_N$ be the N observations.

The likelihood function is built as

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{N} \left\{ (2\pi)^{-\frac{m+n}{2}} |\Sigma(\theta)|^{-\frac{1}{2}} \exp\left[\left(-\frac{1}{2} z' \Sigma^{-1}(\theta) z \right) \right] \right\} \\
&= (2\pi)^{-\frac{N(m+n)}{2}} |\Sigma(\theta)|^{-\frac{N}{2}} \exp\left[\left(\left(-\frac{1}{2} \right) \sum_{i=1}^{N} z' \Sigma^{-1}(\theta) z \right) \right]
\end{aligned}
\tag{2.9}
$$

After taking log of both sides of Equation (2.9), we have

$$
\begin{aligned}
\log L(\theta) &= -\frac{N(m+n)}{2} \log(2\pi) - \frac{N}{2} \log|\Sigma(\theta)| - \frac{1}{2} \sum_{i=1}^{N} z_i' \Sigma^{-1}(\theta) z_i \\
&= -\frac{N(m+n)}{2} \log(2\pi) - \frac{N}{2} \log|\Sigma(\theta)| - \frac{1}{2} \sum_{i=1}^{N} tr\left[z_i' \Sigma^{-1}(\theta) z_i \right] \\
&= -\frac{N(m+n)}{2} \log(2\pi) - \frac{N}{2} \log|\Sigma(\theta)| - \frac{N}{2} \sum_{i=1}^{N} tr\left[\frac{z_i z_i' \Sigma^{-1}(\theta)}{N} \right] \\
&= -\frac{N(m+n)}{2} \log(2\pi) - \frac{N}{2} \log|\Sigma(\theta)| - \frac{N}{2} tr\left[S^* \bullet \Sigma^{-1}(\theta) \right] \\
&= \text{constant} - \frac{N}{2} \left\{ \log|\Sigma(\theta)| + tr\left[S^* \bullet \Sigma^{-1}(\theta) \right] \right\}
\end{aligned}
\tag{2.10}
$$

Here $S^* = \frac{N-1}{N} S$, where S is the sample covariance matrix of z, as mentioned before, and in a large sample, we have $S^* \approx S$.

We can rewrite Equation (2.10) after removing the constant term, which will not affect the estimation. We also use S instead of S^*:

$$
\log L(\theta) = -\frac{N}{2} \left\{ \log|\Sigma(\theta)| + tr\left[S\Sigma^{-1}(\theta) \right] \right\} \tag{2.11}
$$

Comparing to the discrepancy rules described earlier, the fitting function based on the likelihood function can be set up as:

$$
F_{ML} = F(\theta) = \log|\Sigma(\theta)| + tr\left[S\Sigma^{-1}(\theta) \right] - \log|S| - (m+n) \tag{2.12}
$$

It is equivalent to minimize $F(\theta)$ and to achieve the maximum likelihood. A necessary condition for minimizing of $F(\theta)$ is to choose the estimates $\hat{\theta}$ in that the partial derivatives of $F(\theta)$ with respect to each of $\hat{\theta}_i$ are zero. That is,

$$
\frac{\partial F(\theta)}{\partial \theta_i} = 0, \quad i=1 \text{ to the dimension of } \theta. \tag{2.13}
$$

Based on the normality assumption, one can then explicitly test the significance of each path coefficient (i.e., whether the path coefficient equals to zero) using the likelihood ratio test. Numerical methods for these procedures are described in (Bollen, 1989) and adopted almost universally by LISREL, EQS and SAS. In the case of non-normal distributions, one can always resort to the bootstrap resampling methods and other nonparametric procedures such as permutation tests etc. (Bollen, 1989; Bentler and Wu, 2002; Kim et al., 2007; Zhang, 2007). Similarly, the likelihood ratio test and its large sample chi-square approximation can be used to test the goodness-of-fit, and to compare nested models.

Chapter 3. ARMA-based SEM

The use of Structural Equation Modeling for autoregressive (AR), moving average (MA) and autoregressive-moving average data has been increasingly studied and discussed in the past ten years. There are advantages of using SEM for ARMA time series data: SEM has varies of powerful software support; and it is also straightforward in SEM to extend the ARMA model into multivariate case with both testable longitudinal and contemporaneous pathways (Kim et al., 2007; Zhang 2007).

There are two ways in which SEM has been used to fit ARMA models, based on different input time series data. The first application is used when the time steps T is small while the number of observations on each series is large. An "overall" SEM is used in this case. Du Toit and Browne (2007) showed the model specifications and parameter estimations with an example analyzing the South African school children test data.

The second application of SEM was introduced to analyze ARMA data when there is only one observation on each series, which is the standard problem in time series modeling. A "sectional" SEM approach was developed and the properties of the estimates thoroughly studied in both theory and data simulation (vanBuuren, 1997; Molenaar, 1999; Hamaker et al, 2002). This method is also applicable if there are N observations, but N is too small to have a positive definite sample covariance matrix.

It is noticeable that different researchers used the term "fitting ARMA data by SEM" to represent different meanings in previews work as stated above. Although both of the two approaches employ SEM tools for ARMA series, the essential ideas differ as Hamaker and colleagues (Hamaker et al, 2003) have pointed out. The terminology "overall" and "sectional" SEM are coined in this thesis to stress such difference, as it directly relates to the requirement on the input data and the process of the pathway modeling. In addition, it is also shown that the transformation from the overall SEM to the sectional SEM can be done following several principles.

3.1 Basics of ARMA

Autoregressive moving average (ARMA) models, also referred to as the Box-Jenkins models after the iterative Box-Jenkins estimation method, are a general class of time series data (Box and Jenkins, 1976). The model consists of two parts, an autoregressive (AR) part representing the influence of the previous observations to the future ones, and a moving average (MA) part representing independent 'shocks' to the times series. The model is often denoted as the ARMA(p, q) model where p is the order of the

autoregressive part and q is the order of the moving average part (as defined below).

The notation AR(p) refers to an autoregressive model of order p written as:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \tag{3.1}$$

where $\varphi_1, \cdots, \varphi_p$ are the parameters of the model, c is a constant and ε_t is white noise. The constant term is usually omitted for simplicity since data can be centered to make c = 0. Some constraints are necessary on the values of the parameters of this model in order that the model remains stationary.

The term MA(q) refers to the moving average model of order q:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{p} \eta_i \varepsilon_{t-i} \tag{3.2}$$

where $\eta_1, \ldots, \eta_q$ are the parameters of the model, μ is the expectation of X_t (often assumed to equal 0 as c in the AR(p) model), and the ε_t, ε_{t-1},... are again, white noise error terms.

ARMA(p, q), refers to the model with p autoregressive terms and q moving average terms, can thus be written as:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{j=1}^{p} \eta_j \varepsilon_{t-j} \tag{3.3}$$

Using the backward shift operator B, which is defined by $B^s X_t = X_{t-s}$, the ARMA(p,q) model can be rewritten as

$$\varphi(B) X_t - c = \eta(B) \varepsilon_t \tag{3.4}$$

where $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$ and $\eta(B) = 1 + \eta_1 B + \eta_2 B^2 + \cdots + \eta_q B^q$.

Finding appropriate values of p and q in the ARMA(p, q) model can be facilitated by plotting the partial autocorrelation functions for an estimate of p, and likewise using the autocorrelation functions for an estimate of q. Further information can be gleaned by considering the same functions for the residuals of a model fitted with an initial selection of p and q (Box and Jenkins, 1976; Mills, 1990; Pandit and Wu, 1983).

After choosing p and q, ARMA models in general can be fitted by different methods, such as least squares, maximum likelihood function, etc. For a pure AR model the Yule-Walker equations may be used to provide a fit. It is generally considered a good practice to find the smallest values of p and q that provide an acceptable fit to the data. In this dissertation, although the theoretical results are derived for general ARMA(p,q) models, the analyses on real data focus on ARMA model with small p and q values, as indicated by our previous work (Kim et al., 2007; Zhang, 2007).

3.2 ARMA-based Overall SEM

As described in Chapter 2, structural equation modeling, as the name itself implies, is developed based on a system of equations. On the other hand, ARMA time series is modeled by linear equations, too. It is straightforward to apply SEM approach to ARMA series if it is a stationary series.

For the univariate case (Van Buuren 1997), we illustrate the modeling by a simple AR(p) time series. Let $(X_0, X_1, X_2, ..., X_T)$ be the time series to be analyzed. Then X_p is the p-th lagged variable of X_0.

The linear equations can be expressed in the matrix form as the following:

$$\begin{pmatrix} X_0 \\ X_1 \\ \vdots \\ X_{T-p} \end{pmatrix} = \begin{pmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_p & 0 & \cdots & 0 \\ 0 & \varphi_1 & \varphi_2 & \cdots & \varphi_p & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & \varphi_1 & \varphi_2 & \cdots & \varphi_p \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_{T-p} \end{pmatrix} \tag{3.5}$$

The variance-covariance matrix can be easily established, i.e. for the variables involved in the first row, $(X_1, X_2, ..., X_p, \varepsilon_0)$, the variance-covariance matrix is

$$\Phi = \begin{pmatrix} \tau_0 & \tau_1 & \cdots & \tau_{p-1} & 0 \\ \tau_1 & \tau_2 & \cdots & \tau_{p-2} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \tau_{p-1} & \tau_{p-2} & \cdots & \tau_0 & 0 \\ 0 & 0 & \cdots & 0 & \sigma^2 \end{pmatrix} \tag{3.6}$$

in which τ_i is the i-th order autocovariance in AR(p), and σ^2 is the variance of error ε_0. Notice τ_i are not free parameters but determined by the coefficients in Equation (3.5). Therefore, the SEM estimation process can be used as we described in Chapter 2. Figure 3.1 shows the path diagram for an AR(2) time series. One can analyze this model in the usual SEM approach with N independent observations if the sample size N is big enough. This is what we referred to as, the overall SEM approach.

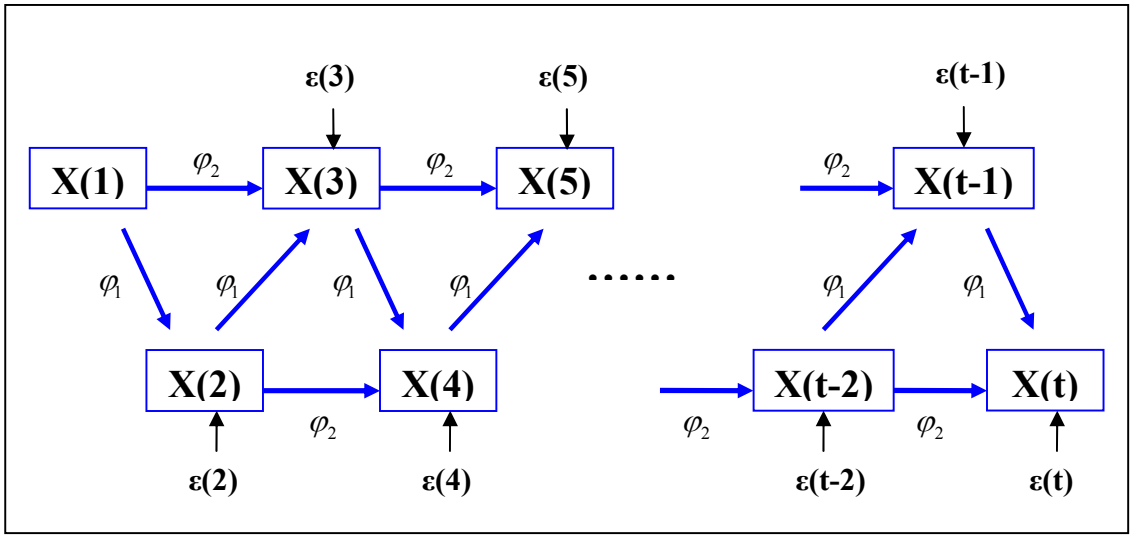


Figure 3.1 Path diagram for a univariate overall SEM for AR(2) time series

For a general multivariate ARMA(p, q) case, the notations are more complicated as discussed below. Let $\boldsymbol{Y_t}$ be an infinite multivariate ARMA(p, q) Gaussian time series:

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + \boldsymbol{u}_t + \sum_{j=1}^{q} B_j \boldsymbol{u}_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots \tag{3.7}$$

Notice it is a k×1 vector variant and u_t are noise vectors mutually independently normal distributed as $\boldsymbol{u}_t \sim N(\boldsymbol{0}, \boldsymbol{\Psi})$. Negative t values mean the status before time 0. Denote s = max(p, q). There are ks×1 observations before time 0, called the "initial state" vector $\boldsymbol{x_1}$. Let $\Theta = \mathrm{cov}(x_1, x_1')$.

Besides, suppose that there are n_T available observations on the time series, denote the following n_T×n_T matrices:

$$T_{-A} = \begin{pmatrix} I_k & 0 & 0 & 0 & \cdots & 0 \\ -A_1 & I_k & 0 & \ddots & \ddots & \vdots \\ \vdots & -A_1 & I_k & \ddots & 0 & 0 \\ -A_p & \vdots & -A_1 & \ddots & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & I_k & 0 \\ 0 & 0 & -A_p & \cdots & -A_1 & I_k \end{pmatrix} \tag{3.8}$$

$$T_B = \begin{pmatrix} I_k & 0 & 0 & 0 & \cdots & 0 \\ B_1 & I_k & 0 & \ddots & \ddots & \vdots \\ \vdots & B_1 & I_k & \ddots & 0 & 0 \\ B_q & \vdots & B_1 & \ddots & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & I_k & 0 \\ 0 & 0 & B_q & \cdots & B_1 & I_k \end{pmatrix} \tag{3.9}$$

We also denote the following $kn_T \times ks$ matrix

$$I_{n_T|s} = \begin{pmatrix} I_k & 0 & 0 \\ 0 & I_k & 0 \\ 0 & 0 & I_k \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ if } n_T = 6,\ s = 3. \tag{3.10}$$

Then, the covariance structure of $\boldsymbol{y}$ is given by (du Toit and Browne, 2007)

$$\Sigma = \mathrm{cov}(y, y') = T_{-A}^{-1}(I_{n_T|s}\Theta I_{n_T|s}^{'} + T_B(I_{n_T} \otimes \Psi)T_B^{'})T_{-A}^{-1\,'} \tag{3.11}$$

Comparing to Equation (2.5): $\Sigma_{yy}(\theta) = (\mathrm{I}-\mathrm{B})^{-1}(\Gamma\Phi\Gamma' + \Psi)(\mathrm{I}-\mathrm{B})^{-1'}$, the SEM tools can be directly applied to VARMA data based on this 1-1 mapping (Table 3.1).

Table 3.1 Relationship between SEM and VARMA parameters (du Toit and Browne 2007)

SEM	VARMA
$\mathrm{I}-\mathrm{B}$	T_{-A}
Γ	T_B
Φ	$I_{n_T} \otimes \Psi$
Ψ	$\begin{bmatrix} \Theta & 0 \\ 0 & 0 \end{bmatrix}$

3.3 ARMA-Based Sectional SEM

The limitation of the overall SEM is obvious: it requires a sample of independent

observations at each time point (the sample size requirement depends on the number of variables and parameters). However, in many cases, we only observe one series over time. This is also the typical data format for time series analysis. Therefore, another technique, which we refer to as the "sectional SEM" in this dissertation, was developed in parallel.

We discuss the univariate case first. From Equation (3.5), we will notice that the same equation repeats T-p+1 times with only increased subscripts. Hence, (3.5) can be rewritten as:

$$X_t = \sum_{i=t+1}^{t+p} \varphi_i X_i, \quad t = 0,1,...,T-p \tag{3.12}$$

Figure 3.2 shows the path diagram for the equation with p=1 and 2 (3.12). In other words, we have T-p+1 "observations" for this pathway model instead of a single time series. Similarly, for pure MA(q) or mixed ARMA(p, q) models, the "sectional" SEM can be constructed (See Figure 3.3).

In sectional SEM, the artificial observations: $(X_t, X_{t-1}, ..., X_{t-p}), (X_{t-1}, X_{t-2}, ..., X_{t-p-1}), ..., (X_{p+1}, X_p, ..., X_1)$ are not independent and the covariance matrix, which is formed of the autocorrelations, usually cannot be assumed to follow Wishart Distribution. However, theoretical and simulation results support the use of SEM for the "sectional" model. (Van Buuren 1997, Molenaar 1999, Hamaker et al., 2002). Van Buuren (1997) reported that the estimates are approximately unbiased and as efficient as those of specialized recursive maximum likelihood estimators for pure AR and mixed ARMA models, but biased and less efficient for pure MA models. Molenaar (1999) fixed the problem of MA models by explicit implementation of exact invertibility. Furthermore, the nature of sectional SEM output was discussed in Hamaker et al. (2002), in which the authors concluded that estimates of ARMA parameters obtained with SEM software are identical to those obtained by univariate stochastic model preliminary estimation (USPE) and are not true maximum likelihood estimators (MLE). For pure AR models, the estimates from SEM software have the same asymptotic properties as MLE, and the log-likelihood ratio is reliable. However, the behavior is not good for pure MA model and varies for mixed ARMA model.

It is not difficult to expand the modeling process above to the multivariate case. Van Buuren (1997) applied this method to a binary time series example. However, it is a danger to include redundant arrows during the modeling, which will lead to wrong results. Van Buuren (1997) has already pointed out this danger in a univarite AR(2) case, but in the multivariate case, the situation is more complicated. Figure 3.4 shows an example of redundant pathways in ARMA (1,1) model. In all of the figures in this dissertation, arrows in different colors represent different links: black for contemporaneous relations, blue for AR effects and green for MA effects. The rules of constructing a sectional SEM correctly from an overall SEM are part of this thesis's achievement, which will be introduced in Chapter 4.

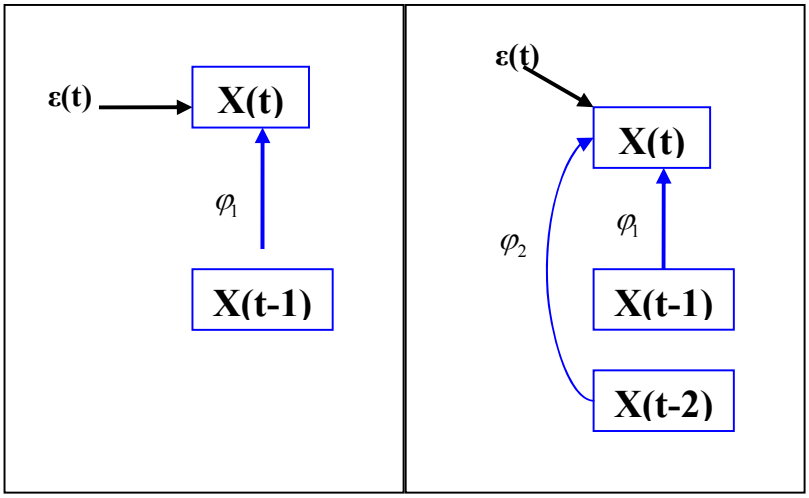


Figure 3.2 Path diagram for a univariate sectional SEM for the AR(1) time series (left) and the AR(2) time series (right).

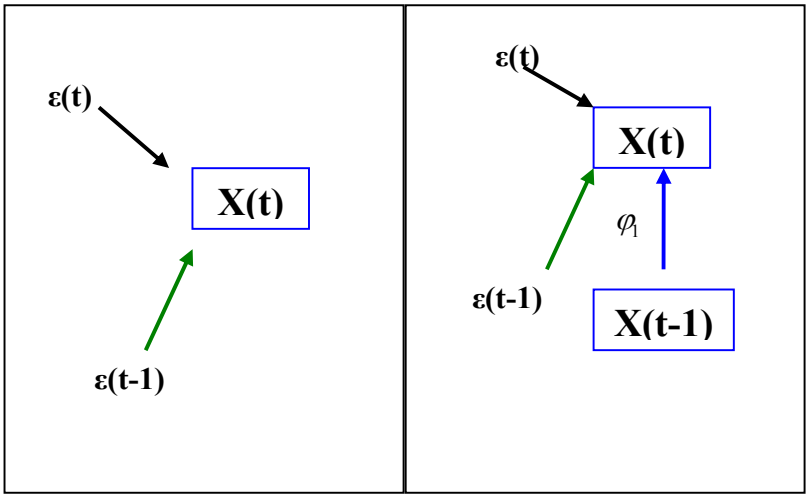


Figure 3.3 Path diagram for a univariate sectional SEM for the MA (1) time series (left) and the ARMA(1,1) time series (right).

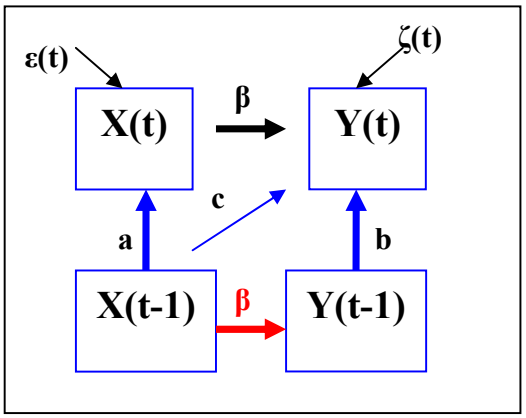


Figure 3.4 Path diagram for a bivariate sectional SEM for MARMA (1, 0) time series. (The red arrow denotes a redundant link.)

3.4 Discussion and Limitation

As stated above, the overall SEM requires multiple observations: generally speaking, the sample size N should be much larger than the time step T ($N >> T$). On the other hand, sectional SEM, which was applied for ARMA parameters estimation at the beginning, has been developed mainly for a single time series ($N = 1$ and $T > 50$). Besides, Hamaker et al (2003) also introduced a software, **Mx**, for directly estimating the model parameters through likelihood function for cases with $1 \leq N < T$. In this dissertation, we will apply the sectional SEM to multiple subject time series ($N > 1$) with three different approaches: 1) summarize then analyze; 2) analyze then summarize; 3) simultaneous analysis (Zhang 2007). We compare these different approaches by simulation in Chapter 4.

There is still room for discussions on the statistical inference for sectional SEM. As we stated before, the autocorrelation matrix usually does not follow the Wishart distribution. The hypothesis test and goodness of fit statistic we introduced in Chapter 2 may fail, especially for MA process (Hammaker et al., 2002; du Toit and Browne 2007). In practice, for *f*MRI study using the traditional experimental setting where the stimulus was applied in a predictable fashion, we found that the AR(1) model is usually sufficient for the inherent longitudinal effects and does not make the model too complicated meanwhile. Hence, we can still use the χ^2 value as the measurement of goodness-of-fit since the asymptotic property has been proven for pure AR models. However, for the increasingly popular new fMRI experimental paradigm with random shocks (i.e. the stimulus was applied at random in an un-predictable fashion), the MA component will become indispensible. We thus make it part of our future work to further develop the statistical inference for the general sectional SEM with ARMA(p,q) time series data.

Situations of different "initial status" before time 0 were also discussed in du Toit and Browne's paper (2007). They proposed three conditions concerning the initial state covariance matrix, $\boldsymbol{\Theta}$.

(1) $\boldsymbol{\Theta}$-free: No equality constraints on $\boldsymbol{\Theta}$, which means a change in process prior to the first observation is permitted;
(2) $\boldsymbol{\Theta}$=0: $\boldsymbol{\Theta}$ is fixed at 0, which means the process starts with the first observation;
(3) $\boldsymbol{\Theta}$(.): $\boldsymbol{\Theta}$ is estimated to keep the series stationary, which means the process started in the distant past with no change at the first observation.

Simulation study does not show big difference among these three conditions. In the *f*MRI visual attention study analyzed in Chapter 6, we will ignore the change before first observation and set no equality constraints on $\boldsymbol{\Theta}$.

Although several studies have been done on ARMA-based SEM, the majority use only time constant coefficients including our previous works (Kim et al., 2007; Zhang

2007). Two papers discussed the time varying path coefficient problems explicitly. Browne and du Toit (1991) adopted a time related exponential function for a structured latent learning curve – their time varying coefficients were applied exclusively to the longitudinal pathways, and not the contemporaneous pathways. The same team (du Toit and Browne 2007) later discussed the importance of having more general time varying coefficients –however, without proposing any explicit method.

In this thesis, we develop explicit modeling and analysis approaches for SEM with time varying contemporaneous path coefficients in Chapter 5. In Chapter 6, using a moving-window approach, we demonstrated that time varying coefficients do exist for the *f*MRI visual attention study. The TVC component was estimated and the TVC SEM model analyzed. However, first, we derive the guidelines for transforming an ARMA-based overall SEM to a sectional SEM based on a TCC SEM model in the ensuing Chapter 4.

Chapter 4. Transform ARMA-Based Overall SEM to Sectional SEM

Traditional SEM requires independent observations to guarantee a valid inference in general cases. ARMA-based overall SEM approach is a traditional SEM approach where the independent observations refer to the set of independent subjects where each subject has his/her own time series. Thus one marked disadvantage of the overall SEM is it requires big sample size, which is related to the number of variables and time steps. On the other hand, if the sample size is big enough, the overall SEM for multivariate analysis is usually intensely time consuming. For small sample size cases (i.e. $1 \leq N < T$), Hamaker et al (2003) introduced the software **Mx** for Maximum Likelihood Estimation.

In contrast, the ARMA-based sectional SEM can deal with a single time series with auto-correlated observations. The trick is that we first transform the longitudinal correlations into an SEM model, and subsequently combine this "longitudinal SEM pathways" with the contemporaneous SEM pathways into a unified SEM pathway diagram (as discussed in Kim et al., 2007 for AR-based sectional SEM). The single, auto-correlated time series data, after re-arrangement, can thus be plugged in as 'independent observations" where their independence is implied by the unified SEM model, in a traditional SEM software package. In this work, following the same approach, we transform the single- or multiple- time series ARMA-based overall SEM into ARMA based sectional SEM models. Guidelines for the transformation are provided.

Lastly, under the sectional-SEM framework, for multiple- time series data ($N > 1$), we discuss and compare three approaches for their analysis: (1) summarize then analyze; (2) analyze and summarize; (3) simultaneous analysis (Zhang, 2007). This and the above development are intended for TCC SEM only. Similar discussions are carried out in the later chapters for the more general TVC SEM.

4.1 Notations and Definitions

Let $X(t) = (X_1(t), X_2(t), ..., X_m(t))$ be the vector of m variables we are interested in. Our purpose is to confirm or explore a structural equation model involving these variables. Here, each variable $X_i(t)\ (i = 1, 2, ..., m)$ is a time series, representing observations at different time points. Generally speaking, $X_t := X(t)\ (t = 1, 2, ..., T)$ are auto-correlated, and are not independent observations.

We define the overall SEM as SEM based on the entire dataset, treating each time

point as a variable, or a node in the path diagram. In Figure 4.1, Model 1 is an example of an overall SEM when we have only two AR(1) variables.

We define the *p*-sectional SEM as the SEM approach in which:

(1) The time series data and paths in the overall SEM are split into *T-p* sections;

(2) Only $X(t), X(t-1), ..., X(t-p)$ are included in the structural equations in the (*t-p*)-th section;

(3) The path coefficients are time related or remain the same across all sections.

An example of a sectional SEM with two variables (Model 2) is also shown in Figure 1, in which $p = 1$. In a conventional definition of multivariate AR(*p*) (also referred to as the vector autoregressive or VAR) time series, there is no arrow from *X(t)* to *Y(t)* (e.g. (Brockwell and Davis, 1996; Hamilton, 1995). However, models containing contemporaneous links are more reasonable in practice and hence it is widely used in ARMA-based SEM (Van Buuren 1997, du Toit and Browne 2007). Besides, notice that there does not exist an arrow from *X(t-1)* to *Y(t-1)* in Model 2. A warning message was also given in Van Buuren (1997) that the redundant path included in sectional SEM would lead to wrong result. How to eliminate the redundant arrows is a key to rewrite an overall SEM to a sectional SEM. The proposed approach is introduced next.

In analyzing the visual attention pathway based on the same *f*MRI data as re-analyzed in this thesis, Kim et al (2007) and Zhang (2007) adopted a vector auto-regressive (VAR) model. Using a novel unified SEM approach, they assumed that only data with one lag between were involved based on an VARMA(1,0) model, so it is what we refer to as, a 1-sectional SEM (Figure 4.2.2). Figure 4.2.3 shows the overall SEM based on the same pathway assumptions.

The definition of the sectional SEM can be directly expanded to ARMA(*p, q*). Let $s = \max(p, q)$. The s-sectional SEM is defined as:

(1) The time series data and paths in the overall SEM are split into *T-p* sections;

(2) Only $X(t), X(t-1), ..., X(t-p)$ are included in the structural equations in the (*t-p*)-th section;

(3) The path coefficients are time related or remain the same across all sections.

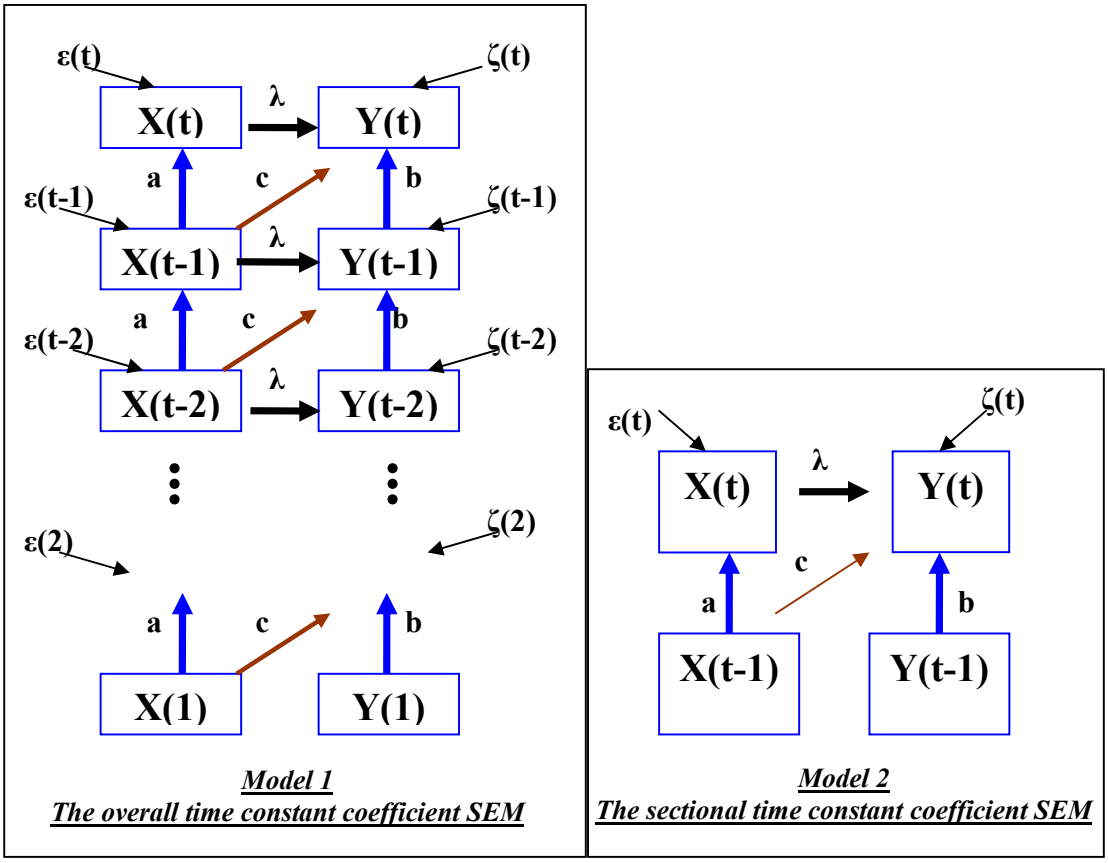
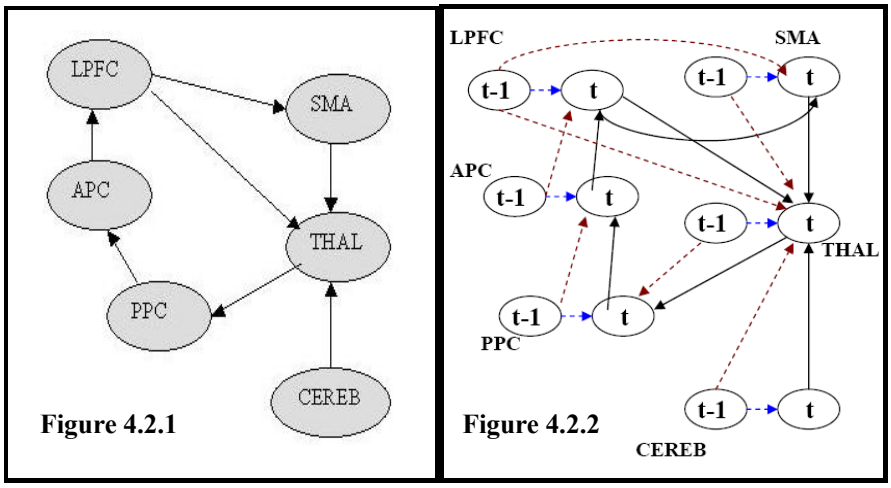


Figure 4.1 Examples of path diagrams for overall- (left) and sectional- (right) SEM with two AR(1) variables.



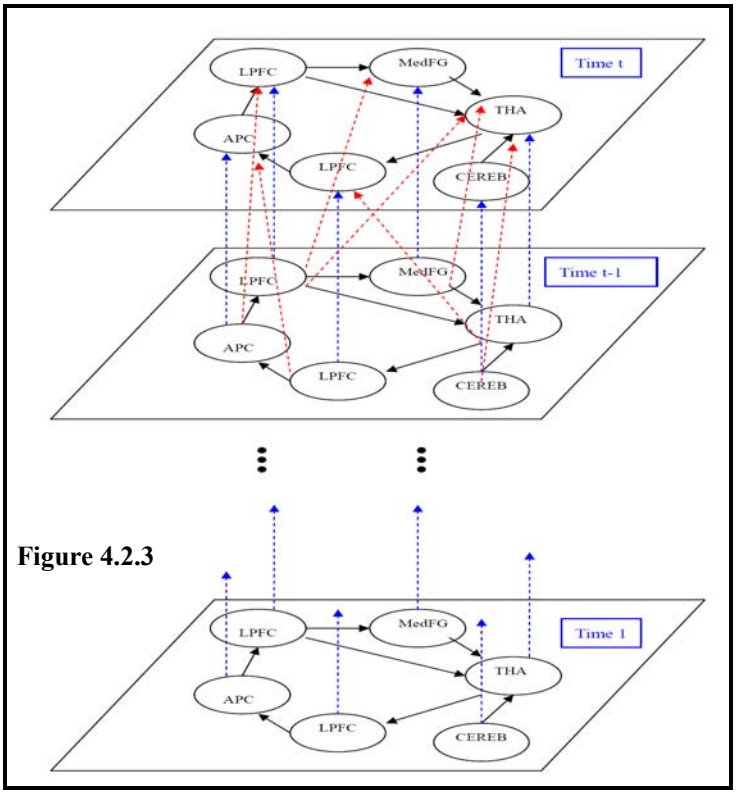


Figure 4.2.3

Figure 4.2 The path diagram in Figure 4.2.1 is the initial model of contemporaneous visual-attention pathways related to six brain regions. The path diagram in Figure 4.2.2 is the 1-sectional SEM (previously referred to as the unified SEM combining the contemporaneous and longitudinal pathways) based on VARMA(1,0) assumption (See Kim et al. 2007 and Zhang 2007). The path diagram in Figure 4.2.3 is the overall SEM based on the entire dataset.

4.2 Rewriting ARMA-Based Overall SEM to Sectional SEM

For the single time series (N = 1), SEM based on the overall path model yields no estimates. Sectional SEM needs to be adopted in this case. However, an assumption behind the p-sectional SEM approach is: we treat

$$(X_t, X_{t-1}, ..., X_{t-p}), (X_{t-1}, X_{t-2}, ..., X_{t-p-1}), ..., (X_{p+1}, X_p, ..., X_1)$$

as a series of "independent observations". We will show that the single time series overall SEM and the sectional SEM will have the same equations if we construct the corresponding sectional SEM correctly. We have the same (or equivalent) equations based on the path diagrams of the overall and the sectional SEM, which subsequently leads to the same point estimates and statistical inference.

Proposition 1. The overall SEM and the p-Sectional SEM have the same parameter estimates if the following requirements are satisfied.

a) Each variable is an AR(p) time series;

b) In the overall SEM, there is no path between variables having a time gap larger than p; in the p-sectional SEM, all variables at time t are endogenous, while all variables at other time points are exogenous;

c) The error terms are mutually independent.

Proof:

A. Univariate case:

If there is only one variable, it is just an AR(p) time series for the overall SEM. Assume the coefficients for AR(p) model are $(a_1, a_2, \ldots a_p)$. Then we have the following equations:

$$\begin{aligned} X(T) &= \sum_{i=1}^{p} \varphi_i \bullet X(T-i) + \varepsilon(T), \\ X(T-1) &= \sum_{i=1}^{p} \varphi_i \bullet X(T-1-i) + \varepsilon(T-1), \\ &\vdots \\ X(p+1) &= \sum_{i=1}^{p} \varphi_i \bullet X(p+1-i) + \varepsilon(p+1) \end{aligned} \tag{4.1}$$

Since the error terms are independent, the likelihood function is:

$$L(\theta) = \prod_{t=p+1}^{t=T} f(\varepsilon(t)) \tag{4.2}$$

For the sectional SEM, we have its path diagram as shown below (Figure 4.3).

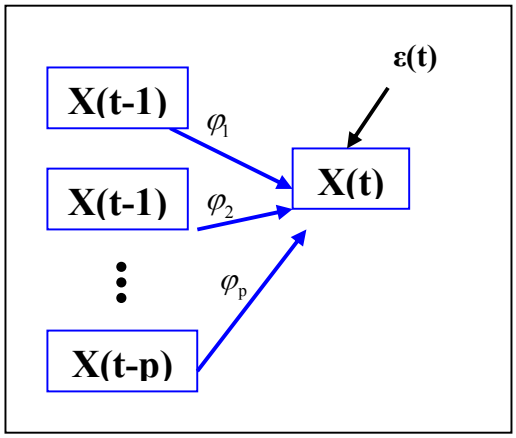


Figure 4.3 Univariate sectional model for AR(p) time series data.

Hence, the equations can be constructed:

$$X(t) = \sum_{i=1}^{p} \varphi_i \bullet X(t-i) + \varepsilon(t), \text{ where } t=p+1, p+2, \ldots T-1, T. \tag{4.3}$$

We treat the data as T-p "independent" observations, thus Equations (4.1) and (4.3) are entirely the same. The MLE of path coefficients are the same as those for the AR(p) coefficients. Our results are in agreement with van Buuren (1997) who has shown the

equivalency of SEM approach and ARMA(p, q) time series analysis for the univariate case.

B. Bivariate case:

Here we show the proof for the two variable case with VAR(1) data (Model 1 and Model 2 in Figure 4.1). The proof can be readily extended to the VAR(p) case.

Both the overall and the sectional SEM can be written in the following equations:

$$\begin{aligned} X(t) &= a \bullet X(t-1) + \varepsilon(t), \ i = 2,3,....,T \\ Y(t) &= b \bullet Y(t-1) + c \bullet X(t-1) + \lambda \bullet X(t) + \varsigma(t), \ t = 2,3,....,T \end{aligned} \tag{4.4}$$

The likelihood functions are the same assuming independent error terms.

C. Multivariate case:

Proposition 1 can be similarly proved in a general multivariable case. Both the overall and sectional SEM can be written in linear equations:

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_m(t) \end{pmatrix} = \mathrm{B} \bullet \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_m(t) \end{pmatrix} + \sum_{i=1}^{p} \Gamma(i) \bullet \begin{pmatrix} X_1(t-i) \\ X_2(t-i) \\ \vdots \\ X_m(t-i) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_m(t) \end{pmatrix}, \ t = p+1, p+2, ..., T \tag{4.5}$$

where B and Γ(i) are coefficient matrix. The likelihood functions are the same assuming independent error terms.

Similarly, for a ARMA(p, q) series, let s=max(p, q). The Proposition 1 could be expanded as follows:

Proposition 2: The overall SEM and the s-Sectional SEM have the same parameter estimates if the following requirements are satisfied.

a) Each variable is an ARMA(p, q) time series;

b) In the overall SEM, there is no path between variables having a time gap larger than p and no error term having a time gap larger than q; in the s-sectional SEM, all variables at time t are endogenous, while all variables at other time points are exogenous;

c) The error terms are mutually independent.

The proof is similar to the AR(p) case.

A. Univariate case:

If there is only one variable, it is just an ARMA(p, q) time series for the overall SEM. Assume the coefficients for ARMA(p, q) model are $(\varphi_1, \varphi_2, ...\varphi_p)$, $(\eta_1, \eta_2, ...\eta_q)$ then we have the following equations:

$$X(T) = \sum_{i=1}^{p} \varphi_i \bullet X(T-i) + \sum_{j=1}^{q} \eta_j \varepsilon(T-j) + \varepsilon(T),$$

$$X(T-1) = \sum_{i=1}^{p} \varphi_i \bullet X(T-1-i) + \sum_{j=1}^{q} \eta_j \varepsilon(T-1-j) + \varepsilon(T-1), \qquad (4.6)$$

$$\vdots$$

$$X(s+1) = \sum_{i=1}^{p} a_i \bullet X(s+1-i) + \sum_{j=1}^{q} \eta_j \varepsilon(s+1-j) + \varepsilon(s+1)$$

Since the error terms are independent, the likelihood function is:

$$L(\theta) = \prod_{t=p+1}^{t=T} f(\varepsilon(t)) \qquad (4.7)$$

For the sectional SEM, we have its path diagram as shown below (Figure 4.4).

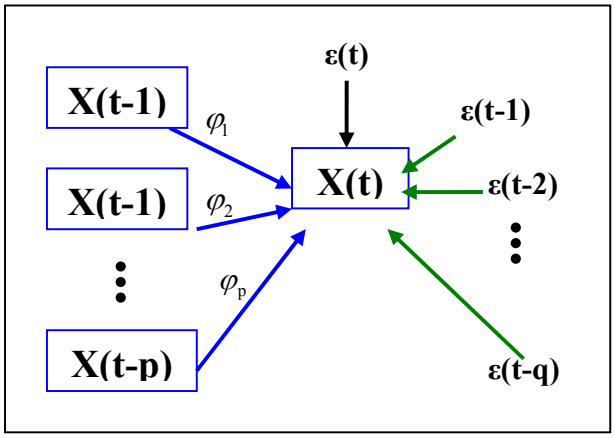


Figure 4.4 Univariate sectional model for ARMA(*p, q*) time series data.

Hence, the equations can be constructed:

$$X(t) = \sum_{i=1}^{p} \varphi_i \bullet X(t-i) + \sum_{j=1}^{q} \eta_j \varepsilon(t-j) + \varepsilon(t) \text{ , where } t=s+1, s+2, \ldots T-1, T. \qquad (4.8)$$

We treat the data as *T-s* "independent" observations, thus Equations (4.6) and (4.8) are entirely the same. The MLE of path coefficients are the same as those for the ARMA(*p, q*) coefficients. Our results are again in agreement with van Buuren (1997) who has shown the equivalency of SEM approach and ARMA(*p*, *q*) time series analysis for the univariate case.

B. Multivariate case:

Similarly, both the overall and sectional SEM for vector ARMA (*p, q*) series can be written in linear equations:

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_m(t) \end{pmatrix} = \mathrm{B}\bullet\begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_m(t) \end{pmatrix} + \sum_{i=1}^{p}\Gamma(i)\bullet\begin{pmatrix} X_1(t-i) \\ X_2(t-i) \\ \vdots \\ X_m(t-i) \end{pmatrix} + \sum_{j=1}^{q}\mathrm{H}(j)\begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_m(t) \end{pmatrix},\ t = s+1, s+2, ..., T \quad (4.9)$$

where B, $\Gamma(i)$ and $\mathrm{H}(j)$ are coefficient matrix. Assuming the error terms are independent and the VARMA (p, q) series are invertible, the likelihood functions will be the same.

As we have proven above, the overall and the sectional SEM yield the same equations for a single time series (N = 1). The sectional SEM also leads to much simpler path diagram. In addition, now that we can assume "independent" observations in the sectional SEM approach, nonparametric procedures such as the bootstrapping resampling method can be readily applied.

4.3 Sectional SEM with Multiple Observations

Since the sectional SEM approach can analyze a single time series data, naturally, it can also be readily applied to multiple time series data. Previously, our group has developed three strategies (Zhang 2007) for applying the sectional SEM to time series data with multiple subjects: (1) summarize and then analyze; (2) analyze and then summarize; (3) simultaneous analysis. The details are described as the following.

With Approach 1, summarize then analyze, the mean time series across all subject is used, leading to N = 1 again. The sectional SEM is then used to fit the average time series data, and the estimates and standard deviation are reported. This is the most straightforward and easiest approach to apply the sectional SEM towards multiple-subject time series data. It is also the most time efficient. However, any secondary analysis such as group comparisons or the analysis of other covariate (age, gender etc.) effects can only be done via bootstrap resampling or other indirect analysis approaches, as direct statistical inference is intangible.

With Approach 2, as was adopted by Kim et al. (2007), the sectional SEM is run for each time series (i.e. each subject) and the resulting path coefficients reported. Taking these path coefficients as input data (dependent variables), one could easily analyze covariate effects (age, gender, treatment groups etc.) through a general linear model approach. This will enable us to compare the pathway connection between different groups as well as examine potential influence of continuous covariates such as age, years of education, etc.

The third approach is the "simultaneous analysis" where the entire multi-subject, multivariate time series data are analyzed simultaneously. Zhang (2007) proposed several methods to analyze data simultaneously with subject-level covariates like gender, age, etc. In this dissertation, since we will not focus on covariate effects, the simplest simultaneous analysis method is adopted for Approach 3. All samples are concatenated,

so N*(T-1) samples in total are used in the sectional SEM at the same time. This concatenation scheme is a common practice in neuroscience (Penny and Holmes, 2004; Mechelli et al., 2002).

Each of these three approaches has its pros and cons. There is no universal winner in all although Approach 1 appears to be less informative than the other two in its dramatic dimension reduction although averaging will increase its robustness and bootstrap resampling can enable secondary analysis. Approach 3 often requires heavy model assumption before the simultaneous analysis can begin and thus is restrictive in that aspect – however, it does have more power thanks to its larger sample size. Approach 2 is informative and less restrictive, albeit it may suffer from the lack of power and robustness. A simulation study is conducted to compare these three approaches numerically for different sample size and time series size combination with rules of thumb provides, as introduced in the following section.

4.4 Simulation Study

Model 2 in Figure 4.1 is used in our simulation. The bivariate autoregressive series $X(t)$ and $Y(t)$ are generated from the formulas below.

$$\begin{aligned} &X(1) = 10 + \varepsilon(1) \\ &Y(1) = 5 + \varsigma(1) \\ &X(t) = a \bullet X(t-1) + \varepsilon(t),\ i = 2,3,....,T \\ &Y(t) = b \bullet Y(t-1) + c \bullet X(t-1) + \lambda \bullet X(t) + \varsigma(t),\ t = 2,3,....,T \end{aligned} \tag{4.10}$$

We set a=0.6, b=0.4, c=0.2, λ=0.5, $\varepsilon(t)$ and $\zeta(t)$ are random numbers following $N(0,1)$. We generate the dataset with T=100 time steps and N=1000 observations. N is set large enough, so the overall SEM could also be used for comparison purposes. SAS Proc Calis was used to estimate the parameters.

The entire simulation process contains four groups: (A) large T with large N; (B) small T with large N; (C) large T with small N; (D) small T with small N. We use T=100 for large T groups and T=50 for small T groups, N=1000 for large sample size groups and N=10 for small sample size groups.

In each group, 6 methods are used if applicable: (1) the overall SEM (T=50 or 100, N=10 or 1000); (2) sectional SEM model with Approach 1 (N=1, which is the mean of 10 or 1000 original samples), denoted as A.1; (3) sectional model with Approach 1 and bootstrapping (T=50 or 100, N=1, B=1000), denoted as Approach 1a or A.1a; (4) sectional model with Approach 1 and another scheme of bootstrapping to enlarge the sample size in each run from resampling: that means if T=50, we will enrich it to T=100 by resampling (T=100, N=1, B=1000), denoted as Approach 1b or A.1b; (5) sectional model with Approach 2 (T=50 or 100, N=10 or 1000), denoted as A.2; (6) sectional

model with Approach 3 (T=50 or 100, N=10 or 1000, so the concatenated sample size is (T-1)*N), denoted as A.3.

For Group(A), large T and large N, the outputs are shown in Table 4.1. All of the five models above yielded unbiased estimates compared to the true values. In addition, we can see bootstrapping will not improve the output significantly for the given data set (std increase in the third column compared to the second column) since the dataset itself is big enough. The sectional SEM with Approach 3 generated the smallest standard errors as expected (large power), however the model fit turns out poor (p-value<0.1) (due to heavy model assumption). Across three approaches for the sectional SEM, Approach 1 appears best for this simulation. This coincides with our suspicion of no clear and overall winner among the three approaches.

Table 4.1 Comparison of point estimates and standard deviation for Group(A) simulation.

Parameter	Point estimate (std)				
	Overall SEM	**Sectional SEM A.1**	**Sectional SEM A.1a**	**Sectional SEM A.2**	**Sectional SEM A.3**
λ	0.4969 (0.0051)	0.5205 (0.0995)	0.5204 (0.1075)	0.4970 (0.1008)	0.4961 (0.0032)
a	0.5998 (0.0041)	0.5962 (0.0027)	0.5954 (0.0083)	0.5828 (0.0581)	0.5986 (0.0018)
b	0.3947 (0.0041)	0.4017 (0.0056)	0.3963 (0.0186)	0.3822 (0.0744)	0.4005 (0.0023)
c	0.2050 (0.0041)	0.1885 (0.0596)	0.1948 (0.0695)	0.2183 (0.1019)	0.2038 (0.0032)

For Gourp (B), small T and large N, we use a truncated dataset with T=50. The six models above are applied, with the results shown in Table 4.2. We denote the biased estimates with yellow background and the insignificant ones with pink. We can see the sectional SEM with Approach 1 gives worse estimates, especially for path coefficients (λ and c) across variable (from X to Y). We can see the estimates of λ and c are biased, and c is even not significant for some methods. Bootrapping (A.1a) makes no improvement again. The other scheme of bootstrapping (A.1b) improves the output a little, but the estimates of λ and c are still biased. The sectional SEM with Approach 3 yielded the best estimates this time.

Table 4.2 Comparison of point estimates and standard deviation for Group(B) simulation.

Parameter	Point estimate (std)					
	Overall SEM	**Sectional SEM A.1**	**Sectional SEM A.1a**	**Sectional SEM A.1b**	**Sectional SEM A.2**	**Sectional SEM A.3**
λ	0.4925 (0.0058)	0.6182 (0.1593)	0.6170 (0.1596)	0.6263 (0.1071)	0.4937 (0.1429)	0.4944 (0.0045)
a	0.5999 (0.0047)	0.5969 (0.0026)	0.5971 (0.0043)	0.5970 (0.0018)	0.5772 (0.0684)	0.5985 (0.0021)
b	0.3934 (0.0048)	0.4003 (0.0058)	0.3889 (0.0251)	0.3975 (0.0086)	0.3647 (0.0981)	0.3976 (0.0031)

c	0.2044 (0.0070)	0.1307 (0.0899)	0.1479 (0.1047)	0.1300 (0.0077)	0.2329 (0.1316)	0.2064 (0.0042)

For Group (C), large *T* and small *N*, we use a dataset with *T*=100 but *N*=10. The overall model is not available this time, and the results of the sectional SEM are shown in Table 4.3. The estimates from the sectional SEM with Approach 1 are a little farther from the true values comparing to the results in Table 4.1, but still good. Bootstrapping (A.1a) does not improve the output. Approach 2 yielded similar results to Approach 1, while Approach 3 provided the best estimates.

Table 4.3 Comparison of point estimates and standard deviation for Group(C) simulation.

Parameter	**Point estimate (std)**			
	Sectional SEM A.1	**Sectional SEM A.1a**	**Sectional SEM A.2**	**Sectional SEM A.3**
λ	0.5439 (0.0989)	0.5427 (0.1055)	0.5362 (0.0960)	0.5375 (0.0301)
a	0.6264 (0.0256)	0.6322 (0.0274)	0.6035 (0.0594)	0.6329 (0.0286)
b	0.3678 (0.0493)	0.3749 (0.0494)	0.3654 (0.0742)	0.3898 (0.0231)
c	0.2062 (0.0800)	0.1913 (0.0901)	0.2321 (0.1044)	0.2113 (0.0332)

For Group (D), small *T* and small *N*, we use a dataset with *T*=50 and *N*=10. The overall SEM is not applicable again and the sectional SEM with Approach 1 generated several insignificant estimates. Bootstrapping (A.1a) improves a little but leads to biased estimates. Bootstrapping (A.1b) performed better than the above two, but the estimate for *c* is still not close to the true value. Approach 2 also yielded insignificant estimate, and Approach 3 has the best output again.

Table 4.4 Comparison of point estimates and standard deviation for Group(D) simulation.

Parameter	**Point estimate (std)**				
	Sectional SEM A.1	**Sectional SEM A.1a**	**Sectional SEM A.1b**	**Sectional SEM A.2**	**Sectional SEM A.3**
λ	0.5320 (0.3209)	0.5300 (0.1431)	0.4802 (0.0879)	0.4896 (0.1317)	0.5011 (0.0409)
a	0.6385 (0.0277)	0.6394 (0.0086)	0.6144 (0.0133)	0.5783 (0.0727)	0.6198 (0.0229)
b	0.3357 (0.0607)	0.3363 (0.0146)	0.3715 (0.0248)	0.3637 (0.0971)	0.3929 (0.0299)
c	0.2309 (0.2109)	0.2320 (0.0894)	0.2652 (0.0622)	0.2478 (0.1303)	0.2315 (0.0419)

Based on the simulation studies on Group (A) to (D), we summarize the model performance in Table 4.5. The one with shadow in each column is the best choice(s) in that case. We can conclude that for data with large *T*, the sectional SEM with Approach 1

is good enough. Approach 3 performs best across all the models (if N=1, it is equivalent to Approach 1). Bootstrapping can improve the output. Especially for data with small T, the scheme to enlarged T from resampling is useful.

Table 4.5 Summary of different SEM approaches for different datasets

Model		**Large T, Large N**	**Small T, Large N**	**Large T, Small N**	**Small T, Small N**
the overall SEM		unbiased estimates, but very time consuming	unbiased estimates, but not time efficient	not applicable	not applicable
the sectional SEM	Approach 1: summarize then analyze	unbiased estimates, small std and good model fit	biased estimates	unbiased estimates, small std and good model fit	big std and insignificant estimates
	Approach 1a: bootstrapping	no improvement	slight improvements	no improvement	biased estimates
	Approach 1b: enlarged T	not necessary	improve the estimates	not necessary	improve the estimates
	Approach 2: analyze then summarize	unbiased estimates, bigger std than Approach 1	unbiased estimates, std are big	unbiased estimates, std are big	unbiased estimates, std are big
	Approach 3: simultaneous analysis	unbiased estimates, small std but poor fit for the entire model	unbiased estimates, smallest std and good model fit	unbiased estimates, smallest std and good model fit	unbiased estimates, smallest std and good model fit

Chapter 5. ARMA-based SEM with Time Varying Coefficients

Most of the previous work on SEM analysis of time series data, both theoretical and applied, assumes time constant (path) coefficients (TCC). However, it is quite commonplace in practice that the causal relationships will change/evolve with time. A straightforward example is the stock market. The relationship between prices of two companies in the same sector may vary with time: negatively correlated sometimes due to competition (e.g. Google Inc. versus Baidu Inc in March 2010), while positively correlated some other times because they are in the same sector and thus share the same fate (e.g. banks in the financial sector during the economic crisis from 2008 till now). In the *f*MRI time series study, it is highly plausible that the path coefficients of the underlying brain functional pathways may change with time as the human brain adapts to the stimuli.

The only work involving time varying coefficients (TVC) to-date has dealt exclusively with the longitudinal pathways and not the contemporaneous pathways (Browne and du Toit, 1991; du Toit and Browne, 2007). No work has been done so far extending the sectional ARMA-based SEM analysis to include time-varying contemporaneous path coefficients. It is our intention to fill this void through this thesis work.

5.1 Model Description

Recall the bivariate sectional AR(1) model (Model 2) described in Chapter 4 and also shown in Figure 5.1 (left). Notice the longitudinal parameters *a*, *b*, *c* are related to the first order vector autoregressive assumption, so only the contemporaneous path coefficient is set as time-varying (See path diagram in Figure 5.1, right). Here $\lambda(t)$ is assumed to be a function of time, i.e., $\lambda(t) = \alpha + \beta t$. Generally speaking, since most functions can be approximated by the polynomials via the Taylor series expansion, we will focus on $\lambda(t) = \sum_k \beta_k t^k$ in this dissertation.

Similarly, for a general ARMA(*p*,*q*) time series, the sectional SEM could be constructed with constant longitudinal parameters and time-varying contemporaneous parameters. Examples of TVC sectional SEM for VMA(1) and VARMA (1,1) series are shown in Figure 5.2 and Figure 5.3.

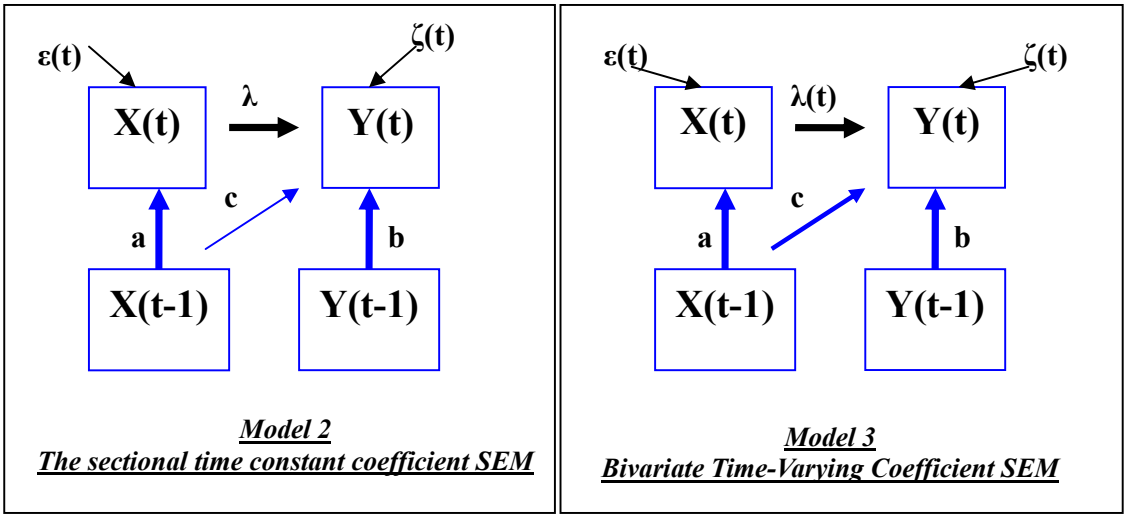


Figure 5.1 Path diagram for a bivariate TCC SEM for VAR(1) time series (left), and path diagram for a bivariate TVC SEM for VAR(1) time series (right).

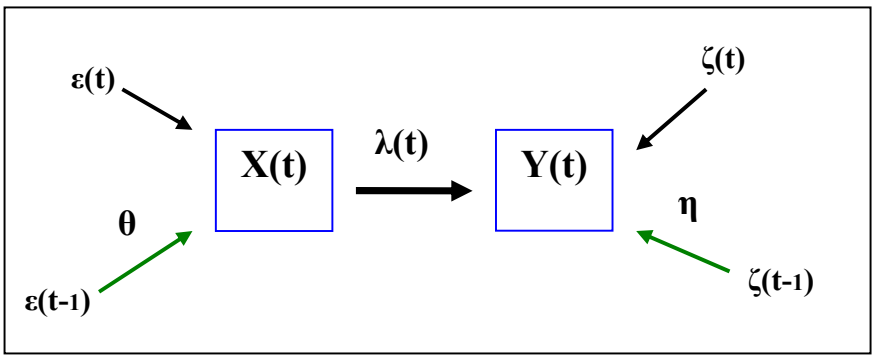


Figure 5.2. Path diagram for a bivariate TVC SEM for VMA(1) time series

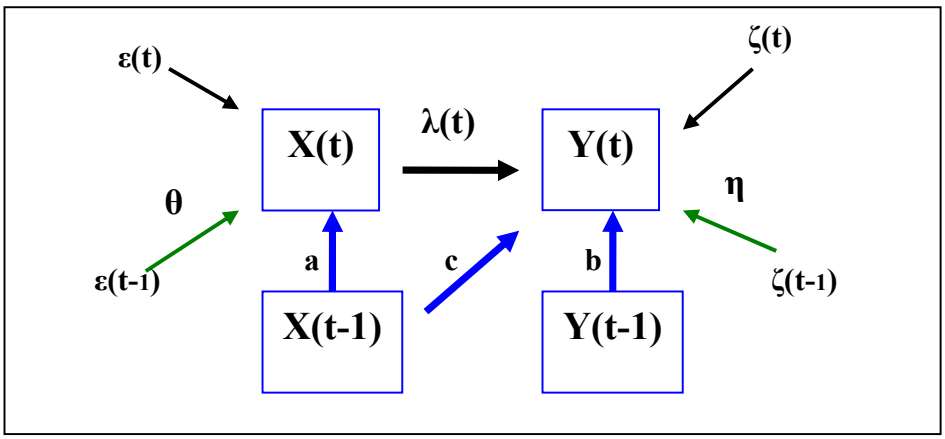


Figure 5.3. Path diagram for a bivariate TVC SEM for VARMA(1,1) time series

5.2 Point Estimation and Statistical Inference

We will start with the easiest case, in which the time series is VAR(1) and $\lambda(t) = \alpha + \beta t$ as shown in Figure 5.1. Assume the error terms for each variable are identically and normally distributed. That is $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, iid; $\varsigma_t \sim N(0, \sigma_\varsigma^2)$, iid. , where σ_ε^2 and σ_ς^2 are also unknown parameters.

From the path diagram, the SEM equations are:

$$\begin{aligned} X(t) &= a \bullet X(t-1) + \varepsilon(t),\ i = 2,3,....,T \\ Y(t) &= b \bullet Y(t-1) + c \bullet X(t-1) + (\alpha + \beta t) \bullet X(t) + \varsigma(t),\ t = 2,3,....,T \end{aligned} \tag{5.1}$$

The likelihood function can be constructed as follows:

$$L(\theta) = L(a, \sigma_\varepsilon^2) \bullet L(b, c, \alpha, \beta, \sigma_\varsigma^2) \tag{5.2}$$

$$\begin{aligned} L(a, \sigma_\varepsilon^2) &= \prod_{t=2}^{T} f(\varepsilon_t) = \prod_{t=2}^{T} \left(\frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\right) \exp\left(-\frac{1}{2\sigma_\varepsilon^2}\varepsilon_t^2\right) \\ &= \left(\frac{1}{2\pi\sigma_\varepsilon^2}\right)^{(T-1)/2} \exp\left[-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=2}^{T}(X_t - aX_{t-1})^2\right] \\ \Rightarrow l(a, \sigma_\varepsilon^2) &= -\frac{T-1}{2}\ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=2}^{T}(X_t - aX_{t-1})^2 \end{aligned} \tag{5.3}$$

$$\begin{aligned} L(b, c, \alpha, \beta, \sigma_\varsigma^2) &= \prod_{t=2}^{T} \left(\frac{1}{\sqrt{2\pi}\sigma_\varsigma}\right) \exp\left(-\frac{1}{2\sigma_\varsigma^2}\varsigma_t^2\right) \\ &= \left(\frac{1}{2\pi\sigma_\varsigma^2}\right)^{(T-1)/2} \exp\left\{-\frac{1}{2\sigma_\varsigma^2}\sum_{t=2}^{T}[Y_t - bY_{t-1} - cX_{t-1} - (\alpha + \beta t)X_t]^2\right\} \\ \Rightarrow l(b, c, \beta, \sigma_\varsigma^2) &= -\frac{T-1}{2}\ln(2\pi\sigma_\varsigma^2) - \frac{1}{2\sigma_\varsigma^2}\sum_{t=2}^{T}[Y_t - bY_{t-1} - cX_{t-1} - (\alpha + \beta t)X_t]^2 \end{aligned} \tag{5.4}$$

Let

$$\begin{aligned} S_1 &= \sum_{t=2}^{T}(X_t - aX_{t-1})^2, \\ S_2 &= \sum_{t=2}^{T}[Y_t - bY_{t-1} - cX_{t-1} - (\alpha + \beta t)X_t]^2 = \sum_{t=2}^{T}[Y_t - bY_{t-1} - cX_{t-1} - \alpha X_t - \beta(tX_t)]^2 \end{aligned} \tag{5.5}$$

To maximize $L(\theta)$, we just need to minimize S_1 and S_2.

Denote

$$\widetilde{X_t} = \begin{bmatrix} X_T \\ X_{T-1} \\ \vdots \\ X_2 \end{bmatrix},\ \widetilde{X_{t-1}} = \begin{bmatrix} X_{T-1} \\ X_{T-2} \\ \vdots \\ X_1 \end{bmatrix},\ \widetilde{Y_t} = \begin{bmatrix} Y_T \\ Y_{T-1} \\ \vdots \\ Y_2 \end{bmatrix},\ \Theta = \begin{bmatrix} Y_{T-1} & X_{T-1} & X_T & TX_T \\ Y_{T-2} & X_{T-2} & X_{T-1} & (T-1)X_{T-1} \\ \vdots & \vdots & \vdots & \vdots \\ Y_1 & X_1 & X_2 & X_2 \end{bmatrix} \tag{5.6}$$

These are least square problems. Let $S_1 = \sum_{t=2}^{T} (X_t - aX_{t-1})^2 = \| \widetilde{X_t} - a\widetilde{X_{t-1}} \|^2$, in which $\| \bullet \|$ is the Euclidean distance. To minimize S_1, $\hat{a}$ should make $(\widetilde{X_t} - a\widetilde{X_{t-1}})$ be orthogonal to $\widetilde{X_{t-1}}$, which means $(\widetilde{X_{t-1}})^T (\widetilde{X_t} - a\widetilde{X_{t-1}}) = 0$.

Suppose $(\widetilde{X_{t-1}})^T (\widetilde{X_{t-1}})$ is non-singular, we will obtain

$$\hat{a} = \left(\widetilde{X_{t-1}}^T \widetilde{X_{t-1}} \right)^{-1} \widetilde{X_{t-1}}^T \widetilde{X_t}, \tag{5.7}$$

To minimize S_2 following the same approach, we will obtain

$$\begin{bmatrix} \hat{b} \\ \hat{c} \\ \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \left(\Theta^T \Theta \right)^{-1} \Theta^T \widetilde{Y}_t \text{, where } \Theta^T \Theta \text{ is assumed non-singular.} \tag{5.8}$$

Hence, we have derived the MLE's of the path coefficients. The results can be readily expanded to vector AR(p) data in p-sectional SEM through the same approach, because only linear equation systems are involved.

For VMA(1) series, suppose $X(t)$ and $Y(t)$ are already centered, then we have the following equations:

$$\begin{pmatrix} X_T \\ X_{T-1} \\ \vdots \\ X_1 \end{pmatrix} = \begin{bmatrix} 1 & \theta & 0 & 0 \\ & 1 & \ddots & 0 \\ & & \ddots & \theta \\ & & & 1 \end{bmatrix} \begin{pmatrix} \varepsilon_T \\ \varepsilon_{T-1} \\ \vdots \\ \varepsilon_1 \end{pmatrix} \tag{5.9}$$

$$\begin{pmatrix} Y_T \\ Y_{T-1} \\ \vdots \\ Y_1 \end{pmatrix} = \lambda(t) \begin{pmatrix} X_T \\ X_{T-1} \\ \vdots \\ X_1 \end{pmatrix} + \begin{bmatrix} 1 & \eta & 0 & 0 \\ & 1 & \ddots & 0 \\ & & \ddots & \eta \\ & & & 1 \end{bmatrix} \begin{pmatrix} \delta_T \\ \delta_{T-1} \\ \vdots \\ \delta_1 \end{pmatrix} \tag{5.10}$$

The likelihood function could be established based on inveritibility assumption and solved with numerical methods.

Similarly, combining Equations (5.1), (5.9) and (5.10), the equation system for VARMA(1,1) is:

$$\begin{pmatrix} X_T \\ X_{T-1} \\ \vdots \\ X_1 \end{pmatrix} = \begin{bmatrix} 0 & a & 0 & 0 \\ & 0 & \ddots & 0 \\ & & \ddots & a \\ & & & 0 \end{bmatrix} \begin{pmatrix} X_T \\ X_{T-1} \\ \vdots \\ X_1 \end{pmatrix} + \begin{bmatrix} 1 & \theta & 0 & 0 \\ & 1 & \ddots & 0 \\ & & \ddots & \theta \\ & & & 1 \end{bmatrix} \begin{pmatrix} \varepsilon_T \\ \varepsilon_{T-1} \\ \vdots \\ \varepsilon_1 \end{pmatrix} \tag{5.11}$$

$$\begin{pmatrix} Y_T \\ Y_{T-1} \\ \vdots \\ Y_1 \end{pmatrix} = \lambda(t) \begin{pmatrix} X_T \\ X_{T-1} \\ \vdots \\ X_1 \end{pmatrix} + \begin{bmatrix} 0 & b & 0 & 0 \\ & 0 & \ddots & 0 \\ & & \ddots & b \\ & & & 0 \end{bmatrix} \begin{pmatrix} Y_T \\ Y_{T-1} \\ \vdots \\ Y_1 \end{pmatrix} + \begin{bmatrix} 1 & \eta & 0 & 0 \\ & 1 & \ddots & 0 \\ & & \ddots & \eta \\ & & & 1 \end{bmatrix} \begin{pmatrix} \delta_T \\ \delta_{T-1} \\ \vdots \\ \delta_1 \end{pmatrix} \tag{5.12}$$

As all of the relationship are linear, it is easy to expand the approach above to general VARMA(p,q) case with polynomial TVC $\lambda(t) = \sum_k \beta_k t^k$.

Based on the property of MLE, the asymptotic variance of the estimated parameter can be calculated:

$$\widehat{Var(\hat{\theta})} = \frac{1}{E[(\frac{\partial}{\partial\theta} \ln L(X;\theta))^2]|_{\theta=\hat{\theta}}}, \text{ and } SE(\hat{\theta}) = \left(\widehat{Var(\hat{\theta})}\right)^{1/2} \tag{5.13}$$

Thus, the 100(1-α)% Wald confidence interval (CI) can be constructed as:
$\hat{\theta} \pm z_{1-\alpha/2} SE(\hat{\theta})$.

5.3 Comparison between TCC and TVC SEM

Comparing the MLE of the TVC model (Model 3) to the time constant coefficient model (Model 2), the only difference one will find is that a column $tX_t, (t = 2,...,T)$ is added up to the matrix Θ. For polynomial time-varying coefficients, we only need to add more columns $t^2X_t, t^3X_t, ...(t = 2,...,T)$ to the matrix Θ. If we introduce an artificial variable $Z_t = tX_t$ and construct Model 4 as illustrated in Figure 5.4, we will have exactly the same results. This equivalency transforms sectional SEM with TVC into sectional SEM with TCC. Based on Proposition 1, we thus obtain the equivalency to the overall SEM with TCC. This helps us to employ the current software for sectional TCC SEM for the TVC SEM and thus readily obtain the data analysis output.

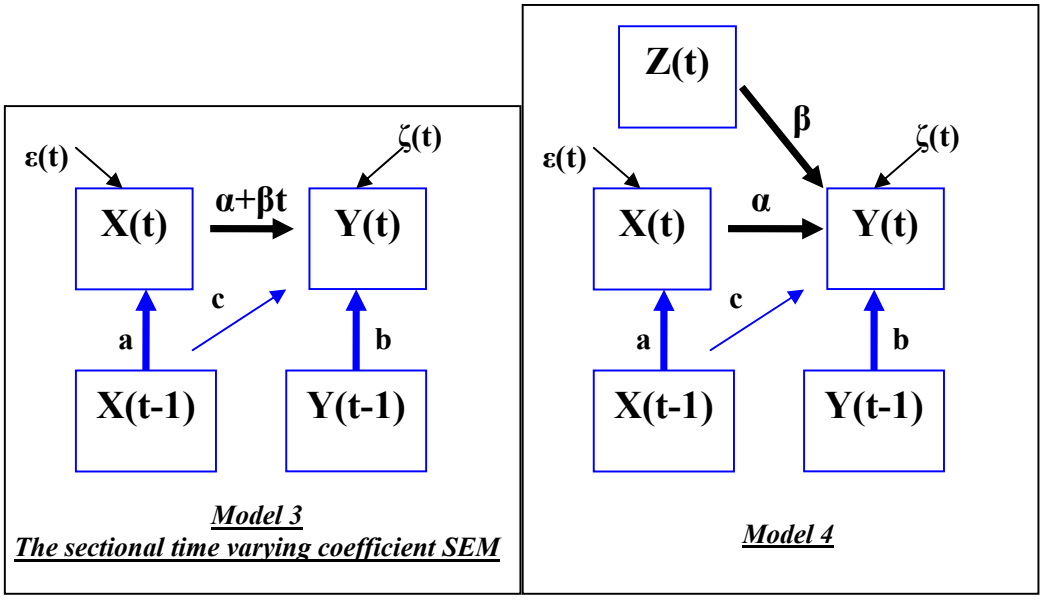


Figure 5.4 The equivalence of Model 3 and Model 4.

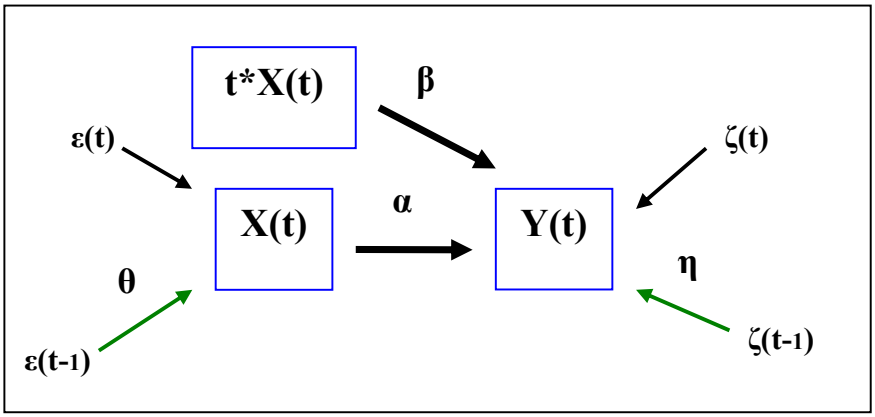


Figure 5.5 TVC SEM with artificial exogenous variable added for VMA(1) series.

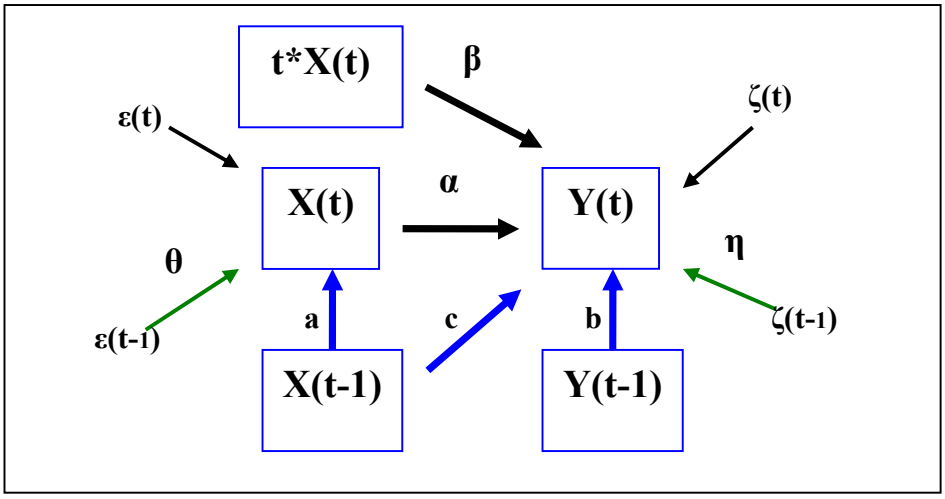


Figure 5.6 TVC SEM with artificial exogenous variable added for VARMA(1,1) series.

For VAMRA(p, q), the conclusion above can be adapted directly. We can also insert a new exogenous variable $t*Xt$ to the path diagram and make all coefficients time-constant (See Figure 5.5 and Figure 5.6). Furthermore, for polynomial TVC $\lambda(t) = \sum_k \beta_k t^k$, k new exogenous variables $tX_t, tX_t^2, ...tX_t^K$ could be inserted with arrows heading to Y(t).

As introduced in Chapter 2, the chi-square statistics is the commonly used goodness of fit index to evaluate an SEM result. However, there is also criticism of it: first, it is highly related to sample size; second, the Wishart distribution assumption on covariance matrix usually fails for real data analysis.

In the TVC model, a covariance matrix on (X_t, X_{t-1}, $t*X_t$) hardly satisfies the Wishart distribution assumption. Hence, the goodness of fit based on chi-square statistics are usually not reliable. Other than chi-square statistics, Kenny (2010) introduced more than ten goodness measures such as Normed Fit Index (NFI), Root Mean Square Error of Approximation (RMSEA), Akaike Information of Criterion (AIC), etc. However, all of these fit indices, except the Root Mean Square Residual (RMR) or Standardized Root Mean Square Residual (SRMR), are calculated based on the chi-square value. The RMR and SRMR are distribution free, but they are less informative at the same time. The model with smaller RMR or SRMR is better, but we have no idea how much better.

To test which model is better, TVC or TCC, the chi-square difference test (Bollen 1989, 1993) could only be used for a sketchy understanding. As we can see, TVC brings new parameters into the model, so decreases the degrees of freedom. Thus, if the difference of chi-square value with the difference of *df* indicates a small p-value, then we will reject the TCC model but use TVC model instead. However, the use of the chi-square difference test should be very careful: first, the TCC model should include also the artificial variable but with constraint on β=0. Second, as we mentioned before, the chi-square values are usually not meaningful since the Wishart distribution fails for TVC model. Hence, in our comparison all through this dissertation, although chi-square values and RMR are compared for certain cases, we focus on comparing the point estimates.

.

5.4 Simulation Study for the TVC Model

Model 3 in Figure 5.1 is used in our simulation. The bivariate autoregressive series $X(t)$ and $Y(t)$ are generated from the formulas below.

$$\begin{aligned} &X(1) = 10 + \varepsilon(1) \\ &Y(1) = 5 + \varsigma(1) \\ &X(t) = a \bullet X(t-1) + \varepsilon(t),\ i = 2,3,....,T \\ &Y(t) = b \bullet Y(t-1) + c \bullet X(t-1) + \lambda(t) \bullet X(t) + \varsigma(t),\ t = 2,3,....,T \end{aligned} \tag{5.14}$$

We set different values for a, b, c and $\lambda(t)$. $\varepsilon(t)$ and $\zeta(t)$ are random numbers following $N(0,1)$. We generate the datasets with T=100 time steps and N=1 observations, since we will focus on the sectional SEM.

The comparison between TCC and TVC models for different parameter settings are shown below:

Group (A): Suppose the contemporaneous path coefficient is linear with time t. That is: a=0.6, b=0.4, c=0.2, $\lambda(t)= \alpha+\beta t$ =0.3+0.05t. The estimates in TCC and TVC models are shown in Table 5.2. The RMR for TCC model is 56.68, for TVC model is 14.74, which indicates an improvement for using TVC SEM.

Table 5.2 Comparison of TCC and TVC SEM with Group (A) simulation dataset.

Parameter	True Value	TCC SEM			TVC SEM		
		Point Estimate	Std	p-value	Point Estimate	Std	p-value
α	0.3	2.8905	0.2008	p<0.01	0.1586	0.1085	p=0.147
β	0.05	0	N/A	N/A	0.0492	0.0017	p<0.01
a	0.6	0.5686	0.0847	p<0.01	0.5686	0.0847	p<0.01
b	0.4	0.5851	0.0633	p<0.01	0.4081	0.0364	p<0.01
c	0.2	-0.4369	0.3003	p=0.149	0.1866	0.1722	p=0.281

Group (B): Suppose the contemporaneous path coefficient is linear of time t with a random noise. That is: a=0.6, b=0.4, c=0.2, $\lambda(t)=\alpha+\beta t+\tau(t)$=0.3+0.05$t$+$\tau(t)$, where $\tau(t)$ follows N(0, 0.01^2). The estimates in TCC and TVC models are shown in Table 5.3. The RMR for TCC model is 49.59, for TVC model is 13.41, which also supports the use of TVC SEM.

Table 5.3 Comparison of TCC and TVC SEM with Group (B) simulation dataset.

Parameter	True Value	TCC SEM			TVC SEM		
		Point Estimate	Std	p-value	Point Estimate	Std	p-value
α	0.3	2.8260	0.1898	p<0.01	0.3117	0.1077	p<0.01
β	0.05	0	N/A	N/A	0.0508	0.0019	p<0.01
a	0.6	0.5876	0.0532	p<0.01	0.5876	0.0532	p<0.01
b	0.4	0.6851	0.0426	p<0.01	0.4222	0.0265	p<0.01
c	0.2	-1.1359	0.1746	p<0.01	0.1540	0.0991	p=0.123

Group (C): Suppose the contemporaneous path coefficient is a non-linear function of time t (e.g. a log function of t) with a random noise. In practice, the researcher may not know the real structure of the TVC function, but a polynomial function could be used for approximation for most common functions. In this case, we still use $\lambda(t)= \alpha+\beta t$ to show the TVC exists. That is: a=0.6, b=0.4, c=0.2, $\lambda(t)$=ln(t)/10+$\tau(t)$, where $\tau(t)$ follows N(0, 0.01). The estimates in TCC and TVC models are shown in Table 5.4. The RMR for TCC model is 14.94, for TVC model is 13.20, which shows no big difference between TCC

and TVC approaches.

Table 5.4 Comparison of TCC and TVC SEM with Group (C) simulation dataset.

Parameter	True Value	TCC SEM			TVC SEM		
		Point Estimate	Std	p-value	Point Estimate	Std	p-value
α	N/A	0.4323	0.1073	p<0.01	0.0723	0.1044	p=0.490
β	N/A	0	N/A	N/A	0.0072	0.0017	p<0.01
a	0.6	0.5897	0.0819	p<0.01	0.5897	0.0819	p<0.01
b	0.4	0.3334	0.0896	p<0.01	0.2996	0.0872	p<0.01
c	0.2	0.1808	0.1207	p=0.137	0.2347	0.1253	p=0.064

Group (D): Suppose the contemporaneous path coefficient is a stationary AR(1) time series itself. That is: a=0.6, b=0.4, c=0.2, $\lambda(t)$=0.9*$\lambda(t-1)$+$\tau(t)$, where $\lambda(1)$=0.3, $\tau(t)$ follows N(0, 0.01). The estimates in TCC and TVC models are shown in Table 5.5. The RMR for TCC model is 13.83, for TVC model is 13.72, which also shows no big difference between TCC and TVC approaches.

Table 5.5 Comparison of TCC and TVC SEM with Group (D) simulation dataset.

Parameter	True Value	TCC SEM			TVC SEM		
		Point Estimate	Std	p-value	Point Estimate	Std	p-value
α	N/A	-0.0761	0.0898	p=0.399	0.0229	0.0895	p=0.799
β	N/A	0	N/A	N/A	-0.0022	0.0016	p=0.175
a	0.6	0.5883	0.0817	p<0.01	0.5883	0.0817	p<0.01
b	0.4	0.2854	0.0923	p<0.01	0.2867	0.0926	p<0.01
c	0.2	0.2279	0.0902	p=0.013	0.2297	0.0997	p=0.023

Some conclusions can be readily drawn upon these four groups of simulation studies. In Group (A) and Group (B), the TVC models yielded better estimates, closer to the true values. For Group (C), although TVC parameter β is tested to be significant, the estimates are not improved, which will remind the researchers to look into the parameter structure and use TVC functions other than a simple linear function. This is further discussed in Chapter 7 where we proposed the nonlinear TVC SEM framework. For Group (D), the coefficients are not varying significantly with time t, so the TCC and TVC yielded similar output. The lesson of the last group is that one should explore potential structure and relationships among the contemporaneous paths along the time series first before deciding upon a suitable starting model. The moving window approach we proposed in Chapter 7 will help researchers in this critical task.

Since we have no prior distribution assumption on variables in the TVC model, the regular power analysis based on Wishart distributed covariance matrix is not valid. A numeric approach is proposed instead as the following: first, we get the estimates from the original data; second, P samples are generate based on the estimates; third, test the coefficient on each new sample and denote P_s as the significant ones through the P

samples. Then we have

$$\text{Power} = (\text{reject } H_0 \mid H_a \text{ holds}) = P_s / P. \tag{5.15}$$

The power analysis has been conducted for Group (B) data as an example. We test the time varying coefficient β through the approach described above (Figure 5.7). We can see 30 samples can already give a power around 0.8 for a bivariate TVC model.

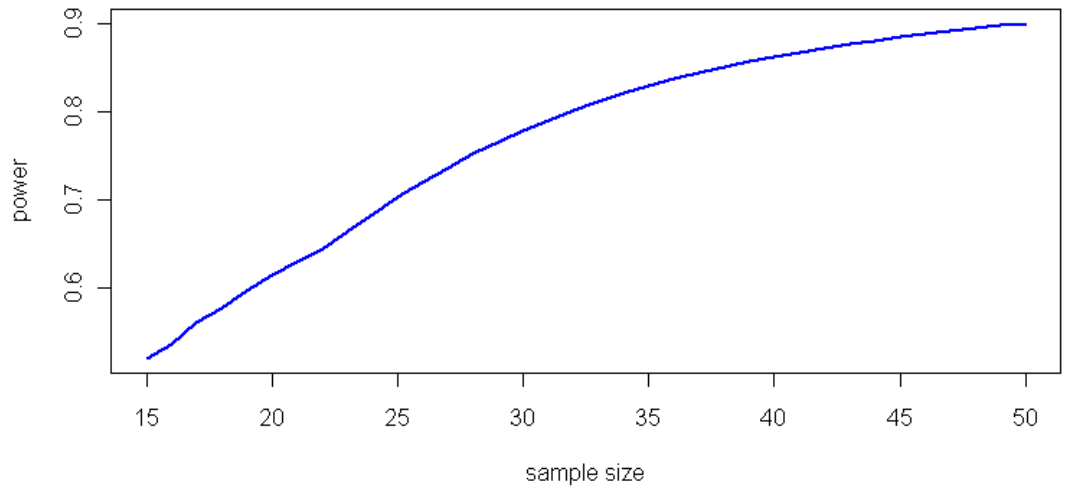


Figure 5.7 Power analysis on TVC model with simulation data Group (B).

Chapter 6. Application to *f*MRI Data

6.1 Experimental Design and Data Preprocess

Twenty-eight volunteers (14 females and 14 males) who had no psychiatric or neurological history participated in this visual attention study with a "three-ball tracking" task (Figure 6.1) (Lange, 1999) conducted on a 4 T Varian MR System at the Brookhaven National Laboratory (BNL). This study was approved by the Medical Research Center at BNL and all subjects provided verbal and written consent.

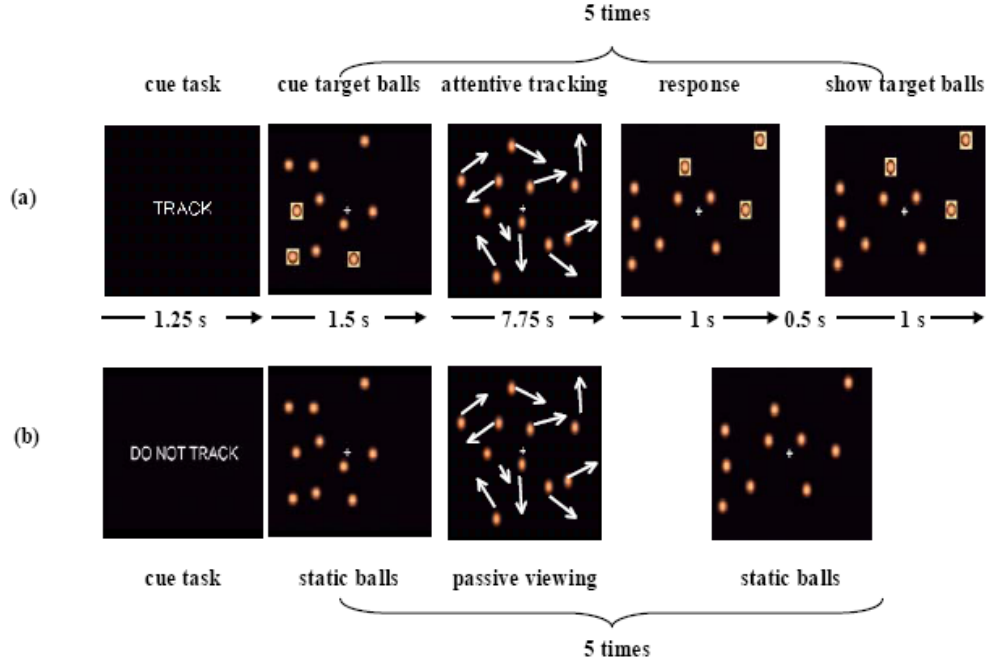


Figure 6.1. A schematic diagram of the visual stimulus used in (a) active tracking and (b) passive viewing trials. Each trial began with a text cue indicating the type of trial. This was followed by a period of static balls (1.5 s), in which the target balls were highlighted with orange squares on active 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.

Preprocessing of *f*MRI 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., 2007).

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) (Buchel and Friston, 1997; Chang et al., 2004; Friston and Buchel, 2000).

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.

6.2 Sectional SEM with Time Constant Coefficients

The initial path model is defined with six regions of interest (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) (Figure 4.2.1). 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. (2007), the longitudinal relations were depicted by the first-order multivariate or vector autoregressive process (VAR(1)). Although the order of VAR for each ROI obtained from the partial autocorrelation function (PACF) analysis was not always 1, due to estimability constraints (Honey et al., 2002), the VAR 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 4.2.2. Equation (6.1) is the matrix form of the unified SEM with its contemporaneous and longitudinal components.

$$\begin{pmatrix} CERCB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & \lambda_{25} & \lambda_{26} \\ 0 & \lambda_{32} & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{43} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{54} & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_{65} & 0 \end{pmatrix} \bullet \begin{pmatrix} CERCB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix}$$
$$+ \begin{pmatrix} \gamma_{11} & 0 & 0 & 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & 0 & 0 & \gamma_{25} & \gamma_{26} \\ 0 & \gamma_{32} & \gamma_{33} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{43} & \gamma_{44} & 0 & 0 \\ 0 & 0 & 0 & \gamma_{54} & \gamma_{55} & 0 \\ 0 & 0 & 0 & 0 & \gamma_{65} & \gamma_{66} \end{pmatrix} \bullet \begin{pmatrix} CERCB(t-1) \\ THAL(t-1) \\ PPC(t-1) \\ APC(t-1) \\ LPFC(t-1) \\ SMA(t-1) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \\ \varepsilon_4(t) \\ \varepsilon_5(t) \\ \varepsilon_6(t) \end{pmatrix} \quad (6.1)$$

Zhang (2007) has further explored three different approaches for analyzing this multiple subject time series data as in (1) summarize and then analyze, (2) analyze and then summarize, and (3) simultaneous analysis. (Please see Chapter 4 for details). In this dissertation, we repeat these three approaches for analyzing multi-subject time series data with time constant coefficients in order to compare between the TCC SEM utilized by Kim et al. (2007) and Zhang (2007), and the newly developed TVC SEM. Table 6.1-6.3 show the results of Approach 1, 2 and 3 using SAS CALIS respectively.

Table 6.1 Estimates of the TCC sectional SEM using the subject-average *f*MRI data (Approach 1). Significant paths are in bold (two sided p-value based on t-test <0.05).

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 (0.091)**	**8.592 (<0.001)**	λ_{21}	-0.041 (0.233)	-0.177 (0.860)
γ_{21}	-0.337 (0.256)	-1.321 (0.193)	λ_{25}	0.364 (0.201)	1.805 (0.077)
γ_{22}	**0.531 (0.111)**	**4.808 (<0.001)**	λ_{26}	0.283 (0.169)	1.676 (0.100)
γ_{25}	-0.001 (0.205)	-0.004 (0.997)	λ_{32}	**-0.282(0.060)**	**-4.688 (<0.001)**
γ_{26}	-0.002 (0.184)	-0.012 (0.990)	λ_{43}	-0.140 (0.126)	-1.112 (0.271)
γ_{32}	**0.252 (0.065)**	**3.907 (<0.001)**	λ_{54}	0.171 (0.180)	0.949 (0.347)
γ_{33}	**0.395 (0.105)**	**3.749 (<0.001)**	λ_{65}	**0.376 (0.155)**	**2.417 (0.019)**
γ_{43}	**0.245 (0.111)**	**2.215 (0.031)**			
γ_{44}	**0.566 (0.111)**	**5.118 (<0.001)**			
γ_{54}	-0.284 (0.178)	-1.595 (0.117)			
γ_{55}	**0.582 (0.113)**	**5.158 (<0.001)**			
γ_{65}	0.281 (0.160)	-1.752 (0.086)			
γ_{66}	**0.590 (0.107)**	**5.514 (<0.001)**			

Table 6.2 Average estimates of the TCC sectional SEM using the *f*MRI data of 28 subjects (Approach 2). Significant paths are in bold (two sided p-value based on t-test <0.05).

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}	**0.612 (0.107)**	**5.720 (<0.001)**	λ_{21}	-0.047 (0.223)	-0.211 (0.834)
γ_{21}	-0.005 (0.234)	-0.021 (0.983)	λ_{25}	0.101 (0.191)	0.529 (0.601)
γ_{22}	**0.591 (0.103)**	**5,738 (<0.001)**	λ_{26}	0.029 (0.192)	0.151 (0.881)
γ_{25}	-0.121 (0.199)	-0.608 (0.548)	λ_{32}	-0.033(0.081)	-0.407 (0.687)
γ_{26}	-0.027 (0.202)	-0.134 (0.894)	λ_{43}	-0.005(0.115)	-0.043 (0.966)
γ_{32}	0.035 (0.081)	0.432 (0.669)	λ_{54}	0.091 (0.211)	0.431 (0.669)
γ_{33}	**0.595 (0.108)**	**5.509 (<0.001)**	λ_{65}	0.228 (0.140)	1.629 (0.113)
γ_{43}	0.031 (0.116)	0.267 (0.791)			
γ_{44}	**0.593 (0.110)**	**5.391 (<0.001)**			
γ_{54}	-0.124 (0.216)	-0.574 (0.570)			
γ_{55}	**0.615 (0.105)**	**5.857 (<0.001)**			
γ_{65}	-0.166 (0.144)	-1.153 (0.258)			
γ_{66}	**0.578 (0.111)**	**5.207 (<0.001)**			

Table 6.3 Estimates of the TCC sectional SEM using the concatenated *f*MRI data of 28 subjects (Approach 3). Significant paths are in bold (two sided p-value based on t-test <0.05).

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 (0.020)**	**29.985 (<0.001)**	λ_{21}	0.073 (0.041)	1.781 (0.071)
γ_{21}	-0.041 (0.041)	-0.992 (0.329)	λ_{25}	**0.086 (0.038)**	**2.287 (0.029)**
γ_{22}	**0.630 (0.020)**	**31.287 (<0.001)**	λ_{26}	0.050 (0.039)	1.293 (0.206)
γ_{25}	-0.075 (0.038)	-1.976 (0.057)	λ_{32}	**-0.040(0.016)**	**-2.528 (0.017)**
γ_{26}	-0.027 (0.039)	-0.689 (0.496)	λ_{43}	0.040(0.020)	1.979 (0.057)
γ_{32}	0.025 (0.016)	1.573 (0.126)	λ_{54}	**0.112 (0.039)**	**2.882 (0.007)**
γ_{33}	**0.618 (0.020)**	**30.371 (<0.001)**	λ_{65}	**0.209 (0.025)**	**8.486 (<0.001)**
γ_{43}	0.006 (0.021)	0.303 (0.764)			
γ_{44}	**0.613 (0.020)**	**29.661 (<0.001)**			
γ_{54}	**-0.133(0.039)**	**-3.448 (0.002)**			
γ_{55}	**0.643 (0.020)**	**32.394 (<0.001)**			
γ_{65}	**-0.180(0.025)**	**-7.247 (<0.001)**			
γ_{66}	**0.580 (0.021)**	**27.354 (<0.001)**			

6.3 Sectional SEM with Time Varying Coefficients

TVC SEM was also applied to the same dataset. As a first attempt, we only set the contemporaneous path parameters to be time-varying linearly ($\lambda(t)= \alpha+\beta t$) as described in Section 3. Hence, we have updated SEM equations as below:

$$\begin{pmatrix} CERCB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{21}+\beta_{21}t & 0 & 0 & 0 & \alpha_{25}+\beta_{25}t & \alpha_{26}+\beta_{26}t \\ 0 & \alpha_{32}+\beta_{32}t & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_{43}+\beta_{43}t & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_{54}+\beta_{54}t & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_{65}+\beta_{65}t & 0 \end{pmatrix} \bullet \begin{pmatrix} CERCB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix} + \begin{pmatrix} \gamma_{11} & 0 & 0 & 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & 0 & 0 & \gamma_{25} & \gamma_{26} \\ 0 & \gamma_{32} & \gamma_{33} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{43} & \gamma_{44} & 0 & 0 \\ 0 & 0 & 0 & \gamma_{54} & \gamma_{55} & 0 \\ 0 & 0 & 0 & 0 & \gamma_{65} & \gamma_{66} \end{pmatrix} \bullet \begin{pmatrix} CERCB(t-1) \\ THAL(t-1) \\ PPC(t-1) \\ APC(t-1) \\ LPFC(t-1) \\ SMA(t-1) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \\ \varepsilon_4(t) \\ \varepsilon_5(t) \\ \varepsilon_6(t) \end{pmatrix} \quad (6.2)$$

Introducing artificial variables as described in Section 3, we transformed this TVC SEM model to a sectional TCC SEM model and used the SAS CALIS procedure to yield the output. The estimated parameters are listed in Table 6.4. The significant paths are shown in Figure 6.1.5. Approach 2 and 3 are not available for the TVC model because the sample covariance matrix is singular after introducing the artificial variable $t*X_t$.

From Table 6.4, we can see no time-varying path is found (because no parameter β is significant). Compared with fixed coefficient model, only 9 paths are significant in TVC model: the path APC(t-1)→LPFC(t) is no longer significant. Besides, the parameters estimated in TCC and TVC models are different (but not changing too much).

Table 6.4 Estimates of the TVC sectional SEM using the subject-average *f*MRI data (Approach 1). Significant paths are in bold (two sided p-value based on t-test <0.05).

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Est. (S. E.)	t-value (p-values)	Path parameters	Est. (S. E.)	t-value (p-value)
γ_{11}	**0.780 (0.091)**	**8.592 (<0.001)**	α_{21}	0.022 (0.230)	0.098 (0.922)
γ_{21}	-0.392 (0.412)	-0.951 (0.346)	β_{21}	0.002 (0.012)	0.191(0.849)
γ_{22}	**0.494 (0.120)**	**4.133 (<0.001)**	α_{25}	0.324 (0. 199)	1.630 (0.109)
γ_{25}	-0.327 (0.396)	-0.826 (0.413)	β_{25}	0.023 (0.012)	1.096 (0.278)
γ_{26}	-0.190 (0.290)	-0.656 (0.515)	α_{26}	0.293 (0.167)	1.755 (0.085)
γ_{32}	**0.329 (0.086)**	**3.840 (<0.001)**	β_{26}	-0.008 (0.009)	-0.926 (0.359)

γ_{33}	**0.393 (0.104)**	**3.779 (<0.001)**	α_{32}	**-0.279(0.060)**	**-4.658 (<0.001)**
γ_{43}	0.213 (0.161)	1.328 (0.190)	β_{32}	-0.003 (0.002)	-1.279 (0.207)
γ_{44}	**0.560 (0.113)**	**4.964 (<0.001)**	α_{43}	-0.144 (0.126)	-1.144 (0.258)
γ_{54}	-0.428 (0.337)	-1.271 (0.209)	β_{43}	0.002 (0.005)	0.294 (0.770)
γ_{55}	**0.577 (0.113)**	**5.108 (<0.001)**	α_{54}	0.181 (0.179)	1.010 (0.317)
γ_{65}	-0.312 (0.265)	-1.174 (0.246)	β_{54}	0.005 (0.010)	0.483 (0.631)
γ_{66}	**0.589 (0.107)**	**5.502 (<0.001)**	α_{65}	**0.375 (0.155)**	**2.414 (0.019)**
			β_{65}	0.001 (0.007)	0.150 (0.881)

6.4 Bootstrap Resampling and Model Comparisons

As we discussed in Chapter 4, we have two bootstrapping schemes (A.1a and A.1b). Here we conduct these two types of bootstrapping for TCC and TVC SEM models respectively. The output results are shown in Table 6.5-6.8.

Table 6.5 Estimates of the TCC sectional SEM using the subject-average *f*MRI data with bootstrapping (Approach 1a, *B*=200). Significant paths are in bold (two sided p-value based on t-test <0.05).

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}	**0.621 (0.099)**	**6.272 (<0.001)**	λ_{21}	-0.049 (0.162)	-0.302 (0.765)
γ_{21}	**0.471 (0.175)**	**2.691 (0.011)**	λ_{25}	0.030 (0.211)	0.142 (0.888)
γ_{22}	**0.469 (0.091)**	**5.153 (<0.001)**	λ_{26}	**0.542 (0.238)**	**2.277 (0.030)**
γ_{25}	-0.035 (0.213)	-0.164 (0.871)	λ_{32}	**-0.232(0.060)**	**-3.867 (<0.001)**
γ_{26}	**-0.755(0.226)**	**-3.340 (0.0.002)**	λ_{43}	0.130(0.067)	1.940 (0.062)
γ_{32}	0.093 (0.062)	1.500 (0.144)	λ_{54}	-0.096 (0.173)	-0.555(0.583)
γ_{33}	**0.531 (0.099)**	**5.509 (<0.001)**	λ_{65}	0.178 (0.102)	1.745 (0.091)
γ_{43}	0.002 (0.073)	0.267 (0.791)			
γ_{44}	**0.573 (0.097)**	**5.364 (<0.001)**			
γ_{54}	-0.084 (0.173)	-0.486 (0.630)			
γ_{55}	**0.644 (0.078)**	**8.256 (<0.001)**			
γ_{65}	**-0.357(0.095)**	**-3.757 (<0.001)**			
γ_{66}	**0.438 (0.092)**	**4.761 (<0.001)**			

Table 6.6 Estimates of the TCC sectional SEM using the subject-average *f*MRI data with bootstrapping (Approach 1b, *B*=200, *T*=100). Significant paths are in bold (two sided p-value based on t-test <0.05).

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}	**0.613 (0.081)**	**7.568 (<0.001)**	λ_{21}	-0.016 (0.141)	-0.113 (0.911)
γ_{21}	0.452 (0.148)	3.054 (0.983)	λ_{25}	0.044 (0.181)	0.243 (0.810)
γ_{22}	**0.471 (0.079)**	**5.963 (<0.001)**	λ_{26}	**0.565 (0.205)**	**2.756 (0.010)**
γ_{25}	-0.042 (0.183)	-0.230 (0.820)	λ_{32}	**-0.218(0.050)**	**-4.360(<0.001)**
γ_{26}	**-0.776(0.192)**	**-4.042 (<0.001)**	λ_{43}	**0.132 (0.059)**	**-2.237(0.033)**
γ_{32}	0.095 (0.052)	1.827 (0.077)	λ_{54}	-0.080 (0.147)	-0.544 (0.590)
γ_{33}	**0.549 (0.083)**	**6.614 (<0.001)**	λ_{65}	**0.177 (0.086)**	**2.058 (0.048)**
γ_{43}	0.009 (0.064)	0.141 (0.889)			
γ_{44}	**0.566 (0.083)**	**6.699 (<0.001)**			
γ_{54}	-0.085 (0.148)	-0.574 (0.570)			
γ_{55}	**0.641 (0.067)**	**9.567 (<0.001)**			
γ_{65}	**-0.360(0.081)**	**-4.444(<0.001)**			
γ_{66}	**0.437 (0.078)**	**5.603 (<0.001)**			

Table 6.7 Estimates of the TVC sectional SEM using the subject-average *f*MRI data with bootstrapping (Approach 1a, *B*=200). Significant paths are in bold (two sided p-value based on t-test <0.05).

	Longitudinal path parameters			Contemporaneous path parameters	
Path parameters	Est. (S. E.)	t-value (p-values)	Path parameters	Est. (S. E.)	t-value (p-value)
γ_{11}	**0.780 (0.092)**	**8.478 (<0.001)**	α_{21}	0.045 (0.208)	0.216 (0.830)
γ_{21}	-0.430 (0.392)	-1.097 (0.277)	β_{21}	0.002 (0.012)	0.167(0.868)
γ_{22}	**0.488 (0.113)**	**4.319 (<0.001)**	α_{25}	0.328 (0. 184)	1.783 (0.080)
γ_{25}	-0.271 (0.379)	-0.715 (0.477)	β_{25}	0.011 (0.011)	1.000 (0.321)
γ_{26}	0.168(0.289)	0.581 (0.563)	α_{26}	0.266 (0.154)	1.727 (0.089)
γ_{32}	**0.329 (0.086)**	**3.825 (<0.001)**	β_{26}	-0.007 (0.009)	-0.778 (0.440)
γ_{33}	**0.413 (0.104)**	**3.971 (<0.001)**	α_{32}	**-0.286(0.058)**	**-4.931 (<0.001)**
γ_{43}	0.228 (0.162)	1.407 (0.165)	β_{32}	-0.002 (0.002)	-1.000 (0.321)
γ_{44}	**0.570 (0.111)**	**5.135(<0.001)**	α_{43}	-0.181 (0.125)	-1.448 (0.153)
γ_{54}	-0.445 (0.336)	-1.324 (0.191)	β_{43}	0.002 (0.005)	0.400 (0.691)
γ_{55}	**0.576 (0.113)**	**5.097 (<0.001)**	α_{54}	0.170 (0.175)	0.971 (0.335)
γ_{65}	-0.268 (0.270)	-0.993 (0.355)	β_{54}	0.006 (0.010)	0.600 (0.551)
γ_{66}	**0.598 (0.109)**	**5.486 (<0.001)**	α_{65}	**0.370 (0.157)**	**2.357 (0.022)**
			β_{65}	0.000 (0.007)	0.000 (1.000)

Table 6.8 Estimates of the TVC sectional SEM using the subject-average *f*MRI data with bootstrapping (Approach 1b, *B*=200， T=100). Significant paths are in bold (two sided

p-value based on t-test <0.05).

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Est. (S. E.)	t-value (p-values)	Path parameters	Est. (S. E.)	t-value (p-value)
γ_{11}	**0.793 (0.065)**	**12.200(<0.001)**	α_{21}	0.055 (0.162)	0.340 (0.735)
γ_{21}	-0.421 (0.299)	-1.408 (0.164)	β_{21}	0.003 (0.009)	0.333 (0.740)
γ_{22}	**0.503 (0.086)**	**5.849 (<0.001)**	α_{25}	**0.334 (0. 141)**	**2.369 (0.021)**
γ_{25}	-0.317 (0.286)	-1.108 (0.272)	β_{25}	0.012 (0.009)	1.500 (0.139)
γ_{26}	0.173 (0.216)	0.801 (0.426)	α_{26}	**0.254 (0.121)**	**2.099 (0.040)**
γ_{32}	**0.329 (0.061)**	**5.393 (<0.001)**	β_{26}	-0.008 (0.007)	-1.143 (0.258)
γ_{33}	**0.427 (0.074)**	**5.770 (<0.001)**	α_{32}	**-0.267(0.042)**	**-6.357(<0.001)**
γ_{43}	**0.251 (0.120)**	**2.092 (0.041)**	β_{32}	-0.003 (0.002)	-1.500 (0.139)
γ_{44}	**0.577 (0.082)**	**7.037 (<0.001)**	α_{43}	-0.180 (0.095)	-1.895 (0.063)
γ_{54}	-0.432 (0.245)	-1.763 (0.083)	β_{43}	0.001 (0.004)	0.250 (0.803)
γ_{55}	**0.572 (0.081)**	**7.062 (<0.001)**	α_{54}	0.177 (0.127)	1.394 (0.168)
γ_{65}	-0.320 (0.190)	-1.684 (0.097)	β_{54}	0.005 (0.007)	0.714 (0.478)
γ_{66}	**0.590 (0.077)**	**7.662 (<0.001)**	α_{65}	**0.373 (0.111)**	**3.360 (0.001)**
			β_{65}	0.001 (0.005)	0.020 (0.842)

A full comparison of different models with different approaches is shown in Figure 6.2. We can see TVC models, generally speaking, return fewer significant arrows than TCC models. No time-varying coefficient is found, but it is high possible that they exist due to the poor fit of the entire model. Although there are several significant paths, the poor goodness of fit implies the original assumption of brain connections is not good.

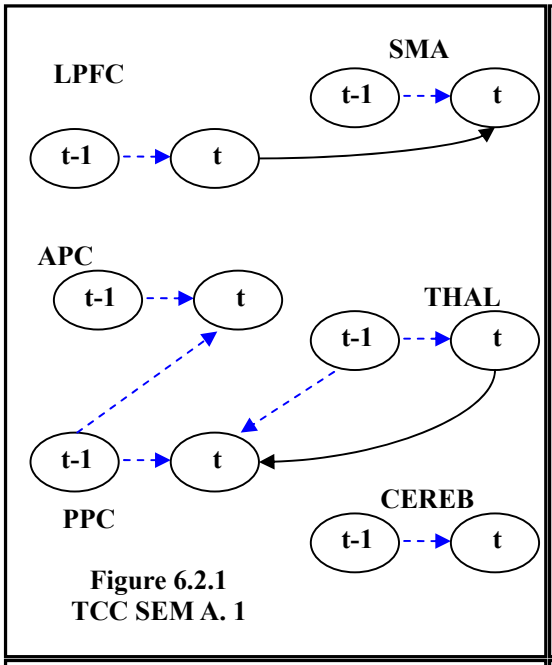


Figure 6.2.1
TCC SEM A. 1

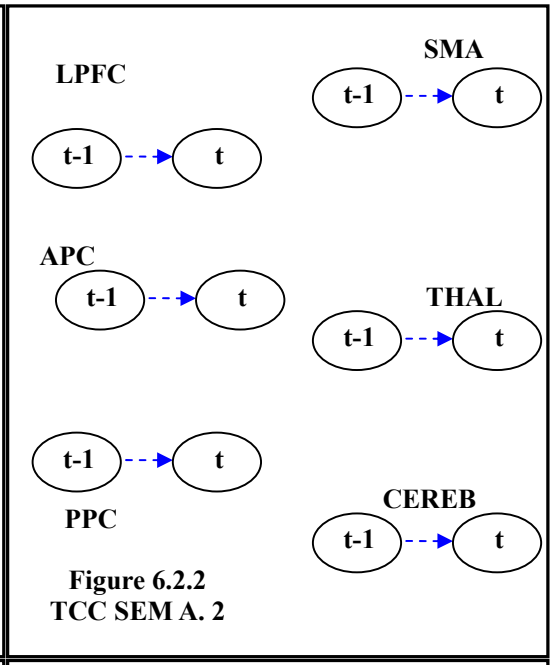


Figure 6.2.2
TCC SEM A. 2

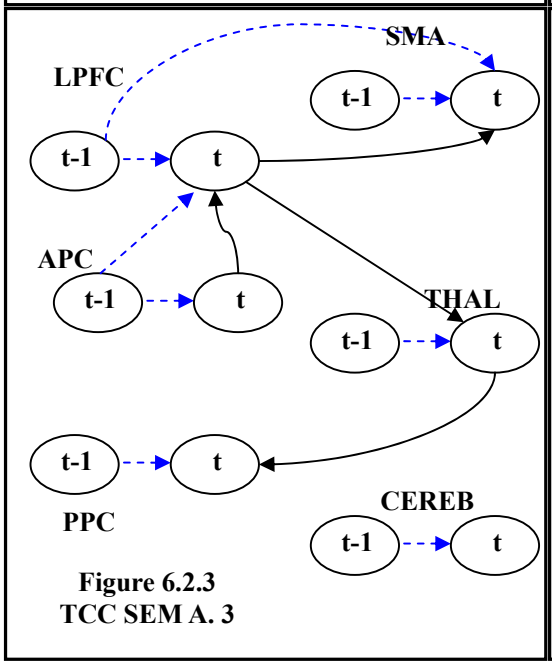


Figure 6.2.3
TCC SEM A. 3

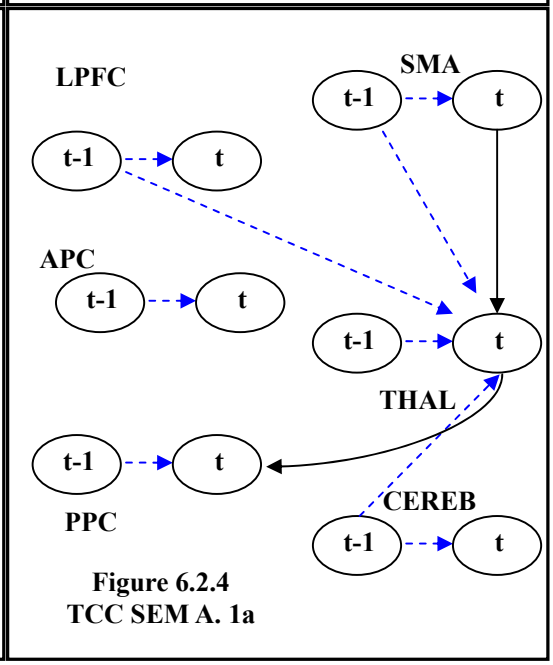


Figure 6.2.4
TCC SEM A. 1a

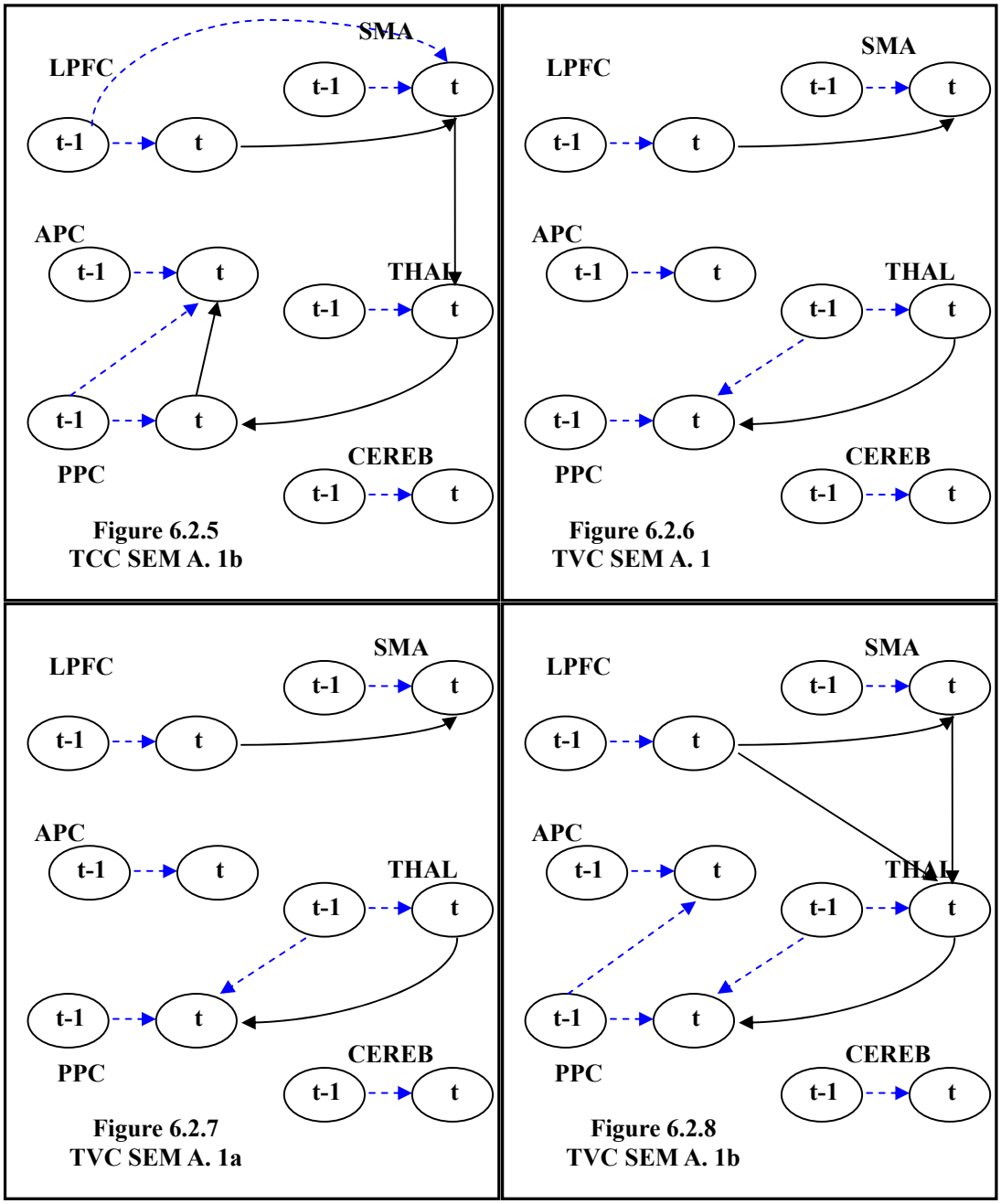


Figure 6.2. Significant paths (p-values < 0.05) in the *f*MRI confirmatory models. Figure 6.2.1-6.2.5 illustrated the output results assuming only time-constant coefficients through different approaches, while Figure 6.2.6-6.2.8 shows the TVC model output.

Chapter 7 Current and Future Work

Our primary goal in this thesis is to expand the current SEM for time series analysis to include the more realistic time varying coefficients (TVC). Our strategy consists of two steps: first, we rewrite the ARMA-based overall SEM to the sectional SEM for time constant path coefficients (TCC) (i.e., Overall-TCC SEM←→ Sectional-TCC SEM); subsequently, we transform the sectional SEM with TVC into the sectional SEM with TCC by introducing artificial variables (i.e., Sectional-TVC SEM ←→ Sectional-TCC SEM). Based on these two steps, we can thus use existing SEM software based on independent observations to analyze TVC SEM based on vector ARMA(p, q) time series data. The above strategy is applicable to both single- and multiple-subject time series data.

For multiple-subject time series data, we further compared the following six different analysis approaches for both TCC and TVC SEM models and provided guidelines on the adoption of these methods:

(1) Overall SEM (only available for $N > T$, where N and T are the total number of subjects and time points respectively);
(2) Sectional SEM using Approach 1 -- summarize then analyze;
(3) Sectional SEM with Approach 1a -- summarize then analyze with the usual bootstrapping method;
(4) Sectional SEM with Approach 1b -- summarize then analyze with enlarged T form bootstrap resampling;
(5) Sectional SEM with Approach 2 -- analyze then summarize;
(6) Sectional SEM with Approach 3 -- simultaneous analysis with concatenated dataset.

Simulation studies demonstrated that the traditional TCC SEM would yield significantly biased estimators when the underlying time series data contain contemporaneous pathways with time varying path coefficients. The TVC SEM we had initially established was based on the Taylor series expansion in that a regular function can usually be represented by a polynomial in time (t). However, we discovered, through the simulation study, that it is perhaps more efficient to directly estimate a non-linear function in t such as the log or the square root of t, rather than approximate it using a truncated polynomial. Subsequently, we have extended our TVC model to include non-linear TVC function as shown in the following section. In addition, we have developed a moving window approach towards the estimation of the TVC structure as is also shown in the following section.

7.1 SEM with Non-linear TVC Functions

Our framework can be extended to scenarios of TVC of non-linear functions in t such as $\lambda(t) = \alpha + \beta t^{1/2}$ and $\lambda(t) = \alpha + \beta \log(t)$. To detect which function to use for better model fit, a "moving window" technique is introduced in this section. Based on this straightforward visualization tool, one can directly estimate the underlying λ(t) structure.

We start with a simulated dataset. A bivariate AR(1) series (T=80) is generated following Equation (5.14), in which a=0.6, b=0.4, c=0.2, $\lambda(t)$=ln(t)/5+$\tau(t)$, where $\tau(t)$ follows N(0, 0.01^2). A series of sample windows are generated for the sectional TCC SEM. To guarantee a positive definite sample covariance matrix, each window has 30 time points, and thus we have 50 windows in total: T2-T31, T3-T32, …, T51-T80. The estimates of these windows are plotted in Figure 7.1. One can see the estimated $\lambda(t)$ curve increases and becomes flat at the end, which indicates that a log-like function would be better than a simple linear function in t. The comparison of TVC model estimates is shown in Table 7.1.

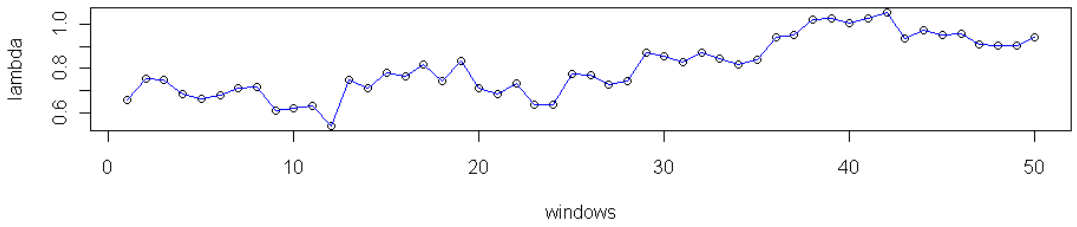


Figure 7.1 Estimates of λ: 50 windows from T2-T80 of simulated data on the path $X(t)$→ $Y(t)$ based on a VAR(1) time series data.

Table 7.1 Comparison of TCC and different TVC functions using the simulated data.

Parameter	**True Value** $\lambda(t) = \alpha + \beta \log(t)$	**TCC SEM** $\lambda(t) = \alpha$		**TVC SEM** $\lambda(t) = \alpha + \beta t$		**TVC SEM** $\lambda(t) = \alpha + \beta \log(t)$	
		Est. (std)	**p-value**	**Est. (std)**	**p-value**	**Est. (std)**	**p-value**
α	0	0.6503 (0.1947)	p<0.01	0.5500 (0.1943)	p<0.01	-0.0142 (0.1861)	p=0.939
β	0.2	0 (N/A)	N/A	0.0039 (0.0055)	p=0.487	0.2083 (0.0505)	p<0.01
a	0.6	0.6096 (0.0070)	p<0.01	0.6096 (0.0070)	p<0.01	0.6096 (0.0070)	p<0.01
b	0.4	0.5154 (0.0355)	p<0.01	0.5154 (0.0355)	p<0.01	0.3771 (0.0514)	p<0.01
c	0.2	-0.1606 (0.1212)	p=0.192	-0.1606 (0.1212)	p=0.192	0.2173 (0.1159)	p=0.069

A real data example is also studied based the *f*MRI time series visual attention

data introduced in Chapter 6. For simplicity, we focus on the pathway between two brain regions only. That is, the pathway from the thalamus (THAL) to the posterior parietal cortex (PPC). To explore the structure of this path coefficient along time, we used the "moving window" method again with a total of 18 windows from T2 to T48. The resulting estimated $\lambda(t)$ curve decreases in time as shown in Figure 7.2. We compared the fit of two potential $\lambda(t)$ functions as the following: $\lambda(t) = \alpha + \beta t$ and $\lambda(t) = \alpha + \beta t^{1/2}$. As shown in Table 7.2, the non-linear TVC function fits better ($p < 0.01$) than the linear TVC function ($p = 0.084$).

Figure 7.2 Estimates of λ: 18 windows from T2-T48 of the *f*MRI visual attention data introduced in Chapter 6, on the path THAL→PPC.

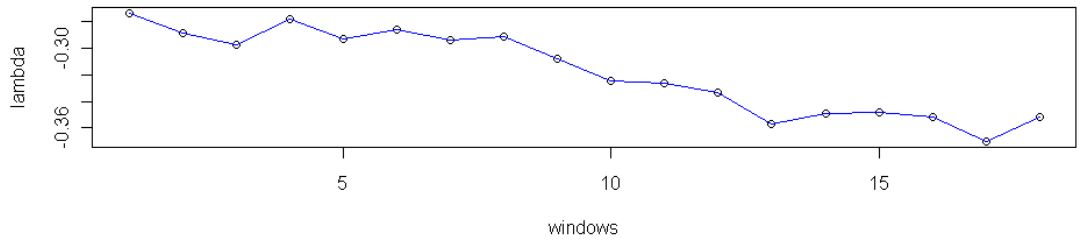


Table 7.2 Comparison of TCC SEM and different TVC SEM functions with the *f*MRI data.

Parameter	TVC SEM T2-T48 $\lambda(t) = \alpha + \beta t$		TVC SEM T2-T48 $\lambda(t) = \alpha + \beta t^{1/2}$	
	Est. (std)	**p-value**	**Est. (std)**	**p-value**
α	-0.2931 (0.0617)	p<0.01	-0.0806 (0.0619)	p=0.199
β	**-0.0057 (0.0032)**	**p=0.084**	**-0.0438 (0.0116)**	**p<0.01**
a	0.5801 (0.1214)	p<0.01	0.5801 (0.1214)	p<0.01
b	0.3786 (0.1072)	p<0.01	0.3916 (0.1081)	p<0.01
c	0.3830 (0.0964)	p<0.01	0.2448 (0.0720)	p<0.01

Both the simulation and the real data example have demonstrated the effectiveness of the "moving window" method for a quick estimation of the underlying TVC structure. Of course, more work has to be done to determine the optimal window length through power and sample size estimations and more simulation studies.

A fundamental draw-back of the SEM approach is that it is a confirmatory analysis tool that depends heavily on the initial hypothesized pathway drawn upon existing field

knowledge. When little is known about the underlying pathway, it is often difficult for the domain expert to arrive at a meaningful and/or realistic starting model. The salvation may lie in modern exploratory data analysis methods, such as the partial correlation network analysis (PCNA) (Peng et al., 2009). The PCNA is a complimentary approach to SEM for it requires no prior knowledge on the underlying pathway, but instead, entirely data-driven. With the increasing availability of time series gene expression data as well as *f*MRI time series data, the extension of partial correlation network analysis to time series data would greatly enhance our ability to uncover the underlying contemporaneous and longitudinal pathways. We have started working in this direction as shown in the next section.

7.2 Exploratory Pathway Analysis

Partial correlation analysis is very powerful tool in exploring possible association between variables. It was discovered and studied in the early 20th century by Pearson, Fisher and others (Fisher, 1924; Goodman and Kruskal, 1979; Isserlis, 1914; Pearson, 1915). Partial correlation is the correlation between two variables while controlling for a third or additional variables. It has been adopted in causal analysis through graphic models (Edwards, 2000; Whittaker, 1990). Researchers would compare the controlled correlation with the original correlation: if there is no difference, the inference is that the control variables have no effect; on the other hand, if the partial correlation approaches 0, it indicates that the original correlation is spurious. Hence, partial correlation can potentially reveal the true relationship between variables excluding the confounding effects from other variables.

A network among all of the variables of interest could then be established based on pairwise partial correlations. If there is a non-zero partial correlation between two variables, an edge is used to link them. For the partial correlation network analysis, traditional methods for estimation require the sample size (n) is larger than the number of variables (p) (e. g. Whittaker, 1990; Edward, 2000).

In recent study, methods have been introduced to estimate partial correlation with $p>n$, even $p>>n$, based on utilization of the sparse property of partial correlation matrix. Meinshausen and Buhlmann (2006) introduced a variable-by-variable approach for neighborhood selection via the lasso regression (Tibshirani, 1996). Yuan and Lin (2007) proposed a penalized maximum likelihood approach which performs model selection and estimation simultaneously and ensures the positive definiteness of the estimated concentration matrix. Friedman et al. (2008) proposed an efficient algorithm to implement this method, such that it can be applied to problems with high dimensions. Schafer and Strimmer (2005) proposed a shrinkage covariance estimation procedure to overcome the ill-conditioned problem of sample covariance matrix when $p > n$. Li and Gui (2006) introduced a threshold gradient descent regularization procedure. Bickel and Levina (2008) proposed to regularize the covariance matrix by hard thresholding for

families of covariance matrices satisfying suitable sparsity assumptions. Peng et al (2009) developed a new algorithm based on the joint sparse regression model (JSRM). The simulation study shows an improvement in performance for $p>>n$ data, and efficiency in identifying network hubs. Applications of PCNA to gene microarray data have yielded many meaningful "hub genes" (Barabasi and Oltvai, 2004). Figure 7.3 shows an example of partial correlation network in microarray study for certain neuron cells (unpublished work of our group). Besides genetics, the existence of hubs is also a well known phenomenon in many other large networks, such as the internet, citation networks, and protein interaction networks (Newman, 2003).

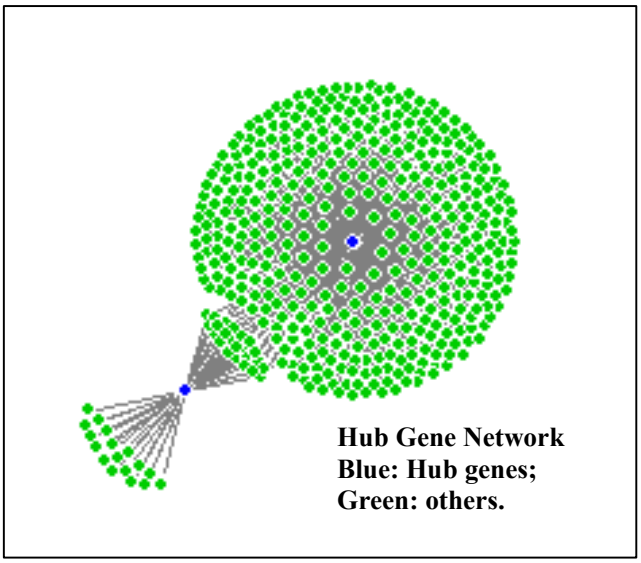


Figure 7.3 A PCNA gene network example. The data set is the genetic expressions for the perkinje (PKJ) cell, a class of GABAergic neurons located in the cerebellar cortex. Each dot denotes a gene. The blue ones have significantly more partial correlation edges and are thus identified as "hub genes".

In the *f*MRI study below, we adopted the joint sparse regression modeling approach, which is implemented with an R package (Peng et al, 2009). We use the subject-average *f*MRI data to establish the PCNA for the six ROIs, which is shown in Figure 7.4.1. Notice the partial correlation itself does not indicate the direction of the linkage between variables. We decide the arrow directions if there is a clear causality referring to the original hypothesis (e. g. APC → LPFC), and keep the two-head arrows for the rest edges (e. g. THAL ←→ APC). The hypothesis for the sectional SEM is then built in Figure 7.4.2, in which the longitudinal pathways are added based on the PCNA output. The equations for the TCC SEM are shown in (7.1). Table 7.3 shows the results of the new TCC model with Approach 1 (summarize then analyze).

$$\begin{pmatrix} CERCB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{41} & \lambda_{42} & \lambda_{43} & 0 & 0 & \lambda_{46} \\ 0 & 0 & 0 & \lambda_{54} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{64} & 0 & 0 \end{pmatrix} \bullet \begin{pmatrix} CERCB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix}$$

$$+ \begin{pmatrix} \gamma_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_{22} & 0 & \gamma_{24} & 0 & 0 \\ 0 & 0 & \gamma_{33} & 0 & 0 & 0 \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & 0 & \gamma_{46} \\ 0 & 0 & 0 & \gamma_{54} & \gamma_{55} & 0 \\ 0 & 0 & 0 & \gamma_{64} & 0 & \gamma_{66} \end{pmatrix} \bullet \begin{pmatrix} CERCB(t-1) \\ THAL(t-1) \\ PPC(t-1) \\ APC(t-1) \\ LPFC(t-1) \\ SMA(t-1) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \\ \varepsilon_4(t) \\ \varepsilon_5(t) \\ \varepsilon_6(t) \end{pmatrix} \quad (7.1)$$

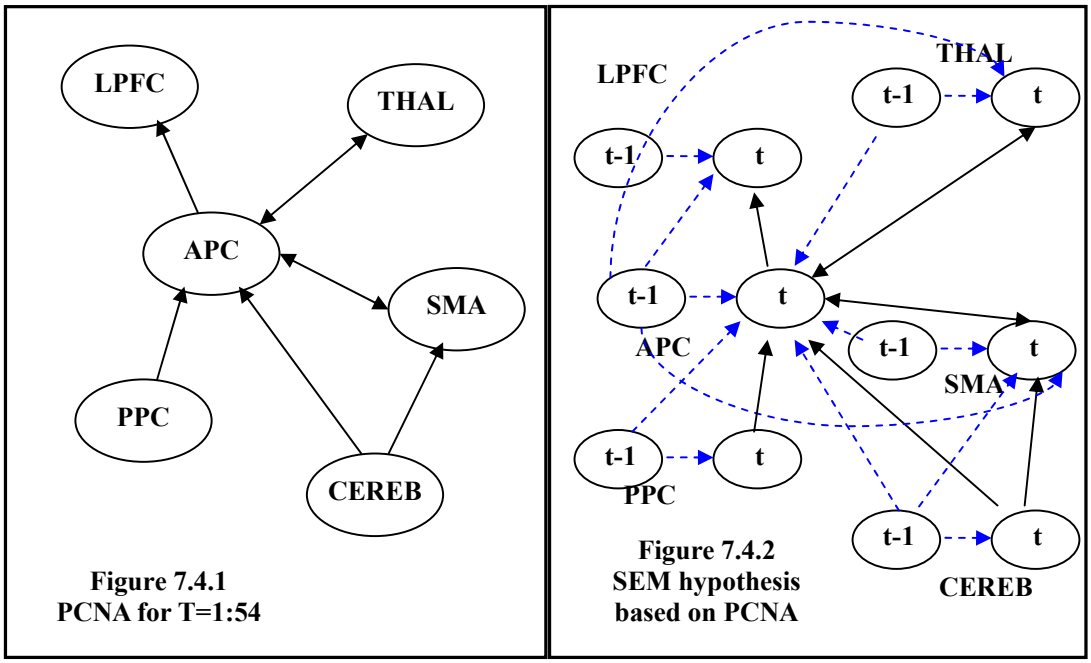


Figure 7.4 The new pathway hypothesis using PCNA with the subject-average *f*MRI data without longitudinal paths (left) and with longitudinal paths (right).

Table 7.3 Estimates of the TCC sectional SEM with the new hypothesis based on PCNA output (see Figure 6.3.2) using the subject-average *f*MRI data (Approach 1). Significant paths are in bold (two sided p-value based on t-test <0.05).

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 (0.091)**	**8.592 (<0.001)**	λ_{24}	-2.786 (1.730)	-1.611 (0.118)

γ_{22}	0.220 (0.322)	0.683 (0.500)	λ_{41}	**-0.287 (0.122)**	**-2.354 (0.026)**
γ_{24}	2.069 (1.193)	1.733 (0.094)	λ_{42}	**0.555 (0.117)**	**4.752 (<0.001)**
γ_{33}	**0.383 (0.114)**	**3.357 (0.002)**	λ_{43}	**0.437 (0.127)**	**3.424 (0.002)**
γ_{41}	-0.128 (0.136)	-0.942 (0.354)	λ_{46}	-0.068 (0.096)	-0.709 (0.484)
γ_{42}	**-0.439 (0.089)**	**-4.927 (<0.001)**	λ_{54}	0.201 (0.174)	1.155 (0.258)
γ_{43}	0.234 (0.122)	1.918 (0.065)	λ_{61}	**0.423 (0.182)**	**2.331 (0.027)**
γ_{44}	0.236 (0.139)	1.698 (0.100)	λ_{64}	**0.497 (0.203)**	**2.444 (0.021)**
γ_{46}	**0.267 (0.104)**	**2.562 (0.016)**			
γ_{54}	-0.300(0.175)	-1.717 (0.097)			
γ_{55}	**0.584 (0.113)**	**5.161 (<0.001)**			
γ_{61}	-0.329 (0.201)	-1.640 (0.011)			
γ_{64}	**-0.577 (0.208)**	**-2.766 (0.010)**			
γ_{66}	**0.515 (0.113)**	**4.558 (<0.001)**			

As we have stated above, PCNA is an exploratory tool to detect the variable relations. We also used the JSRM technique to generate the networks of the six ROIs within different time period. We divided the entire dataset into three subsets along the time steps (T=1:18, 19:36 and 37:54) and the partial correlation networks in different time periods are shown in Figure 7.5, which obviously indicate that the pathways vary with time.

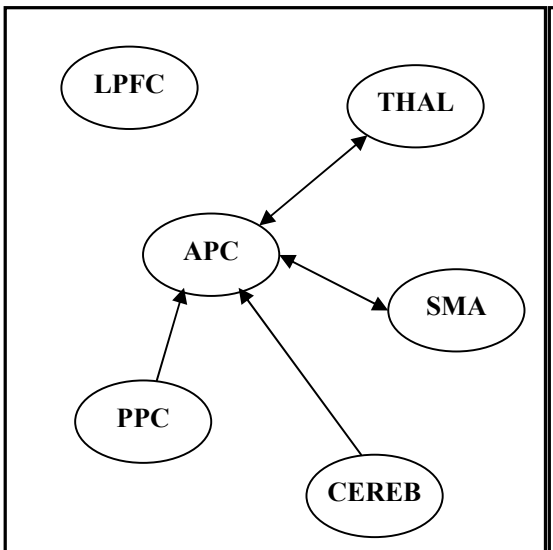


Figure 7.5.1
PCNA for T=1:18

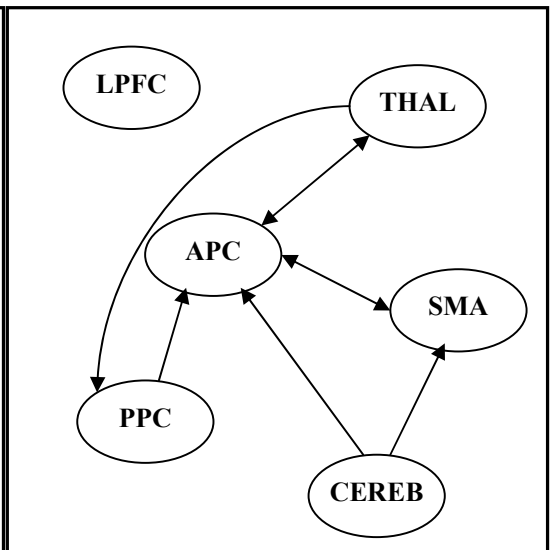


Figure 7.5.2
PCNA for T=19:36

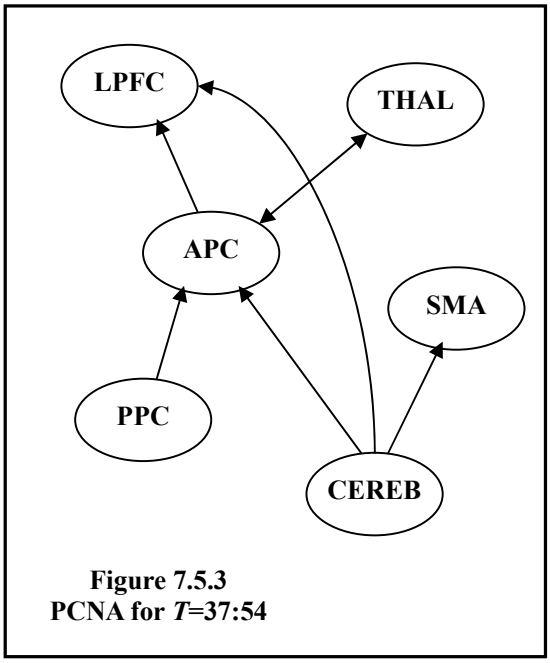


Figure 7.5.3
PCNA for *T*=37:54

Figure 7.5

Different pathways at different time periods. The subject-average *f*MRI data is used with the JSRM method (**R** *space* package).

The equations for the TVC SEM are shown in (7.2). Table 7.4 tabulates the results of the new TVC model with Approach 1. The confirmed TCC and TVC models are displayed in Figure 7.6. We found there are significant time-varying coefficients (β_{42}, β_{46}, β_{61}) in the TVC SEM at the significance level of 0.05.

$$\begin{pmatrix} CERCB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_{24}+\beta_{24}t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{41}+\beta_{41}t & \alpha_{42}+\beta_{42}t & \alpha_{43}+\beta_{43}t & 0 & 0 & \alpha_{46}+\beta_{46}t \\ 0 & 0 & 0 & \alpha_{54}+\beta_{54}t & 0 & 0 \\ 0 & 0 & 0 & \alpha_{64}+\beta_{64}t & 0 & 0 \end{pmatrix} \bullet \begin{pmatrix} CERCB(t) \\ THAL(t) \\ PPC(t) \\ APC(t) \\ LPFC(t) \\ SMA(t) \end{pmatrix}$$

$$+ \begin{pmatrix} \gamma_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_{22} & 0 & \gamma_{25} & 0 & 0 \\ 0 & 0 & \gamma_{33} & 0 & 0 & 0 \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & 0 & \gamma_{46} \\ 0 & 0 & 0 & \gamma_{54} & \gamma_{55} & 0 \\ 0 & 0 & 0 & \gamma_{64} & 0 & \gamma_{66} \end{pmatrix} \bullet \begin{pmatrix} CERCB(t-1) \\ THAL(t-1) \\ PPC(t-1) \\ APC(t-1) \\ LPFC(t-1) \\ SMA(t-1) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \\ \varepsilon_4(t) \\ \varepsilon_5(t) \\ \varepsilon_6(t) \end{pmatrix} \tag{7.2}$$

Table 7.5 Estimates of the TVC sectional SEM with the new hypothesis based on PCNA output (see Figure 7.3.2) using the subject-average *f*MRI data (Approach 1). Significant paths are in bold (two sided p-value based on t-test <0.05).

Longitudinal path parameters			Contemporaneous path parameters		
Path parameters	Est. (S. E.)	t-value (p-values)	Path parameters	Est. (S. E.)	t-value (p-value)
γ_{11}	**0.780 (0.091)**	**8.592 (<0.001)**	α_{24}	-0.824 (0.582)	-1.416 (0.163)
γ_{22}	**0.440 (0.162)**	**2.725 (0.009)**	β_{24}	-0.018 (0.020)	-0.915 (0.364)
γ_{24}	1.350 (0.778)	1.736 (0.089)	α_{41}	**-0.265 (0. 089)**	**-2.987 (0.004)**
γ_{33}	**0.383 (0.114)**	**3.357 (0.001)**	β_{41}	-0.006 (0.005)	-1.152 (0.255)
γ_{41}	-0.105 (0.183)	-0.576 (0.567)	α_{42}	**0.326 (0.061)**	**5.348 (<0.001)**
γ_{42}	-0.118 (0.101)	-1.172 (0.247)	β_{42}	**-0.007 (0.003)**	**-2.385 (0.021)**
γ_{43}	0.125 (0.135)	0.926 (0.359)	α_{43}	**0.225 (0.083)**	**2.723 (0.009)**
γ_{44}	**0.270 (0.106)**	**2.538 (0.014)**	β_{43}	0.003 (0.005)	0.551 (0.584)
γ_{46}	-0.168 (0.132)	-1.270 (0.210)	α_{46}	0.012 (0.080)	0.148 (0.883)
γ_{54}	-0.448 (0.340)	-1.319 (0.193)	β_{46}	**0.015 (0.004)**	**3.384 (0.001)**
γ_{55}	**0.579 (0.114)**	**5.090 (<0.001)**	α_{54}	0.209 (0.170)	1.229 (0.225)
γ_{61}	**-0.972 (0.265)**	**-3.676 (<0.001)**	β_{54}	0.005 (0.010)	0.498 (0.621)
γ_{64}	-0.274 (0.358)	-0.766 (0.447)	α_{61}	**0.465 (0.168)**	**2.761 (0.008)**
γ_{66}	**0.450 (0.104)**	**4.324 (<0.001)**	β_{61}	**0.18 (0.006)**	**3.064 (0.003)**
			α_{64}	0.261 (0.205)	1.271 (0.209)
			β_{64}	-0.011 (0.010)	-1.104 (0.275)

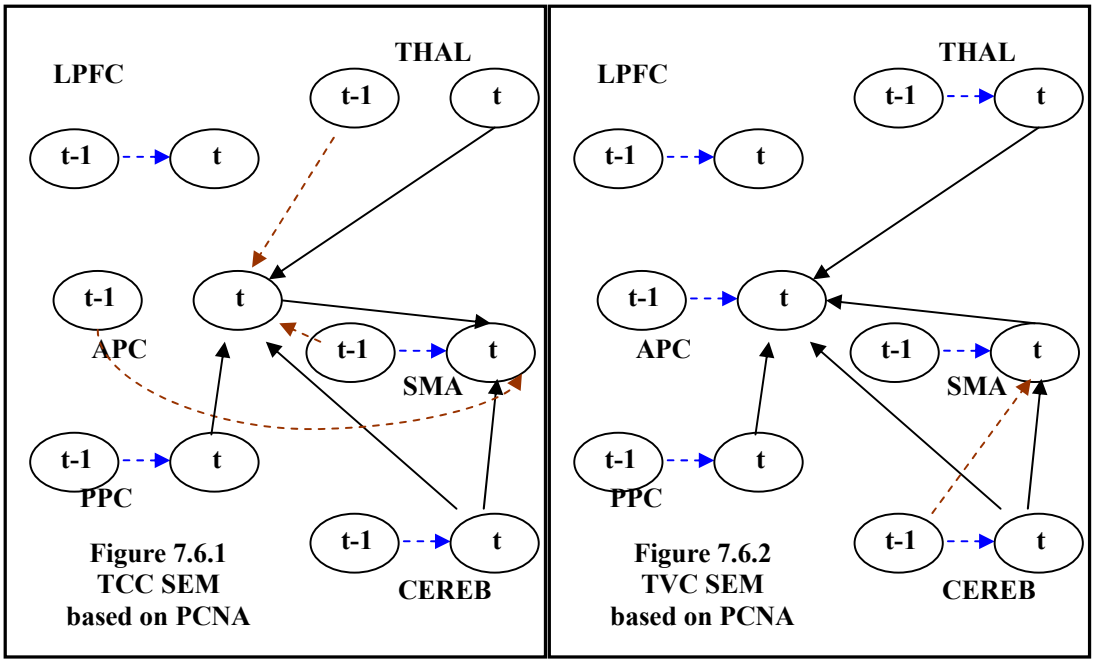


Figure 7.6 The confirmed pathways in TCC SEM (left) and TVC SEM (right). The hypotheses are updated with the PCNA output. The significance level is 0.05

References

Arbuckle, J. L. (1995). Amos 6.0 User's Guide: Amos Development Corporation.

Bandettini, P. A., Jesmanowicz, A., Wong, E. C. and Hyde, J. S. (1993). Processing Strategies for Time-Course Data Sets in Functional Mri of the Human Brain. *Magnetic Resonance in Medicine* **30**, 161-173.

Barabasi, A. L. and Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics* **5**, 101-U15.

Bentler, P. M. (1983). Simultaneous Equation Systems as Moment Structure Models - with an Introduction to Latent Variable Models. *Journal of Econometrics* **22**, 13-42.

Bentler, P. M. (1995). EQS sturctural equations program manual. Encino, CA: Multivariate Software.

Bentler, P. M., Wu, E. J. C. (2002). EQS 6 for windows user's guide. Encino CA.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Annals of Statistics* **36**, 199-227.

Bock, R. D. and Bargmann, R. E. (1966). Analysis of Covariance Structures. *Psychometrika* **31**, 507-507.

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

Bollen, K. A. Long, J. S. (1993). Testing structural equation modeling. Thousand Oaks, CA: Sage Publications, Inc.

Bonvento, G., Cholet, N. and Seylaz, J. (2000). Sustained attenuation of the cerebrovascular response to a 10 min whisker stimulation following neuronal nitric oxide synthase inhibition. *Neuroscience Research* **37**, 163-166.

Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and control. San Francisco, CA: Holden-Day.

Brockwell, P. J. and Davis, R. A. (1996). Introduction to time series and forecasting. New York: Springer-Verlag.

Browne, M. W. (1982). Covariance structures. Cambridge: Cmabridge University Press.

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

Browne, M. W., du Toit, S. H. C. (1991). Models for learning data. Newbury Park, CA: Sage.

Buchel, C. and Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex* **7**, 768-778.

Buchel, C., Coull, J. T. and Friston, K. J. (1999). The predictive value of changes in effective connectivity for human learning. *Science* **283**, 1538-1541.

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

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

du Toit, S. H. C., Browne, M. W. (2001). The covariance structure of a vector time series. Chicago, IL: Scinetific Software International Inc.

du Toit, S. H. C., Browne, M. W. (2007). Structural equation modeling of multivariate time series. *Multivariate Behavioral Research* **42**, 67-101.

Edwards, D. (2000). Introduction to graphical modelling. New York: Springer.

Fletcher, P., Buchel, C., Josephs, O., Friston, K. and Dolan, R. (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex* **9**, 168-178.

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

Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432-441.

Friston, K. J. and Buchel, C. (2000). Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7591-7596.

Glodberger, A. S. (1972). Structural equation models in the social sciences. *Econometrica* **40**, 979-1001.

Goodman, L. A. and Kruskal, W. H. (1979). Measures of Association for Cross-classifications. New York: Springer-Verlag.

Grafton, S. T., Sutton, J., Couldwell, W., Lew, M., Waters, C. (2004). Network analysis of motro system connectivity in Parkinson's disease: Modulation of thalamocortical interactions after pallidotomy. *Human Brain Mapping* **2**, 45-55.

Gusnard, D. A. and Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience* **2**, 685-694.

Hamaker, E. L. D., Conor V. ; Molenaar, Peter, C. M. (2002). On the nature of SEM estimates of ARMA parameters. *Structural Equation Modeling* **9**, 347-368.

Hamaker, E. L. D., Conor V. ; Molenaar, Peter, C. M. (2003). ARMA-based SEM when the number of time points T exceeds the number of cases N: raw data maximum likelihood. *Structural Equation Modeling* **10**, 352-379.

Hamilton, J. D. (1995). Time Series Analysis. Princeton, NJ: Princeton University Press.

Honey, G. D., Fu, C. H. Y., Kim, J., Brammer, M. J., Croudace, T. J., Suckling, J., Pich, E. M., Williams, S. C. R. and Bullmore, E. T. (2002). Effects of verbal working memory load on corticocortical connectivity modeled by path analysis of functional magnetic resonance imaging data. *Neuroimage* **17**, 573-582.

Isserlis, L. (1914). On the partial correlation ratio - Part I Theoretical. *Biometrika* **10**, 391-411.

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

Jöreskog, K. G. (1971). Estimation and testing of simplex models. *The British Journal of Mathematical and Statistical Psychology* **23**, 121-145.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. New York: Academic Press.

Jöreskog, K. G., Sörbom, D. (1996). LISREL 8. user's reference guide. Chicago, IL: Scientific Software International.

Keesling, J. W. (1972). Maximum likelihood approaches to causal analysis. In *Department of Education*, vol. Ph. D. : Universtiy of Chicago.

Kenny, D. A. (2010). Measuring Model Fit: http://davidakenny.net/cm/fit.htm.

Kim, D. S., Duong, T. Q. and Kim, S. G. (2000). High-resolution mapping of iso-orientation columns by fMRI. *Nature Neuroscience* **3**, 164-169.

Kim, J. (2004). Path Analysis fo the visual attention network for functional MRI data. In *Department of Applied Mathematics and Statistics*, vol. Ph. D. Stony Brook: Stony Brook Universtiy.

Kim, J., Zhu, W., Chang, L., Bentler, P. M. and Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Human Brain Mapping* **28**, 85-93.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R. et al. (1992). Dynamic Magnetic-Resonance-Imaging of Human Brain Activity during Primary Sensory Stimulation. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5675-5679.

Lange, N. (1999). Statistical procedures for functional MRI. New York: Springe Medicine.

Li, H. Z. and Gui, J. (2006). Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. *Biostatistics* **7**, 302-317.

Loehlin, J. C. (2004). Latent variable models: an introduction to factor, path, and structural equation analysis. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150-157.

Malonek, D. and Grinvald, A. (1996). Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: Implications for functional brain mapping. *Science* **272**, 551-554.

McIntosh, A. R. and Gonzalezlima, 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.

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

McIntosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I. and Horwitz, B. (1994). Network Analysis of Cortical Visual Pathways Mapped with Pet. *Journal of Neuroscience* **14**, 655-666.

Meinshausen, N. and Buhlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics* **34**, 1436-1462.

Mills, T. C. (1990). Time Series Techniques for Economists. Cambridge: Cambridge University Press.

Molenaar, P. C. M. G., Jan G. DE ; Schmitz, Bernhard. (1992). Dynamic Factor Analysis of Nonstationary Multivariate Time Series. *Psychometrika* **57**, 333-349.

Molenaar, P. C. M. (1999). Comment on fitting MA time series by structural equation models. *Psychometrika* **64**, 91-94.

Neale, M. C., Boker, S. M., Xie, G., Maes, H. H. (1999). Mx: statistical modeling. Richmond, VA: Department of Psychiatry.

Nesselroade, J. R., McArdle, J. J., Aggen, S. H. and Meyers, J. M. (2002). Alternative Dynamic Factor Models for Multivariate Time Series Analyses. Mahwah, NJ: Lawrence Erlbaum Associates.

Newman, M. E. J. (2003). The structure and function of complex networks. *Siam Review* **45**, 167-256.

Pandit, S. M. and Wu, S. (1983). Time Series and System Analysis with Applications. New York: John Wiley & Sons Inc.

Pearson, K. (1915). On the partial correlation ratio. *Proceedings of the Royal Society of London Series a-Containing Papers of a Mathematical and Physical Character* **91**, 492-498.

Peng, J., Wang, P., Zhou, N. F. and Zhu, J. (2009). Partial Correlation Estimation by Joint Sparse Regression Models. *Journal of the American Statistical Association* **104**, 735-746.

Raichle, M. E. and Mintun, M. A. (2006). Brain work and brain imaging. *Annual Review of Neuroscience* **29**, 449-476.

Rogowska, J. and Wolf, G. L. (1992). Temporal Correlation Images Derived from Sequential Mr Scans. *Journal of Computer Assisted Tomography* **16**, 784-788.

SAS (2008). SAS/STAT 9.2 User's Guide. Cary, NC: SAS Institute Inc.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological* **58**, 267-288.

vanBuuren, S. (1997). Fitting ARMA time series by structural equation models. *Psychometrika* **62**, 215-236.

Whittaker, J. (1990). Graphical models in applied mathematical multivariate statistics. New York: John Wiley & Sons.

Wiley, D. E. (1973). The identification problem for sturctural equation models with unmeasured variables. New York: Academic Press.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research* **20**, 0557-0585.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94**, 19-35.

Zhang, Y. (2007). Path Analysis o fMultivariate Time Series fMRI Data with Subject-level Covariates. In *Department of Applied Mathematics and Statistics*, vol. Ph. D. . Stony Brook: Stony Brook University.